Support Tensor Machines for Classification of Hyperspectral Remote Sensing Imagery

Xian Guo, Xin Huang, *Senior Member, IEEE*, Lefei Zhang, *Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—In recent years, the support vector machines (SVMs) have been very successful in remote sensing image classification, particularly when dealing with high-dimensional data and limited training samples. Nevertheless, the vector-based feature alignment of the SVM can lead to an information loss in representation of hyperspectral images, which intrinsically have a tensor-based data structure. In this paper, a new multiclass support tensor machine (STM) is specifically developed for hyperspectral image classification. Our newly proposed STM processes the hyperspectral image as a data cube and then identifies the information classes in tensor space. The multiclass STM is developed from a set of binary STM classifiers using the *one-against-one* parallel strategy. As a part of our tensor-based processing chain, a multilinear principal component analysis (MPCA) is used for preprocessing, in order to reduce the tensorial data redundancy and, at the same time, preserve the tensorial structure information in sparse and high-order subspaces. As a result, the contributions of this work are twofold: a new multiclass STM model for hyperspectral image classification is developed, and a tensorial image interpretation framework is constructed, which provides a system consisting of tensor-based feature representation, feature extraction, and classification. Experiments with four hyperspectral data sets, covering agricultural and urban areas, are conducted to validate the effectiveness of the proposed framework. Our experimental results show that the proposed STM and MPCA-STM can achieve better results than traditional SVM-based classifiers.**

***Index Terms*—Classification, dimensionality reduction, feature extraction, hyperspectral, support tensor machine (STM), support vector machine (SVM), tensor.**

Manuscript received December 15, 2014; revised June 7, 2015 and November 15, 2015; accepted December 31, 2015. Date of publication January 21, 2016; date of current version April 27, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 91338111, by the China National Science Fund for Excellent Young Scholars under Grant 41522110, by the Foundation for the Author of National Excellent Doctoral Dissertation of China under Grant 201348, and by the China Postdoctoral Science Foundation under Grant 2015M581162. *(Corresponding author: Xin Huang.)*

X. Guo is with the State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: guoxian@lreis.ac.cn).

X. Huang is with the School of Remote Sensing and Information Engineering and also with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: huang_whu@163.com).

L. Zhang is with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong, and also with the School of Computer, Wuhan University, Wuhan 430072, China (e-mail: zhanglefei@whu.edu.cn).

L. Zhang is with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: zlp62@whu.edu.cn).

A. Plaza is with the Department of Technology of Computers and Communications, University of Extremadura, 10003 Cáceres, Spain (e-mail: aplaza@unex.es).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik 101, Iceland (e-mail: benedikt@hi.is).

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

Digital Object Identifier 10.1109/TGRS.2016.2514404

I. Introduction

In THE past decades, hyperspectral imaging has attracted extensive research efforts since such data provide continuous bands throughout the electromagnetic spectrum, as well as exhibit the potential to precisely discriminate between different land cover types using a wealth of spectral information [1], [2]. Supervised hyperspectral image classification where pixels are labeled to one of the predefined classes, is among the most active research areas in hyperspectral analysis [3]. Many supervised classification methods have been developed for hyperspectral imagery [4]–[8], including random forest [9], [10], spectral unmixing [11], [12], subpixel mapping [13], neural networks [14], decision trees [15], and several others [16]–[18]. As a supervised machine learning technique, support vector machine (SVM) has achieved great success in various applications, including military problems [19], biological analysis [20], and pattern recognition [21]. It aims to acquire a classification hyperplane that maximizes the margin between the samples [22]. Taking advantage from its ability in dealing with high-dimensional data sets and limited training samples, SVM has been shown to be successful for hyperspectral image classification [23]–[25].

However, when the ratio between the number of spectral bands and the number of available training samples is unbalanced, hyperspectral image classification becomes an ill-posed problem that can reduce the interpretation accuracy [26], [27]. To alleviate the problem of information redundancy in hyperspectral spaces, SVM is often implemented together with the dimensionality reduction preprocessing [28]. Meanwhile, semisupervised SVMs have been also developed for solving the problem of limited training samples in high-dimensional feature spaces [29]. In addition, variations of the SVM-based algorithms include postprocessing based on decision rules [30], consideration of spatial information [31], [32], composite kernels [33], and active learning [34], [35].

It should be noted that the vector-based processing mode and feature alignment traditionally used by the SVM can lead to the loss of structural information for hyperspectral imagery, which has an intrinsic tensor-based data structure. Specifically, the spatial relationship between pixels cannot be effectively exploited in a vector-based approach for feature representation and interpretation. Inspired by the tensor learning methods

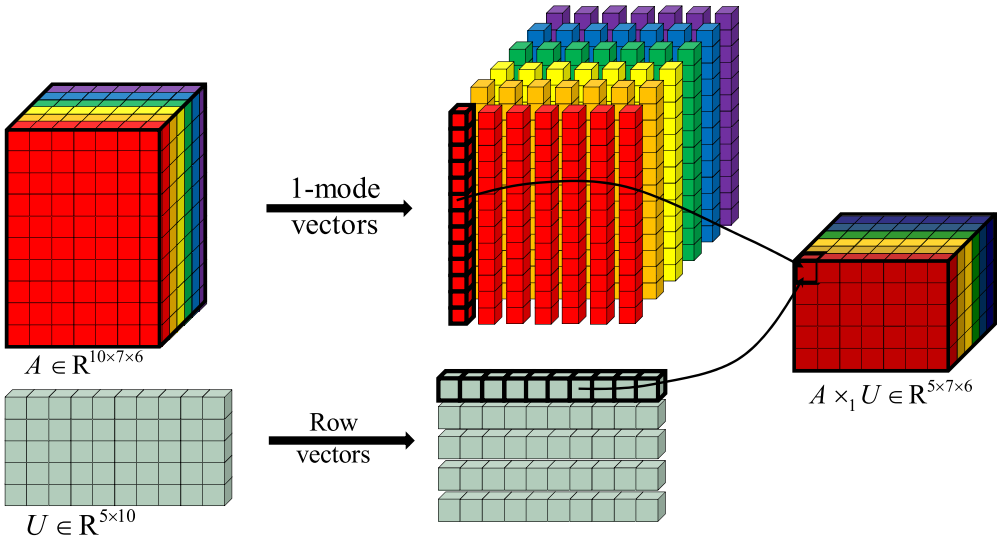


Fig. 1. Visual illustration of the mode-d tensor product operation.

developed in the area of signal processing, a support tensor machine (STM) [36] can be achieved by generalizing the SVM to its tensorial version based on multilinear algebra principles. The goal is to keep the robust performance of the classic SVM formulation while, at the same time, preserving the manifold data structure in sparse and high-dimensional feature spaces. STM has been applied to multiple applications, such as data retrieval [37], face recognition [38], and gait recognition [39]. However, in the existing literature, few studies have been reported for applying STM to remote sensing problems in general and hyperspectral image classification problems in particular. In this context, here, we develop a new multiclass STM and a tensor-based hyperspectral image processing chain to solve the multiclass classification for remote sensing imagery. Our systematic framework includes tensor-based feature extraction, dimensionality reduction, and classification. Specifically, in this paper, a multilinear principal component analysis (MPCA) [40] is used to reduce data redundancy due to the sparse and redundant high-dimensional tensor feature space for the hyperspectral data. The main contribution of this work is to introduce and adapt STM into hyperspectral remote sensing image classification. The effectiveness of the proposed method is validated by a set of experiments conducted on four hyperspectral data sets covering various land cover types. In addition, the parameters (e.g., window size, number of training samples, penalty coefficient, and feature dimensionality) are analyzed in detail in experiments.

The remainder of this paper is organized as follows. A brief review of multilinear algebra is given in Section II. The proposed STM and its multiclass version are described in Section III, followed by the tensor-based dimensionality reduction framework, which is described in Section IV. Experimental results and a comparison with state-of-the-art methods for hyperspectral image classification are provided in Section V. Section VI analyzes the impact of parameters in the developed processing chain. The last section concludes this paper with some remarks and hints at plausible future research lines.

II. Tensor and Multilinear Algebra

The proposed STM classification framework was based on multilinear algebra theory [41], [42] and processes the tensorial input data in a multidimensional manner. In the related literature, a multidimensional array is defined as a tensor, which is represented as $\mathsf{A} \in \mathrm{R}^{L_1\times L_2\times\cdots\times L_N}$. The order of the tensor A is N, and each order is called the ith mode. An arbitrary element of A is a scalar, which is denoted by $\mathbf{a}_{l_1,l_2,\ldots,l_i,\ldots,l_N}$, where $1 \le l_i \le L_i$, and $1 \le i \le N$, with l_i being the location of this element in the ith mode. According to this definition, a tensor is the higher order equivalent of a vector (one-order tensor) and a matrix (two-order tensor). In this context, a hyperspectral data cube can be regarded as a three-order tensor $\mathsf{A} \in \mathrm{R}^{L_1\times L_2\times L_3}$, in which orders 1 and 2 represent the spatial domain and order 3 represents the spectral domain. Some definitions related to basic tensor algebra [43] are given as follows.

Definition 2.1 (Mode-d Matricizing): Also known as mode-d matrix unfolding, the mode-d matricizing of an N-order tensor $\mathsf{A} \in \mathrm{R}^{L_1\times L_2\times\cdots\times L_N}$ is denoted by $\mathrm{mat}_d\mathsf{A} \in \mathrm{R}^{L_d\times(L_1L_2,\ldots,L_{d-1}L_{d+1},\ldots,L_N)}$, which represents an ensemble of vectors in the dth mode obtained by keeping index L_d fixed and varying the other indices.

Definition 2.2 (Mode-d Product): The mode-d product is an operation between a tensor $\mathsf{A} \in \mathrm{R}^{L_1\times L_2\times\cdots\times L_N}$ and a matrix $\mathbf{U} \in \mathrm{R}^{L_d\times L_{d'}}$, which is defined as

$$(\mathsf{A}\times_d \mathbf{U}^{\mathrm{T}})_{l_1\times l_2\times\cdots\times l_{d-1}\times l_{d'}\times l_{d+1}\times\cdots\times l_N} = \sum_{l_{d'}} \left(\mathsf{A}_{l_1\times l_2\times\cdots\times l_{d-1}\times l_d\times l_{d+1}\times\cdots\times l_N}\mathbf{U}_{l_{d'}\times l_d}\right) \quad (1)$$

where $l_1, l_2, \ldots, l_N$ represent the indices of each mode. It should be emphasized that the dth mode size of tensor A is equal to the second mode size of matrix $\mathbf{U}$. The mode-d product can be regarded as incorporating the information of the input tensor and matrix. Meanwhile, as the first mode size of $\mathbf{U}$ varies, dimensionality reduction and data filtering can be implemented via this operation. A visual illustration of mode-1 product between a three-order tensor and a matrix is presented in Fig. 1.

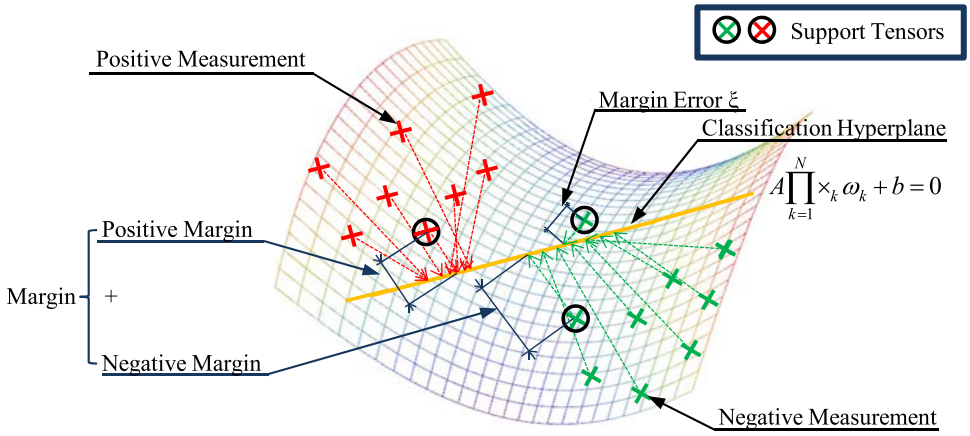


Fig. 2. Two-class problem in a high-dimensional input space solved with STMs.

