ROBUST JOINT SPARSITY MODEL FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Shaoguang Huang[1], *Hongyan Zhang*[2], *Wenzhi Liao*[1] *and Aleksandra Pižurica*[1]

[1]Department of Telecommunications and Information Processing, Ghent University, Belgium
[2]The State Key Lab. of Inform. Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, China

ABSTRACT

Sparsity-based classification methods have been widely used in hyperspectral image (HSI) classification. These methods typically assumed Gaussian noise, neglecting the fact that HSIs are often corrupted by different types of noise in practice. In this paper, we develop a robust super-pixel level joint sparse representation classification model (RSJSRC) to address the mixed noise problem in sparsity-based HSI classification. Our method takes into account both Gaussian and sparse noise. Experimental results on simulated and real data demonstrate the efficiency of the proposed method and clear benefits from the introduced mixed-noise model.

***Index Terms*—** Robust classification, hyperspectral image, super-pixel segmentation, sparse representation

1. INTRODUCTION

Hyperspectral images (HSIs) can provide detailed spectral information about the image objects in hundreds of narrow bands, allowing this way differentation between materials that are often visually indistinguishable. Numerous application areas include agriculture [1], defense and security [2] and environmental monitoring [3].

Classification of HSIs gains currently lots of attention in remote sensing community. The objective of supervised hyperspectral classification is to group pixels into different classes with the classifiers trained by the given training samples. A large number of HSI classification methods have been proposed, based on artificial neural networks [4], multinomial logistic regression [5], [6] and support vector machines (SVM) [7], just to name a few. With the target of exploiting spatial information in the classification task, spatial-spectral classification approaches have been developed, including SVM with composite kernels [8], methods based on mathmatical morphology [9–12] and image segmentation [13].

In recent years, sparse representation classification (SRC) emerged as another effective classification approach for HSI [14–18]. It assumes that each test sample can be sparsely represented as a linear combination of atoms from a dictionary, which is constructed or learned from training samples [14]. Chen *et al*. [14] first applied the joint sparse representation classification (JSRC) in HSI classification by incorporating spatial information. The model was based on the observation that the pixels in a patch share similar spectral characteristics and can be represented by a common set of atoms but with different sparse coefficients. Zhang *et al* [15] proposed a nonlocal weighted joint sparse representation (NLW-JSRC) to further improve the classification accuracy. They enforced a weigh matrix on the pixels of a patch in order to discard the invalid pixels whose class was different from that of the central pixel. In addition, other improved JSRC models [16, 19, 20] also have been proposed for the HSI classification and achieved good results.

However, previous sparsity-based methods for HSI classification only take into account Gaussian noise. In real applications, HSIs are inevitably corrupted by different kinds of noise, including Gaussian noise, impulse noise, dead lines and strips [21]. Here, *sparse noise* is defined as the noise of arbitrary magnitude that only affects certain bands or pixels. It may arise due to the defective pixels and poor imaging conditions such as water vapor and atmospheric effect [22]. While this effect hinders the classification performance, we are not aware of any classification method that takes it explicitly into account. Therefore, it is desirable to develop a classification method which accounts for these degradations and validate its performance on real data.

We propose here a robust classification method for HSI in the presence of Gaussian noise and sparse noise, by extending and generalizing the JSRC model [14]. The key idea of our model is to integrate a prior for sparse noise together with the prior on the spatial distribution of class labels in the HSI within the same unified framework, and to derive accordingly an elegant classification method, alleviating effectively the influence of sparse noise. In order to exploit the available spatial information, we perform classification on a super-pixel level. We derive an optimization algorithm for our objective function, based on the alternating minimization strategy. We name the overall method robust super-pixel level joint sparse representation classification (RSJSRC) and val-

This work was supported by the FWO project: G.OA26.17N Dictionary Learning and Distributed Inference for the Processing of Large-scale Heterogeneous Image Data (DOLPHIN). Wenzhi Liao is a postdoctoral researcher of the Fund for Scientific Research in Flanders (FWO).

idate it on simulated and real data. The results demonstrate improved performance in comparison to related recent methods and a clear benefit resulting from the introduced noise model.

