Generalized Composite Kernel Framework for Hyperspectral Image Classification

Jun Li, Prashanth Reddy Marpu, *Member, IEEE*, Antonio Plaza, *Senior Member, IEEE*, José M. Bioucas-Dias, *Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—This paper presents a new framework for the development of generalized composite kernel machines for hyperspectral image classification. We construct a new family of generalized composite kernels which exhibit great flexibility when combining the spectral and the spatial information contained in the hyperspectral data, without any weight parameters. The classifier adopted in this work is the multinomial logistic regression, and the spatial information is modeled from extended multiattribute profiles. In order to illustrate the good performance of the proposed framework, support vector machines are also used for evaluation purposes. Our experimental results with real hyperspectral images collected by the National Aeronautics and Space Administration Jet Propulsion Laboratory's Airborne Visible/Infrared Imaging Spectrometer and the Reflective Optics Spectrographic Imaging System indicate that the proposed framework leads to state-of-the-art classification performance in complex analysis scenarios.**

***Index Terms*—Extended multiattribute morphological profiles (MPs), generalized composite kernel, hyperspectral imaging, multinomial logistic regression (MLR), supervised classification.**

I. Introduction

THE rich spectral information available in remotely sensed hyperspectral images allows for the possibility to distinguish between spectrally similar materials [1]. However, supervised classification of hyperspectral images (which assumes prior knowledge in the form of class labels for some spectral signatures) is a very challenging task due to the generally unfavorable ratio between the (large) number of spectral bands and the (limited) number of training samples available *a priori*, which results in the Hughes phenomenon [2]. The application of methods originally developed for the classification of lower dimensional data sets (such as multispectral images) generally provides poor results when applied to hyperspectral images, particularly in the case of small training sets [3].

On the other hand, the collection of reliable training samples is very expensive in terms of time and finance, and the possibility to exploit large ground truth information is not common [4]. To address this issue, kernel methods [5] have been widely used due to their insensitivity to the curse of dimensionality [6]. A relevant aspect of these methods in the context of supervised classification techniques is their ability to perform with limited training sets [6], [7]. The good generalization capability of machine learning techniques such as the support vector machine (SVM) [6] and multinomial logistic regression (MLR) [8]–[12] can still be enhanced by reducing the input data dimensionality [13], [14], which can help in addressing ill-posed problems based on limited training samples [7], [15], [16]. On the other hand, recent studies have shown that composite kernels [17] and multiple kernels [18]–[22] can provide enhanced classification accuracy by incorporating the spatial information in addition to the spectral information. This approach exhibits flexibility to balance between the spatial and spectral information and computational efficiency.

The introduction of composite kernels opened a wide field for future developments in which spatial and spectral information can be easily integrated. However, standard composite kernels and multiple kernel learning (MKL) methods based on SVM classifiers generally require convex combinations of kernels. This introduces some practical limitations. First, the kernels are bounded by convex combinations. Second, it is difficult to optimize the parameters involved in the learning process. In order to overcome these limitations, in this paper, we introduce a new framework for the development of generalized composite kernel machines intended to perform hyperspectral image classification, which proposes several innovative contributions with regards to previously developed approaches.

1) First and foremost, our multiple kernels can be linearly combined without any restriction of convexity. This introduces a different approach with regards to standard composite kernels and MKL methods, in which composite kernels need to be convex combinations of kernels. Compared with our approach, these methods are less flexible when modeling the classifier boundaries and further depend on weight parameters which need to be carefully adjusted in order to control the ratio of spatial to spectral information in the final classification.
2) Another innovative contribution of this work is to use the MLR classifier in order to implement the proposed framework. Unlike SVMs—which require convex

Manuscript received March 19, 2012; revised August 10, 2012 and October 28, 2012; accepted November 20, 2012. This work was supported in part by the European Community's Marie Curie Research Training Networks Programme through Hyperspectral Imaging Network under Contract MRTN-CT-2006-035927, by the Portuguese Science and Technology Foundation under Project PEst-OE/EEI/LA0008/2011, by the Spanish Ministry of Science and Innovation through the Calibration of Earth Observation Systems-SPAIN Project under Reference AYA2011-29334-C02-02, by the Icelandic Research Fund, and by the University of Iceland Research Fund.

J. Li and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, 10071 Cáceres, Spain (e-mail: jun@lx.it.pt; aplaza@unex.es).

P. R. Marpu is with the Masdar Institute of Science and Technology, Abu Dhabi 54224, United Arab Emirates (e-mail: prashanthreddym@gmail.com).

J. M. Bioucas-Dias is with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade Técnica de Lisboa, 1049-1 Lisbon, Portugal (e-mail: bioucas@lx.it.pt).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, 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.2012.2230268

combinations of kernels—the MLR classifier has a lot of flexibility in the construction of nonlinear kernels. As a consequence, our logistic regression functions do not even need to be kernels. Since we are working under a probabilistic framework, the MLR is a natural approach to derive probabilities, and this allows a perfect blend with composite kernels while allowing us, at the same time, to control the generalization capacity by using logistic regressors [11]. By adopting a Laplacian prior [8], [11], our optimization problem is convex, and we can provide powerful algorithms to solve the problem with ensured convergence [9].

3) Finally, the proposed approach models the spatial information using extended multiattribute morphological profiles (MPs) [23], defined as a generalization of the well-known MPs [24], [25] that provide a multilevel characterization of an image created by the sequential application of morphological filters that can be used to model different kinds of spatial structures in the scene.

The aforementioned features conform an innovative framework rooted on solid fundamental principles and which brings the composite kernel idea to a new domain. At the same time, the proposed approach allows for a more flexible generalized formulation (not necessarily based on kernels) in order to naturally integrate the spatial and the spectral information in hyperspectral image classification, without the need for complex parameter settings. The remainder of this paper is organized as follows. Section II briefly outlines the MLR classifier used to model the posterior class probabilities and the EMAPs used to model the spatial information. Section III presents the proposed generalized composite kernel framework. Section IV reports classification results using real hyperspectral images collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) [26] and Reflective Optics Spectrographic Imaging System (ROSIS) [25] imaging spectrometers. Comparisons with standard composite kernels implemented using the SVM classifier are also included. Section V concludes this paper with some remarks and hints at plausible future research lines.

II. Modules Used for Building the Proposed Approach

A. Classification Using MLR

First of all, we define the notation that will be adopted throughout this paper. Let $\mathcal{K} := \{1, \ldots, K\}$ denote a set of K class labels, let $\mathcal{S} := \{1, \ldots, n\}$ denote a set of integers indexing the n pixels of a hyperspectral image, let $\mathbf{x} := (\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathbb{R}^d$ denote such hyperspectral image made up of d-dimensional feature vectors, let $\mathbf{y} := (y_1, \ldots, y_n)$ denote an image of labels, and let $\mathcal{D}_L := \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_L, y_L)\}$ be the labeled training set with L being the number of samples in $\mathcal{D}_L$. In this work, we model the posterior class probabilities using MLR [27]

$$p(y_i = k|\mathbf{x}_i, \boldsymbol{\nu}) := \frac{\exp\left(\boldsymbol{\nu}^{(k)^{\mathrm{T}}}\mathbf{h}(\mathbf{x}_i)\right)}{\sum_{k=1}^{K}\exp\left(\boldsymbol{\nu}^{(k)^{\mathrm{T}}}\mathbf{h}(\mathbf{x}_i)\right)} \tag{1}$$

where $\mathbf{h}(\mathbf{x}_i) := [h_1(\mathbf{x}_i), \ldots, h_l(\mathbf{x}_i)]^{\mathrm{T}}$ is a vector of l fixed functions of the input data, often termed features, $\boldsymbol{\nu}$ denotes the regressors, and $\boldsymbol{\nu} := [\boldsymbol{\nu}^{(1)^{\mathrm{T}}}, \ldots, \boldsymbol{\nu}^{(K-1)^{\mathrm{T}}}]^{\mathrm{T}}$. Since the density (1) does not depend on translations on the regressors $\boldsymbol{\nu}^{(k)}$, in this work, we take $\boldsymbol{\nu}^{(K)} = \mathbf{0}$. It should be noted that the function $\mathbf{h}$ may be linear (i.e., $\mathbf{h}(\mathbf{x}_i) := [1, x_{i,1}, \ldots, x_{i,d}]^{\mathrm{T}}$, where $x_{i,j}$ is the jth component of $\mathbf{x}_i$) or nonlinear. A kernel is some symmetric function which offers a mechanism to deal with the nonlinear case, i.e.,