Definition 2.3 (Outer Product): Operator "$\circ$" refers to outer product. $\mathbf{a}_1 \circ \cdots \circ \mathbf{a}_N = \mathsf{A}$ constructs a new tensor $\mathsf{A} \in \mathsf{R}^{L_1 \times L_2 \times \cdots \times L_N}$, with each element being the product of the corresponding vector elements [44], i.e.,

$$\mathbf{a}_{l_1,l_2,\ldots,l_N} = \mathbf{a}_1(l_1) \cdot \mathbf{a}_2(l_2), \ldots, \mathbf{a}_N(l_N) \tag{2}$$

where $\mathbf{a}_N(l_N)$ denotes the $l_N (1 \le l_N \le L_N)$ element from $\mathbf{a}_N$.

III. STM

A. Binary STM

As aforementioned, SVM employs optimization algorithms to find the optimal separation boundaries between classes [45]. These optimal boundaries are calculated by minimizing the errors among all possible boundaries estimated by the available samples for each class. The decision function of the classic SVM is $y(\mathbf{a}) = \text{sign}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{a} + b)$, where $\mathbf{a} \in \mathsf{R}^{L\times 1}$ is the unlabeled vector input, $y \in \{-1, +1\}$ refers to the predicted label, b is the bias term, and $\boldsymbol{\omega} \in \mathsf{R}^{L\times 1}$ is the projection matrix. The model parameters can be calculated by solving the following optimization problem:

$$\begin{bmatrix} \min\limits_{\omega,b,\xi} & f(\boldsymbol{\omega}, b, \xi) = \frac{1}{2}\|\boldsymbol{\omega}\|^2 + c\sum_{i=1}^{N}\xi_i \\ \text{s.t.} & y_i[\boldsymbol{\omega}^{\mathrm{T}}\mathbf{a}_i + b] \ge 1 - \xi_i,\ 1 \le i \le N \\ & \xi_i \ge 0 \end{bmatrix}. \tag{3}$$

As a tensorial generalization of SVM, STM aims at finding a series of hyperplanes that separate the tensorial data set into a set of predefined classes in a way that is consistent with the training samples. The multilinear decision function for STM is modeled to classify the unlabeled tensorial input $\mathsf{A} \in \mathsf{R}^{L_1\times L_2\times\cdots\times L_N}$ in the following high-order form:

$$y(\mathsf{A}) = \text{sgn}\left(\mathsf{A}\prod_{k=1}^{N}\times_k\boldsymbol{\omega}_k + b\right) \tag{4}$$

where $y \in \{-1, +1\}$, b is the bias term, and $\boldsymbol{\omega}_k \in \mathsf{R}^{L_k\times 1}$ is the weight calculated by

$$\begin{bmatrix} \min\limits_{\omega_k|_{k=1}^{N},b,\xi} & f\left(\boldsymbol{\omega}_k|_{k=1}^{N}, b, \xi\right) = \frac{1}{2}\|\boldsymbol{\omega}_1 \circ \cdots \circ \boldsymbol{\omega}_N\|^2 + c\sum\limits_{i=1}^{N}\xi_i \\ \text{s.t.} & y_i\left[\mathsf{A}_i\prod\limits_{k=1}^{N}\times_k\boldsymbol{\omega}_k + b\right] \ge 1 - \xi_i, 1 \le i \le M \\ & \xi_i \ge 0 \end{bmatrix} \tag{5}$$

where $\boldsymbol{\omega}_k|_{k=1}^{N}$ represents $\boldsymbol{\omega}_k$ $(k = 1, 2, \ldots, N)$ for short, and $\boldsymbol{\omega}_k$ corresponds to the weight matrix on the kth mode. In the equation, ξ is the slack variable vector used to deal with the linearly nonseparable cases, and c is a regularization coefficient for penalizing misclassification errors. M is the number of total training samples, and A_i $(1 \le i \le M)$ corresponds to the ith training tensor.

Based on the objective function defined in (5), the Lagrangian for STM is given by

$$L\left(\boldsymbol{\omega}_k|_{k=1}^{N}, b, \xi, \alpha\right) = f\left(\boldsymbol{\omega}_k|_{k=1}^{N}, b, \xi\right) - \sum_{i=1}^{N}\alpha_i\left(y_i c_i\left(\mathsf{A}_i\prod_{k=1}^{N}\times_k\omega_k + b\right) - \xi_i\right) \tag{6}$$

where $\alpha_i \ge 0\ (i = 1, 2, \ldots, M)$ are the Lagrangian multipliers. The derivations of $L(\boldsymbol{\omega}_k|_{k=1}^{N}, b, \xi, \alpha)$ with respect to variables $\boldsymbol{\omega}_k$ $(k = 1, 2, \ldots, N)$ and b should be set to 0 in the optimization process, which can be solved by

$$\partial_{\omega_j} f = \sum_{i=1}^{M}\alpha_i y_i\frac{dc_i}{dz}(\mathsf{A}_i\times_j\boldsymbol{\omega}_j) \tag{7}$$

$$\partial_b f = \sum_{i=1}^{M}\alpha_i y_i\frac{dc_i}{dz} \tag{8}$$

with $z = \mathsf{A}_i\prod_{k=1}^{N}\times_k\boldsymbol{\omega}_k + b$. However, the weight matrix $\boldsymbol{\omega}j$ in (7) depends on matrix $\boldsymbol{\omega}_i$ $(1 \le i \le N,\ i \ne j)$ and hence cannot be solved independently. An alternating projection optimization [46] is therefore adopted to obtain the local optimal solution. Specifically, the parameters associated with $\boldsymbol{\omega}_i$ are iteratively solved while keeping the other parameters fixed. After achieving convergence, with the parameters $\boldsymbol{\omega}$ and b resolved, the classification hyperplane obtained is given by

$$\mathsf{A}\prod_{k=1}^{N}\times_k\boldsymbol{\omega}_k + b = 0. \tag{9}$$

A graphical illustration of the scenario corresponding to a two-class problem in a high-dimensional input space solved with STMs is given in Fig. 2.

B. Multiclass STM

The STM is intrinsically a binary classifier. It can be adapted to multiclass classification problems, however, with a

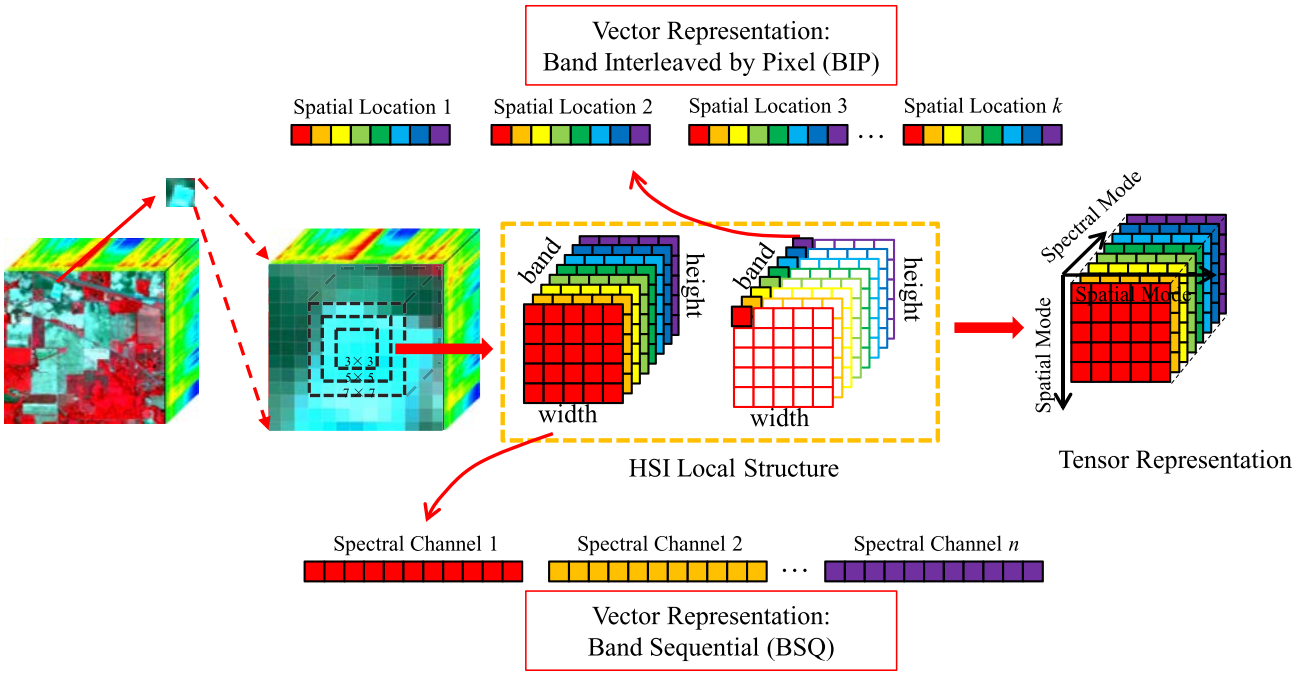


Fig. 3. Comparison between vector and tensor representations for a hyperspectral image (HSI).

one-against-one (*OAO*) parallel strategy [47]. Similarly, the multiclass STM can be extended from the binary STM by finding a series of hyperplanes which pairwisely maximize the margin between positive and negative samples from different classes, as well as predict the class label by majority voting. This way, the outputs of all binary classifiers are taken into account for constructing the multiclass STM.

Let us consider a multiclass problem defined by a set $\Omega = \{\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_S\}$ consisting of S information classes; the decision function $y_{i,j}(\mathsf{A})$ between each possible pair of classes $\mathbf{C}_i$ and $\mathbf{C}_j$ $(i \neq j)$ can be obtained. The OAO strategy constructs $S(S-1)/2$ binary STMs to model all the possible pairwise classifications. Subsequently, an unlabeled sample A is classified by applying each pairwise classifier and computing the score function which sums the number that this sample is assigned to that class label as follows:

$$\text{Score}_i = \sum_{i=1, i\neq j}^{S} y_{i,j}(\mathsf{A}). \tag{10}$$

The class with the highest score is then considered as the predicted label for the unlabeled sample A.

C. STM-Based Classification

To obtain the tensorial inputs corresponding to the labeled and unlabeled samples, a local neighborhood around the central pixel is extracted to represent a spatial–spectral tensor. For example, the tensor representation with a series of window sizes is graphically illustrated in Fig. 3 considering two types of vector representation: band interleaved by pixel (BIP) and band sequential (BSQ). After the tensor feature representation is obtained, the labeled training tensors are used for generating the classification hyperplanes, which can be used to interpret the unlabeled ones.

It should be noted that the conventional vector-based analysis approach needs to unfold the cubic features into vectors for feature description and classification since traditional classifiers can only process vector inputs. However, the proposed STM is able to directly model the spectral–spatial tensor, which is naturally aligned as a cube for image interpretation. Compared with the traditional SVM, this characteristic presents a main methodological advantage of STM.

IV. Redundancy Reduction for STM

Although the tensor-based processing approach described in the previous section can fully exploit the image spectral–spatial data structure and preserve the data manifold, the tensor-based data representation will necessarily lead to information redundancy due to the fact that a sparse hyperspectral feature space is derived from hundreds of continuous spectral channels. Moreover, due to the spatial similarity, the spectral signals of neighboring pixels are likely to be quite correlated, which is also a source of redundancy. The sparse and high-dimensional feature space exists in both spectral and spatial domains, particularly when the neighboring pixels (in a moving window) in a hyperspectral image are simultaneously taken into consideration for classification. Therefore, the direct use of tensor inputs will lead to a large amount of data redundancy, which poses problems to computational storage and efficiency, particularly for hyperspectral data with limited training samples. In this regard, reducing the redundancy is a crucial step for the proposed tensor processing and classification framework.

