Spectral–Spatial Preprocessing Using Multihypothesis Prediction for Noise-Robust Hyperspectral Image Classification

Chen Chen, *Student Member, IEEE*, Wei Li, *Member, IEEE*, Eric W. Tramel, *Member, IEEE*, Minshan Cui, *Student Member, IEEE*, Saurabh Prasad, *Member, IEEE*, and James E. Fowler, *Senior Member, IEEE*

Abstract—**Spectral–spatial preprocessing using multihypothesis prediction is proposed for improving accuracy of hyperspectral image classification. Specifically, multiple spatially collocated pixel vectors are used as a hypothesis set from which a prediction for each pixel vector of interest is generated. Additionally, a spectral-band-partitioning strategy based on inter-band correlation coefficients is proposed to improve the representational power of the hypothesis set. To calculate an optimal linear combination of the hypothesis predictions, a distance-weighted Tikhonov regularization to an ill-posed least-squares optimization is used. The resulting predictions effectively integrate spectral and spatial information and thus are used during classification in lieu of the original pixel vectors. This processed hyperspectral image dataset has less intraclass variability and more spatial regularity as compared to the original dataset. Classification results for two hyperspectral image datasets demonstrate that the proposed method can enhance the classification accuracy of both maximum-likelihood and support vector classifiers, especially under small sample size constraints and noise corruption.**

Index Terms—**Hyperspectral image classification, multihypothesis (MH) prediction, spectral–spatial analysis, Tikhonov regularization.**

I. Introduction

HYPERSPECTRAL IMAGERY (HSI) captures reflectance values over a wide range of electromagnetic spectra for each pixel in the image. This rich spectral information

C. Chen is with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: chenchen870713@gmail.com).

W. Li (corresponding author) is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: liwei089@ieee.org).

E. W. Tramel is with École Supérieure de Physique et de Chimie Industrielles (ESPCI), Paris 75005, France (e-mail: eric-william.tramel@espci.fr).

M. Cui and S. Prasad are with the Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77004 USA (e-mail: minshan@ieee.org; saurabh.prasad@ieee.org).

J. E. Fowler is with the Department of Electrical and Computer Engineering, Mississippi State University, Starville, MS 39762 USA (e-mail: fowler@ece.msstate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2013.2295610

potentially provides information useful for image classification and target recognition. HSI classification plays an important role in many remote-sensing applications, being a theme common to environmental mapping, crop analysis, plant and mineral identification, and abundance estimation, among others [1]. In such applications, the users are generally interested in assigning each pixel of a hyperspectral image to one of a number of given classes.

The traditional approach is to treat the spectral measurements of every pixel in the image as a separate signal and use pattern recognition techniques to label it without considering the correlations between the pixel of interest and its neighboring pixels [2]. Numerous supervised classification techniques for hyperspectral data have been developed based on pixel-level processing, i.e., techniques that assign each pixel to one of the classes based on its spectral values alone (so-called *pixel-wise* classifiers) [3]. Classical examples include the k-nearest-neighbor (k-NN) classifier [4], [5], maximum-likelihood estimation (MLE) [6], artificial neural networks [7], [8], fuzzy rule-based techniques [9], [10], and kernel-based techniques [11]–[14]. In particular, support vector machines (SVMs) have demonstrated excellent performance for classifying hyperspectral data when a relatively low number of labeled training samples are available [15]–[17] as well as when noise is present in the samples [11], [18].

However, besides the spectral variations in pixels, the spatial-context information contained in the image is also useful for accurate scene interpretation. Considering both spectral and spatial information for HSI classification has become a primary research focus in recent years. The ECHO classifier [19] was one of the first that incorporated both spectral- and spatial-context information. Segmentation techniques such as watershed, partitional clustering, and hierarchical segmentation (HSEG) have been explored to distinguish spatial structures in the hyperspectral image [3], [20] with a majority-voting rule applied on the segmentation, resulting in pixel-wise classification with improved performance. Additionally, the morphological-profile method [21]–[23] uses morphological transformations to build morphological profiles with the potential drawback that a large feature set is generated by a series of morphological transforms applied on the original data. Finally, wavelet transforms have been recently investigated for HSI classification, e.g., in [24] and [25], a 3-D discrete wavelet transform (3-D DWT)—which effectively captures the spatial information in various scales, frequencies, and orientations—is employed to extract spectral–spatial features from the HSI. The noise robustness of extracted wavelet features for HSI classification is reported in [25] as well.

In this paper, we propose a spectral–spatial preprocessing algorithm driven by the idea that each pixel vector in a hyperspectral image can be represented as a linear combination of the neighboring pixel vectors, since its neighboring pixel vectors will likely share similar spectral characteristics due to piecewise-continuous nature of the image. The focus is on the use of multihypothesis (MH) prediction [26] in which multiple hypotheses are created and then combined to yield a composite prediction. Central to our discussion is a formulation of the MH-prediction procedure as well as hypothesis generation based on spectral-band partitioning, as this formulation results in an ill-posed optimization that we resort to Tikhonov regularization [27], which is widely used to yield tractable solutions to such ill-posed problems.

The motivation to use MH prediction for HSI data preprocessing is our earlier success at applying it in compressed-sensing image and video reconstruction [28], [29], single-image super-resolution [30], and hyperspectral image reconstruction from random projections [31]. Here, we find that the HSI dataset processed by our MH prediction has less intraclass variability and greater spatial smoothness as compared to the original dataset. In particular, in experimental results, we compare the classification performance resulting from applying the same classifiers on both the original HSI datasets as well as the preprocessed datasets based on our proposed Tikhonov-regularized MH prediction. We find that our proposed approach yields superior classification accuracy. In addition, we also show that the proposed preprocessing algorithm is robust to white Gaussian noise, allowing for noise-robust classification.

The rest of this paper is organized as follows. In Section II, we provide a brief review of relevant parametric classification methods. Then, in Section III, we employ the general strategy of MH prediction as preprocessing to the problem of HSI classification, focusing as well on the generation of hypotheses for HSI data. In Section IV, experimental comparison of various algorithms is presented. Finally, in Section V, some concluding remarks are made.

II. PARAMETRIC CLASSIFICATION

The Gaussian MLE classifier [6]—arguably one of the most commonly employed parametric classifiers for remote-sensing tasks—assumes Gaussian class-conditional statistics and relies on the first- and second-order statistics of the data. Gaussian-MLE classification is simple (few parameters must be estimated) and has attractive convergence properties as the amount of training data increases [32]. A particularly popular parametric classification approach is to use an MLE classifier after dimensionality reduction based on Fisher's linear discriminant analysis (LDA) [33].

In contrast to the statistical-modeling approach of MLE classifiers, given a training dataset $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$ in $\mathbb{R}^d$ with class labels $y_i \in \{+1, -1\}$ and a nonlinear kernel mapping $\phi(\cdot)$, the SVM classification method solves

$$\min_{\omega, \xi_i, b} \left\{ \frac{1}{2} \|\omega\|^2 + \varsigma \sum_{i=1}^n \xi_i \right\} \qquad (1)$$

subject to the constraints

$$y_i(\langle \phi(\omega, \mathbf{x}_i) \rangle + b) \geq 1 - \xi_i \qquad (2)$$

for $\xi_i \geq 0$ and $i = 1, \ldots, n$, where ω is normal to the optimal decision hyperplane (i.e., $\langle \omega, \phi(\mathbf{x}) \rangle + b = 0$), n denotes the number of samples, b is the bias term, ς is the regularization parameter that controls the generalization capacity of the machine, and ξ_i is the positive slack variable allowing one to accommodate permitted errors appropriately. The aforementioned problem is solved by maximizing its Lagrangian dual form [34]

$$\max_\alpha \left\{ \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\} \qquad (3)$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are nonzero Lagrange multipliers constrained to $0 \leq \alpha_i \leq \varsigma$ and $\sum_i \alpha_i y_i = 0$, for $i = 1, \ldots, n$. In this paper, a radial basis function (RBF) kernel [11] is considered such that

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right) \qquad (4)$$

where σ is a width parameter characterizing the RBF. Finally, the decision function is

$$f(\mathbf{x}) = \mathrm{sgn}\left(\sum_{i=1}^n y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + b \right). \qquad (5)$$

III. SPECTRAL–SPATIAL PREPROCESSING USING MH PREDICTION

In this section, we propose a spectral–spatial preprocessing algorithm based on MH prediction. The algorithm is driven by the idea that, for each pixel vector in a hyperspectral image, its neighboring pixel vectors will likely share similar spectral characteristics since HSI commonly exhibits piecewise continuity. Therefore, each pixel vector in an HSI is likely to lie close to some linear combination of its neighboring pixel vectors. In [31], MH prediction was employed for HSI reconstruction. Specifically, multiple predictions drawn for a pixel vector of interest were made from spatially surrounding pixel vectors within an initial (nonpredicted) reconstruction. These predictions were then combined to yield a composite prediction that was close to the original pixel vector in the projection domain. In this paper, we adopt the same MH prediction approach but modify the procedure to obtain a prediction that approximates each pixel vector of interest. The processed image dataset has less intraclass variability and greater spatial regularity as compared to the original dataset. In addition, the proposed preprocessing algorithm exhibits robustness to white Gaussian noise.