The rest of this paper is organized as follows. Section 2 introduces the classical sparsity-based models in HSI classification. Section 3 describes our proposed model and optimization algorithm. Section 4 presents experimental results with simulated and real data and Section 5 concludes the paper.

2. SPARSITY-BASED MODELS IN HSI CLASSIFICATION

2.1. Sparse representation classification

Let $\mathbf{x} \in \mathbb{R}^B$ be a test sample and $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, ..., \mathbf{D}_C] \in \mathbb{R}^{B\times d}$ a structured dictionary constructed from training samples, where B is the number of bands in the HSI; d is the number of training samples; C is the number of classes, and $\mathbf{D}_i$ (i=1,2,...,C) is the sub-dictionary in which each column is a training sample of i-th class. The goal of sparse representation is to represent each test sample as

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{R}^B$ is Gaussian noise and $\boldsymbol{\alpha} \in \mathbb{R}^d$ are sparse coefficients, satisfying

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{x} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 \quad s.t. \quad \|\boldsymbol{\alpha}\|_0 \le K. \tag{2}$$

$\|\boldsymbol{\alpha}\|_0$ denotes the number of non-zero elements in $\boldsymbol{\alpha}$ and K is the sparsity level, i.e. the largest number of atoms in dictionary $\mathbf{D}$ needed to represent any input sample $\mathbf{x}$. Problem (2) is typically solved with a greedy algorithm, such as Orthogonal Matching Pursuit (OMP) [23].

The class of the test sample is identified by calculating the class-specific residuals r_i [24]:

$$\begin{aligned} class(\mathbf{x}) &= \arg\min_{i=1,2,...,C} r_i(\mathbf{x}) \\ &= \arg\min_{i=1,2,...,C} \|\mathbf{x} - \mathbf{D}_i\boldsymbol{\alpha}_i\|_2, \end{aligned} \tag{3}$$

where $\boldsymbol{\alpha}_i$ are the sparse coefficients associated with class i.

2.2. Joint sparse representation classification

An effective method to exploit the spatial information of the HSI is using joint sparse representation of neighbouring pixels. The assumption is that the pixels in a small patch often belong to the same class and could share the same sparsity pattern, which means that all the neighbouring pixels can be represented by the same set of atoms but with different sets of coefficients [14]. In JSRC model, a square window is used to find the spatial neighbourhood for the central pixel and all the neighbouring pixels are stacked as the input matrix $\mathbf{X}^{pa} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T] \in \mathbb{R}^{B\times T}$, where $\mathbf{x}_i$ are the spectral signatures of pixels in one patch of size $\sqrt{T} \times \sqrt{T}$. $\mathbf{X}^{pa}$ is approximated by dictionary $\mathbf{D}$ and row-sparse matrix $\mathbf{A}^{pa} \in \mathbb{R}^{d\times T}$ as follows:

$$\mathbf{X}^{pa} = \mathbf{D}\mathbf{A}^{pa} + \mathbf{N}. \tag{4}$$

The sparse matrix $\mathbf{A}^{pa}$ can be obtained by solving the following problem [25]:

$$\begin{aligned} \hat{\mathbf{A}}^{pa} &= \arg\min_{\mathbf{A}^{pa}} \|\mathbf{X}^{pa} - \mathbf{D}\mathbf{A}^{pa}\|_F^2 \\ &s.t. \ \|\mathbf{A}^{pa}\|_{row,0} \le K_0, \end{aligned} \tag{5}$$

where $\|\mathbf{A}^{pa}\|_{row,0}$ denotes the number of non-zero rows of $\mathbf{A}^{pa}$ and K_0 is the row-sparsity level. In a similar way to SRC, the central test pixel of the patch is labeled by calculating the class-specific reconstruction errors:

$$class(\mathbf{x}_{central}) = \arg\min_{i=1,2,...,C} \|\mathbf{X}^{pa} - \mathbf{D}_i\mathbf{A}_i^{pa}\|_F, \tag{6}$$

where $\mathbf{A}_i^{pa}$ is the portion of sparse matrix $\mathbf{A}^{pa}$ associated with class i.