To deal with this problem, dimensionality reduction and feature extraction methods have been commonly applied. Here, we consider a tensor dimensionality reduction method and adapt it to the proposed tensor-based processing framework. Specifically, we use the MPCA [40], [48] as preprocessing for the STM classifier. This way, a novel tensor-based image interpretation framework is proposed by effectively

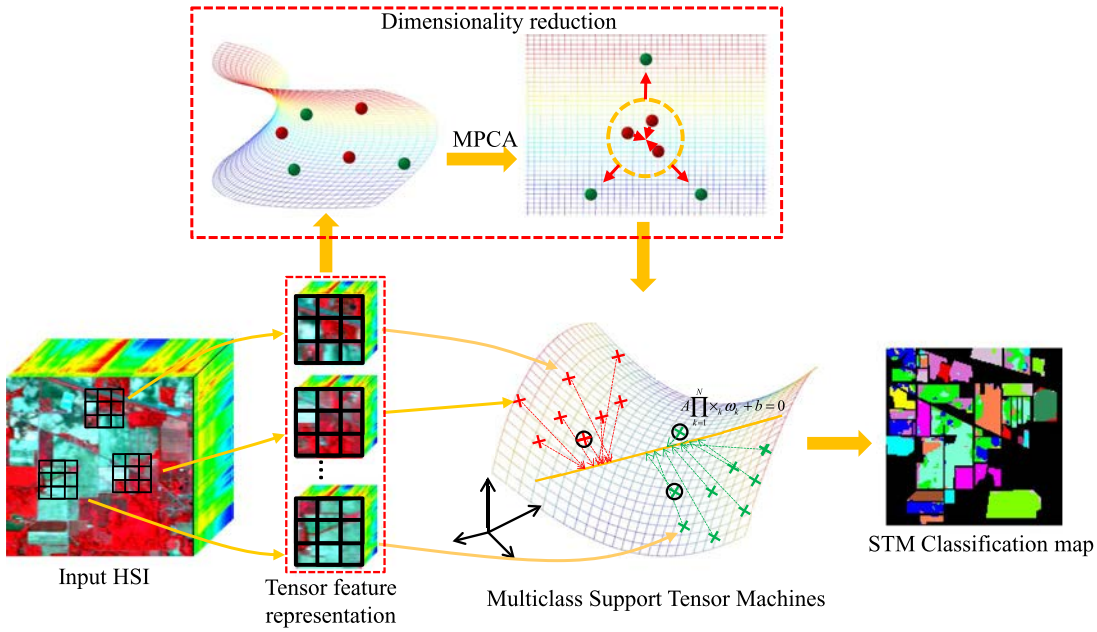


Fig. 4. Flowchart of the proposed tensor-based classification framework.

integrating tensor-based dimensionality reduction and classification. Please note that, in the proposed framework, MPCA can be replaced by other tensor-based dimensionality reduction methods, e.g., higher order singular value decomposition (HOSVD) [49]–[51].

A. MPCA

The basic idea of the MPCA is to extract the most significant signal components for all the tensor modes, while retaining as much as possible the data variation in the original data set. The potential of the MPCA-based tensor feature extraction has been shown in high-order computer vision and pattern recognition applications, such as face [52], handwriting [53], and gait recognition [54], [55]. However, few studies have been devoted to the use of MPCA for hyperspectral image feature extraction in the existing literature.

MPCA formulation involves the acquisition of projection matrices and low-dimensional tensors. Given a set of N-mode tensors $\{\mathsf{A}_m \in \mathsf{R}^{L_1 \times L_2 \times \cdots \times L_N}\}$, $m = 1, 2, \ldots, M$, we attempt to derive the projection matrices $\mathbf{U}_d \in \mathsf{R}^{L_d \times D_d}$, $d = 1, 2, \ldots, N$, with which the low-dimensional representation of input tensor $\mathsf{B} \in \mathsf{R}^{L_1 \times L_2 \times \cdots \times L_N}$ can be obtained by $\mathsf{B}' = \mathsf{B} \times_1 \mathbf{U}_1^{\mathrm{T}} \times \cdots \times_d \mathbf{U}_d^{\mathrm{T}} \times \cdots \times_N \mathbf{U}_N^{\mathrm{T}}$, with $\mathsf{B} \in \mathsf{R}^{D_1 \times D_2 \times \cdots \times D_N}$. However, it is an NP-hard problem to obtain all N projection matrices simultaneously. For this purpose, the alternating least square (ALS) algorithm can be used to solve the optimization of projection matrices approximately[56]. The procedure of MPCA can be summarized by the following steps:

1) Data centralization: $\tilde{\mathsf{A}}_m = \mathsf{A}_m - \overline{\mathsf{A}}$, $\overline{\mathsf{A}} = (1/M) \sum_{m=1}^{M} \mathsf{A}_m$.
2) At dth mode, unfold tensor $\tilde{\mathsf{A}}_m$ into matrix $\mathrm{mat}_d \tilde{\mathsf{A}}_m$ and perform eigendecomposition using $\phi_d = \sum_{m=1}^{M} (\mathrm{mat}_d \tilde{\mathsf{A}}_m) \cdot (\mathrm{mat}_d \tilde{\mathsf{A}}_m)^{\mathrm{T}}$. The projection matrix $\mathbf{U}_d$ is initialized as the eigenvectors associated with the most significant D_d eigenvalues.
3) Optimization of the projection matrices with ALS. The initial projection is formed by calculating $\tilde{\mathsf{P}}_m = \tilde{\mathsf{A}}_m \times_1 \mathbf{U}_1^{\mathrm{T}} \times \cdots \times_d \mathbf{U}_d^{\mathrm{T}} \times \cdots \times_N \mathbf{U}_N^{\mathrm{T}}$ and the scatter matrix $\psi_0 = \sum_{m=1}^{M} \|\tilde{\mathsf{P}}_m\|_{\mathrm{F}}^2$. Perform eigendecomposition for the covariance matrix $\phi_d = \sum_{m=1}^{M} (\mathrm{mat}_d \tilde{\mathsf{P}}_m) \cdot (\mathrm{mat}_d \tilde{\mathsf{P}}_m)^{\mathrm{T}}$ and define $\mathbf{U}_d$ as the eigenvectors corresponding to the most significant D_d eigenvalues, $d = 1, 2, \ldots, N$. In an iterative way, the projection $\tilde{\mathsf{P}}_m = \tilde{\mathsf{A}}_m \times_1 \mathbf{U}_1^{\mathrm{T}} \times \cdots \times_d \mathbf{U}_d^{\mathrm{T}} \times \cdots \times_N \mathbf{U}_N^{\mathrm{T}}$ and the corresponding scatter ψ_k can be updated. If the difference of the scatter matrix $\psi_k - \psi_{k-1}$ is smaller than a predefined threshold, the optimization of projection matrices is completed.
4) The low-dimensional tensor can be obtained by applying the mode-d product between the input tensor and the optimized projection matrix: $\mathsf{B}' = \mathsf{B} \times_1 \mathbf{U}_1^{\mathrm{T}} \times \cdots \times_d \mathbf{U}_d^{\mathrm{T}} \times \cdots \times_N \mathbf{U}_N^{\mathrm{T}}$.

B. MPCA-STM

The traditional vector-based image processing framework unfolds high-order features into vectors, which are subsequently fed into a vector-based classifier. This vector-based image processing strategy cannot sufficiently exploit the intrinsic tensor-based data structure. However, the tensor-based data representation is necessarily subject to the problem of data redundancy. As aforementioned, the redundancy is caused by the sparse hyperspectral feature space in both spectral (hundreds of continuous spectral channels) and spatial (similar pixels in the neighborhood) domains. In this context, in order to guarantee the efficiency and successful application of the STM to hyperspectral image classification, redundancy reduction is needed for preprocessing. In this paper, a tensor-based dimensionality reduction method (e.g., MPCA) is used for preprocessing of the STM classification. The whole procedure of the proposed tensor-based classification framework is therefore summarized in Fig. 4. Here, the tensor-based redundancy reduction refers to both the spectral and spatial dimensions. In the following,

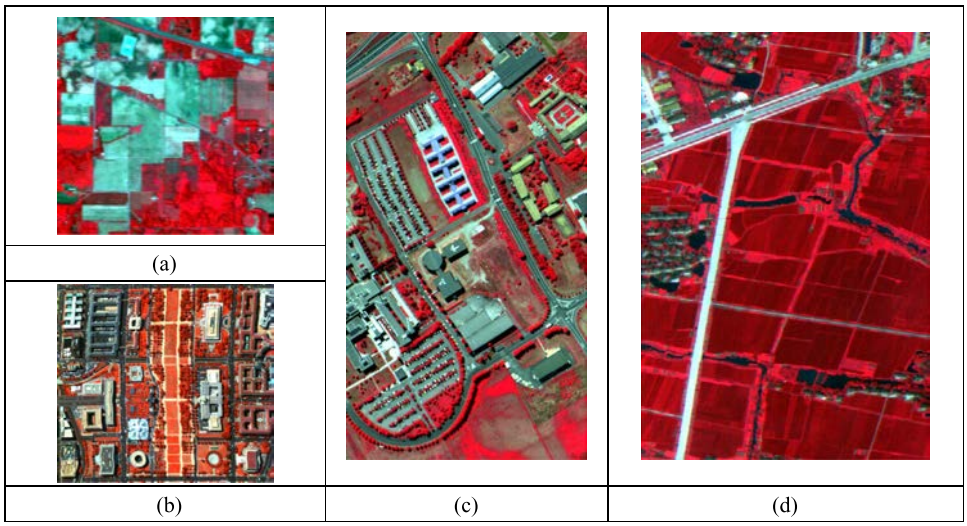


Fig. 5. False-color compositions of the four hyperspectral images used in our experiments. (a) AVIRIS Indian Pines. (b) HYDICE Washington DC Mall. (c) ROSIS-03 Pavia University. (d) PHI Xiaqiao.

$\{D_{\rm spa}, D_{\rm spa}, D_{\rm spe}\}$ are used to denote the reduced dimensions for a hyperspectral image, where $D_{\rm spa}$ and $D_{\rm spe}$ are the spatial and spectral dimensions after reduction, respectively. The same value of dimensionality (i.e., $D_{\rm spa}$) is defined for the x (horizontal) and y (vertical) directions in the spatial domain, since no prior information is available to determine which direction is the most relevant. In the following, we provide a set of experimental results to evaluate the classification accuracy of the framework presented in Fig. 4.

V. Experiments

The proposed classification framework has been validated using four real hyperspectral images acquired from the instruments: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), Reflective Optics System Imaging Spectrometer (ROSIS-03), Hyperspectral Digital Imagery Collection Experiment (HYDICE), and Pushbroom Hyperspectral Imager (PHI). The considered images are presented in Fig. 5.

To verify the effectiveness of the proposed tensor-based classification framework, the proposed STM and MPCA-STM are compared with SVM with various kernel functions (e.g., the linear kernel and the Radial Basis Function (RBF) kernel). Furthermore, in this study, PCA-SVM is also carried out to consider dimensionality reduction in the SVM classification for comparison with MPCA-STM. Specifically, analogous to MPCA-STM, the pixels in a window are stacked as a long feature vector, and PCA is then used to reduce the information redundancy in the vector. The reduced features are then input to an SVM for classification. The model parameters (i.e., penalty coefficient c and bandwidth γ for the RBF kernel) of the SVM are selected based on fivefold cross validation. The classification accuracy is assessed by using the overall accuracy (OA) and the κ coefficient based on a confusion matrix. Each classification scenario is repeated ten times with different randomly selected training sets, and the mean and standard deviation of the classification results are reported for the assessment of classification performance.