To demonstrate the benefit of using MH prediction as a preprocessing for HSI classification, we use the Fisher discriminant ratio [33] to determine the effect of MH preprocessing on the separability of two HSI datasets as compared to three filtering

Method	Indian pines	University of Pavia
MH	$\mathbf{6.75 \times 10^9}$	$\mathbf{2.05 \times 10^4}$
LM	3.09×10^9	1.03×10^3
Wiener	8.64×10^5	6.17×10^0
Bilateral	6.17×10^3	1.19×10^2
Original	1.09×10^2	1.34×10^{-2}

approaches. The Fisher discriminant ratio is calculated as the ratio of interclass scatter to intraclass scatter, i.e., larger ratios indicate tighter class distributions with greater separation between classes and, therefore, easier class discrimination. First, we compare against localmean (LM) filtering that substitutes each pixel vector with the mean of its neighboring pixels within a spatial window. We also compare to the separability performance of Wiener and bilateral filtering applied band-by-band on the the two HSI datasets. The resulting Fisher discriminant ratios of the preprocessed datasets are reported in Table I. As shown, both the datasets preprocessed using MH prediction exhibit much higher Fisher discriminant ratios than the original datasets. Next, we describe our proposed MH prediction process in detail.

A. MH Prediction

Let us consider the input hyperspectral image as a set of M pixel vectors $\mathbf{X} = \{\mathbf{x}_m \in \mathbb{R}^N, m = 1, 2, \ldots, M\}$, where N is the number of spectral bands. Suppose that $\mathbf{x}$ is a pixel vector (hyperspectral signature) from the image. We aim to find an optimal linear combination of all possible predictions, or *hypotheses*, to represent $\mathbf{x}$. The optimal representation can be formulated as

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{Hw}\|_2^2 \tag{6}$$

where $\mathbf{H}$ is a matrix of dimensionality $N \times L$, whose columns are L possible hypotheses, and $\widehat{\mathbf{w}} = [\widehat{w}_1, \ldots, \widehat{w}_L]^T$ is an $L \times 1$ vector of coefficients corresponding to all the columns of $\mathbf{H}$. However, because $N < L$ for large window sizes, the ill-posed nature of the problem requires some kind of regularization to differentiate among the infinite number of possible linear combinations that lie in the solution space of (6).

The most common approach to regularizing a least-squares problem is Tikhonov regularization [27], which imposes an ℓ_2 penalty on the norm of $\widehat{\mathbf{w}}$

$$\widehat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \|\mathbf{x} - \mathbf{Hw}\|_2^2 + \lambda \|\Gamma \mathbf{w}\|_2^2 \tag{7}$$

where Γ is the Tikhonov matrix and λ is the regularization parameter; this strategy for MH prediction was initially proposed in [28]. The Γ term allows the imposition of prior knowledge on the solution; in some contexts, it might make sense to use a high-pass operator or a difference operator to enforce smoothness on the solution, and in other cases, to set $\Gamma = I$ to impose an energy constraint on the solution. In our case, we take the approach proposed in [28] and [29], where the hypotheses that are most dissimilar from the original pixel vector should be given less

weight than the hypotheses that are most similar. Specifically, a diagonal Γ takes the form of

$$\Gamma = \begin{bmatrix} \|\mathbf{x} - \mathbf{h}_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|\mathbf{x} - \mathbf{h}_L\|_2 \end{bmatrix} \tag{8}$$

where $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_L$ are the columns of $\mathbf{H}$. With this structure, Γ penalizes weights of large magnitude assigned to those hypotheses that have a significant distance from $\mathbf{x}$. For each pixel vector, $\widehat{\mathbf{w}}$ can then be calculated directly by the usual Tikhonov solution

$$\widehat{\mathbf{w}} = \left(\mathbf{H}^T \mathbf{H} + \lambda \Gamma^T \Gamma\right)^{-1} \mathbf{H}^T \mathbf{x}. \tag{9}$$

Therefore, a predicted pixel vector that approximates $\mathbf{x}$ is calculated as

$$\bar{\mathbf{x}} = \mathbf{H} \widehat{\mathbf{w}} \tag{10}$$

and the predicted dataset $\bar{\mathbf{X}} = \{\bar{\mathbf{x}}_m \in \mathbb{R}^N, m = 1, 2, \ldots, M\}$ is generated by replacing each pixel vector in $\mathbf{X}$ with its corresponding predicted pixel vector. Furthermore, once we have the predicted dataset $\bar{\mathbf{X}}$ through MH prediction, we can use the predicted dataset as the current input dataset to repeat the MH prediction process in an iterative fashion.

Note that the LM filtering is a special case of MH prediction where the weights for the L hypotheses are identical and equal to $1/L$, such that $\bar{\mathbf{x}}$ becomes the mean vector of the hypothesis matrix $\mathbf{H}$

$$\bar{\mathbf{x}} = \frac{1}{L} \sum_{l=1}^{L} \mathbf{h}_l. \tag{11}$$

B. Generating the Hypothesis Set

An HSI dataset usually exhibits some degree of spatial piecewise continuity: for each sample, its spatially neighboring pixel vectors will likely share similar spectral characteristics. For the present approach, multiple hypotheses are generated for a pixel vector of interest by considering all neighboring pixel vectors within a spatial search window of size $\omega \times \omega$. This hypothesis search procedure is illustrated in Fig. 1. These neighboring pixel vectors are then placed as columns of the hypothesis matrix $\mathbf{H}$.

Since the spectral bands of a hyperspectral image are also correlated, they can be partitioned into several groups based on the correlation coefficients between bands such that the bands in each group are highly correlated with one another [35], [36]. For example, Fig. 2 illustrates the matrix of correlation coefficients between bands of the Indian Pines dataset as well as the resulting four spectral-band partitions. More generally, assume an N-dimensional pixel vector is divided into J partitions according to the cross-band correlation-coefficient matrix. Then, J distinct hypothesis vectors are created by keeping only one of the J partitions while replacing the other partitions with zeros (i.e., "zero padding") to form an N-dimensional vector. This process for hypothesis generation based on spectral-band partitioning is illustrated in Fig. 3 for $J = 3$.

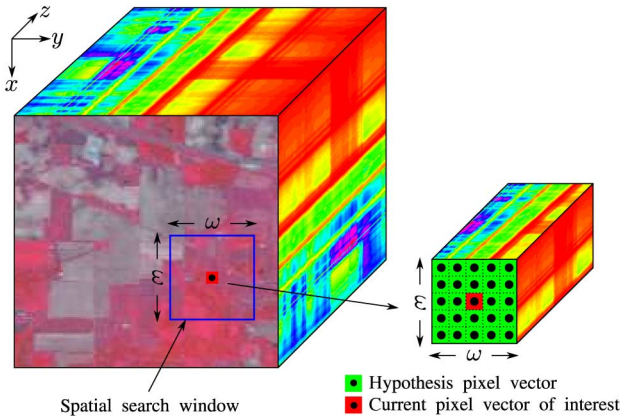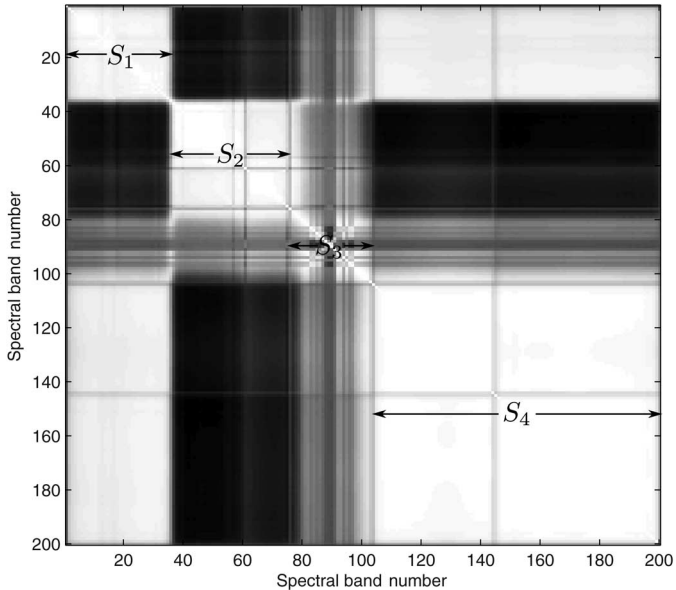Fig. 1. Generation of multiple hypotheses from a spatial search window with window size ω.

Fig. 2. About 200×200 matrix of cross-band correlation coefficients of the Indian Pines dataset with partitions labeled as S_1, S_2, S_3, and S_4. White, ± 1 and black, 0.
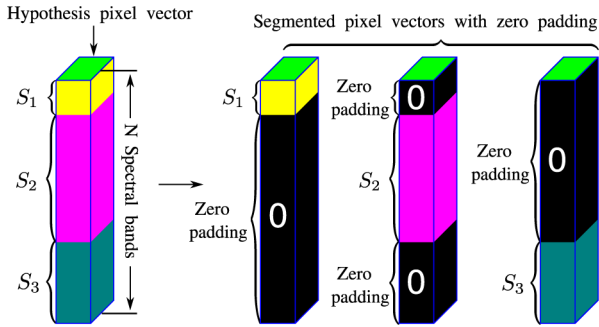

Fig. 3. Generation of a hypothesis set through spectral-band partitioning and zero padding for $J = 3$ partitions. S_1, S_2, and S_3 are spectral-band partitions.

If L hypotheses are drawn from the search window and J partitions are used for spectral-band partitioning, then the total number of hypotheses in $\mathbf{H}$ is JL. The motivation for this hypothesis generation from the partitioned spectral bands is such

Fig. 4. Formation of a prediction using multiple hypotheses and the corresponding weights for $J = 3$ partitions. $\mathbf{H}_{S_j} \in \mathbb{R}^{N \times L}$ $(j = 1, 2, 3)$ is the hypothesis set generated using the jth partition of the hypotheses with zero padding. $\widehat{\mathbf{w}}_{\mathbf{H}_{S_j}}$ are the corresponding set of weights for hypothesis set $\mathbf{H}_{S_j}$.

that the weights calculated for the hypotheses become adjustable for the different spectral partitions. The details are illustrated in Fig. 4 with $J = 3$ spectral-band partitions.