$$\mathbf{h}(\mathbf{x}_i) := [1, K(\mathbf{x}_i, \mathbf{x}_1), \ldots, K(\mathbf{x}_i, \mathbf{x}_l)]^{\mathrm{T}}, \quad \text{where } K(\mathbf{x}_i, \mathbf{x}_j) := \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \tag{2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product and $\phi(\cdot)$ is a nonlinear mapping function [5]. Kernels have been largely used in this context since they tend to improve data separability. In this work, we use the Gaussian radial basis function kernel, i.e., $K(\mathbf{x}_i, \mathbf{x}_j) := \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$ kernel, which is widely used in hyperspectral image classification [6]. In the present scenario, learning the class densities amounts to estimating the logistic regressors $\boldsymbol{\nu}$. Following previous works [8], [10], [11], [27], we compute $\boldsymbol{\nu}$ by calculating the *maximum a posteriori* estimate

$$\widehat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} \; \ell(\boldsymbol{\nu}) + \log p(\boldsymbol{\nu}) \tag{3}$$

where $\ell(\boldsymbol{\nu})$ is the log-likelihood function given by

$$\begin{aligned} \ell(\boldsymbol{\nu}) &:= \log \prod_{i=1}^{L} p(y_i|\mathbf{x}_i, \boldsymbol{\nu}) \\ &:= \sum_{i=1}^{L}\left(\mathbf{h}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\nu}^{(y_i)} - \log\sum_{k=1}^{K}\exp\left(\mathbf{h}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\nu}^{(k)}\right)\right) \end{aligned} \tag{4}$$

and $\log p(\boldsymbol{\nu})$ is a prior over $\boldsymbol{\nu}$ which is independent from the observation $\mathbf{x}$. In order to control the machine complexity and, thus, its generalization capacity, we model $\boldsymbol{\nu}$ as a random vector with Laplacian density $p(\boldsymbol{\nu}) \propto \exp(-\lambda\|\boldsymbol{\nu}\|_1)$, where λ is the regularization parameter controlling the degree of sparsity [8], [11].

Therefore, problem (3) turns to

$$\begin{aligned} \widehat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} &\sum_{i=1}^{L}\left(\mathbf{h}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\nu}^{(y_i)} - \log\sum_{k=1}^{K}\exp\left(\mathbf{h}^{\mathrm{T}}(\mathbf{x}_i)\boldsymbol{\nu}^{(k)}\right)\right) \\ &+ \log p(\boldsymbol{\nu}) \end{aligned} \tag{5}$$

$$= \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{L}\nu_1^{(y_i)} + \log p(\boldsymbol{\nu}) \tag{6}$$

$$+ \sum_{i=1}^{L}\sum_{j=1}^{L}\left(K(\mathbf{x}_i, \mathbf{x}_j)\nu_{j+1}^{(y_i)} - \log\sum_{k=1}^{K}\exp\left(\nu_1^{(k)} + K(\mathbf{x}_i, \mathbf{x}_j)\nu_{j+1}^{(k)}\right)\right) \tag{7}$$

where the term in (6) is independent from the observation data $\mathbf{x}$; therefore, it is also independent from the kernel function.

The optimization problem (5) can be solved by sparse MLR (SMLR) [8] and fast SMLR [12]. However, most hyperspectral data sets are beyond the reach of these algorithms as their analysis becomes unbearable as the kernel size increases. This is even more critical in our framework in which we use multiple kernels. In order to address this issue, we take advantage of the logistic regression via variable splitting and augmented Lagrangian (LORSAL) algorithm [9] with overall complexity $O(L^2K)$ (recall that L is the number of training samples and K is the number of classes), which is able to deal with large kernel sizes. LORSAL plays a central role in this work, as in previous contributions [10], [11].

B. Modeling Spatial Information With EMAPs

MPs [25] and attribute profiles (APs) [23], [28] have been successfully employed for combining the spectral and spatial information while classifying remote sensing imagery. APs are obtained by applying a sequence of attribute filters (AFs) to a gray-level image [23]. AFs are operators defined in the mathematical morphology framework which operate by merging connected components at different levels in the image [29]. Extended APs (EAPs) [28] are an extension of APs for the analysis of multi-/hyperspectral images. The EAP is a stack of APs obtained with different features. When the EAPs obtained using different types of attributes are stacked together, the resulting profile is called as extended multiattribute profile (EMAP) [28].

The filtering operation is based on the evaluation of how an attribute A, computed for every region, compares to a given reference value λ. For a connected component of the image C_i, if $A(C_i) > \lambda$, then the region is kept unaltered; otherwise, it is set to the grayscale value of the adjacent region with closer value, thereby merging the connected components. When the region is merged to the adjacent region of a lower (or greater) gray level, the operation performed is a thinning (or thickening). Given a sequence of thresholds $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, an AP is obtained by applying a sequence of attribute thinning and attribute thickening operations as follows [30]:

$$\mathrm{AP}\left(f_j(\mathbf{x}_i)\right) := \{\Phi_n\left(f_j(\mathbf{x}_i)\right), \ldots, \Phi_1\left(f_j(\mathbf{x}_i)\right), f_j(\mathbf{x}_i), \gamma_1\left(f_j(\mathbf{x}_i)\right), \ldots, \gamma_n\left(f_j(\mathbf{x}_i)\right)\} \quad (8)$$

where Φ_i and γ_i denote the thickening and thinning transformations, respectively, and $f_j(\mathbf{x}_i)$ denotes a feature extracted from the original pixel information $\mathbf{x}_i$. This is because the aforementioned formulation refers to a single feature (or spectral band) extracted from the hyperspectral data, and therefore, the full spectral information contained in $\mathbf{x}$ is not considered. In [25], it was suggested to use several principal components (PCs) [31] of the hyperspectral data to address this issue, but any feature reduction technique on the hyperspectral image could also be used [23]. In this way, the EAP is obtained by generating an AP on each of the first q PCs (or any other features retained after applying feature selection on the hyperspectral image), thus building a stacked vector using the AP on each feature, as illustrated graphically in Fig. 1. This leads to the following definition of the EAP for the pixel $\mathbf{x}_i$:

$$\mathrm{EAP}(\mathbf{x}_i) := \{\mathrm{AP}\left(f_1(\mathbf{x}_i)\right), \mathrm{AP}\left(f_2(\mathbf{x}_i)\right), \ldots, \mathrm{AP}\left(f_q(\mathbf{x}_i)\right)\} \quad (9)$$

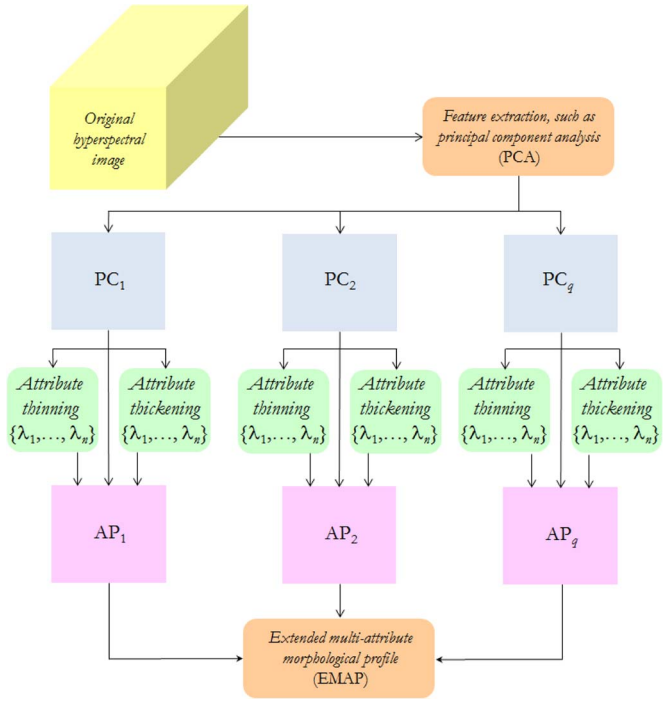

Fig. 1. Graphical illustration of the procedure adopted in order to construct an EAP from a hyperspectral image. The EMAP is a combination of EAPs obtained with different attributes.

where q is the number of retained features. In our experiments, we have used $q = 3$, i.e., we retain the first three PCs which account for most of the variance present in the considered data sets. From the EAP definition in (9), the consideration of multiple attributes leads to the concept of EMAP which improves the capability in extracting the spatial characteristics of the structures in the scene. However, a significant increase of the dimensionality of the data is also obtained. In this work, only area and standard deviation attributes are considered as they can be more easily calculated and are well related to the object hierarchy in the images [32].