3. PROPOSED METHOD

In practice, HSI is often corrupted by multiple noises. Next to the Gaussian noise in (4), degradation like impulse noise, dead lines and strips are typically also present. We call these degradations sparse noise because they only corrupt relatively few pixels in HSI. We extend the model of $\mathbf{X}^{pa}$ in (4) as:

$$\mathbf{X}^{pa} = \mathbf{D}\mathbf{A}^{pa} + \mathbf{S}^{pa} + \mathbf{N}, \tag{7}$$

where $\mathbf{S}^{pa} \in \mathbb{R}^{B\times T}$ is the sparse noise of $\mathbf{X}^{pa}$. In order to better exploit the spatial information of HSI, we perform the HSI classification on a super-pixel level. The efficiency of super-pixel level analysis for HSI has been reported in recent works [16, 17].

3.1. Robust super-pixel level JSRC

Suppose that a HSI is segmented into p non-overlapping super-pixels [26], and each super-pixel is regarded as a homogeneous region with adaptive shape and size. It is assumed that all the pixels in one super-pixel can be represented by the same set of training samples as in the JSRC model. If we vectorize the super-pixel of size n_s into a matrix $\mathbf{X}^s \in \mathbb{R}^{B\times n_s}(s = 1, 2, ...p)$, the approximation for each super-pixel could be formulated by

$$\mathbf{X}^s = \mathbf{D}\mathbf{A}^s + \mathbf{S}^s + \mathbf{N}^s, \tag{8}$$

where $\mathbf{N}^s \in \mathbb{R}^{B\times n_s}$ is the Gaussian noise and $\mathbf{S}^s \in \mathbb{R}^{B\times n_s}$ is the sparse noise. The optimization problem with respect to

$\mathbf{A}^s$ and $\mathbf{S}^s$ becomes

$$\min_{\mathbf{A}^s,\mathbf{S}^s} \|\mathbf{X}^s - \mathbf{D}\mathbf{A}^s - \mathbf{S}^s\|_F^2 + \lambda\|\mathbf{S}^s\|_1$$
$$s.t.\ \|\mathbf{A}^s\|_{row,0} \leq K_0. \tag{9}$$

We define a new matrix $\mathbf{X} \in \mathbb{R}^{B\times N} = [\mathbf{X}^1, \mathbf{X}^2, ..., \mathbf{X}^p]$, which is stacked by all the super-pixels, where $N = \sum_{i=1}^{p} n_i$ is the number of pixels in the HSI. Also all the $\mathbf{A}^s$ and $\mathbf{S}^s$ are stacked as $\mathbf{A} \in \mathbb{R}^{d\times N}$=$[\mathbf{A}^1, \mathbf{A}^2, ..., \mathbf{A}^p]$ and $\mathbf{S} \in \mathbb{R}^{B\times N} = [\mathbf{S}^1, \mathbf{S}^2, ..., \mathbf{S}^p]$. Now we can formulate a unified classification framework as follows:

$$\min f(\mathbf{A}, \mathbf{S}) = \min_{\mathbf{A},\mathbf{S}} \|\mathbf{X} - \mathbf{D}\mathbf{A} - \mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_1$$
$$s.t.\ \|\mathbf{A}_i\|_{row,0} \leq K_0, i = 1, 2, ..., p, \tag{10}$$

where $\|\mathbf{S}\|_1$ is a norm defined as $\|\mathbf{S}\|_1 = \sum_{i,j} |S_{i,j}|$ and λ is a positive parameter used to control the tradeoff between reconstruction term and the sparse noise term.

The objective function (10) can be solved by an alternating minimization algorithm which will be described in detail next. Once sparse coefficient matrix $\mathbf{A}$ and sparse noise $\mathbf{S}$ are obtained, we can label the class for each super-pixel by

$$class(\mathbf{X}^s) = \arg\min_{i=1,2,...,C} \|\mathbf{X}^s - \mathbf{D}_i\mathbf{A}_i^s - \mathbf{S}^s\|_F, \tag{11}$$

where $\mathbf{A}_i^s$ denotes the sparse matrix of $\mathbf{A}^s$ corresponding to class i.