IV. EXPERIMENTAL RESULTS

A. Data Description and Experimental Setup

The first HSI dataset in our tests was acquired using NASA's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor and was collected over northwest Indiana's Indian Pines test site in June 1992. This scene represents a vegetation classification scenario with 145×145 pixels in the 0.4- to 2.45-μm region of the visible and infrared spectrum with a spatial resolution of 20 m. For this dataset, spectral bands $\{104–108\}$, $\{150–163\}$, and 220 correspond to water-absorption bands and are removed, resulting in 200 spectral bands. The original Indian Pines dataset consists of 16 ground-truthed land-cover classes. Some of the classes contain a small number of samples. For example, the *Oats* class has only 20 samples. Since MLE is employed as one of the classification methods, the probability distribution obtained from such a small number of training samples cannot well represent the statistical characteristics of the class. Therefore, we sort the 16 classes according to the number of samples in each class in ascending order and conduct a separate set of experiments with the last 9 classes, allowing for more training samples from a statistical viewpoint [16]. The class numbers of the 9 classes are highlighted in boldface in Table II.

The second dataset used in our experiments, University of Pavia, is an urban scene acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) [37]. The ROSIS sensor generates 115 spectral bands ranging from 0.43 to 0.86 m, has a spatial resolution of 1.3 m per pixel, and contains 610×340 pixels. The dataset consists of 103 spectral bands with the 12 noisiest bands removed. The labeled ground truth of this dataset is comprised of 9 classes. The class descriptions and sample distributions for both the Indian Pines and University of Pavia datasets are given in Tables II and III. Both datasets, and their corresponding groundtruths, are obtained from the publicly available Web site[1] of the Computational Intelligence Group from the Basque University (UPV/EHU).

TABLE II
PER-CLASS SAMPLES FOR THE INDIAN PINES DATASET

Class		Number of samples
No.	Name	
1	Alfalfa	46
2	**Corn-notill**	1428
3	**Corn-mintill**	830
4	Corn	237
5	**Grass-pasture**	483
6	**Grass-trees**	730
7	Grass-pasture-mowed	28
8	**Hay-windrowed**	478
9	Oats	20
10	**Soybean-notill**	972
11	**Soybean-mintill**	2455
12	**Soybean-clean**	593
13	Wheat	205
14	**Woods**	1265
15	Building-grass-trees-drives	386
16	Stone-steel-towers	93
Total		10 249

TABLE III
PER-CLASS SAMPLES FOR THE UNIVERSITY OF PAVIA DATASET

Class		Number of samples
No.	Name	
1	Asphalt	6631
2	Meadows	18 649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare soil	5029
7	Bitumen	1330
8	Self-blocking bricks	3682
9	Shadows	947
Total		42 776

For the Indian Pines dataset using 9 classes, we randomly chose a certain number of samples from each class for training and use the rest for testing. In all the experiments, we chose 5%, 7.5%, 10%, 12.5%, 15%, and 17.5%, respectively, of the samples from each class as a function of training samples. Each classification experiment is repeated for 20 trials with different training and testing samples and overall classification accuracy is averaged over the 20 repeated trials. The University of Pavia dataset is processed similarly, the only difference being that we first chose 900 samples at random from each class to form the total sample set of 8100 samples for each trial. Then the training and testing samples are chosen randomly from each class of the total sample set for classification. This procedure is used since some classes of the University of Pavia dataset contain significantly more samples than other classes, which might bias the accuracy. In order to have a fair comparison, the number of samples per class should be equal or similar.

All experiments are carried out using MATLAB (except SVM, which is implemented in C++) on an Intel i7 quadcore 2.63-GHz machine with 6 GB of RAM.

B. Optimizing MH Prediction Preprocessing Algorithm

As aforementioned, we generate hypotheses for the MH prediction procedure using spectral-band partitioning. To validate the benefit of this spectral partitioning, we also employ a hypothesis-generation procedure using the whole spectral band,
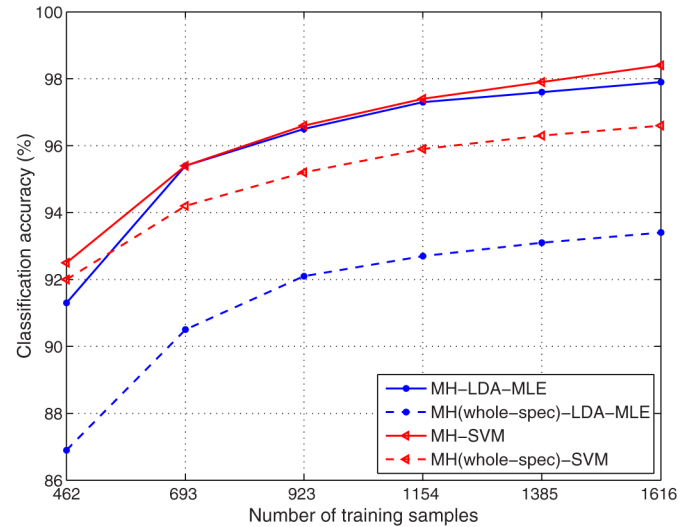

Fig. 5. Overall classification accuracy as a function of training samples for the Indian Pines dataset for MH prediction both with and without spectral-band partitioning (the latter denoted "whole-spec").

i.e., without spectral-band partitioning, denoted by MH(whole-spec). For an original HSI dataset, two processed HSI datasets are generated using MH prediction and MH(whole-spec) prediction, respectively. We then compare the classification performance by applying two conventional classifiers on those two processed HSI datasets. Specifically, LDA-MLE (denoting LDA followed by the MLE classifier) and SVM with an RBF kernel are the two classifiers used in our experiments. Therefore, MH-LDA-MLE, MH-SVM, MH(whole-spec)-LDA-MLE, and MH(whole-spec)-SVM are the four classification algorithms compared. The Indian Pines dataset is used as the experimental data. Although both the MH prediction and MH(whole-spec) prediction procedures can be applied iteratively, here in this experiment, we use only one iteration. The search window size ω is set to 9 for hypothesis generation based on spectral-band partitioning and 17 for hypothesis generation based on the whole spectral band in order to achieve optimal classification performance. The optimal parameters for SVM are chosen as those that maximize the training accuracy by means of fivefold cross-validation. It can be seen in Fig. 5 that MH prediction outperforms MH(whole-spec) prediction. Therefore, hypothesis generation based on spectral-band partitioning is used for MH prediction in all our experiments. Specifically, the spectral-band partitions for the Indian Pines dataset and the University of Pavia dataset are {1–35, 36–75, 76–105, 106–200} and {1–75, 76–103}, respectively. We note that the spectral-band partitions we considered here are manually chosen. A more sophisticated process—such as in [38], which uses edge detection on the correlation coefficient map—could be employed to remove the need for manual selection of band partitions.

An important parameter involved in MH prediction is the search-window size ω used in hypothesis generation. We analyze the effect of the search-window size in terms of the overall classification accuracy as well as the execution time of the algorithm. A set of window sizes, $\omega \in \{3, 5, 7, 9, 11, 13, 15\}$, is used for testing. Two iterations of MH prediction are used. From Fig. 6, we conclude that larger search-window size does
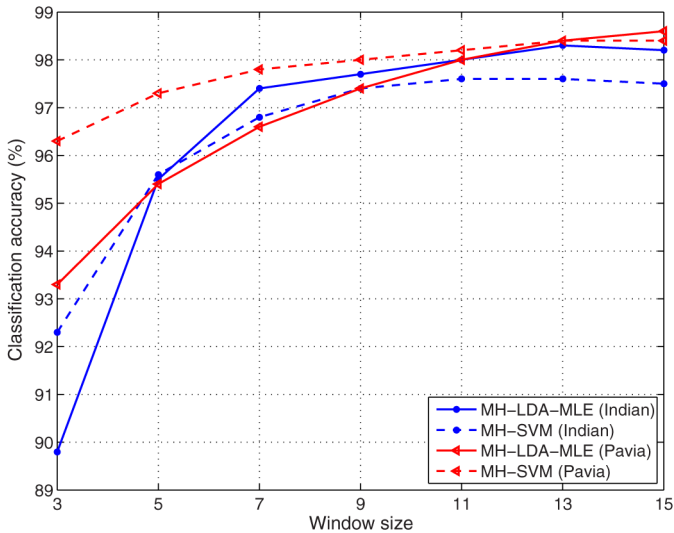
Fig. 6. Overall classification accuracy for Indian Pines and University of Pavia datasets as a function of search-window size ω.