III. Generalized-Composite-Kernel-Based Framework

In this section, we present a new generalized composite kernel framework for spectral–spatial classification of hyperspectral images using the modules described in the previous section. Let us redefine a pixel entity $\mathbf{x}_i$ as $\mathbf{x}_i^\omega$ when its original spectral information is considered and as $\mathbf{x}_i^s$ when its associated spatial-contextual information, obtained using (9), is considered. Hence, we can refer to a spectral kernel as $K^\omega(\mathbf{x}_i^\omega, \mathbf{x}_j^\omega)$ and to a spatial kernel as $K^s(\mathbf{x}_i^s, \mathbf{x}_j^s)$. The proposed method stacks the spectral and spatial kernels as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) := \left[K^\omega\left(\mathbf{x}_i^\omega, \mathbf{x}_j^\omega\right), K^s\left(\mathbf{x}_i^s, \mathbf{x}_j^s\right)\right]^{\mathrm{T}}. \quad (10)$$

With these definitions in mind, the input function $\mathbf{h}(\mathbf{x}_i)$ becomes

$$\begin{aligned}\mathbf{h}(\mathbf{x}_i) &:= \left[1, K^{\mathrm{T}}(\mathbf{x}_i, \mathbf{x}_1), \ldots, K^{\mathrm{T}}(\mathbf{x}_i, \mathbf{x}_L)\right]^{\mathrm{T}} \\ &:= [1, K^\omega\left(\mathbf{x}_i^\omega, \mathbf{x}_1^\omega\right), \ldots, K^\omega\left(\mathbf{x}_i^\omega, \mathbf{x}_L^\omega\right), K^s\left(\mathbf{x}_i^s, \mathbf{x}_1^s\right) \\ &\quad \ldots, K^s\left(\mathbf{x}_i^s, \mathbf{x}_L^s\right)]^{\mathrm{T}}. \end{aligned} \quad (11)$$

Therefore, problem (5) changes to

$$\widehat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{L} \nu_1^{(y_i)} + \log p(\boldsymbol{\nu}) + \sum_{i=1}^{L}\sum_{j=1}^{L} \times \left(\nu_{j+1}^{(y_i)} K^{\omega}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{\omega}\right) + \nu_{j+L+1}^{(y_i)} K^{s}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{s}\right) - \log \sum_{k=1}^{K} \exp \times \left(\nu_1^{(k)} + \nu_{j+1}^{(k)} K^{\omega}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{\omega}\right) + \nu_{j+L+1}^{(k)} K^{s}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{s}\right) \right) \right). \quad (12)$$

Notice that, in (12), the logistic weights ν_j^k and ν_{j+L}^k for the spectral and spatial kernels are independent from each other. This linear combination of multiple kernels is more flexible than the convex combination used in MKL, where the relative kernel weights are fixed. Therefore, our approach provides more freedom to balance the spectral and spatial information.

The kernels introduced in (10) can be conveniently modified to account for the cross-information between the spatial and the spectral information as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) := \left[K^{\omega}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{\omega}\right), K^{s}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{s}\right), K^{\omega s}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{s}\right), K^{s\omega}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{\omega}\right)\right]^{\mathrm{T}} \quad (13)$$

where $K^{\omega s}(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{s}) := \langle \phi(\mathbf{x}_i^{\omega}), \phi(\mathbf{x}_j^{s}) \rangle$ and $K^{s\omega}(\mathbf{x}_i^{s}, \mathbf{x}_j^{\omega}) := \langle \phi(\mathbf{x}_i^{s}), \phi(\mathbf{x}_j^{\omega}) \rangle$ are the spectral–spatial and spatial–spectral cross-information kernels, respectively. Notice that, in most cases, the spectral vectors and the EMAPs have different dimensions, which brings difficulty for building the cross-information kernels. In order to build such kernels, we apply a two-step procedure. Let d_1 and d_2 denote the dimensionalities of the spectral- and spatial-based features, respectively. If $d_1 > d_2$, we obtain the first d_2 PCs of the spectral features [31]. If $d_2 > d_1$, we obtain the first d_1 components of the spatial features. In this way, spectral information and spatial information are always equally balanced, and we can obtain a cross-information kernel. We are aware that this procedure may be suboptimal. Nevertheless, as shown in Section IV, this strategy is shown to provide good results. By including the cross-information, the input function $\mathbf{h}(\mathbf{x}_i)$ becomes

$$\begin{aligned}\mathbf{h}(\mathbf{x}_i) &:= \left[1, K^{\mathrm{T}}(\mathbf{x}_i, \mathbf{x}_1), \ldots, K^{\mathrm{T}}(\mathbf{x}_i, \mathbf{x}_L)\right]^{\mathrm{T}} \\ &:= [1, K^{\omega}(\mathbf{x}_i^{\omega}, \mathbf{x}_1^{\omega}), \ldots, K^{\omega}(\mathbf{x}_i^{\omega}, \mathbf{x}_L^{\omega}), K^{s}(\mathbf{x}_i^{s}, \mathbf{x}_1^{s}), \ldots, \\ &\quad K^{s}(\mathbf{x}_i^{s}, \mathbf{x}_L^{s}), K^{\omega s}(\mathbf{x}_i^{\omega}, \mathbf{x}_1^{s}), \ldots, K^{\omega s}(\mathbf{x}_i^{\omega}, \mathbf{x}_L^{s}), \\ &\quad K^{s\omega}(\mathbf{x}_i^{s}, \mathbf{x}_1^{\omega}), \ldots, K^{s\omega}(\mathbf{x}_i^{s}, \mathbf{x}_L^{\omega})]^{\mathrm{T}}. \end{aligned} \quad (14)$$

In this context, problem (5) changes to

$$\widehat{\boldsymbol{\nu}} = \arg\max_{\boldsymbol{\nu}} \sum_{i=1}^{L} \nu_1^{(y_i)} + \log p(\boldsymbol{\nu}) + \sum_{i=1}^{L}\sum_{j=1}^{L} \times \left(\nu_{j+1}^{(y_i)} K^{\omega}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{\omega}\right) + \nu_{j+L+1}^{(y_i)} K^{s}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{s}\right) + \nu_{j+2L+1}^{(y_i)} K^{\omega s}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{s}\right) + \nu_{j+3L+1}^{(y_i)} K^{s\omega}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{\omega}\right) - \log \sum_{k=1}^{K} \exp \times \left(\nu_1^{(k)} + \nu_{j+1}^{(k)} K^{\omega}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{\omega}\right) + \nu_{j+L+1}^{(k)} K^{s}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{s}\right) + \nu_{j+2L+1}^{(k)} K^{\omega s}\left(\mathbf{x}_i^{\omega}, \mathbf{x}_j^{s}\right) + \nu_{j+3L+1}^{(k)} K^{s\omega}\left(\mathbf{x}_i^{s}, \mathbf{x}_j^{\omega}\right) \right) \right). \quad (15)$$

Optimization problems (12) and (15) have the same structure as problem (5). Therefore, we also adopt the LORSAL algorithm [9] to compute the logistic regressors $\boldsymbol{\nu}$. The overall complexity of the LORSAL algorithm is $O(L^2K)$, which depends on the number of training samples and classes. In our framework, the overall complexity depends on the stacked-kernel size, where the overall complexities are $O((2L)^2K)$ and $O((4L)^2K)$ for problems (12) and (15), respectively. In the following section, we provide an experimental evaluation of the proposed generalized composite kernel framework using two well-known hyperspectral data sets that have been widely used to evaluate the classification accuracy of techniques for remotely sensed hyperspectral data interpretation.