A. Experiments With the AVIRIS Indian Pines Data Set

The Indian Pines data were collected by the AVIRIS sensor in 1992 over an agricultural area in northern Indiana, USA. This hyperspectral image contains 16 classes with 145 $\times$ 145 pixels and 220 spectral bands between 0.4 and 2.5 μm. The classes and the corresponding numbers of training and test samples are listed in Table I. For each class, we randomly select a very limited number of 15 pixels for each class for training and use the remaining samples for accuracy assessment. The window size for the STM classification is set empirically to 9×9, and the reduced dimensionality in MPCA-STM is set as $\{1,1,40\}$ according to our cross-validation experiments. Please note that the first principal component in the spatial domain is sufficient to represent the local image cube, as will be discussed in Section VI-D.

The classification maps and accuracy assessment for the AVIRIS Indian Pines are shown in Fig. 6 and Table I. By analyzing the classification accuracies, it is observed that the SVM with the RBF kernel (RBF-SVM) outperforms the linear SVM (Lin-SVM) since the former can effectively discriminate samples that are not linearly separable. It can be seen that the proposed STM outperforms the SVM in terms of accuracies, as the SVM neglects the local correlation of neighboring pixels, which also leads to salt and pepper noise. Compared with SVM with linear and RBF kernel, the proposed multiclass STM increases the accuracy of OA by 11.8% and 5.0%, respectively. It can be also seen that the highest accuracy, i.e., OA $= 80.6\%$, is achieved by MPCA-STM, which significantly outperforms STM. This demonstrates that large amount of redundancy exists in the tensor-based hyperspectral image representation, and a spectral–spatial dimensionality reduction approach is necessary for the tensor-based image interpretation. As shown in Table I, the proposed STM and MPCA-STM methods outperform PCA-SVM in terms of improvements of OA (0.2% and 18.2%, respectively). The better performance of the STM methods can be attributed to the fact that they process an image as a cube, but the PCA-SVM realigns the cube into a

TABLE I
ACCURACY COMPARISON FOR THE AVIRIS INDIAN PINES DATA SET (MEAN AND STANDARD DEVIATIONS OF TEN DIFFERENT EXPERIMENTS ARE GIVEN)

class	Test	Lin-SVM	RBF-SVM	PCA-SVM	STM	MPCA-STM
Alfalfa	54	81.4±4.8	95.3±8.2	100±0	94.8±6.1	98.7±2.3
Corn-notill	1434	41.2±8.0	51.3±7.3	49.4±4.9	49.7±9.4	71.9±3.5
Corn-min	834	41.1±6.3	45.6±4.2	47.3±4.5	43.1±6.8	78.2±7.5
Corn	234	65.6±7.3	65.2±8.3	86.9±7.0	82.1±4.4	93.2±4.7
Grass/Pasture	497	65.7±4.4	75.4±6.9	87.2±9.1	83.9±2.7	89.7±4.0
Grass/Trees	747	71.1±4.4	82.4±8.6	80.3±2.9	89.2±3.4	93.5±3.0
Grass/pasture-mowed	26	92.3±6.6	94.6±4.3	100±0	100±0	100±0
Hay-windrowed	489	75.8±3.5	92.0±5.5	94.6±1.7	94.8±8.6	98.4±3.4
Oats	20	86.0±7.3	97.0±3.1	100±0	100±0	100±0
Soybeans-notill	968	48.0±8.2	55.9±7.6	54.0±5.4	61.3±11.5	75.9±6.1
Soybeans-min	2468	36.1±8.0	39.4±6.5	53.6±3.0	42.0±11.6	66.9±7.1
Soybeans-clean	614	38.8±7.5	42.8±7.6	59.2±3.6	44.8±10.4	76.9±9.7
Wheat	212	90.8±2.9	95.7±4.0	97.6±4.0	96.8±2.4	99.3±0.8
Woods	1294	71.2±6.2	78.6±7.8	61.5±5.8	87.8±3.8	94.4±3.1
Bldg-Grass-Tree-Drives	380	41.4±5.4	49.7±5.8	79.2±4.6	74.0±7.5	94.5±3.7
Stone-steel towers	95	93.1±4.3	92.4±4.5	100±0	99.0±0.5	99.6±0.6
OA(%)		50.8±1.9	57.6±2.1	62.4±1.2	62.6±2.6	80.6±1.9
κ		0.45±0.02	0.53±0.02	0.58±0.01	0.58±0.03	0.78±0.02

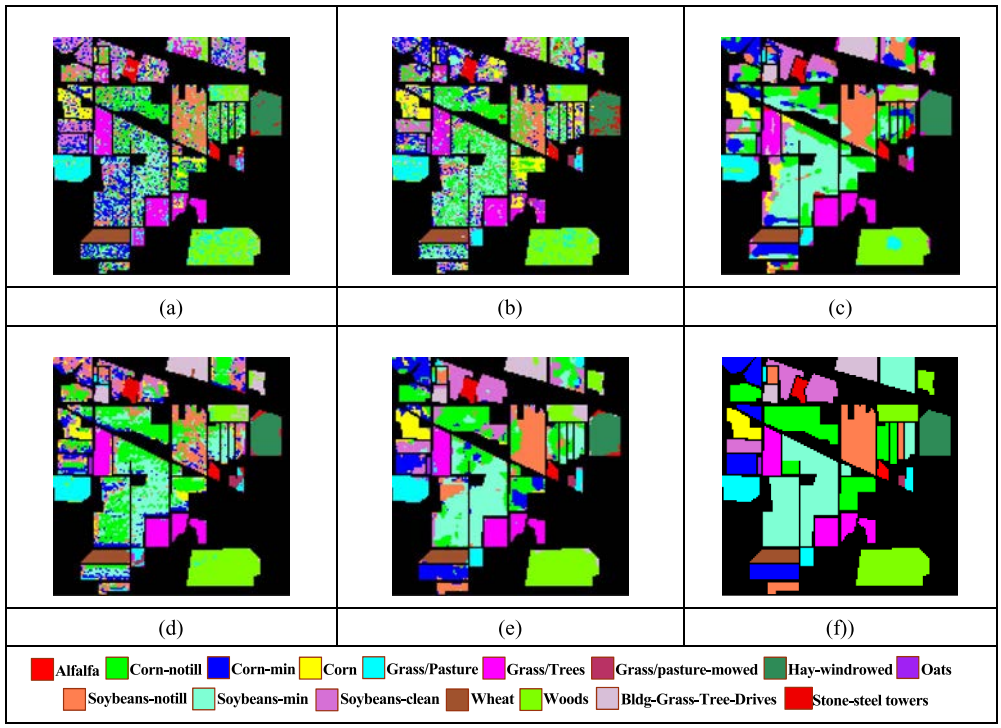


Fig. 6. Classification results obtained by different methods for the AVIRIS Indian Pines data. (a) SVM (linear kernel). (b) SVM (RBF kernel). (c) PCA-SVM. (d) STM. (e) MPCA-STM. (f) Ground-truth reference.

vector, which does not sufficiently exploit the spatial correlation between pixels.

B. ROSIS-03 Pavia University Data Set

The data set considered in this experiment is a high-resolution hyperspectral image (610 × 340 pixels) acquired by the ROSIS-03 sensor on the campus of the University of Pavia, Italy. A total of 115 spectral bands was acquired in the range 0.43–0.86 μm of the visible and infrared spectrum with a 1.3-m spatial resolution. The geospatial objects in this image refer to nine thematic classes, consisting of urban, soil, and vegetation features. In total, 42 776 labeled pixels are used for validation, as listed in Table II. The window size of the proposed tensor-based classifier and PCA-SVM is set empirically to 7 × 7, and a limited number of 20 pixels per class are randomly selected as

TABLE II
ACCURACY COMPARISON FOR THE ROSIS PAVIA UNIVERSITY DATA (MEAN AND STANDARD DEVIATIONS OF TEN DIFFERENT EXPERIMENTS ARE GIVEN)

Class	Test	Lin-SVM	RBF-SVM	PCA-SVM	STM	MPCA-STM
Asphalt	6631	69.7±3.2	73.6±2.8	73.4±6.1	72.6±4.7	88.5±1.9
Meadows	18649	76.1±4.6	76.3±5.9	91.6±8.9	88.0±2.5	90.1±3.1
Gravel	2099	80.4±3.3	77.5±3.3	70.2±6.3	65.6±4.2	80.6±3.8
Trees	3064	87.3±1.3	84.7±3.1	92.5±2.8	86.5±3.6	88.9±2.9
Metal_sheets	1345	99.3±0.4	99.0±0.3	100±0	98.2±0.9	99.0±0.7
Bare_soil	5029	74.0±4.9	73.7±4.3	39.1±9.6	75.2±9.3	93.7±2.3
Bitumen	1330	90.4±3.6	86.9±5.3	83.1±4.2	84.4±4.6	96.2±4.1
Bricks	3682	73.3±11.2	76.6±5.2	58.8±8.8	70.8±10.8	78.8±6.8
Shadows	947	99.8±0.2	99.9±0.1	98.7±1.1	96.7±1.0	98.4±0.5
OA(%)		77.3±2.9	77.8±1.9	79.0±3.2	79.5±1.4	89.4±0.5
κ		0.71±0.03	0.72±0.02	0.72±0.03	0.76±0.02	0.86±0.01

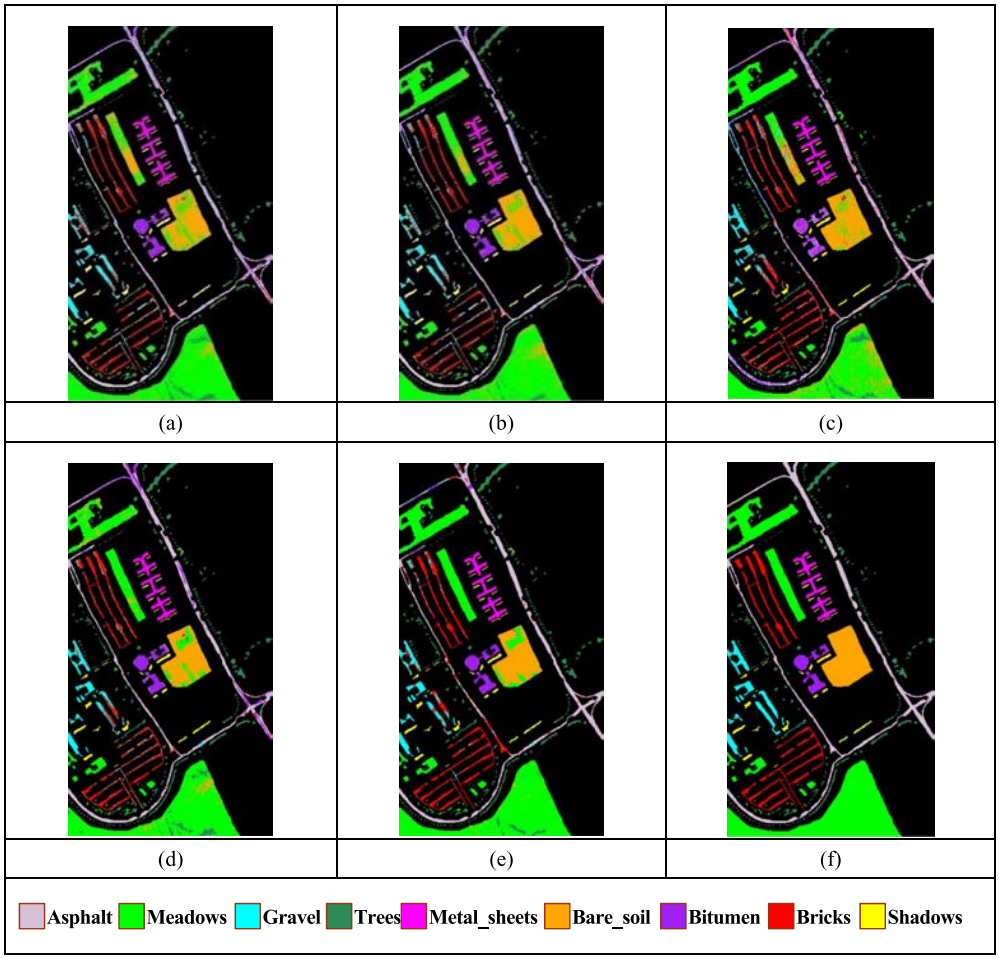


Fig. 7. Classification maps obtained by different methods for the ROSIS Pavia University data. (a) SVM (linear kernel). (b) SVM (RBF kernel). (c) PCA-SVM. (d) STM. (e) MPCA-STM. (f) Ground-truth reference.

training samples. The reduced dimensionality for MPCA-STM is chosen as $\{1,1,20\}$. This aspect is discussed in detail in Section VI-D.