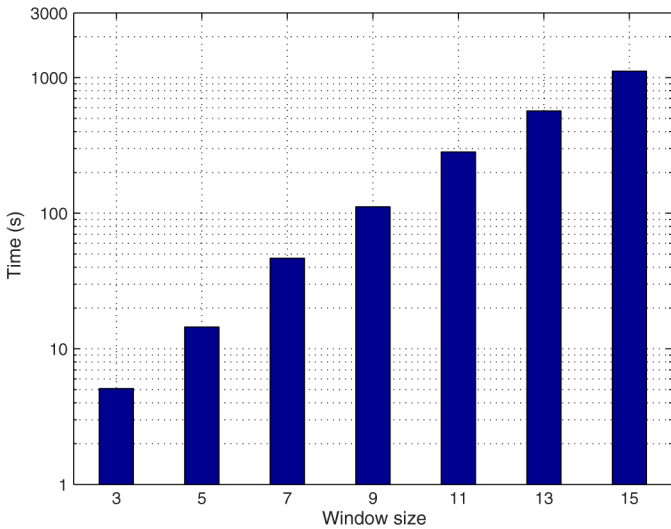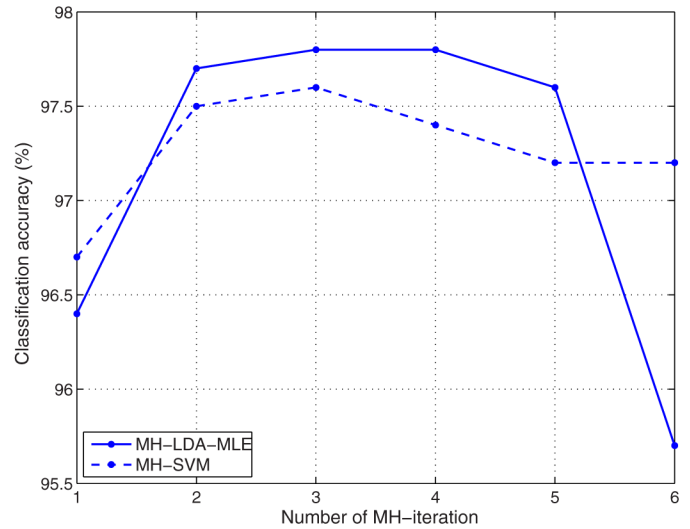Fig. 8. Overall classification accuracy for Indian Pines dataset as a function of MH prediction iterations.

Fig. 7. Execution time for one iteration of MH prediction for the Indian Pines dataset as a function of search-window size ω. A quadcore 2.63-GHz machine is used.
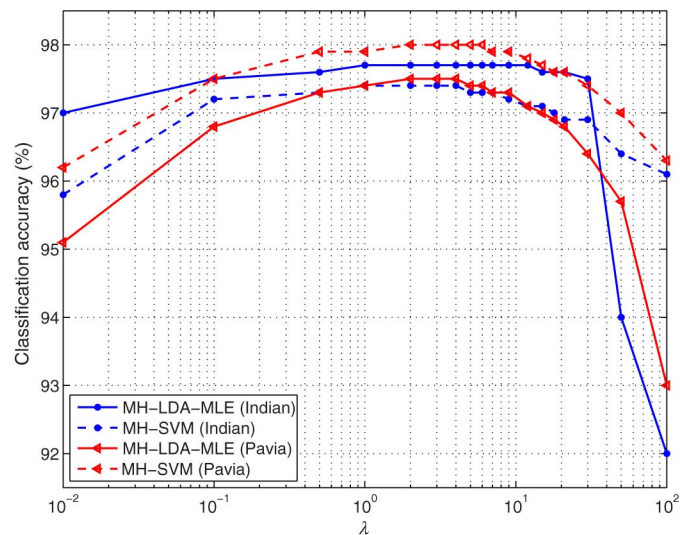
Fig. 9. Overall classification accuracy for Indian Pines and University of Pavia datasets as a function of the MH-prediction regularization parameter λ.

not necessary lead to higher classification accuracy. For example, classification accuracy starts to decrease when the window size is larger than 13×13 for the Indian Pines dataset. We also find that using $\omega = 11$ takes almost three times the execution time of $\omega = 9$ but does not yield any significant gains in classification accuracy.

Specifically, Fig. 7 shows the execution time of one iteration of MH prediction for various search-window sizes. In addition, we test a number of iterations for MH prediction and show the corresponding classification performance in Fig. 8. To balance between the runtime and the classification accuracy, we fix $\omega = 9$ and use two iterations for MH prediction in all subsequent experiments.

Another important parameter is λ that controls the relative effect of the Tikhonov regularization term in the optimization of (7). Many approaches have been presented in the literature—such as L-curve [39], discrepancy principle, and generalized

cross-validation (GCV)—for finding an optimal value for such regularization parameters. Here, we find an optimal λ by examining a set of values as shown in Fig. 9, which presents the overall classification accuracy with different values of λ for MH prediction. One can see that the classification accuracy is quite stable over the interval $\lambda \in [1, 4]$. As a result, in all the experiments reported here, we use $\lambda = 2$.

C. Experiments

In order to quantify the efficacy of the proposed spectral–spatial preprocessing algorithm, we compare three methods considering both spectral and spatial information for hyperspectral image classification. The first method is LM filtering, which is a special case of our proposed MH prediction, as illustrated in (11). The processed dataset, which has the same dimensionality as the original dataset, is then used for classification experiments

TABLE IV
OVERALL CLASSIFICATION ACCURACY (±STD) FOR THE INDIAN PINES DATASET (9 CLASSES) OVER 20 TRIALS AS A FUNCTION OF TRAINING SAMPLES

Method	Number of training samples					
	462	693	923	1154	1385	1616
MH-LDA-MLE	**94.7** ± 1.02	**96.8** ± 0.63	**97.7** ± 0.46	**98.2** ± 0.39	98.4 ± 0.25	98.7 ± 0.22
LM-LDA-MLE(15)	92.8 ± 1.22	95.9 ± 0.71	96.9 ± 0.44	97.5 ± 0.42	97.7 ± 0.39	98.0 ± 0.37
LM-LDA-MLE(9)	90.5 ± 1.05	94.5 ± 0.74	95.9 ± 0.57	96.6 ± 0.46	97.0 ± 0.40	97.3 ± 0.31
Wiener-LDA-MLE	88.1 ± 1.27	91.8 ± 0.82	93.3 ± 0.41	93.9 ± 0.49	94.4 ± 0.37	94.6 ± 0.35
LDA-MLE	60.8 ± 1.16	71.2 ± 1.04	75.3 ± 0.83	77.4 ± 0.79	78.7 ± 0.89	79.5 ± 0.76
MH-SVM	94.4 ± 0.99	96.4 ± 0.62	97.4 ± 0.42	98.0 ± 0.38	**98.5** ± 0.24	**98.8** ± 0.21
LM-SVM(15)	94.2 ± 0.94	96.3 ± 0.51	97.4 ± 0.33	97.9 ± 0.35	98.4 ± 0.27	98.7 ± 0.19
LM-SVM(9)	93.6 ± 0.92	95.7 ± 0.64	96.9 ± 0.45	97.6 ± 0.37	98.0 ± 0.34	98.4 ± 0.23
Wiener-SVM	91.8 ± 0.62	94.2 ± 0.77	95.6 ± 0.54	96.4 ± 0.58	96.9 ± 0.39	97.3 ± 0.34
3-D-DWT-SVM	83.5 ± 1.05	88.2 ± 0.81	90.7 ± 0.58	92.4 ± 0.47	93.6 ± 0.54	94.4 ± 0.45
SVM	78.3 ± 0.95	82.1 ± 0.77	84.2 ± 0.78	85.8 ± 0.61	86.7 ± 0.57	87.5 ± 0.51

TABLE V
OVERALL CLASSIFICATION ACCURACY (±STD) FOR THE UNIVERSITY OF PAVIA DATASET OVER 20 TRIALS AS A FUNCTION OF TRAINING SAMPLES

| Method	Number of training samples					
	405	608	810	1012	1215	1418
MH-LDA-MLE	95.5 ± 0.63	96.9 ± 0.61	97.5 ± 0.35	97.8 ± 0.49	98.1 ± 0.25	98.3 ± 0.36
LM-LDA-MLE(9)	90.8 ± 0.75	93.0 ± 0.62	93.7 ± 0.43	94.3 ± 0.36	94.6 ± 0.41	94.8 ± 0.33
LM-LDA-MLE(7)	91.0 ± 0.58	93.1 ± 0.47	93.8 ± 0.34	94.3 ± 0.41	94.6 ± 0.43	94.9 ± 0.38
Wiener-LDA-MLE	88.5 ± 1.05	91.3 ± 0.65	92.2 ± 0.57	92.7 ± 0.61	93.0 ± 0.47	93.2 ± 0.56
LDA-MLE	76.1 ± 1.13	80.6 ± 0.94	82.9 ± 0.73	84.4 ± 0.77	85.4 ± 0.66	86.2 ± 0.42
MH-SVM	**96.7** ± 0.52	**97.5** ± 0.31	**98.0** ± 0.29	**98.4** ± 0.28	**98.6** ± 0.25	**98.8** ± 0.22
LM-SVM(9)	93.3 ± 0.78	94.8 ± 0.52	95.6 ± 0.48	96.4 ± 0.29	96.8 ± 0.34	97.3 ± 0.23
LM-SVM(7)	93.7 ± 0.87	95.1 ± 0.54	95.8 ± 0.40	96.4 ± 0.28	96.8 ± 0.24	97.2 ± 0.27
Wiener-SVM	94.1 ± 0.67	95.2 ± 0.45	95.6 ± 0.56	96.1 ± 0.43	96.4 ± 0.36	96.7 ± 0.28
3-D-DWT-SVM	93.0 ± 0.71	94.4 ± 0.73	95.4 ± 0.49	96.0 ± 0.46	96.4 ± 0.52	96.8 ± 0.39
SVM	87.9 ± 0.91	89.0 ± 0.74	89.7 ± 0.76	90.1 ± 0.53	90.5 ± 0.63	90.8 ± 0.44
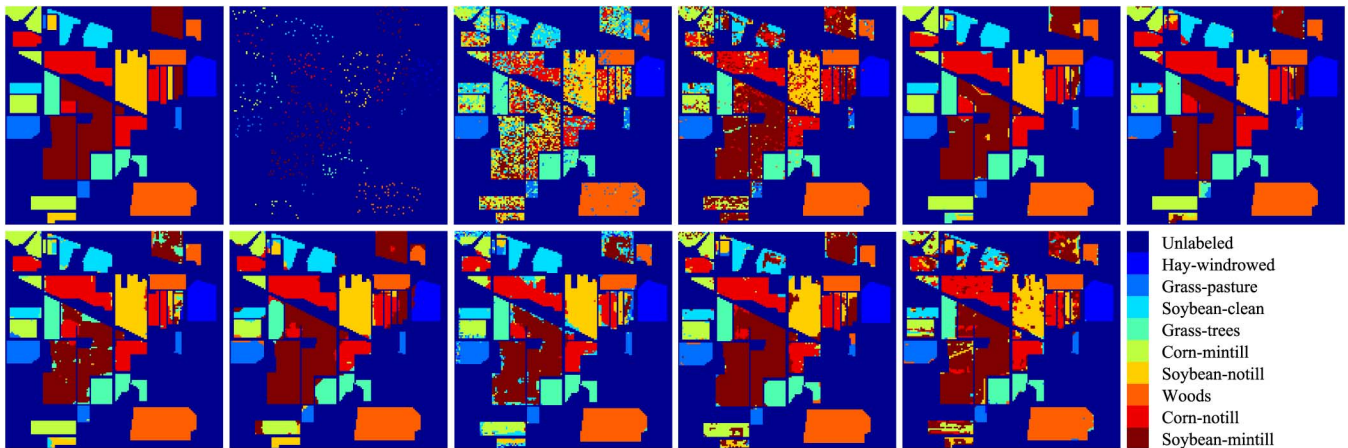
Fig. 10. Thematic maps resulting from classification using 462 training samples for the Indian Pines dataset using 9 classes. Overall classification accuracies are reported in parentheses. Top row, left to right: ground truth, training samples, LDA-MLE (60.6%), SVM (78.6%), MH-LDA-MLE (94.7%), and MH-SVM (94.5%). Bottom row, left to right: LM-LDA-MLE(15) (92.8%), LM-SVM(15) (94.2%), Wiener-LDA-MLE (87.5%), Wiener-SVM (92.0%), and 3-D-DWT-SVM (83.4%) class labels.