IV. Experimental Results

In this section, we evaluate the proposed approach using two real hyperspectral data sets. In our experiments, K^s refers to the kernel-based classification obtained using only the spatial information provided by the EMAPs, i.e., $\mathbf{x}_i^s$ is obtained by applying an EMAP-based approach to the original hyperspectral image. On the other hand, K^{ω} refers to the classification obtained using the full spectral information contained in the hyperspectral image, i.e., $\mathbf{x}_i^{\omega} := \mathbf{x}_i$. Similarly, $K^{\omega,s}$ refers to the stacked-feature approach, i.e., the input feature vector is stacked as $\left[\mathbf{x}_i^{s^{\mathrm{T}}}, \mathbf{x}_i^{\omega^{\mathrm{T}}}\right]^{\mathrm{T}}$. $K^{(\omega+s)} := K^{\omega} + K^{s}$ refers to the direct summation kernel, $\mu K^{(\omega+s)} := \mu K^{\omega} + (1-\mu)K^{s}$ refers to the weighted summation kernel, and $K^{(\omega+s+\omega s+s\omega)} := K^{\omega} + K^{s} + K^{\omega s} + K^{s\omega}$ refers to the cross-information kernel, all defined in the original composite kernel framework described in [17]. Finally, $K^{[\omega,s]} := [K^{\omega}, K^{s}]^{\mathrm{T}}$ and $K^{[\omega,s,\omega s,s\omega]} := [K^{\omega}, K^{s}, K^{\omega s}, K^{s\omega}]^{\mathrm{T}}$ refer to the stacked and cross-information kernels defined in our proposed generalized composite framework (see (10) and (13), respectively). It should be noted that the parameters used for building the EMAPs were not optimized but selected empirically based on previous experiences [23]. Regarding the parameters involved in the MLR (LORSAL) method, we performed a parameter search between minima and maxima values according to our experiments with the same hyperspectral scenes in previous contributions [11], [14], [33] and selected the optimal parameter values in those ranges. We would like to emphasize that it has been found empirically that the proposed framework is not very sensitive to parameter settings. Concerning the SVM classifier, we optimized the related parameters using tenfold cross-validation. The reported figures of overall accuracy (OA), average accuracy (AA), κ statistic, and individual classification accuracies are obtained by averaging the results obtained after conducting ten independent Monte Carlo runs with respect to the training set $\mathcal{D}_L$. In our experiments, we will report both the average and the standard deviation obtained after ten Monte Carlo runs. The remainder of this section is organized as follows. In Section IV-A, we introduce the data sets used for evaluation, which comprise the AVIRIS Indian Pines and ROSIS Pavia University data sets (two widely used benchmarks

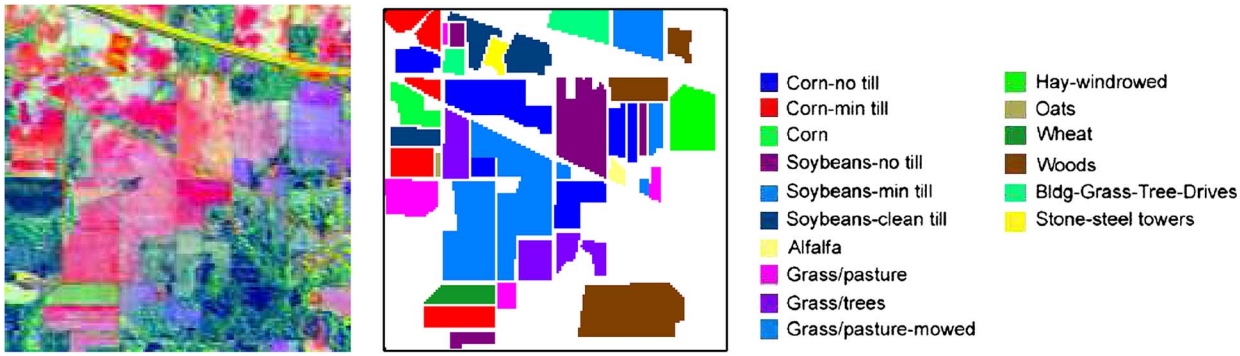

Fig. 2. (a) False-color composition of the AVIRIS Indian Pines scene. (b) Reference map containing (right) 16 mutually exclusive land-cover classes.

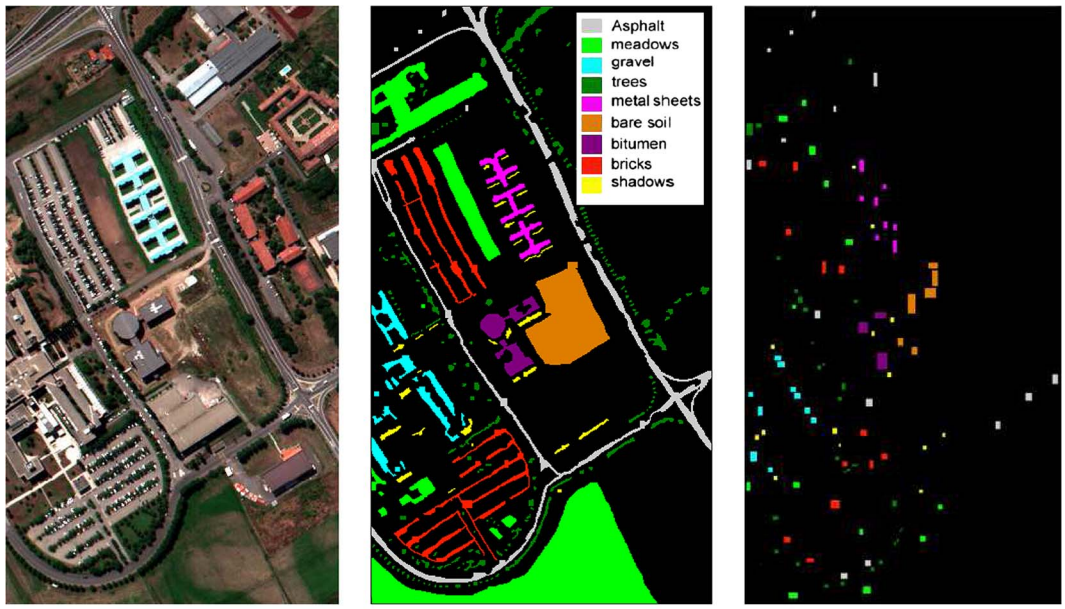

Fig. 3. (a) False-color composition of the ROSIS University of Pavia scene. (b) Reference map containing nine mutually exclusive land-cover classes. (c) Training set used in experiments.

for hyperspectral image classification). Section IV-B describes the experiments with the AVIRIS Indian Pines data set. Finally, Section IV-C conducts experiments using the ROSIS Pavia University data set.

A. *Hyperspectral Data Sets*

Two hyperspectral data sets collected by two different instruments are used in our experiments.