The classification maps obtained by the considered algorithms are shown in Fig. 7, and the corresponding accuracy scores are presented in Table II. Misclassifications in this case lie in spectrally similar objects such as meadows–bare soil, gravel–bricks, which implies that spectral information alone is inadequate for classification purposes. The accuracies can be improved using the proposed STM, which actually considers the intrinsic spectral–spatial data structure. For instance, the classification accuracies of bare soil and meadows classes are increased by 2% and 12%, respectively, compared with RBF-SVM. Furthermore, when redundancy reduction is considered, MPCA-STM yields 12.1% and 11.6% improvements in terms of OA, compared with Lin-SVM and RBF-SVM,

TABLE III
ACCURACY COMPARISON FOR THE HYDICE WASHINGTON DC DATA (MEAN AND STANDARD DEVIATIONS OF TEN DIFFERENT EXPERIMENTS ARE GIVEN)

class	test	Lin-SVM	RBF-SVM	PCA-SVM	STM	MPCA-STM
Roof	8806	69.7±5.2	78.3±4.0	74.8±7.1	84.6±5.3	89.9±2.6
Road	1402	93.3±3.4	84.8±5.9	89.0±2.0	90.8±2.6	92.1±0.6
Trails	1446	96.1±3.0	92.4±5.1	85.4±5.2	91.7±3.6	95.6±1.9
Grass	1790	98.1±4.0	91.0±3.1	95.6±1.2	96.0±1.2	95.8±2.7
Shadow	2423	96.6±1.0	97.2±0.6	91.1±3.9	84.5±4.8	91.3±3.7
Tree	1194	97.0±0.9	94.4±1.9	96.1±0.9	95.6±2.0	98.0±0.4
OA(%)		81.2±2.4	84.4±6.8	82.0±3.8	87.6±2.7	91.7±1.6
κ		0.74±0.03	0.78±0.08	0.74±0.05	0.83±0.04	0.87±0.02

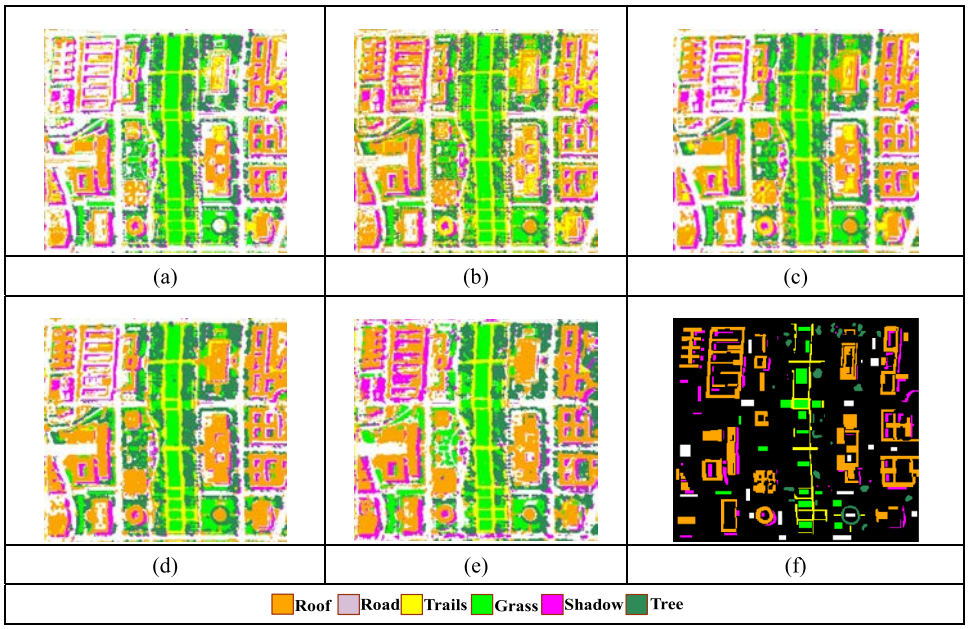


Fig. 8. Classification maps obtained by different methods for the HYDICE Washington DC data. (a) SVM (linear kernel). (b) SVM (RBF kernel). (c) PCA-SVM. (d) STM. (e) MPCA-STM. (f) Ground-truth reference.

respectively. Compared with PCA-SVM, STM and MPCA-STM achieve an accuracy increase of 0.5% and 10.4%, respectively, benefitting from the tensor-based data processing. In this experiment, it can be also observed that the proposed MPCA-STM not only exploits the implicit spectral–spatial information of hyperspectral data via tensor representation but also overcomes the issues caused by redundancy by incorporating STM with MPCA.

C. HYDICE Data Set

The HYDICE airborne hyperspectral data flight over the Washington DC Mall has been also considered in experiments. These data consist of 210 spectral bands from 0.4 to 2.4 μm, among which 19 bands were discarded due to water absorption. The test image contains 280 $\times$ 307 pixels with a 5-m spatial resolution. The reference data consist of 19 332 test samples, as listed in Table III, from which a limited number of 20 pixels per class are randomly selected as training samples. The classification results using Lin-SVM, RBF-SVM, STM, and MPCA-STM are given in Table III. The corresponding classification maps are shown in Fig. 8. In this experiment, the window size is chosen as 3 $\times$ 3 for PCA-SVM, STM, and MPCA-STM. The reduced dimensionality for MPCA-STM is set to $\{1,1,20\}$.

As shown in Table III, the proposed tensor-based classifiers can give substantially more accurate results than vector-based approaches. Compared with Lin-SVM and RBF-SVM, the improvements of OA achieved by STM are 6.4% and 3.2%, respectively, whereas the increments obtained by MPCA-STM are 10.6% and 7.3%, respectively. It can be also observed from Fig. 8 that Lin-SVM is not very effective for discrimination between bare soil, roads, and buildings due to their similar spectral reflectance. The classification of these similar classes is slightly improved with RBF-SVM. In the case of STM, where tensors are considered as the basic processing unit and, hence, the contextual spectral spatial relationship can be considered adequately, the classification accuracy for the roof is improved by 14.8% and 6.3%, respectively, compared with Lin-SVM and RBF-SVM. In this experiment, the proposed MPCA-STM achieves an increment of 5.3% compared with STM, by reducing the redundancy in the tensor-based spectral–spatial domain.

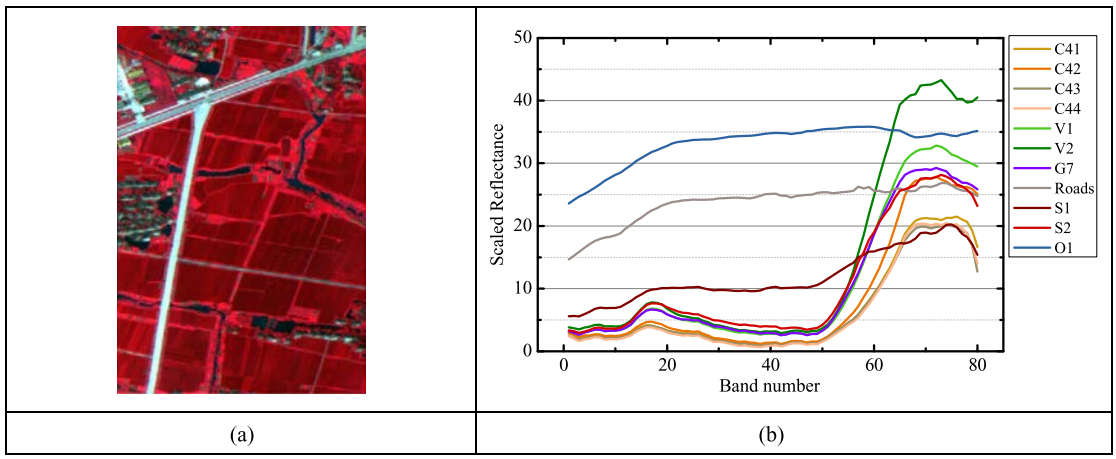


Fig. 9. (a) PHI Xiaqiao image RGB (70,40,10). (b) Reflectance curves for the classes: C, V, S, G, and O refer to corn, vegetation, soil, grass, and other classes (including lime and rocks).

TABLE IV
ACCURACY COMPARISON FOR THE PHI DATA SET (MEAN AND STANDARD DEVIATIONS OF TEN DIFFERENT EXPERIMENTS ARE GIVEN)

class	test	Lin-SVM	RBF-SVM	PCA-SVM	STM	MPCA-STM
C41	345	83.2±1.2	88.0±4.4	91.8±4.4	97.2±2.8	98.9±1.9
C42	150	96.4±3.6	99.2±0.9	99.0±1.3	99.5±1.0	100±0
C43	262	66.0±4.2	64.9±0.2	88.7±8.3	96.5±2.6	98.5±1.3
C44	155	86.7±5.9	83.7±5.9	98.4±4.2	99.3±1.1	99.5±0.9
V1	218	79.3±2.8	75.2±4.2	91.2±10.0	92.4±7.9	95.1±3.9
V2	253	85.4±10.1	94.4±9.9	90.5±5.7	90.1±4.1	97.4±1.9
G	229	67.0±5.5	62.7±12.1	86.4±6.7	90.1±3.8	99.8±0.3
Roads	221	90.0±0.5	90.0±0.3	91.0±7.1	91.1±5.7	87.1±5.5
S1	112	97.5±1.3	97.8±2.5	100±0	100±0	100±0
S2	280	88.6±3.8	93.8±7.4	85.6±6.3	90.5±9.1	98.1±1.0
O	147	74.5±5.6	84.2±2.1	99.6±1.0	83.5±9.0	98.4±2.7
OA(%)		81.9±2.2	83.9±1.6	91.6±1.5	93.4±2.2	97.4±2.8
κ		0.80±0.02	0.82±0.02	0.91±0.02	0.93±0.02	0.97±0.04

D. PHI Data Set

The PHI is a Chinese hyperspectral sensor based on push-broom imaging technique with a focal plain detector, which has a field of view of 21° and a spatial resolution of 1.5 mrad. The data set used here (346 × 512 pixels) was acquired by the PHI sensor over a mixed agricultural area of Xiaqiao in Jiangsu Province, China. The wavelength range for the data set is from 0.417 to 0.854 μm, and 80 spectral bands are available. Eleven land cover classes are labeled with 2372 pixels used as reference based on a field survey. The challenge for this data set is to discriminate multiple crop types, which have very similar spectral properties, as shown in Fig. 9. The figure presents the spectral reflectance curves, where C, V, S, G, and O refer to corn, vegetation, soil, grass, and other classes (including lime and rocks), respectively. The classes C41, C42, C43, and C44 represent different types of corn. In experiments, a limited number of 20 pixels are randomly selected from each class as training samples.

In this experiment, STM and MPCA-STM are implemented with a 9 × 9 window, and the reduced dimensionality for the MPCA is {1, 1, 30}. From the classification results in Table IV, it can be seen that (compared with Lin-SVM and RBF-SVM) the improvements of OA achieved by STM are 11.4% and 9.4%, respectively, when the local contextual information is included in the classification. Similarly, in this experiment, MPCA-STM further improves the classification accuracy of the original STM by removing the data redundancy in the sparse and hyperspectral tensor space, resulting in the highest accuracy $(\mathrm{OA} = 97.35\%)$.