using two classical classifiers, e.g., MLE and SVM. Therefore, the corresponding methods are denoted as LM-LDA-MLE and LM-SVM. To achieve the best classification performance, we chose a spatial window of sizes 15×15 and 7×7 for the Indian Pines and the University of Pavia datasets, respectively, based on the results of testing a series of window sizes. The window size used for the Indian Pines dataset is larger than that used for the University of Pavia dataset because there exist larger homogeneous regions in the Indian Pines dataset. Since LM filtering is a special case of MH prediction, we repeat the experiments using the same window size as chosen for MH prediction, i.e., 9×9,

for LM filtering and report the classification performance as well. To distinguish the different window sizes used for LM-filtering methods, the window size is noted in the results, e.g., LM-LDA-MLE(15) and LM-LDA-MLE(9).

We additionally report the classification performance gains when the HSI datasets are preprocessed with band-by-band Wiener filtering. To obtain the optimal classification performance, the spatial window size used for Wiener filtering the Indian Pines and the University of Pavia datasets are 10×10 and 6×6, respectively. The Wiener-filtering classification methods are denoted by Wiener-LDA-MLE and Wiener-SVM.

TABLE VI
OVERALL CLASSIFICATION ACCURACY (±STD) FOR THE INDIAN PINES DATASET (9 CLASSES) OVER 20 TRIALS AS A FUNCTION OF SNR

Method	SNR (dB)					
	29.9	20.4	15.9	13.0	10.8	7.6
MH-LDA-MLE	**97.7** ± 0.43	97.5 ± 0.47	97.4 ± 0.48	97.3 ± 0.47	97.6 ± 0.38	97.3 ± 0.49
LM-LDA-MLE(15)	96.7 ± 0.41	96.3 ± 0.44	96.3 ± 0.61	96.4 ± 0.54	96.3 ± 0.37	96.1 ± 0.47
LM-LDA-MLE(9)	95.6 ± 0.52	94.7 ± 0.43	94.3 ± 0.46	94.1 ± 0.53	93.4 ± 0.56	92.9 ± 0.69
Wiener-LDA-MLE	92.2 ± 0.57	91.6 ± 0.53	90.9 ± 0.65	90.8 ± 0.55	89.7 ± 0.53	87.3 ± 0.58
LDA-MLE	63.6 ± 0.81	54.3 ± 0.84	50.4 ± 0.67	47.6 ± 0.75	44.9 ± 0.70	41.4 ± 0.91
MH-SVM	97.5 ± 0.38	**98.1** ± 0.30	**98.3** ± 0.32	**98.4** ± 0.34	**98.5** ± 0.33	**98.5** ± 0.36
LM-SVM(15)	97.6 ± 0.35	98.0 ± 0.32	98.1 ± 0.36	98.3 ± 0.28	98.3 ± 0.33	98.2 ± 0.37
LM-SVM(9)	97.3 ± 0.47	97.5 ± 0.51	97.5 ± 0.47	97.6 ± 0.48	97.6 ± 0.38	97.5 ± 0.42
Wiener-SVM	95.6 ± 0.51	95.7 ± 0.43	95.6 ± 0.53	95.9 ± 0.57	95.8 ± 0.58	95.4 ± 0.46
3-D-DWT-SVM	89.5 ± 0.56	85.3 ± 0.53	82.6 ± 0.61	79.4 ± 0.74	76.5 ± 0.67	71.1 ± 0.62
SVM	78.6 ± 0.51	70.1 ± 0.45	64.1 ± 0.67	60.0 ± 0.65	57.6 ± 0.46	55.3 ± 0.37

TABLE VII
OVERALL CLASSIFICATION ACCURACY (±STD) FOR THE UNIVERSITY OF PAVIA DATASET OVER 20 TRIALS AS A FUNCTION OF SNR

Method	SNR (dB)					
	27.3	18.5	13.4	9.0	5.4	2.5
MH-LDA-MLE	97.0 ± 0.45	96.7 ± 0.50	96.2 ± 0.49	95.6 ± 0.51	95.4 ± 0.53	94.9 ± 0.45
LM-LDA-MLE(9)	92.9 ± 0.54	91.7 ± 0.48	90.5 ± 0.57	89.8 ± 0.68	89.2 ± 0.73	88.3 ± 0.80
LM-LDA-MLE(7)	93.1 ± 0.48	91.5 ± 0.44	90.6 ± 0.56	89.5 ± 0.59	88.5 ± 0.81	87.0 ± 0.52
Wiener-LDA-MLE	91.2 ± 0.57	89.9 ± 0.74	89.0 ± 0.59	87.3 ± 0.87	86.1 ± 0.93	84.4 ± 0.86
LDA-MLE	81.1 ± 0.85	77.3 ± 0.76	73.9 ± 0.92	70.7 ± 0.98	68.6 ± 0.91	66.5 ± 0.74
MH-SVM	**98.0** ± 0.30	**98.1** ± 0.33	**98.3** ± 0.25	**98.2** ± 0.27	**98.0** ± 0.41	**97.9** ± 0.28
LM-SVM(9)	95.6 ± 0.43	95.6 ± 0.47	95.2 ± 0.41	94.4 ± 0.43	94.1 ± 0.45	93.4 ± 0.54
LM-SVM(7)	95.8 ± 0.40	95.5 ± 0.38	95.1 ± 0.43	94.5 ± 0.41	93.9 ± 0.47	92.8 ± 0.45
Wiener-SVM	95.6 ± 0.53	95.4 ± 0.54	95.1 ± 0.31	94.4 ± 0.36	93.3 ± 0.46	91.9 ± 0.48
3-D-DWT-SVM	95.4 ± 0.47	95.0 ± 0.49	94.2 ± 0.52	92.7 ± 0.58	90.7 ± 0.46	88.7 ± 0.62
SVM	89.3 ± 0.65	87.0 ± 0.67	84.0 ± 0.84	81.3 ± 0.59	78.0 ± 0.56	75.6 ± 0.42

In [25], 3-D DWT coefficients are extracted as spectral–spatial features for classification. We follow their experimental setup and use an 8×8 window for the windowed 3-D DWT. SVM is employed as the classifier. The resulting algorithm is denoted by 3-D-DWT-SVM. LDA-MLE and SVM using spectral information alone serve as the baseline classification algorithms. We refer to the MH versions of LDA-MLE and SVM as MH-LDA-MLE and MH-SVM, respectively. Note that the LIBSVM toolbox[2] with an RBF kernel is used for SVM and the optimal parameters for SVM are chosen as those that maximize the training accuracy by means of fivefold cross-validation. We study the classification accuracy as a function of training samples over a range from a small number to a reasonably high number.