1) The first hyperspectral image used in experiments was collected by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145 lines by 145 samples, was acquired over a mixed agricultural/forest area, early in the growing season. The scene comprises 202 spectral channels in the wavelength range from 0.4 to 2.5 μm, with a nominal spectral resolution of 10 nm, a moderate spatial resolution of 20 m by pixel, and a 16-b radiometric resolution. After an initial screening, several spectral bands were removed from the data set due to noise and water absorption phenomena, leaving a total of 164 radiance channels to be used in the experiments. For illustrative purposes, Fig. 2(a) shows a false-color composition of the AVIRIS Indian Pines scene, while Fig. 2(b) shows the reference map available for the scene, displayed in the form of a class assignment for each labeled pixel, with 16 mutually exclusive reference classes, i.e., 10 366 samples in total. These data, including reference information, are available online,[1] a fact which has made this scene a widely used benchmark for testing the accuracy of hyperspectral data classification algorithms. This scene constitutes a challenging classification problem due to the presence of mixed pixels in all available classes and because of the unbalanced number of available labeled pixels per class.
2) The second hyperspectral data set was collected by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The flight was operated by the Deutschen Zentrum für Luft- und Raumfahrt (German Aerospace Center) in the framework of the HySens project, managed and sponsored by the European Union. The image size in pixels is 610 $\times$ 340, with a very high spatial resolution of 1.3 m per pixel. The number of

[1] Available online: http://dynamo.ecn.purdue.edu/biehl/MultiSpec.

TABLE I
OVERALL CLASSIFICATION ACCURACIES (IN PERCENT) AND κ STATISTIC (IN THE PARENTHESES) OBTAINED FOR DIFFERENT CLASSIFICATION METHODS WHEN APPLIED TO THE AVIRIS INDIAN PINES HYPERSPECTRAL DATA SET USING DIFFERENT NUMBERS OF TRAINING SAMPLES PER CLASS. THE STANDARD DEVIATION OF THE TEN CONDUCTED MONTE CARLO RUNS IS ALSO REPORTED IN EACH CASE

# Samples per class	Classification methods							
	K^{ω}	K^{s}	Composite kernels				Generalized composite kernels	
			$K^{\omega,s}$	$K^{(\omega+s)}$	$\mu K^{(\omega+s)}$	$K^{(\omega+s+\omega s+s\omega)}$	$K^{[\omega,s]}$	$K^{[\omega,s,\omega s,s\omega]}$
5	56.05±1.87 (50.87)	67.16±2.19 (63.28)	66.53±2.92 (62.49)	67.80±2.55 (63.93)	68.68±2.84 (64.90)	66.53±3.02 (62.49)	69.15±3.30 (65.39)	69.63±3.22 (65.93)
10	64.12±2.94 (59.69)	77.95±2.62 (75.18)	76.68±2.24 (73.72)	78.97±2.27 (76.29)	79.74±2.19 (77.14)	76.68±2.44 (73.72)	80.43±1.79 (77.89)	80.79±1.69 (78.29)
15	68.55±2.58 (64.63)	82.70±2.15 (80.44)	81.87±2.65 (79.51)	83.65±2.22 (81.50)	84.21±2.16 (82.13)	81.87±2.01 (79.51)	85.14±1.98 (83.16)	85.32±1.88 (83.36)
20	71.67±1.59 (68.13)	85.46±1.42 (83.55)	85.19±0.59 (83.22)	86.30±1.14 (84.48)	86.80±1.06 (85.05)	85.19±1.43 (83.22)	87.98±1.06 (86.36)	88.19±0.93 (86.59)
25	73.14±1.87 (69.73)	87.26±1.26 (85.55)	86.51±1.42 (84.69)	87.77±1.33 (86.11)	87.99±1.41 (86.36)	86.51±1.79 (84.69)	89.17±1.53 (87.69)	89.21±1.46 (87.73)
30	75.50±1.22 (72.30)	89.59±0.89 (88.16)	88.81±1.05(87.26)	89.98±0.89 (88.60)	90.21±0.95 (88.86)	88.81±1.03 (87.26)	91.29±0.83 (90.07)	91.35±1.00 (90.14)
35	76.46±0.99 (73.41)	90.64±0.68 (89.33)	89.58±0.97 (88.13)	91.05±0.78 (89.79)	91.04±0.97 (89.78)	89.58±1.12 (88.13)	92.04±0.85 (90.91)	92.41±0.93 (91.33)
40	76.36±1.41 (73.28)	91.31±1.11 (90.10)	90.95±1.08 (89.67)	91.89±1.20 (90.75)	92.08±1.19 (90.96)	90.95±1.35 (89.67)	92.71±0.92 (91.67)	92.90±0.97 (91.89)
45	77.44±1.12 (74.45)	91.73±0.98 (90.56)	91.25±1.02 (90.00)	92.11±0.94 (90.98)	92.15±0.95 (91.03)	91.25±0.91 (90.00)	93.13±0.86 (92.14)	93.18±0.90 (92.19)
50	77.71±0.68 (74.74)	92.55±0.92 (91.49)	91.96±0.71 (90.81)	92.93±1.03 (91.92)	92.99±1.03 (91.98)	91.96±1.07 (90.81)	93.66±0.81 (92.75)	94.08±0.80 (93.23)

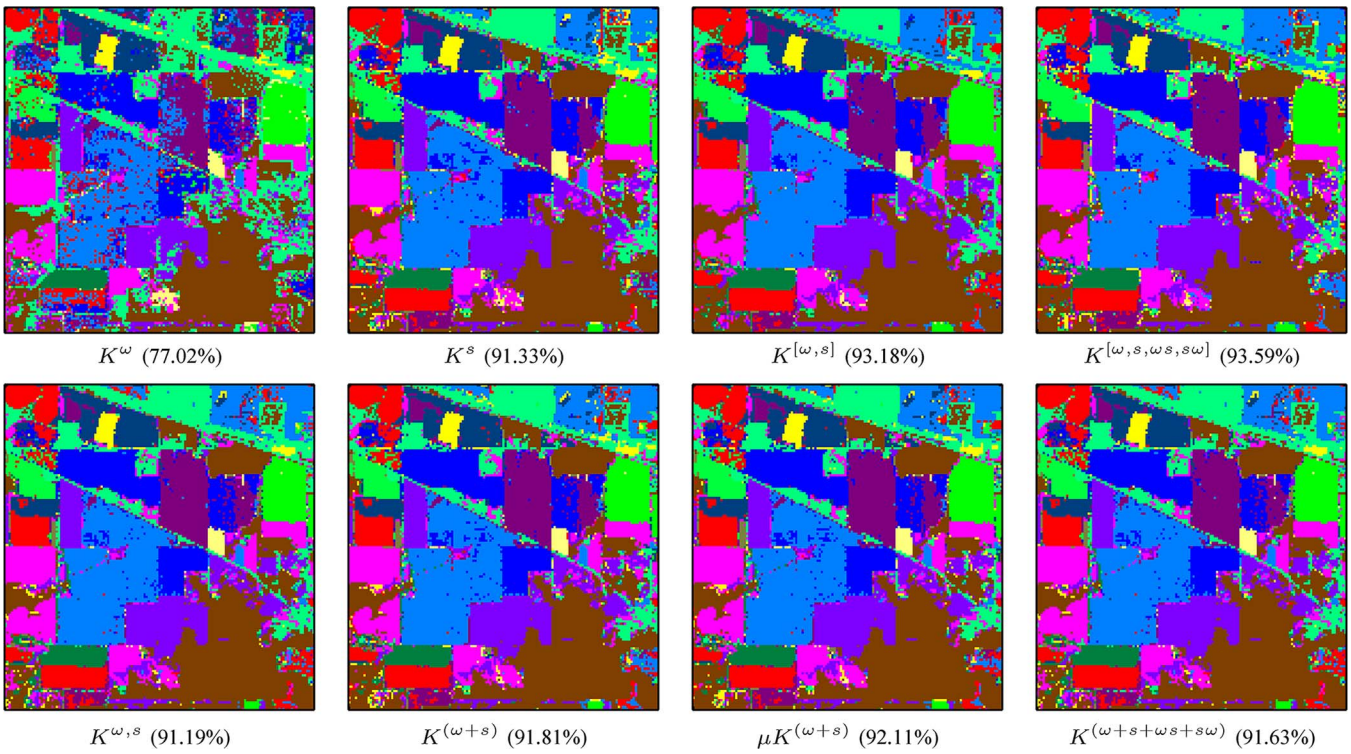

Fig. 4. Classification maps and overall classification accuracies (in the parentheses) obtained for the AVIRIS Indian Pines data set.

data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Fig. 3(a) shows a false-color composite of the image, while Fig. 3(b) shows nine reference classes of interest, which comprise urban features, as well as soil and vegetation features. Finally, Fig. 3(c) shows a fixed training set available for the scene which comprises 3921 training samples (the remaining 42 776 samples were used for testing in our experiments).