VI. DISCUSSION

The experimental results in the previous section show that the proposed STM and MPCA-STM can achieve a satisfactory performance for hyperspectral remote sensing image classification. Moreover, they provide more accurate results than the traditional SVM by constructing a new tensor-based image feature representation and interpretation framework, which is particularly appropriate for the intrinsic tensor data structure of hyperspectral data.

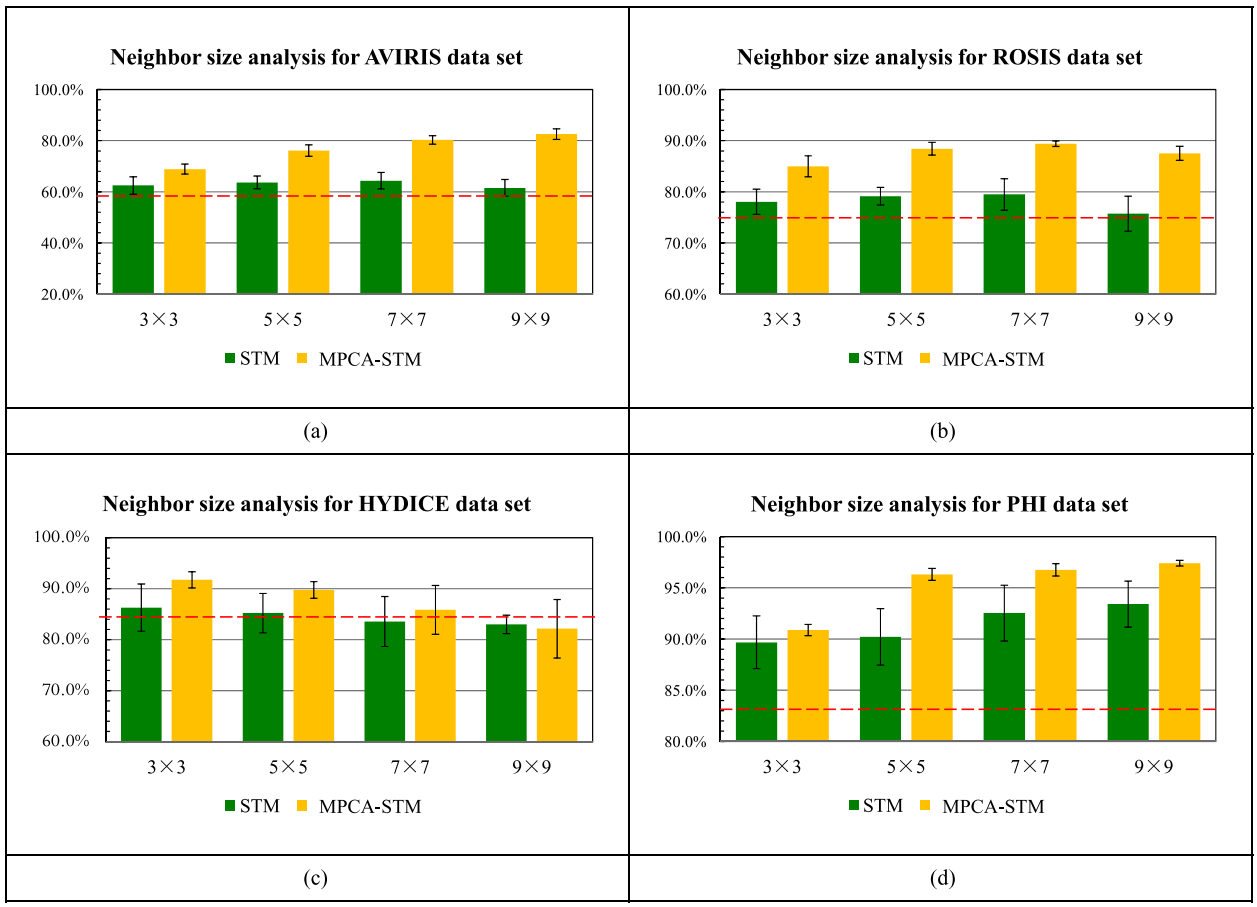


Fig. 10. Overall classification accuracy (in percentage) of STM and MPCA-STM with different window sizes for tensor representation for (a) AVIRIS, (b) ROSIS-03, (c) HYDICE, and (d) PHI data sets.

Here, the parameters for the proposed STM and MPCA-STM, namely, the window size, training samples, penalty coefficient, and tensor dimensionality, are discussed in detail in order to better understand the tensor-based image processing system.

A. *Window Size for the Tensorial Feature Representation*

Fig. 10 shows the classification accuracies of STM and MPCA-STM with different window sizes for the four considered experimental data sets. STM and MPCA-STM are implemented using window sizes varying from 3×3 to 9×9 by randomly choosing 20 pixels (15 pixels for AVIRIS data set) from each class for training. In the figure, the dash lines represent the accuracies obtained by RBF-SVM with the same training samples. The reduced dimensionality for MPCA is set to $\{1,1,40\}$ (AVIRIS), $\{1,1,20\}$ (ROSIS-03), $\{1,1,20\}$ (HYDICE), and $\{1,1,30\}$ (PHI), respectively.

From the perspective of image classification, small windows may not contain sufficient spatial information, whereas large windows can complicate the analysis with the possibility that multiple objects or classes are included in the window. In our experiments, it was found that the window size does not show a significant influence on the results obtained, demonstrating the robustness of the tensor image processing system. In addition, it can be seen that the tensor-based classifiers obtain a better performance than SVM for different window sizes. Meanwhile, it can be also seen that MPCA-STM significantly outperforms the original STM in nearly all the cases since redundancy reduction is very important for the tensor-based processing.

B. *Influence of the Number of Training Samples*

To investigate the impact of the number of training samples on the proposed tensor-based classification framework, different numbers of training samples are used in this experiment. Specifically, for each data set, training samples are randomly generated from the reference data using a number of training samples per class that ranges from 5 to 95 pixels and using the remaining ones for testing. As shown in Fig. 11, the OA values increase rapidly with the increment of the number of training samples, and then, the accuracy curves become stable when 20–40 samples per class are used. In addition, it can be observed that the proposed tensor-based classifiers outperform the classic SVM and contextual SVM (i.e., PCA-SVM) in all cases, particularly for small numbers of training samples. This reveals that the proposed tensor feature presentation has a potential for addressing the classification uncertainty with small numbers of training samples, since tensor-based classifiers yield much better results than vector-based methods when

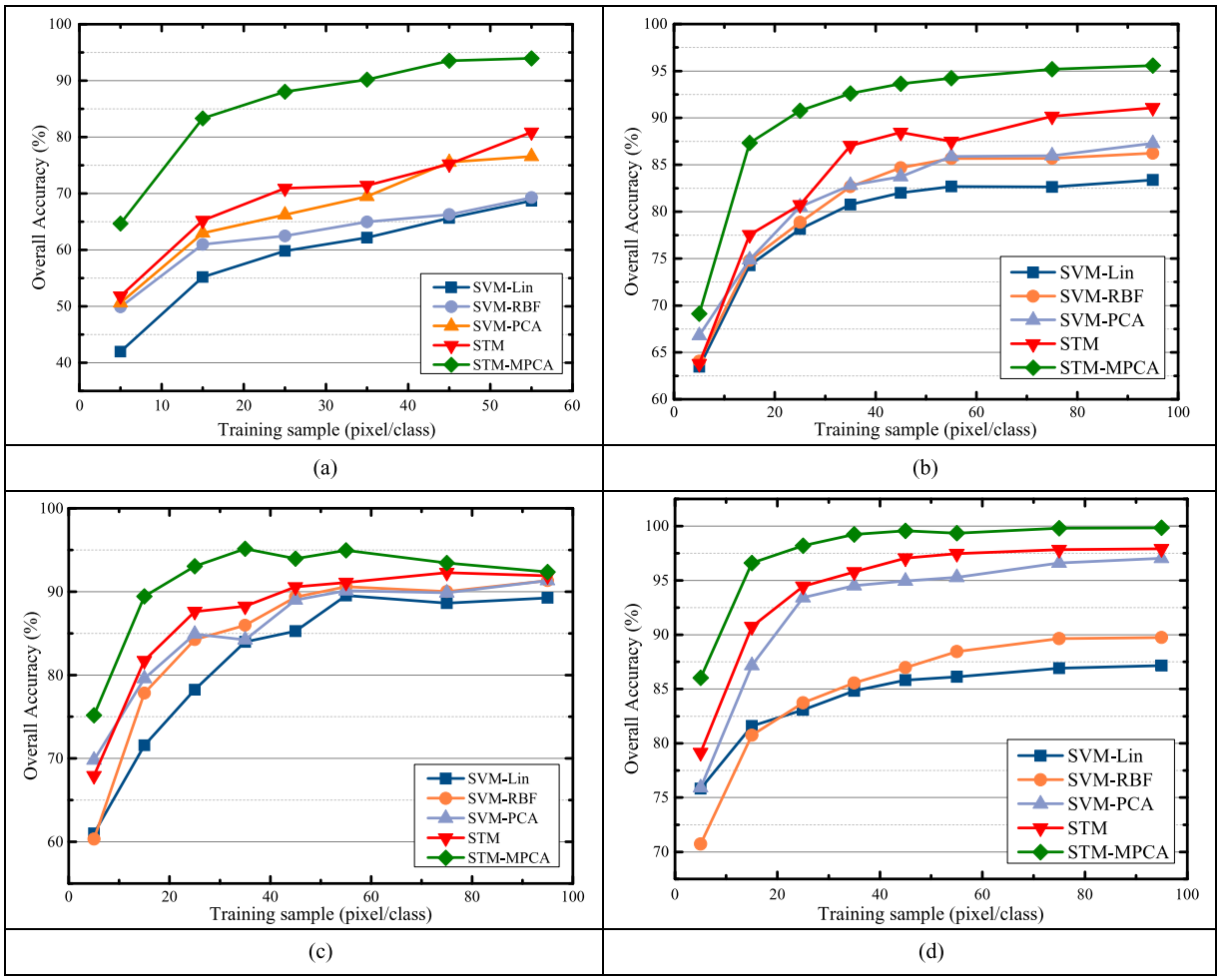


Fig. 11. Impact of numbers of training samples for different classifiers in (a) AVIRIS, (b) ROSIS-03, (c) HYDICE, and (d) PHI data sets.

5–10 samples/class are used. In this experiment, it can be also seen that MPCA-STM outperforms STM no matter how many training samples are used for training.

C. Penalty Coefficient of the STM

The penalty coefficient c determines the degree by means of which the classifier penalizes the negative samples in the training. It is an important parameter for STM since it controls the generalization of the classifier. Here, a series of experiments are conducted to investigate how the penalty coefficient c affects the performance of STM and MPCA-STM. Fig. 12 shows the obtained classification accuracies as a function of the coefficient $c \in [2^{-16}, 2^{16}]$. From the figure, it can be seen that the accuracy curves are parabolic. This phenomenon can be attributed to the fact that a small value means giving less penalty to misclassified samples, which can mix up margin support tensors and nonmargin support tensors (i.e., fall on the "wrong" side of this margin) in a problem that is not linearly separable. On the other hand, a large value can shrink the classification margin and result in performance degradation. The optimal value of the penalty parameter can be determined by cross validation.