The classification accuracy of the tested methods are studied over a range of training samples: 462 to 1616 for the Indian Pines dataset using 9 classes and 405 to 1418 for the University of Pavia dataset. Tables IV and V show that the proposed preprocessing algorithm significantly improves the overall classification accuracy at all training sample sizes compared with the accuracy of classifying on the original datasets without MH prediction, i.e., the baseline algorithms LDA-MLE and SVM. For the case of classification on the Indian Pines dataset using the MLE classifier, there is more than a 30% improvement in overall classification accuracy under a small training sample size (462 training samples). For both the Indian Pines and University of Pavia datasets, MH-LDA-MLE and MH-SVM consistently yield higher performance compared with the other methods. We note that for the Indian Pines dataset, MH-SVM and LM-SVM have
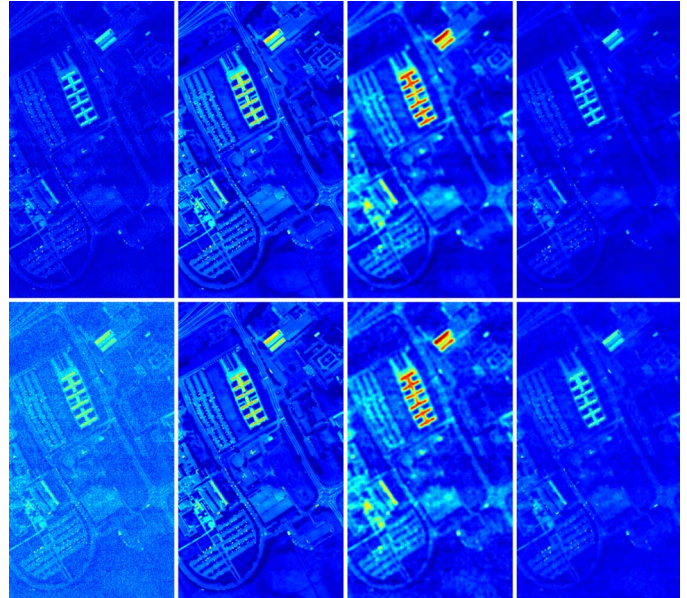

Fig. 11. Single spectral-band images (band 1) from the University of Pavia dataset preprocessed by various methods. Top row (noise-free case): original dataset, MH prediction, LM filtering, and Wiener filtering. Bottom row (noisy case, SNR = 2.5 dB): original dataset with additive noise, MH prediction, LM filtering, and Wiener filtering.

similar performance across training sample sizes. The reason might be the fact that applying equal weights to the neighboring pixels within a spatial window is good for a dataset consisting of large homogeneous regions, e.g., the Indian Pines dataset. In other words, pixels inside the window are most likely to belong to the same class as the pixel of interest. Hence, nonuniform
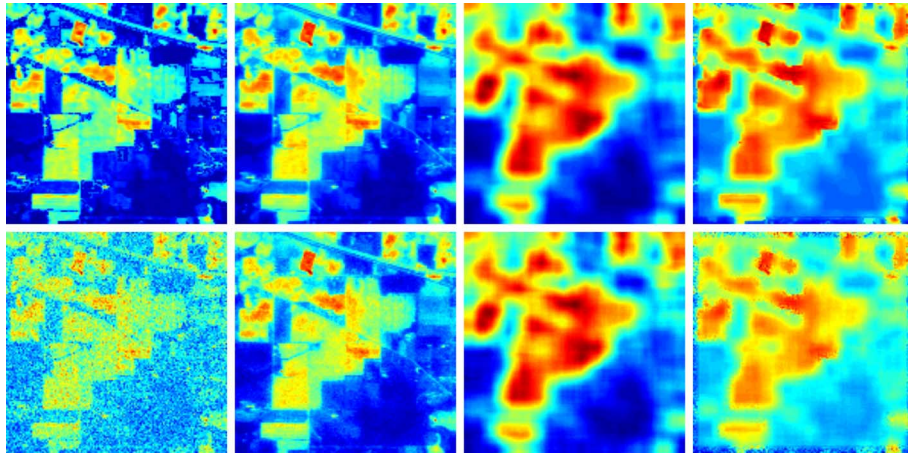
Fig. 12. Single spectral-band images (band 26) from the Indian Pines dataset preprocessed by various methods. Top row (noise-free case): original dataset, MH prediction, LM filtering, and Wiener filtering. Bottom row (noisy case, $\mathrm{SNR} = 7.6$ dB): original dataset with additive noise, MH prediction, LM filtering, and Wiener filtering.
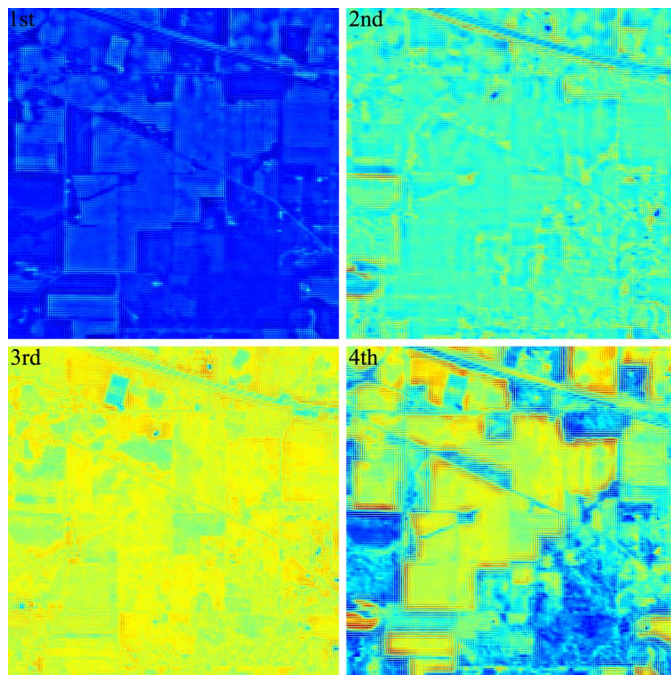

Fig. 13. Visual representation of four weight maps for the corresponding four spectral-band partitions of the Indian Pines dataset.
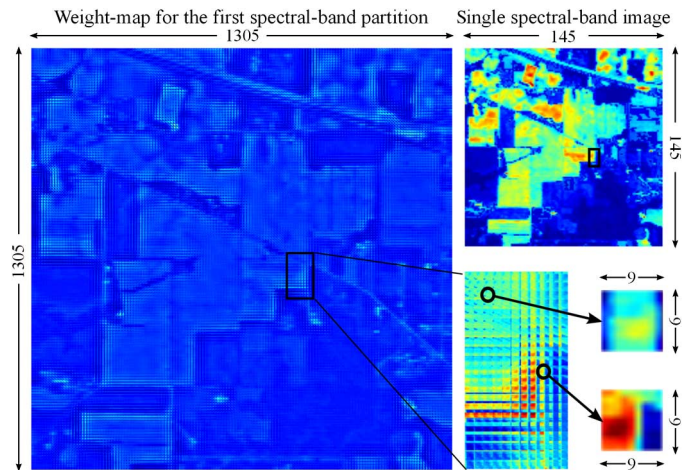

Fig. 14. Visual representation of the weight map for the first spectral-band partition of the Indian Pines dataset and a single spectral-band image (band 25) from the first spectral-band partition.

TABLE VIII
SNR After Preprocessing of the Indian Pines Dataset for Varying Levels of Noise Corruption

Method	Noise level (SNR)						
	$+\infty$	29.9	20.4	15.9	13.0	10.8	7.6
MH	8.8	8.8	8.7	8.7	8.7	8.6	8.5
LM	17.5	17.5	17.5	17.5	17.4	17.3	17.2
Wiener	19.4	19.4	19.0	18.7	18.3	18.0	17.3

TABLE IX
SNR After Preprocessing of the University of Pavia Dataset for Varying Levels of Noise Corruption

Method	Noise level (SNR)						
	$+\infty$	27.3	18.5	13.4	9.0	5.4	2.5
MH	10.4	10.4	10.4	10.3	10.2	9.9	9.5
LM	7.8	7.8	7.8	7.7	7.6	7.4	7.1
Wiener	10.4	10.4	10.3	10.0	9.6	9.0	8.2

weighting for the pixels in a window does not improve the filtering when compared with uniform weighting. Fig. 10 provides a visual inspection of the classification maps generated using the whole HSI scene for the Indian Pines dataset (145×145, including unlabeled pixels). As shown in Fig. 10, our proposed techniques show the best spatial homogeneity of the tested approaches. This homogeneity is observable within almost every labeled area.

We also conducted an experiment to estimate the noise robustness of our proposed algorithm. Since additive Gaussian white noise is often used to model electronic noise introduced during the hyperspectral data-acquisition process, only additive Gaussian white noise is considered in our experiments ([38], [40], e.g., do similarly). We assume that noise is independent of

signal information and add varying amounts of white Gaussian noise to the original HSI datasets. The signal-to-noise ratio (SNR) is used to measure the quality of noisy HSI dataset

TABLE X
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES DATASET (16 CLASSES)

Class	Samples		LDA-MLE	MH-LDA-MLE	LM-LDA-MLE(15)	LM-LDA-MLE(9)	Wiener-LDA-MLE	SVM	MH-SVM	LM-SVM(15)	LM-SVM(9)	Wiener-SVM	3-D-DWT-SVM
	Train	Test											
Alfalfa	36	10	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Corn-notill	114	1314	66.21	95.66	90.79	90.56	90.56	79.45	**97.11**	95.21	94.06	95.59	81.96
Corn-mintill	66	764	60.21	96.73	95.03	**98.17**	82.46	61.39	92.15	94.24	92.54	83.38	81.02
Corn	47	190	74.21	**100**	**100**	**100**	96.32	80.53	99.47	99.47	**100**	92.63	92.11
Grass-pasture	38	445	78.65	**95.96**	82.92	82.7	92.36	89.44	93.48	90.11	85.39	89.89	91.69
Grass-trees	58	672	95.83	**100**	99.4	99.7	99.55	98.96	99.55	96.58	94.2	98.66	97.62
Grass-pasture-mowed	22	6	66.67	83.33	83.33	**100**	**100**	83.33	**100**	**100**	**100**	**100**	**100**
Hay-windrowed	38	440	90	**100**	**100**	**100**	**100**	92.73	97.73	**100**	99.55	99.32	98.86
Oats	16	4	0	75	0	25	0	**100**	**100**	**100**	**100**	**100**	**100**
Soybean-notill	77	895	55.87	**98.1**	97.21	96.54	89.16	69.61	95.64	96.87	**98.1**	93.74	81.12
Soybean-mintill	196	2259	55.38	92.47	91.06	93.09	84.29	83.31	97.57	96.41	**97.83**	95.71	89.42
Soybean-clean	47	546	60.81	89.74	93.04	92.31	87.18	68.86	**94.14**	79.67	82.97	91.76	80.59
Wheat	41	164	96.95	**100**	**100**	96.95	**100**	98.17	99.39	98.17	**100**	99.39	97.56
Woods	101	1164	94.24	**99.91**	99.23	99.4	99.31	93.64	98.97	98.63	96.56	99.23	97.59
Building-grass-trees-drives	30	356	38.48	**100**	97.19	99.72	85.39	38.48	99.44	91.29	99.16	85.11	61.52
Stone-steel-towers	46	47	**100**	**100**	**100**	91.49	97.87	97.87	97.87	**100**	97.87	95.74	97.87
OA			68.97	96.27	94.3	94.97	90.42	80.54	**96.95**	95.2	95.22	94.4	87.72
AA			70.84	95.43	89.32	91.6	87.78	83.49	**97.66**	96.04	96.14	95.01	90.56