B. Experiments With AVIRIS Indian Pines Data Set

In this set of experiments, we first evaluated the classification accuracy of the proposed approach using the AVIRIS Indian Pines data set in Fig. 2(a). This set of experiments is designed to analyze the impact of the type of kernel used for classification (as a function of the number of training samples used per class). The EMAPs in this particular scene were built using threshold values in the range of 2.5%–10% with respect to the mean of the individual features, with a step of 2.5%, were chosen for the standard deviation attribute and thresholds of 200, 500, and 1000, and were selected for the area attribute. Let L_k be the total number of labeled samples in the reference map for a single class k, and let l_k be the number of selected training samples per class. Table I shows the overall classification accuracies (in percent) and the κ statistic obtained for different kernel-based classification methods for different values of l_k which are kept very small on purpose. Here, if $l_k < L_k$ (i.e., the number of labeled samples per class is very

TABLE II
OAS, AAS, AND INDIVIDUAL CLASSIFICATION ACCURACIES (IN PERCENT) OBTAINED FOR DIFFERENT CLASSIFICATION METHODS WHEN APPLIED TO THE AVIRIS INDIAN PINES HYPERSPECTRAL DATA SET WITH A BALANCED TRAINING SET (5% OF THE LABELED SAMPLES PER CLASS USED FOR TRAINING (A TOTAL OF 515 SAMPLES) AND THE REMAINING LABELED SAMPLES USED FOR TESTING)

Class	# Samples		Classification methods							
					Composite kernels				Generalized composite kernels	
	Train	Test	K^{ω}	K^{s}	$K^{\omega,s}$	$K^{(\omega+s)}$	$\mu K^{(\omega+s)}$	$K^{(\omega+s+\omega s+s\omega)}$	$K^{[\omega,s]}$	$K^{[\omega,s,\omega s,s\omega]}$
			MLR classifier							
Alfalfa	3	51	47.06±15.41	80.78±6.52	66.67±13.27	81.37±6.42	84.12±3.63	84.31±3.58	86.08±0.62	86.08±0.62
Corn-no till	71	1363	78.24±3.01	87.77±2.11	86.98±2.13	88.12±2.11	90.07±2.11	90.37±2.09	91.78±1.76	91.80±1.74
Corn-min till	41	793	64.17±3.01	88.02±2.42	88.08±2.06	88.12±2.41	89.27±2.69	89.53±2.72	92.04±2.19	92.02±2.24
Corn	11	223	48.21±11.76	69.24±12.53	64.26±11.88	69.82±12.07	77.44±11.74	77.04±12.31	84.08±8.99	84.93±8.69
Grass/pasture	24	473	87.76±2.27	90.47±3.29	91.14±2.35	90.30±3.43	90.53±3.80	90.66±3.69	93.04±2.43	93.04±2.54
Grass/tree	37	710	95.13±1.40	97.21±1.41	97.34±1.14	97.24±1.39	97.58±1.29	97.58±1.31	98.35±1.07	98.41±1.15
Grass/pasture-mowed	3	23	53.04±11.74	84.78±7.17	75.22±8.71	85.22±7.16	90.00±6.17	90.00±6.17	94.35±2.93	94.35±2.93
Hay-windrowed	24	465	98.84±0.61	99.01±0.25	99.29±0.10	99.01±0.27	99.27±0.11	99.25±0.15	99.33±0.12	99.33±0.12
Oats	3	17	68.82±17.33	81.76±18.64	81.18±17.49	82.94±17.69	87.65±10.80	88.24±11.19	94.12±4.52	95.29±3.61
Soybeans-no till	48	920	68.42±5.22	86.50±4.69	85.60±5.17	86.53±4.86	88.96±4.14	89.09±4.15	90.00±4.57	89.91±4.60
Soybeans-min till	123	2245	82.56±1.26	94.55±0.85	94.01±1.20	94.62±0.95	94.70±1.15	94.83±1.24	96.65±0.62	96.70±0.69
Soybeans-clean till	30	584	74.52±5.35	81.64±5.43	84.95±4.71	82.28±5.50	83.05±5.42	83.54±5.54	89.33±5.71	89.61±5.80
Wheat	10	202	99.36±0.52	98.61±0.45	99.21±0.48	98.61±0.45	98.86±0.52	99.01±0.57	99.01±0.49	99.06±0.47
Woods	64	1230	95.46±1.53	96.62±2.49	95.63±2.50	96.61±2.57	97.31±2.34	97.32±2.37	97.76±1.99	97.80±2.03
Bldg-grass-tree-drives	19	361	50.75±3.49	81.58±5.31	80.80±5.74	81.94±5.37	83.43±4.86	83.96±4.72	89.14±4.18	89.22±4.15
Stone-steel towers	4	91	62.09±6.95	92.53±5.07	77.91±5.09	87.69±6.81	70.88±8.26	76.26±7.37	62.31±9.90	62.53±10.27
Overall accuracy			80.16±0.73	90.93±0.62	90.38±0.53	91.03±0.69	91.93±0.71	92.13±0.67	93.87±0.52	93.93±0.57
Average accuracy			73.40±1.26	88.19±1.41	85.52±1.35	88.15±1.45	88.94±1.57	89.44±1.40	91.09±1.34	91.26±1.42
κ statistic			77.26±0.84	89.69±0.70	89.03±0.61	89.80±0.76	90.82±0.80	91.05±0.76	93.01±0.59	93.07±0.65
Time (Seconds)			5.88	6.82	8.33	7.06	8.25	6.11	9.19	32.50
			SVM classifier							
Overall accuracy			76.95±0.85	90.52±0.75	90.36±0.86	90.61±0.62	90.85±0.79	91.17±0.65	-	-
Average accuracy			73.18±1.63	86.44±1.27	85.82±1.47	87.28±1.27	87.37±1.22	88.55±1.07	-	-
κ statistic			73.65±0.96	89.18±0.86	88.99±0.98	89.29±0.71	89.56±0.90	89.93±0.74	-	-
Time (Seconds)			9.12	8.66	18.94	13.03	13.18	19.13	-	-

small), we take $l_k = L_k/2$. For the classes which are very small, i.e., *Alfalfa*, *Grass/pasture-mowed*, and *Oats*, the number of training samples was set to $l_k = 10$. As shown by Table I, the proposed generalized composite kernel framework exhibits the potential to improve the classification results provided by using the spectral and the spatial information alone and also by the different kernels introduced in [17] for the composite kernel approach. It can be observed that the proposed generalized framework can improve the obtained classification results in terms of accuracies in all cases, with the cross-information kernel providing better results than the stacked-kernel approach as the number of training samples per class increases. For illustrative purposes, Fig. 4 shows some of the classification maps obtained for the experiment with $l_k = 50$ training samples in Table I. These classification maps correspond to one of the ten Monte Carlo runs that were averaged in order to generate the classification scores reported in Table I, which also reports the standard deviation of the ten conducted Monte Carlo runs in each case. The advantages obtained by adopting the proposed generalized composite kernel approach with regards to other considered methods can be visually appreciated in the maps displayed in Fig. 4, which also reports the overall classification accuracies obtained for each considered kernel-based classifier in the parentheses.

In order to show the performance of our proposed approach under different training conditions and scenarios, in the second experiment, we evaluated the classification accuracy of the proposed approach using a balanced training set per class in which around 5% of the labeled samples per class have been used for training (a total of 515 samples) and the remaining labeled samples are used for testing. For very small classes, we take a minimum of three training samples per class. Table II shows the OAs, AAs, and individual classification accuracies (in percent) and the κ statistic obtained for different kernel-based classification methods when applied to this scene. In the table, we also include the results obtained by the SVM classifier with the same kernels and inputs as our proposed approach based on the MLR. The processing times in seconds, measured in a desktop PC equipped with an Intel Core 2 Duo CPU (at 2.33 GHz) and 2 GB of RAM, are also included for reference.