D. Reduced Dimensionality for MPCA-STM

Here, the impact of using reduced spatial and spectral dimensions (i.e., D_{spa} and D_{spe}) for the proposed MPCA-STM is analyzed. The overall accuracies of MPCA-STM for different numbers of dimensions (in both spectral and spatial directions) are shown in Fig. 13, where the x- and y-axes refer to the spatial and spectral dimensions after reduction, respectively; and the z-axis denotes the OA (%). The highest accuracy achieved by the original STM with the optimal parameters is also indicated with a dot dashed line. Fig. 13 shows that reduced dimensionality is an important parameter, which significantly influences the performance of MPCA-STM. Specifically, we have found that the 1-D or 2-D principal components are adequate for the tensor feature representation in the spatial domain. In particular, when the first spatial principal component is considered, the performances of the MPCA-STM are much better than the original STM regardless of the spectral dimensionality. An exception is the HYDICE DC mall experiment, where the high classification accuracy (between 80% and 90%) corresponds to wide spatial dimension (between 1 and 5), but relatively narrow spatial dimension (between 1 and 50). This suggests that the spatial redundancy is relatively smaller in this study area.

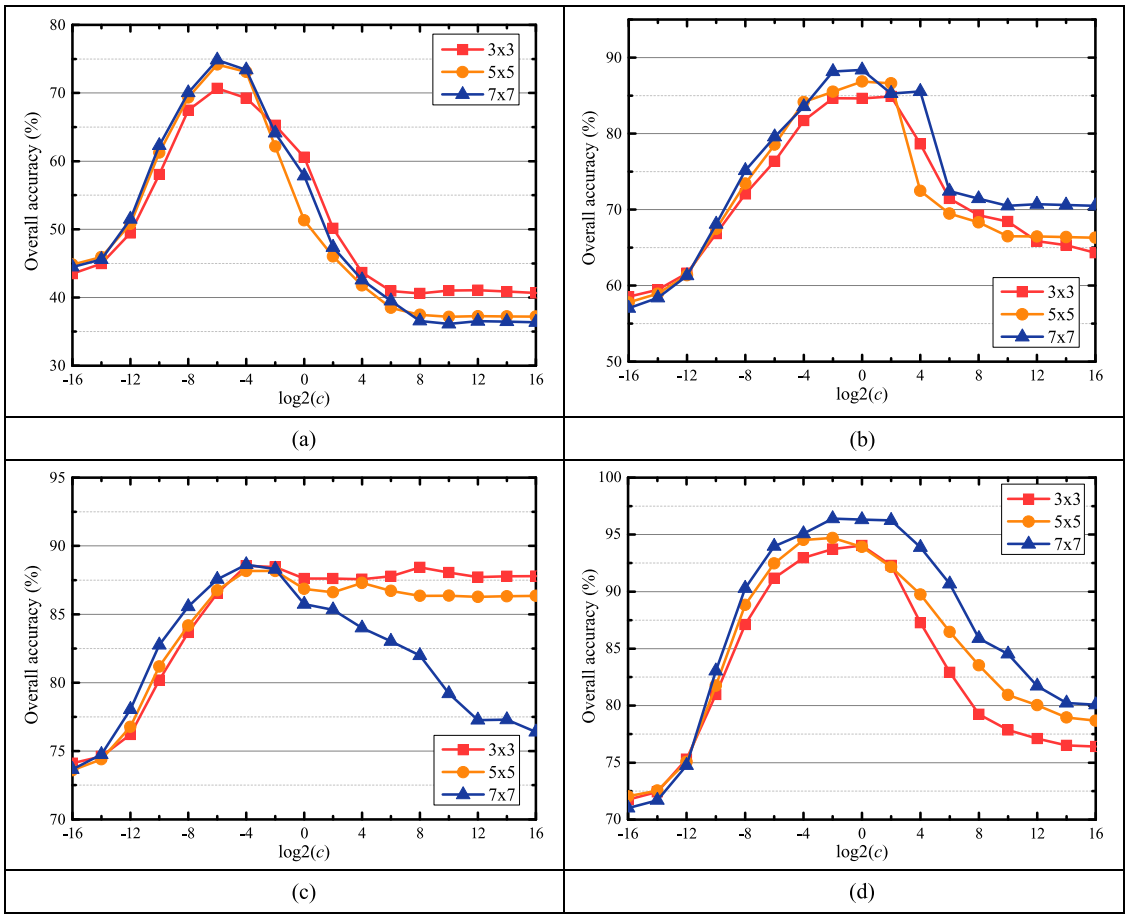


Fig. 12. Analysis of the sensitivity of the penalizing parameter c for the STM on four considered data sets. (a) AVIRIS. (b) ROSIS. (c) HYDICE. (d) PHI.

E. Computational Load

Here, the computational cost for different methods is compared in Table V. From the table, we can conclude that the moving-window-based classifiers are subject to more processing time than the traditional vector-based ones. It should be noted that PCA-SVM leads to large computational burden due to the pixel-by-pixel alignment of the local window. Additionally, the use of MPCA can remarkably alleviate the computational burden of STM.

VII. Conclusion and Future Lines

In this paper, we have presented a new tensor-based framework for classification of hyperspectral remote sensing imagery. Specifically, we have developed a new multiclass STM, which extends the traditional vector-based feature representation and classification strategy such as the SVM to a tensor-based version. In particular, tensor-based processing is naturally appropriate for hyperspectral remote sensing imagery, which has an intrinsic tensor data structure. Furthermore, our work reveals that information redundancy exists in the sparse and high-dimensional feature space for hyperspectral data. As a result, redundancy reduction becomes a crucial issue, particularly when the tensor data representation is considered, since redundancy can exist in both spectral and spatial directions. Accordingly, we introduced an MPCA, which is a tensor-based dimensionality reduction algorithm, and then constructed a novel MPCA-STM classifier. MPCA-STM takes advantage of tensor feature extraction and at the same time solves the problem of information redundancy.

In order to validate the effectiveness of the newly developed STM and MPCA-STM, a series of experiments were conducted using four representative hyperspectral data sets, covering various land cover types (e.g., rural, urban, and agriculture). In all the experiments, the proposed STM outperformed the SVM in terms of both quantitative accuracy scores and visual inspection. Moreover, the interpretation accuracy can be further improved when MPCA-STM is carried out, since data redundancy is reduced simultaneously in both the spectral and spatial domains. These results demonstrate that our proposed methods exhibit potential for hyperspectral classification.

Hyperspectral data play an important part in quantitative remote sensing, precision agriculture, detailed land cover mapping, and military applications. The tensor-based system presented in this paper represents a novel method for processing and understanding hyperspectral imagery, by fully considering its intrinsic tensor data structure. In the future, we will conduct further experiments to fully substantiate the potential of the proposed tensor-based system for replacing the traditional vector-based data representation, providing more accurate analysis

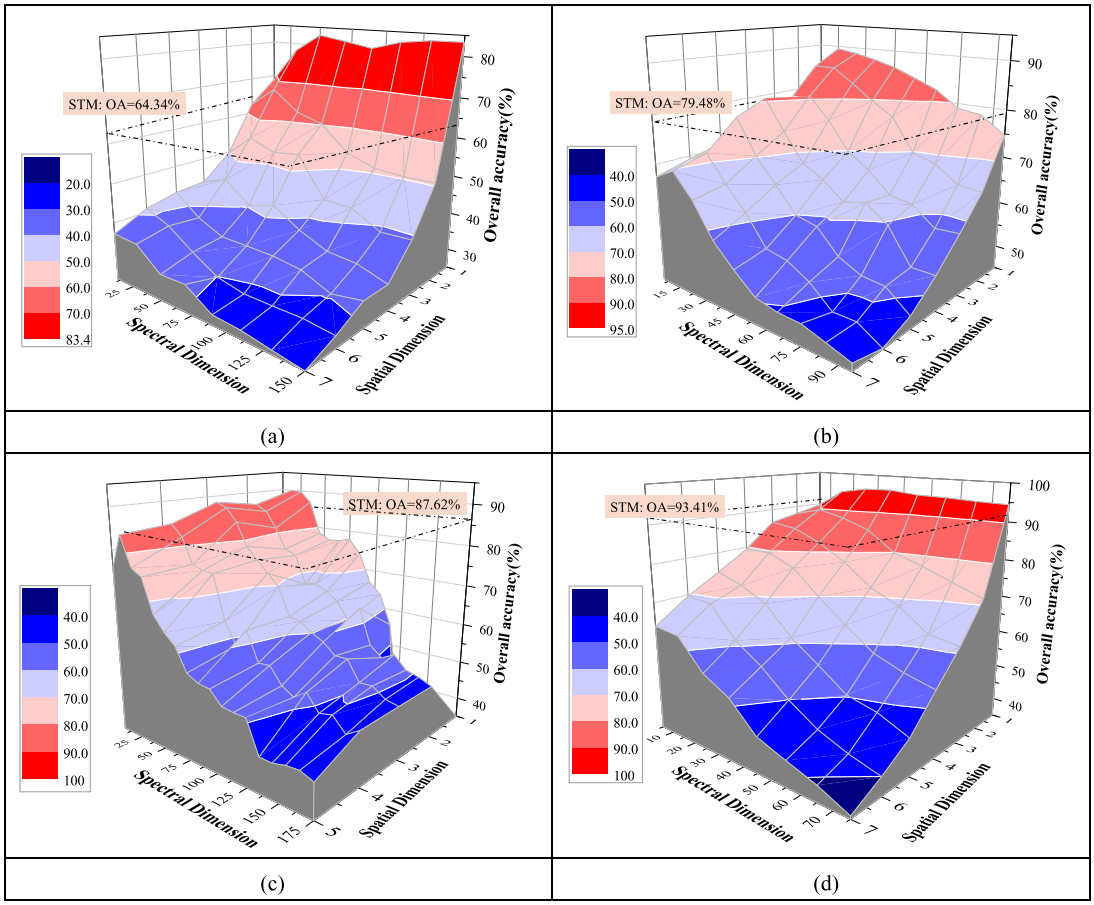


Fig. 13. Effect of using reduced spatial and spectral dimensionality for MPCA-STM. (a) AVIRIS Indian Pines. (b) ROSIS-03 Pavia University. (c) HYDICE Washington DC Mall. (d) PHI Xiaqiao.

TABLE V
COMPUTATIONAL TIME (IN SECONDS) FOR VARIOUS METHODS

Computational time(s)	SVM-Lin	SVM-RBF	PCA-SVM	STM	MPCA-STM
AVIRIS	3.10	7.20	43.42	350.52	45.93
ROSIS	1.38	3.37	150.52	143.27	43.79
HYDICE	0.63	3.31	9.28	28.67	13.55
PHI	1.95	2.13	104.78	58.12	26.91

for remote sensing applications such as classification, target detection, regression, and change detection. In addition, the spatial or structural features [57], [58] can be also represented in a tensor framework.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe (Purdue University, West Lafayette, IN, USA) for providing the free downloads of the AVIRIS data set over Indian Pines and Prof. P. Gamba (University of Pavia, Italy) for making the ROSIS-03 data set over Pavia, Italy, along with the training and test set, available to the community. The authors would also like to thank the anonymous reviewers for the insightful suggestions, which significantly improved the quality of this paper.

REFERENCES

[1] 6. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 110–122, Sep. 2009.

[2] L. G. Olmanson, P. L. Brezonik, and M. E. Bauer, "Airborne hyperspectral remote sensing to assess spatial distribution of water quality characteristics in large rivers: The Mississippi river and its tributaries in Minnesota," *Remote Sens. Environ.*, vol. 130, pp. 254–265, Mar. 2013.

[3] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis*, 4th ed. Berlin, Germany: Springer-Verlag, 2006.

[4] M. S. Moran, Y. Inoue, and E. M. Barnes, "Opportunities and limitations for image-based remote sensing in precision crop management," *Remote Sens. Environ.*, vol. 61, no. 3, pp. 319–346, Sep. 1997.

[5] J. Stuckens, P. R. Coppin, and M. E. Bauer, "Integrating contextual information with per-pixel classification for improved land cover classification," *Remote Sens. Environ.*, vol. 71, no. 3, pp. 282–296, Mar. 2000.

[6] F. Yuan, K. E. Sawaya, B. C. Loeffelholz, and M. E. Bauer, "Land cover classification and change analysis of the Twin Cities (Minnesota) metropolitan area by Multitemporal Landsat remote sensing," *Remote Sens. Environ.*, vol. 98, no. 2/3, pp. 317–328, Oct. 2005.