3.2. Optimization algorithm

In this section, we present an optimization algorithm for problem (10) by an alternating minimization strategy. The main idea is to split a difficult problem into two easy solvable ones by fixing one variable as the parameter in the other subproblem, and alternating the process iteratively, as it is done in [16, 27]. In the $(k + 1)$th iteration, we update $\mathbf{A}$ and $\mathbf{S}$ as follows:

$$\mathbf{A}^{(k+1)} = \arg\min_{\|\mathbf{A}_i\|_{row,0}\leq K_0, i=1,2,...,p} f(\mathbf{A}, \mathbf{S}^{(k)}) \tag{12}$$

$$\mathbf{S}^{(k+1)} = \arg\min_{\mathbf{S}} f(\mathbf{A}^{(k+1)}, \mathbf{S}) \tag{13}$$

Problem (12) can be separated into p sub-problems with respect to $\mathbf{A}^s$, as follows:

$$\min_{\mathbf{A}^s} \|\mathbf{X}^s - \mathbf{D}\mathbf{A}^s - \mathbf{S}^{s^{(k)}}\|_F^2$$
$$s.t.\quad \|\mathbf{A}^s\|_{row,0} \leq K_0, s = 1, 2, ...p, \tag{14}$$

which is similar to the JSRC model discussed in section 2.2 and also could be solved by the SOMP algorithm [25].

For problem (13), the optimization with respect to $\mathbf{S}^{(k+1)}$ is formulated by

$$\min_{\mathbf{S}} \|\mathbf{X} - \mathbf{D}\mathbf{A}^{(k+1)} - \mathbf{S}\|_F^2 + \lambda||\mathbf{S}||_1, \tag{15}$$

which is the well-known shrinkage problem. By introducing the following soft-thresholding operator:

$$\Re_\Delta(x) = \begin{cases} sgn(x)(|x| - \Delta) & if\ |x| \geq \Delta \\ 0 & if\ |x| < \Delta, \end{cases} \tag{16}$$

the solution of (15) could be given by

$$\mathbf{S}^{(k+1)} = \Re_{\lambda/2}(\mathbf{X} - \mathbf{D}\mathbf{A}^{(k+1)}). \tag{17}$$

The update of $\mathbf{A}$ and $\mathbf{S}$ is executed until the stop criterion is satisfied.

4. EXPERIMENTS

The performance of our RSJSRC method is tested on both simulated and real hyperspectral images, in comparison with SVM with radical basis function (RBF) kernel [28], SRC [24], JSRC [14] and NLW-JSRC [15]. The commonly used index measurements, such as overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) are adopted as the quantitative assessment of classification performances. All results are reported by the average of ten runs.

4.1. Results on simulated HSI experiment

The Washington DC image was collected by the Hyperspectral Digital Image Collection Experiment (HYDICE) as shown in Fig. 1. Due to its high quality, this image was commonly used to simulate corrupted data with different kinds of noise. We also generate our simulated data this way. The image is of size $280 \times 307 \times 210$ with the spectrum ranging from 0.4 to 2.4 μm and has six classes in total. In this experiment, we reduce the number of bands to 191 by removing the opaque bands. 5% of labeled samples were randomly selected as training samples and the reminder as test samples as shown in Table 1.

Four kinds of noise are added as follows: (1) Zero-mean Gaussian noise in all bands with SNR value for each band varying from 10 to 20 dB. (2) Impulse noise in bands 30-40 with 20% of corrupted pixels in each band. (3) Dead lines in bands 70-73 with width ranging from one line to three lines. (4) Strips in bands 101-104 with width ranging from one line to three lines.

The optimal parameters of our method were determined empirically as $p = 7000, \lambda = 0.02$ and $K_0 = 30$. For other classification methods in Table 1, all the parameters were tuned to give the best results. The super-pixel level joint sparse representation classification (SJSRC) method was also implemented with the same segmentation map as RSJSRC. The results in Table 1 show a superior performance of our method in terms of OA, AA and Kappa coefficient. With the exploitation of spatial information from super-pixels, the OA of SJSRC was at least improved by 6.8% over SRC, JSRC and NLW-JSRC. With the introduction of sparsity prior of mixed

Table 1. Results for simulated data with different classifiers.