TABLE XI
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES DATASET (16 CLASSES) WITH ADDITIVE WHITE GAUSSIAN NOISE (NOISY DATASET SNR = 13 DB)

Class	Samples		LDA-MLE	MH-LDA-MLE	LM-LDA-MLE(15)	LM-LDA-MLE(9)	Wiener-LDA-MLE	SVM	MH-SVM	LM-SVM(15)	LM-SVM(9)	Wiener-SVM	3-D-DWT-SVM
	Train	Test											
Alfalfa	36	10	60	**100**	**100**	**100**	**100**	30	**100**	**100**	**100**	**100**	**100**
Corn-notill	114	1314	30.75	93.46	93.46	87.82	89.88	44.6	**99.39**	95.97	95.13	93.99	60.27
Corn-mintill	66	764	21.86	94.11	**97.51**	95.42	77.49	0.79	95.55	97.12	95.55	87.7	56.28
Corn	47	190	27.37	**100**	98.95	99.47	93.16	26.32	**100**	**100**	**100**	96.84	78.42
Grass-pasture	38	445	24.72	86.74	88.76	86.52	86.52	0.22	**94.16**	86.74	85.84	92.58	86.74
Grass-trees	58	672	68.6	**99.7**	98.51	97.92	99.11	90.63	99.4	97.77	94.94	98.66	94.35
Grass-pasture-mowed	22	6	0	**100**	**100**	83.33	83.33	66.67	83.33	**100**	**100**	**100**	**100**
Hay-windrowed	38	440	42.73	**100**	**100**	**100**	**100**	88.18	99.77	**100**	**100**	99.32	96.14
Oats	16	4	0	0	50	50	75	0	**100**	**100**	**100**	**100**	**100**
Soybean-notill	77	895	29.27	98.32	94.75	94.97	88.72	20.67	97.54	**99.33**	98.21	91.73	71.84
Soybean-mintill	196	2259	23.55	94.38	92.34	90.75	83.93	82.96	98.58	**99.11**	97.88	95.22	79.68
Soybean-clean	47	546	18.13	92.86	93.59	92.49	88.1	7.33	**95.05**	76.56	84.62	89.19	53.3
Wheat	41	164	77.44	**100**	97.56	97.56	**100**	87.8	99.39	**100**	**100**	**100**	95.12
Woods	101	1164	85.91	99.74	99.66	98.88	98.54	98.28	**99.91**	98.45	98.28	99.66	93.99
Building-grass-trees-drives	30	356	10.11	95.51	91.57	95.79	66.01	0	**99.16**	94.94	98.6	85.39	70.51
Stone-steel-towers	46	47	93.62	95.74	93.62	80.85	**100**	93.62	97.87	**100**	**100**	95.74	97.87
OA			37.6	95.71	94.98	93.43	88.65	54.74	**98.26**	96.34	96.03	94.36	76.7
AA			38.38	90.66	93.14	90.74	89.36	46.13	**97.45**	96.62	96.82	95.38	83.41

$\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1 \ldots \widehat{\mathbf{x}}_M]$ as was done in [31]. To quantify the noise level, we use a vector-based SNR measured in dB, i.e.,

$$\mathrm{SNR}(\mathbf{x}_m, \widehat{\mathbf{x}}_m) = 10 \log_{10} \frac{\mathrm{var}(\mathbf{x}_m)}{\mathrm{MSE}(\mathbf{x}_m, \widehat{\mathbf{x}}_m)} \qquad (12)$$

where $\mathbf{x}_m$ is an original pixel vector, $\widehat{\mathbf{x}}_m$ is the noisy pixel vector, $\mathrm{var}(\mathbf{x}_m)$ is the variance of the components of vector $\mathbf{x}_m$, and the mean squared error (MSE) is

$$\mathrm{MSE}(\mathbf{x}_m, \widehat{\mathbf{x}}_m) = \frac{1}{M} \|\mathbf{x}_m - \widehat{\mathbf{x}}_m\|_2^2. \qquad (13)$$

The average SNR is then the vector-based SNR of (13) averaged over all vectors of the dataset.

The two HSI data sets are corrupted by white Gaussian noise added to each spectral band over varying levels of noise corruption: from 7.6 to 29.9 dB for the Indian Pines data set and from 2.5 to 27.3 dB for the University of Pavia data set. In all the noisy-case experiments, we use 10% of the total samples from the HSI data sets for training and the rest for testing (923 training samples for the Indian Pines data set using 9 classes and 810 training samples for the University of Pavia data set). As Tables VI and VII indicate, the proposed MH-based classifiers (MH-LDA-MLE and MH-SVM) achieve higher classification accuracies than the other methods and are more robust to noise. As the noise level increases for both HSI data sets, we see that classification accuracy of MH-LDA-MLE decreases only slightly. Therefore,

our proposed MH-based classifiers exhibit strong noise robustness.

To better illustrate the noise-robustness feature of MH prediction, we compare single spectral-band images from the original datasets with the corresponding bands from the datasets preprocessed with MH prediction in Figs. 11 and 12. As can be seen in Fig. 11, a single spectral-band image (e.g., the first spectral band) from the original dataset appears noisy but looks very clear after preprocessing with MH prediction. For the case of the original dataset with additive white Gaussian noise, MH prediction is able to smooth out highpass noise effectively. We display the weight maps for each spectral-band partition of the Indian Pines dataset in Fig. 13 as well as detailed weights for one pixel vector in Fig. 14. Since the spatial dimensionality of the Indian Pines dataset is 145×145 and the spatial window size is 9×9 for MH prediction, the resulting weight map for each spectral-band partition is a matrix of size 1305×1305.

Furthermore, we list the SNR of the preprocessed noisy dataset using different methods in Tables VIII and IX (SNRs are used to indicate different noise levels that are added to the original dataset and $+\infty$ indicates the original dataset without additive noise). We observe that preprocessed datasets with higher SNR do not necessarily lead to higher classification accuracy. Therefore, we observe that more accurate denoising prior to classification does not necessarily yield better classification performance as the denoising algorithm not only removes the noise but also removes some of the discriminant features of the

TABLE XII
TRAINING AND VALIDATION EXECUTION TIME FOR THE INDIAN PINES DATASET
(9 CLASSES) USING 923 TRAINING AND 8311 TESTING SAMPLES

Algorithm	Time (s)
LDA-MLE	31
Wiener-LDA-MLE	33
LM-LDA-MLE	35
MH-LDA-MLE	252
SVM	3
Wiener-SVM	5
LM-SVM	7
MH-SVM	226
3-D-DWT-SVM	837

dataset. Our proposed MH prediction preprocessing is able to reduce the intraclass variations and enhance the class separability. It is agnostic to noise level and, in fact, performs better than techniques, such as Wiener filtering, which are specifically developed to remove noise prior to dimensionality reduction and classification.

In addition to using 9 classes in the Indian Pines dataset for classification, we conduct experiments using all 16 classes. These results are reported in Tables X and XI for both the original dataset as well as the original dataset with additive white Gaussian noise. Note that OA and AA in the tables indicate overall classification accuracy and average accuracy, respectively.

Finally, we compare the computational complexity of the classification methods. As an example, the execution times (in seconds) to train and validate with the Indian Pines dataset is shown in Table XII. As can be seen in Table XII, in terms of execution time, classification with MH-LDA-MLE and MH-SVM is, as expected, slower than LDA-MLE and SVM due to the iterated MH-prediction procedure, but both are faster than 3-D-DWT-SVM. The reason for 3-D-DWT-SVM running slow is that it not only has a time-consuming spatial feature-extraction process but also has the extracted spectral–spatial features with an increased dimensionality. However, our proposed method results in a feature set that has the dimensionality identical to the number of spectral bands. To speed up our method, techniques such as parallel computing can be employed.

V. CONCLUSION

In this paper, we proposed a preprocessing method that makes use of the spatial piecewise-continuous nature of HSI to improve the accuracy of supervised classification techniques. For each pixel vector in an HSI dataset, multiple predictions were drawn from the spatially surrounding pixel vectors. To take advantage of the fact that the spectral bands of a hyperspectral image are correlated, we proposed a hypothesis-generation procedure based on partitioning the spectral bands according to the correlation coefficients between bands. We formed MH prediction using a distance-weighted Tikhonov regularization to find the best linear combination of hypotheses. The MH prediction was then used to create a prediction that integrates spectral and spatial information. The processed dataset using MH prediction showed less intraclass variability and higher spatial smoothness as compared to the original dataset. The experimental results demonstrated that conventional classifiers such as LDA-MLE and SVM witnessed significantly improved classification

accuracy when applied to the datasets processed using MH prediction. In addition, MH prediction also provided robustness in the presence of additive white Gaussian noise.