[7] R. A. Schowengerdt, *Remote Sensing: Models and Methods for Image Processing*, 3rd ed. Burlington, VT, USA: Academic, 2007.

[8] A. Brenning, "Benchmarking classifiers to optimally integrate terrain analysis and multispectral remote sensing in automatic rock glacier detection," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 239–247, Jan. 2009.

[9] J. Ham, C. Yangchi, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

[10] R. L. Lawrence, S. D. Wood, and R. L. Sheley, "Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest)," *Remote Sens. Environ.*, vol. 100, no. 3, pp. 356–362, Feb. 2006.

[11] J. B. Adams *et al.*, "Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon," *Remote Sens. Environ.*, vol. 52, no. 2, pp. 137–154, May 1995.

[12] U. Heiden, K. Segl, S. Roessner, and H. Kaufmann, "Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data," *Remote Sens. Environ.*, vol. 111, no. 4, pp. 537–552, Dec. 2007.

[13] D. Vikhamar and R. Solberg, "Subpixel mapping of snow cover in forests by optical remote sensing," *Remote Sens. Environ.*, vol. 84, no. 1, pp. 69–82, Jan. 2003.

[14] A. M. Filippi and J. R. Jensen, "Fuzzy learning vector quantization for hyperspectral coastal vegetation classification," *Remote Sens. Environ.*, vol. 100, no. 4, pp. 512–530, Feb. 2006.

[15] S. Delalieux *et al.*, "Heathland conservation status mapping through integration of hyperspectral mixture analysis and decision tree classifiers," *Remote Sens. Environ.*, vol. 126, pp. 222–231, Nov. 2012.

[16] 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.

[17] J. A. Benediktsson and I. Kanellopoulos, "Classification of multisource and hyperspectral data based on decision fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1367–1377, May 1999.

[18] 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, Oct. 2011.

[19] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[20] T. S. Furey *et al.*, "Support vector machine classification and validation of cancer tissue samples using microarray expression data," *Bioinformatics*, vol. 16, no. 10, pp. 906–914, Oct. 2000.

[21] A. K. Jain, R. P. W. Duin, and M. Jianchang, "Statistical pattern recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 1, pp. 4–37, Jan. 2000.

[22] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm.*, vol. 66, pp. 247–259, May 2011.

[23] G. M. Foody and A. Mathur, "Toward intelligent training of supervised image classifications: Directing training data acquisition for SVM classification," *Remote Sens. Environ.*, vol. 93, no. 1/2, pp. 107–117, Oct. 2004.

[24] 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.

[25] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

[26] A. Okujeni, S. van der Linden, L. Tits, B. Somers, and P. Hostert, "Support vector regression and synthetically mixed training data for quantifying urban land cover," *Remote Sens. Environ.*, vol. 137, pp. 184–197, Oct. 2013.

[27] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. New York, NY, USA: Wiley, 2003.

[28] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[29] 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, Nov. 2006.

[30] 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.

[31] 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.

[32] M. Khodadadzadeh *et al.*, "Spectral–spatial classification of hyperspectral data using local and global probabilities for mixed pixel characterization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6298–6314, Oct. 2014.

[33] 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.

[34] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, Sep. 2011.

[35] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

[36] D. Tao, X. Li, X. Wu, W. Hu, and S. J. Maybank, "Supervised tensor learning," *Knowl. Inf. Syst.*, vol. 13, no. 1, pp. 1–42, 2007.

[37] Z. Zhang and T. W. S. Chow, "Maximum margin multisurface support tensor machines with application to image classification and segmentation," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 849–860, Jan. 2012.

[38] Z. Hao, L. He, B. Chen, and X. Yang, "A linear support higher-order tensor machine for classification," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2911–2920, Jul. 2013.

[39] I. Kotsia, W. Guo, and I. Patras, "Higher rank support tensor machines for visual recognition," *Pattern Recognit.*, vol. 45, no. 12, pp. 4192–4203, Dec. 2012.

[40] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear principal component analysis of tensor objects," *IEEE Trans. Neural Netw.*, vol. 19, no. 1, pp. 18–39, Jan. 2008.

[41] L. D. Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Departement Elektrotechniek, Katholike Univ. Leuven, Leuven, Belgium, 1997.

[42] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Apr. 2000.

[43] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, Sep. 2009.

[44] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognit.*, vol. 44, no. 7, pp. 1540–1551, Jul. 2011.

[45] C. Huang, L. S. Davis, and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 5, pp. 725–749, Jan. 2002.

[46] J. Bezdek and R. Hathaway, "Some notes on alternating optimization," in *Proc. AFSS Int. Conf. Fuzzy Syst.*, London, U.K., 2002, pp. 288–300.

[47] C. Cortes and V. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, pp. 273–297, 1995.

[48] D. Xu *et al.*, "Concurrent subspaces analysis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, 2005, vol. 2, pp. 203–208.

[49] A. Karami, M. Yazdi, and A. Z. Asli, "Noise reduction of hyperspectral images using kernel non-negative Tucker decomposition," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 487–493, Jun. 2011.

[50] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.

[51] N. Renard and S. Bourennane, "Dimensionality reduction based on tensor modeling for classification methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1123–1131, Apr. 2009.

[52] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear subspace analysis of image ensembles," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Madison, WI, USA, 2003, vol. 2, pp. 93–99.

[53] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Multilinear principal component analysis of tensor objects for recognition," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Hongkong, 2006, vol. 2, pp. 776–779.

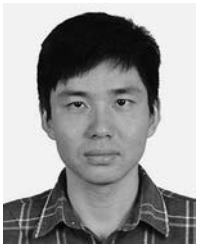
[54] D. Xu *et al.*, "Human gait recognition with matrix representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 7, pp. 896–903, Jul. 2006.

[55] S. Yan *et al.*, "Multilinear discriminant analysis for face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 1, pp. 212–220, Jan. 2007.

[56] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[57] F. Tsai and J.-S. Lai, "Feature extraction of hyperspectral image cubes using three-dimensional gray-level cooccurrence," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3504–3513, Jun. 2013.
[58] X. Huang, Q. Lu, and L. Zhang, "A multi-index learning approach for classification of high-resolution remotely sensed images over urban areas," *ISPRS J. Photogramm.*, vol. 90, pp. 36–48, Apr. 2014.

Xian Guo received the B.S. degree from Central South University, Changsha, China, in 2010, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2015.

He is currently a Postdoctoral Researcher with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. His current research interests focus on high-resolution image processing, machine learning, and remote sensing applications.



Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009, working in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing.

He is currently a Full Professor with Wuhan University, where he teaches remote sensing, photogrammetry, image interpretation, etc. He is the Founder and the Director of the Institute of Remote Sensing Information Processing, School of Remote Sensing and Information Engineering at Wuhan University. He has authored or coauthored over 60 peer-reviewed articles in international journals. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications.

Prof. Huang has served as an Associate Editor of the IEEE Geoscience and Remote Sensing Letters since 2014. He was the lead Guest Editor of the special issue on information extraction from high spatial resolution optical remotely sensed imagery of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He was a recipient of the Top Ten Academic Star of Wuhan University in 2009, the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing in 2010, the New Century Excellent Talents in University from the Ministry of Education of China in 2011, the National Excellent Doctoral Dissertation Award of China in 2012, and the China National Science Fund for Excellent Young Scholars in 2015. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as a Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS. He was the winner of the IEEE GRSS 2014 Data Fusion Contest.

Lefei Zhang (S'11–M'14) received the B.S. and Ph.D. degrees from Wuhan University, Wuhan, China, in 2008 and 2013, respectively.

From August 2013 to July 2015, he was a Postdoctoral Researcher with the School of Computer, Wuhan University. In 2015, he was a Visiting Scholar with the State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China. He is currently a Lecturer with the School of Computer, Wuhan University, and a Hong Kong Scholar with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong. His research interests include pattern recognition, image processing, and remote sensing.

Dr. Zhang is a Reviewer for over 20 international journals, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a Chang Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has over 450 research papers and five books. He is the holder of 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of The Institution of Engineering and Technology, an Executive Member (Board of Governor) of the China National Committee of International Geosphere-Biosphere Programme, and an Executive Member of the China Society of Image and Graphics and others. He is the founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He was the General Chair for the 4th IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing. He regularly serves as a Cochair of the series SPIE conferences on multispectral image processing and pattern recognition, conference on Asia remote sensing, and many other conferences. He is currently serving as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, and the *Journal of Remote Sensing* and the Guest Editor of the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He has been a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN EARTH OBSERVATIONS AND APPLIED REMOTE SENSING (JSTARS). He was a recipient of the 2010 best paper Boeing award and the 2013 best paper ERDAS award from the American Society of Photogrammetry and Remote Sensing. He was also a recipient of best reviewer awards from IEEE GRSS for his service to JSTARS in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2014. His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium student paper contest in recent years.

Antonio Plaza (M'05–SM'07–F'15) was born in Caceres, Spain, in 1975. He received the B.Sc., M.S. and Ph.D. degrees in computer engineering from the University of Extremadura (UEX), Cáceres, Spain, in 1997, 1999 and 2002, respectively.

He is currently an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Caceres, where he is the Head of the Hyperspectral Computing Laboratory, one of the most productive research groups working on remotely sensed hyperspectral data processing worldwide. He has been the advisor of 12 Ph.D. dissertations and over 30 M.Sc. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with a total funding of 2.8 million Euro. He has authored over 500 publications, including 166 journal papers (115 in IEEE journals), 22 book chapters, and over 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited *High-Performance Computing in Remote Sensing* for CRC Press/Taylor and Francis and guest edited nine special issues on hyperspectral remote sensing for different journals. His main research interests consist of hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza is a Fellow of the IEEE "for contributions to hyperspectral data processing and parallel computing of Earth observation data." He served as the Director of Education Activities of the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012. He has served as the President of the Spanish Chapter of IEEE GRSS since November 2012. He has served as a Proposal Evaluator for the European Commission, the National Science Foundation, the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He has reviewed over 500 manuscripts for over 50 different journals. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal. He is also an Associate Editor of IEEE ACCESS. He was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He was a recipient of the recognition of Best Reviewers of the IEEE Geoscience and Remote Sensing Letters in 2009 and of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING in 2010, a journal for which he served as an Associate Editor in 2007–2012. He was also a recipient of the most highly cited paper (2005–2010) in the Journal of Parallel and Distributed Computing, the 2013 Best Paper Award of the JSTARS journal, and the Best Column Award of the IEEE SIGNAL PROCESSING MAGAZINE in 2015. He has been a recipient of the best paper awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology. He has been also a recipient of the Best Ph.D. Dissertation award at the University of Extremadura, a recognition also received by six of his Ph.D. students.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently a Rector and a Professor of electrical and computer engineering with the University of Iceland. He is a Cofounder of the biomedical start-up company Oxymap. His research interests are in remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing. He has published extensively in those fields.

Prof. Benediktsson is a Fellow of The International Society of Optics and Photonics and a member of the Association of Chartered Engineers in Iceland (VFI) and Societas Scinetiarum Islandica. He was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He has served as an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and IEEE ACCESS since 2013. He was an Editor of TGRS from 2003 to 2008. He was on the International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2007–2010. He was a recipient of the Stevan J. Kristof Award from Purdue University as outstanding graduate student in remote sensing in 1991, the Icelandic Research Council's Outstanding Young Researcher Award in 1997, and the IEEE Third Millennium Medal in 2000. He was a Corecipient of the University of Iceland's Technology Innovation Award in 2004, the 2012 TGRS Paper Award, and the IEEE GRSS Highest Impact Paper Award in 2013. He was also a recipient of the yearly research award from the Engineering Research Institute of the University of Iceland in 2006, the Outstanding Service Award from the IEEE GRSS in 2007, and the IEEE/VFI Electrical Engineer of the Year Award in 2013. He is also a member of Tau Beta Pi.