Several conclusions can be obtained from Table II. First of all, the proposed approach obtained the best results in terms

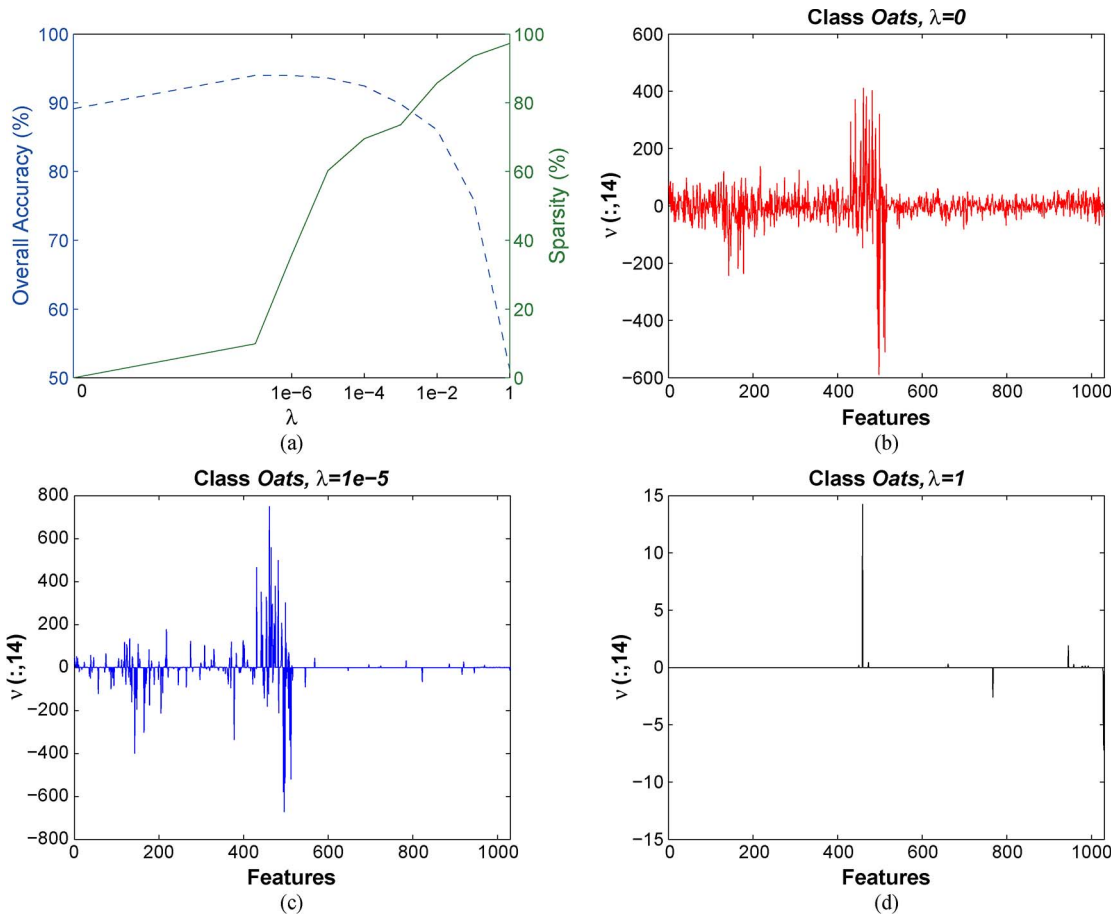

Fig. 5. (a) Graphical representation of the *sparsity* (in percent) and the overall classification accuracy (OA) (in percent) obtained by the proposed method as functions of parameter λ in the classification experiment using the AVIRIS Indian Pines scene, reported in Table II. (b) and (c) Regressors for a specific class in the AVIRIS Indian Pines scene (*Oats*) with different values of λ.

of accuracies in this scenario with balanced training sample selection per class. In fact, the proposed approach is more accurate with a balanced training set as the large classes can be better modeled using the additional training information available. These results indicate that our proposed approach performs very accurately under different training conditions and scenarios. Furthermore, it is noticeable that the other type of kernels defined in the generalized approach (stacked and cross-information) produced very competitive results. Finally, it is observed that the obtained results involving the spatial information are much better in terms of accuracies than those obtained only with the spectral information. The experiment demonstrates that the EMAP method can accurately model spatial-contextual information in all cases.

Finally, in order to show that the proposed solution is effectively sparse and investigate the relationships with parameter λ, we simply define the *sparsity* as $n_0/n \times 100\%$, where n is the number of parameters in the regressors $\boldsymbol{\nu}$ and n_0 is the number of small values, i.e., $|\nu_i| \leq 1e-3$ for any i in $\boldsymbol{\nu}$. Fig. 5(a) represents both the overall classification accuracy and the *sparsity* as functions of parameter λ in the experiment reported in Table II with the AVIRIS Indian Pines scene. We recall that, in this experiment, 5% of the reference data are used for training purposes. From Fig. 5(a), it can be observed that, without using the sparse prior (i.e., $\lambda = 0$), the obtained OA is below 90%. In turn, by enforcing sparsity, we can increase the OA to more than 93% (with above 50% sparsity), which corresponds to the best generalization capacity. For illustrative purposes, Fig. 5(b) and (c) shows the regressors for a specific class in the AVIRIS Indian Pines scene (*Oats*) with different values of λ.

C. Experiments With ROSIS University of Pavia Data Set

In this experiment, we evaluate the classification accuracy of the proposed approach using the ROSIS University of Pavia data set in Fig. 3(a). The EMAPs for this particular scene were built using threshold values in the range of 2.5%–10% with respect to the mean of the individual features, with a step of 2.5%, and were chosen for the definition of the criteria based on the standard deviation attribute. Values of 100, 200, 500, and 1000 were selected as references for the area attribute. The threshold values considered for the area attribute were chosen according to the resolution of the data and, thus, the size of the objects present in the scene. For illustrative purposes, Fig. 6 shows some of the filtered images obtained after using the area attribute (top row) and the standard deviation attribute (bottom row) using different thresholds.

Table III shows the OAs, AAs, and individual classification accuracies (in percent) and the κ statistic obtained for different kernel-based classification methods when applied to this scene, using the fixed training set in Fig. 3(c). In addition, Table III