REFERENCES

[1] J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[2] J. A. Richardson and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed., New York, NY, USA: Springer-Verlag, 2006.

[3] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[4] L. Samaniego, A. Bárdossy, and K. Schulz, "Supervised classification of recovery sensed image using a modified k-NN technique," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2112–2125, Jul. 2008.

[5] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.

[6] S. D. Zenzo, P. Bernstein, S. D. Degloria, and H. G. Kolsky, "Gaussian maximum likelihood and contextual classification algorithms for multicrop classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 25, no. 6, pp. 805–814, Nov. 1987.

[7] P. M. Atkinson and A. R. L. Tatnall, "Introduction neural networks in remote sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, Mar. 1997.

[8] D. Stathakis and A. Vasilakos, "Comparison of computational intelligence based classification techniques for remotely sensed optical image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2305–2318, Aug. 2006.

[9] A. Bárdossy and L. Samaniego, "Fuzzy rule-based classification of remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 2, pp. 362–374, Feb. 2002.

[10] D. G. Stavrakoudis, G. N. Galidaki, I. Z. Gitas, and J. B. Theocharis, "A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 130–148, Jan. 2012.

[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, Jun. 2005.

[12] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[13] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Customizing kernel functions for SVM-based hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 622–629, Apr. 2008.

[14] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 4, no. 1, pp. 65–74, Mar. 2011.

[15] Y. Bazi and F. Melgani, "Toward an optimal SVM classification system for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3374–3385, Nov. 2006.

[16] 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, Aug. 2004.

[17] G. M. Foody and A. Mathur, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.

[18] G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla, J. D. Martín-Guerrero, E. Soria-Olivas, L. Alonso-Chordá *et al.*, "Robust support vector method for hyperspectral data classification and knowledge discovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1530–1542, Jul. 2004.

[19] R. L. Kettig and D. A. Landgrebe, "Classification of multispectral image data by extraction and classification of homogeneous objects," *IEEE Trans. Geosci. Electron.*, vol. 14, no. 1, pp. 19–26, Jan. 1976.

[20] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel, "Segmentation and classification of hyperspectral data using watershed," in *Proc. Int. Geosci. Remote Sens. Symp.*, Boston, MA, USA, Jul. 2008, vol. 3, pp. 652–655.

[21] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Salt Lake City, UT, USA, May 2001, vol. 3, pp. 1793–1796.

[22] 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, Mar. 2005.

[23] 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. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[24] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[25] Z. Ye, S. Prasad, W. Li, J. E. Fowler, and M. He, "Classification based on 3-D DWT and decision fusion for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 173–177, Jan. 2014.

[26] G. J. Sullivan, "Multi-hypothesis motion compensation for low bit-rate video coding," in *Proc. Int. Conf. Acoust. Speech Signal Process.*, Minneapolis, MN, USA, Apr. 1993, vol. 5, pp. 437–440.

[27] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems*, Washington, DC, USA: V. H. Winston & Sons, 1977.

[28] E. W. Tramel and J. E. Fowler, "Video compressed sensing with multihypothesis," in *Proc. Data Compress. Conf.*, J. A. Storer and M. W. Marcellin, Eds., Snowbird, UT, USA, Mar. 2011, pp. 193–202.

[29] C. Chen, E. W. Tramel, and J. E. Fowler, "Compressed-sensing recovery of images and video using multihypothesis predictions," in *Proc. 45th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2011, pp. 1193–1198.

[30] C. Chen and J. E. Fowler, "Single-image super-resolution using multihypothesis prediction," in *Proc. 46th Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2012, pp. 608–612.

[31] C. Chen, W. Li, E. W. Tramel, and J. E. Fowler, "Reconstruction of hyperspectral imagery from random projections using multihypothesis prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 365–374, Jan. 2014.

[32] S. Tadjudin and D. A. Landgrebe, "Robust parameter estimation for mixture model," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 439–445, Jan. 2000.

[33] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., New York, NY, USA: Wiley, 2001.

[34] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classifications," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[35] Q. Du, W. Zhu, H. Yang, and J. E. Fowler, "Segmented principal component analysis for parallel compression of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 713–717, Oct. 2009.

[36] S. Prasad and L. M. Bruce, "Limitations of principal component analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, Oct. 2008.

[37] P. Gamba, "A collection of data for urban area characterization," in *Proc. Int. Geosci. Remote Sens. Symp.*, Anchorage, AK, USA, Sep. 2004, vol. 1, pp. 69–72.

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

[39] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487–1503, Nov. 1993.

[40] G. Martin, A. Plaza, and M. Zortea, "Noise-robust spatial preprocessing prior to endmember extraction from hyperspectral data," in *Proc. Int. Geosci. Remote Sens. Symp.*, Vancouver, BC, Canada, Jul. 2011, pp. 1287–1290.

Wei Li (S'11–M'13) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent 1 year as a Postdoctoral Researcher at the University of California, Davis, CA, USA. Currently, he is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.

Dr. Li is an Active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Eric W. Tramel (S'08–M'13) received the B.S. and Ph.D. degrees in computer engineering from Mississippi State University, Starkville, MS, USA, in 2007 and 2012, respectively.

In 2011, he served as a Research Intern at Canon USA, Inc. From 2009 to 2012, he also served as a Research Associate with the Geosystems Research Institute (GRI), Mississippi State University. Currently, he is a Postdoctoral Research Fellow with the Physico-Chimie Théorique Lab at École Supérieure de Physique et de Chimie Industrielles (ESPCI), Paris, France, investigating the intersection between statistical physics and signal processing. His research interests include compressed sensing, image and video coding, image and video multiview systems, data compression, and pattern recognition.

Dr. Tramel is an Active Reviewer for IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS.

Minshan Cui (S'12) received the B.E. degree in computer, electronics, and telecommunications from Yanbian University of Science and Technology, Yanbian, China, in 2008, and the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2011. He is currently working toward the Ph.D. degree in electrical and computer engineering from the University of Houston, Houston, TX, USA.

He is a Graduate Research Assistant with the Electrical and Computer Engineering Department, Geosensing Systems Engineering Research Center, University of Houston. His research interests include sparse representation theory, hyperspectral image analysis, and pattern recognition.

Chen Chen (S'10) received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2009 and the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2012. Currently, he is working toward the Ph.D. degree with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA.

His research interests include compressed sensing, signal and image processing, pattern recognition, and computer vision.

Dr. Chen is an Active Reviewer for the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Saurabh Prasad (S'05–M'09) received the B.S. degree in electrical engineering from Jamia Millia Islamia, New Delhi, India, in 2003, the M.S. degree in electrical engineering from Old Dominion University, Norfolk, VA, USA, in 2005, and the Ph.D. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2008.

Currently, he is an Assistant Professor with the Electrical and Computer Engineering Department, University of Houston (UH), Houston, TX, USA, and is also affiliated with the Geosensing Systems

Engineering Research Center and the National Science Foundation-funded National Center for Airborne Laser Mapping. He is the Principal Investigator/ Technical Lead on projects funded by the National Geospatial-Intelligence Agency, National Aeronautics and Space Administration, and Department of Homeland Security. He was the Lead Editor of the book entitled *Optical Remote Sensing: Advances in Signal Processing and Exploitation Techniques*, published in March 2011. His research interests include statistical pattern recognition, adaptive signal processing, and kernel methods for medical imaging, optical, and synthetic aperture radar remote sensing. In particular, his current research work involves the use of information fusion techniques for designing robust statistical pattern classification algorithms for hyperspectral remote sensing systems operating under low-signal-to-noise-ratio, mixed pixel, and small-training-sample-size conditions.

Dr. Prasad is an Active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the Elsevier *Pattern Recognition Letters*. He was awarded the Geosystems Research Institute Graduate Research Assistant of the Year award, in May 2007 and the Office-of-Research Outstanding Graduate Student Research Award, in April 2008 at Mississippi State University. In July 2008, he received the Best Student Paper Award at IEEE International Geoscience and Remote Sensing Symposium 2008 held in Boston, MA, USA. In October 2010, he received the State Pride Faculty Award from Mississippi State University for his academic and research contributions.
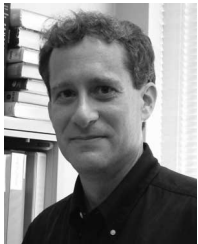
James E. Fowler (S'91–M'96–SM'02) received the B.S. degree in computer and information science engineering, the M.S. and the Ph.D. degrees in electrical engineering from Ohio State University, Columbus, OH, USA, in 1990, 1992, and 1996, respectively.

In 1995, he was an Intern Researcher at AT&T Labs in Holmdel, NJ, USA and in 1997, he held a National Science Foundation-sponsored postdoctoral assignment at the Université de Nice-Sophia Antipolis, France. In 2004, he was a Visiting Professor with the Département Traitement du Signal et des Images, École Nationale Supérieure des Télécommunications, Paris, France. Currently, he is Billie J. Ball Professor and Graduate Program Director of the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA; he is also a Researcher in the Geosystems Research Institute (GRI), Mississippi State University.

Dr. Fowler is an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING and the *EURASIP Journal on Image and Video Processing*; he formerly served as an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA and IEEE SIGNAL PROCESSING LETTERS. He is the Chair of the Image, Video, and Multidimensional Signal Processing Technical Committee of the IEEE Signal Processing Society and a Member of the Strategic Planning Committee of the IEEE Publication Services and Products Board. He is a General Co-Chair of the 2014 IEEE International Conference on Image Processing, Paris, France, as well as the Publicity Chair of the program committee for the Data Compression Conference.