Class	Class name	Train	Test	SRC	JSRC	NLW-JSRC	SJSRC	RSJSRC
1	Roof	146	2770	0.5842	0.7727	0.7790	0.7897	**0.7962**
2	Road	91	1728	0.4100	0.5219	0.5204	0.5122	**0.5425**
3	Trail	64	1200	0.6900	0.7417	0.7543	**0.9110**	0.9099
4	Grass	90	1700	0.7536	0.9463	0.9468	0.9801	**0.9834**
5	Shadow	56	1064	0.4234	0.5778	0.5617	0.8237	**0.8273**
6	Tree	65	1216	0.4792	0.5954	0.5881	0.6846	**0.7160**
OA				0.5650 ± 0.0087	0.7109 ± 0.0142	0.7114 ± 0.0156	0.7792 ± 0.0208	$\mathbf{0.7912 \pm 0.0192}$
AA				0.5567 ± 0.0142	0.6941 ± 0.0144	0.6917 ± 0.0160	0.7836 ± 0.0230	$\mathbf{0.7959 \pm 0.0190}$
κ				0.4623 ± 0.0123	0.6421 ± 0.0174	0.6426 ± 0.0192	0.7284 ± 0.0258	$\mathbf{0.7432 \pm 0.0232}$

noise to the SJSRC model, our method yielded further 1.2% increase over SJSRC, proving its efficiency in handling the mixed noise.

4.2. Results on real HSI experiment

The real data was acquired by the Airborne/Visible Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines region in North-western Indiana in 1992 as shown in Fig. 1 This image has 16 classes and 220 spectral reflectance bands ranging from 0.4 to 2.5μm. In this experiment, 20 water absorption spectral bands in 104-108, 150-163 and 200 are removed, therefore, the real hyperspectral image size is $145 \times 145 \times 200$. 9% of the labeled samples are randomly selected as training samples and the remainder as test samples, which is the same as that in [15].

The optimal parameters of our method were $p = 700$, $K_0 = 50, \lambda = 0.003$. For JSRC, the optimal widow size was 7×7 and sparsity level was 30. In NLW-JSRC, the parameters were chosen from the recommendation of [15]. For SVM and SRC classifiers, we tuned the parameters such to produce the best classification results. The results are listed in Table 2. In most cases, our method RSJSRC yields better results than other classifiers. Based on super-pixel segmentation, SJSRC model had at least 2.7% improvement over other classical methods, such as JSRC and NLW-JSRC, which exploited the spatial information from square window with fixed shape and size. Considering the sparse prior for multiple noise in the HSIs, our proposed RSJSRC improves OA by 1.5% over SJSRC.

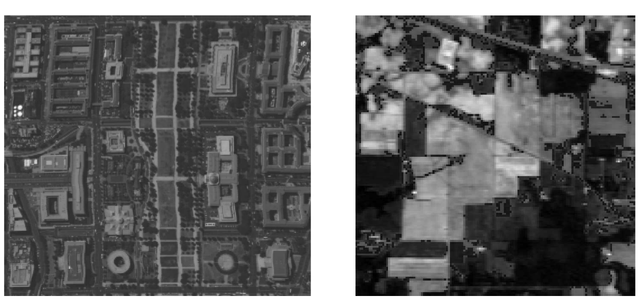

Fig. 1. Washington DC Mall in band 41 (left) and Indian Pines image in band 31 (right).

Table 2. Results for the Indian Pines with different classifiers.

Class	SVM	SRC	JSRC	NLW-JSRC	SJSRC	RSJSRC
1	0.6275	0.4125	0.5625	0.5950	**0.9800**	**0.9800**
2	0.7807	0.6122	0.8570	0,8917	**0.9799**	0.9427
3	0.7106	0.5396	0.8371	0,8617	**0.9601**	0.9426
4	0.5362	0.3286	0.6892	0,7113	**0.9920**	0.8441
5	0.8968	0.8478	0.9159	**0,9366**	0.9172	0.9163
6	0.9534	0.9307	0.9962	0,9976	**1.0000**	0.9976
7	0.8130	0.7565	0.6304	0,6783	**0.9696**	**0.9696**
8	0.9584	0.9170	0.9988	**0,9995**	0.9977	0.9977
9	0.5813	0.5125	0.4125	0,6625	**1.0000**	0.8000
10	0.7506	0.6103	0.8312	0,8665	0.8574	**0.9271**
11	0.8053	0.7000	0.8726	0,9137	0.9099	**0.9508**
12	0.7315	0.5075	0.8384	0,9026	0.9296	**0.9700**
13	0.9544	0.9538	**0.9967**	**0,9967**	0.9951	0.9951
14	0.9308	0.9056	0.9791	**0,9856**	0.9569	0.9818
15	0.5545	0.4596	0.7960	0,8369	0.8939	**0.9677**
16	0.9346	0.8531	0.9840	**0,9938**	0.9790	0.9679
OA	0.8096	0.7015	0.8851	0.9137	0.9407	**0.9547**
std.	0.0066	0.0039	0.0047	0.0064	0.0008	0.0095
AA	0.7825	0.6780	0.8248	0.8644	**0.9574**	0.9469
std.	0.0211	0.0137	0.0226	0.0283	0.0016	0.0271
κ	0.7827	0.6588	0.8690	0.9014	0.9325	**0.9483**
std.	0.0074	0.0043	0.0053	0.0074	0.0009	0.0109

5. CONCLUSION

In this paper, we proposed a robust classification method for HSI by combining a prior for sparse noise and a spatial distribution prior for the class labels within a unified framework. We derived an alternating minimization algorithm to solve the resulting problems, where we update the sparse coefficient matrix and sparse noise alternatively. The experiments on both real and simulated data demonstrated the effectiveness of the proposed approach.

6. REFERENCES

[1] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. Jupp, and J. S. 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.

[2] M. T. Eismann, A. D. Stocker, and N. M. Nasrabadi, "Automated hyperspectral cueing for civilian search and rescue," *Proceedings of the IEEE*, vol. 97, no. 6, pp. 1031–1055, 2009.

[3] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, 2014.

[4] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, 2010.

[5] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, 2010.

[6] R. Roscher, B. Waske, and W. Forstner, "Incremental import vector machines for classifying hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3463–3473, 2012.

[7] 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.

[8] D. Tuia, F. Ratle, A. Pozdnoukhov, and G. Camps-Valls, "Multisource composite kernels for urban-image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 88–92, 2010.

[9] 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. Remote Sens.*, vol. 43, no. 3, pp. 480–491, 2005.

[10] W. Liao, R. Bellens, A. Pizurica, W. Philips, and Y. Pi, "Classification of hyperspectral data over urban areas using directional morphological profiles and semi-supervised feature extraction," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 5, no. 4, pp. 1177–1190, 2012.

[11] W. Liao, A. Pizurica, R. Bellens, S. Gautama, and W. Philips, "Generalized graph-based fusion of hyperspectral and lidar data using morphological features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 552–556, 2015.

[12] W. Liao, M. Dalla Mura, J. Chanussot, R. Bellens, and W. Philips, "Morphological attribute profiles with partial reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1738–1756, 2016.

[13] S. Prasad, M. Cui, W. Li, and J. E. Fowler, "Segmented mixture-of-gaussian classification for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 138–142, 2014.

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

[15] 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. in Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, 2014.

[16] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, 2015.

[17] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, 2015.

[18] R. Roscher and B. Waske, "Shapelet-based sparse representation for landcover classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1623–1634, 2016.

[19] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, 2014.

[20] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 2, pp. 556–567, 2016.

[21] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, 2014.

[22] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, vol. 9, no. 9, pp. 4267–4279, 2016.

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

[24] 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, 2009.

[25] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

[26] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 2097–2104.

[27] D. Pham and S. Venkatesh, "Joint learning and dictionary construction for pattern recognition," in *IEEE conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8.

[28] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive svm for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, 2006.