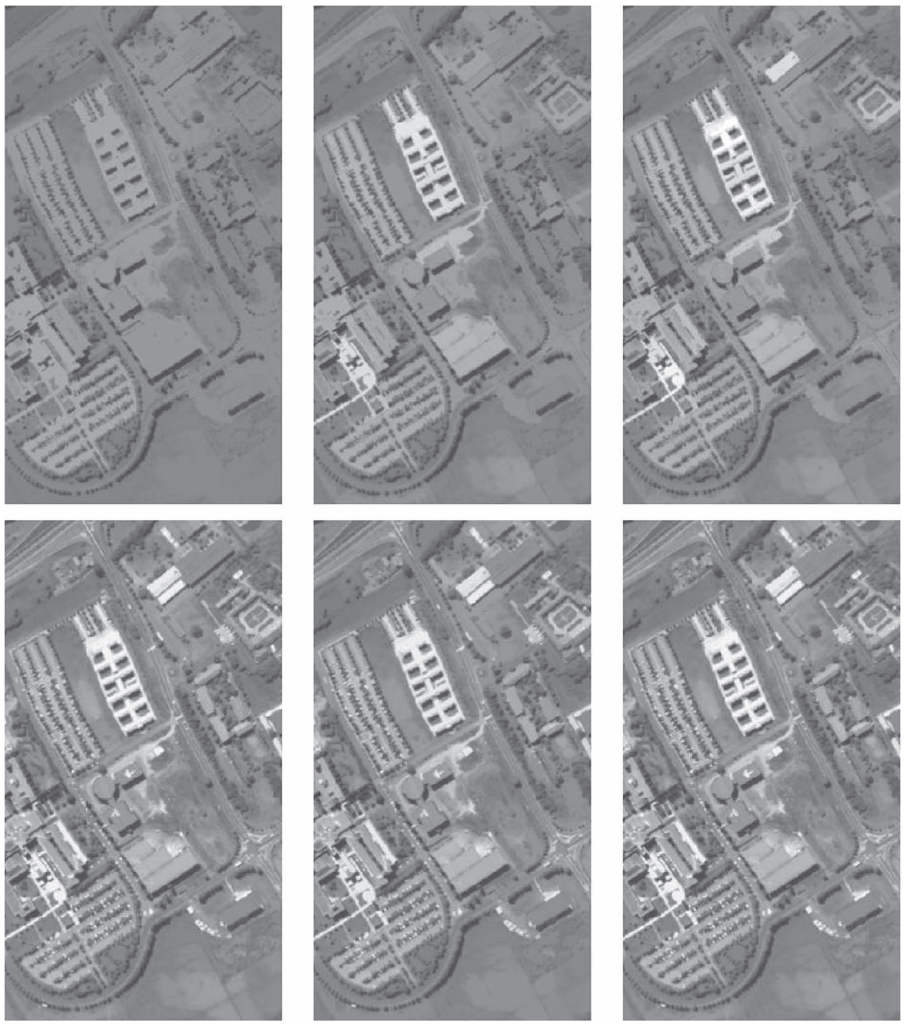

Fig. 6. Filtered images obtained after using (top row) the area attribute and (bottom row) the standard deviation attribute using different thresholds for the ROSIS University of Pavia data set.

includes the results obtained by using the SVM instead of the MLR classifier, with the same set of considered kernels. The table also includes the processing times in seconds, measured for the different methods in a desktop PC equipped with an Intel Core 2 Duo CPU (at 2.33 GHz) and 2 GB of RAM. As shown by Table III, the classification accuracies obtained by the different kernel-based methods are very high for all the considered classes. The higher classification accuracies were always obtained by using the proposed generalized-composite-kernel-based framework, which provides state-of-the-art performance for this scene. For illustrative purposes, Fig. 7 shows some of the classification maps obtained for the ROSIS Pavia University scene using the fixed training set depicted in Fig. 3(c). A very good delineation of complex urban structures can be clearly observed in the obtained results.

In a second experiment, we analyze the relative weight of spatial and spectral information in the proposed spectral and spatial kernels. Let $(\boldsymbol{\nu}^{\omega})^{\mathrm{T}} K^{\omega}$ and $(\boldsymbol{\nu}^{s})^{\mathrm{T}} K^{s}$ denote the spectral and spatial numerators of the MLR in (1). From this expression, we can readily conclude that, for a given pixel $\mathbf{x}_i$, if $(\boldsymbol{\nu}^{\omega})^{\mathrm{T}} K^{\omega} \geq (\boldsymbol{\nu}^{s})^{\mathrm{T}} K^{s}$, then the classification is dominated by the spectral information. Otherwise, the spatial information has more impact in the final classification result. In order to analyze this relevant issue in more detail, Table IV shows an experimental assessment (conducted using the ROSIS Pavia University scene) of the number of samples dominated by spectral and spatial information in the proposed spatial and spectral kernels. In this experiment, we use 20% of the original set of labeled samples (the original set is formed by a total of 3291 samples, and we are using a total of 784 samples for training). Here, we performed ten Monte Carlo runs on the Pavia data set by merging/splitting the testing set. From Table IV, it can be observed that classes *Asphalt*, *Bare soil*, *Bitumen*, and *Meadows* are dominated by the spatial information whereas classes *Bricks* and *Shadows* are dominated by the spectral information. In the classes *Gravel* and *Metal sheet*, the spectral and spatial information had similar impact. A qualitative interpretation for these results is that neither type of information is dominant. Furthermore, the joint use of the two types of information leads to better classification results than that obtained with the spectral or the spatial information alone.

To conclude this section, Table V analyzes the statistical significance of the obtained MLR-based classification results, obtained using different kernels for the Pavia University data set, using McNemar's test [34]. From Table V, we can conclude that the kernels combining the spatial and the spectral information

TABLE III
OAs, AAs, and Individual Class Accuracies (in Percent) Obtained for Different Classification Methods When Applied to the ROSIS Pavia University Hyperspectral Data Set

Class	# Samples		Classification methods							
			K^{ω}	K^{s}	Composite kernels				Generalized composite kernels	
	Train	Test			$K^{\omega,s}$	$K^{(\omega+s)}$	$\mu K^{(\omega+s)}$	$K^{(\omega+s+\omega s+s\omega)}$	$K^{[\omega,s]}$	$K^{[\omega,s,\omega s,s\omega]}$
			MLR classifier							
Asphalt	548	6631	82.64	98.36	97.86	98.48	98.91	98.48	98.88	98.88
Bare soil	540	18649	68.62	97.83	97.45	98.51	98.26	98.52	98.44	98.47
Bitumen	392	2099	75.04	90.85	91.14	89.57	90.23	89.61	90.09	90.23
Bricks	524	3064	97.00	98.79	98.66	98.30	98.69	98.30	98.76	98.76
Gravel	265	1345	99.41	99.85	99.70	99.70	99.93	99.70	99.93	99.93
Meadows	532	5029	93.88	86.88	96.66	94.87	96.20	94.95	96.34	96.46
Metal sheets	375	1330	90.08	99.92	99.70	99.85	99.92	99.85	99.85	99.85
Shadows	514	3682	91.36	99.43	99.54	99.62	99.59	99.62	99.70	99.67
Trees	231	947	97.57	98.52	99.16	97.68	97.99	97.68	97.57	97.57
Overall accuracy			80.34	96.63	97.56	97.78	97.97	97.80	98.05	98.09
Average accuracy			88.40	96.71	97.76	97.40	97.75	97.41	97.73	97.76
κ			75.41	95.52	96.77	97.06	97.31	97.08	97.42	97.46
Time (Seconds)			166.95	175.65	184.58	188.29	187.01	208.93	944.78	3532
			SVM classifier							
Overall accuracy			80.89	90.80	93.07	92.87	92.97	92.44	-	-
Average accuracy			89.09	94.08	94.95	94.92	94.92	90.19	-	-
κ statistic			76.12	88.13	90.99	90.74	90.86	94.63	-	-
Time (Seconds)			121.89	148.80	157.59	277.25	283.22	307.85	-	-

bring significant differences with regards to those only considering either the spatial or the spectral information. Also, Table V reveals that the cross-information kernel cannot significantly improve the results provided by the stacked kernels which are less complex in computational terms. Hence, if users have time restrictions, we recommend building our proposed generalized composite kernel framework with stacked kernels to integrate the spectral and the spatial information.

V. Conclusion and Future Research Lines

In this paper, we have developed a new framework for generalized-composite-kernel-based classification of remotely sensed hyperspectral data. Compared with the original developments in [17], which set the basis for the development of this kind of classifiers, we have introduced several distinctive features. First and foremost, we rely on an SMLR classifier and in the LORSAL algorithm [9] instead of the SVM adopted in [17] to produce the final classification results. Second, our proposed approach equally balances the spectral and the spatial information contained in the hyperspectral data without any weight parameters. Finally, the proposed approach models the spatial information using EMAPs, thus addressing one of the main future directions identified in [17]. The aforementioned features are used to define a new framework for kernel-based classification that provides state-of-the-art results with two widely used hyperspectral images representing very challenging classification scenarios. Although the results obtained are very encouraging, further experiments with additional scenes and comparison methods should be conducted. Based on our previous work [10], [11], [33], we also envisage the following future perspectives.

1) As the MLR classifier provides a probabilistic output, in future work, we will consider including a Markov random field multilevel logistic prior in order to achieve even smoother classification results by modeling spatial information, following our previous developments in [10] and [11]. It is expected that such framework could address the somewhat noisy classification output in Figs. 4 and 7.

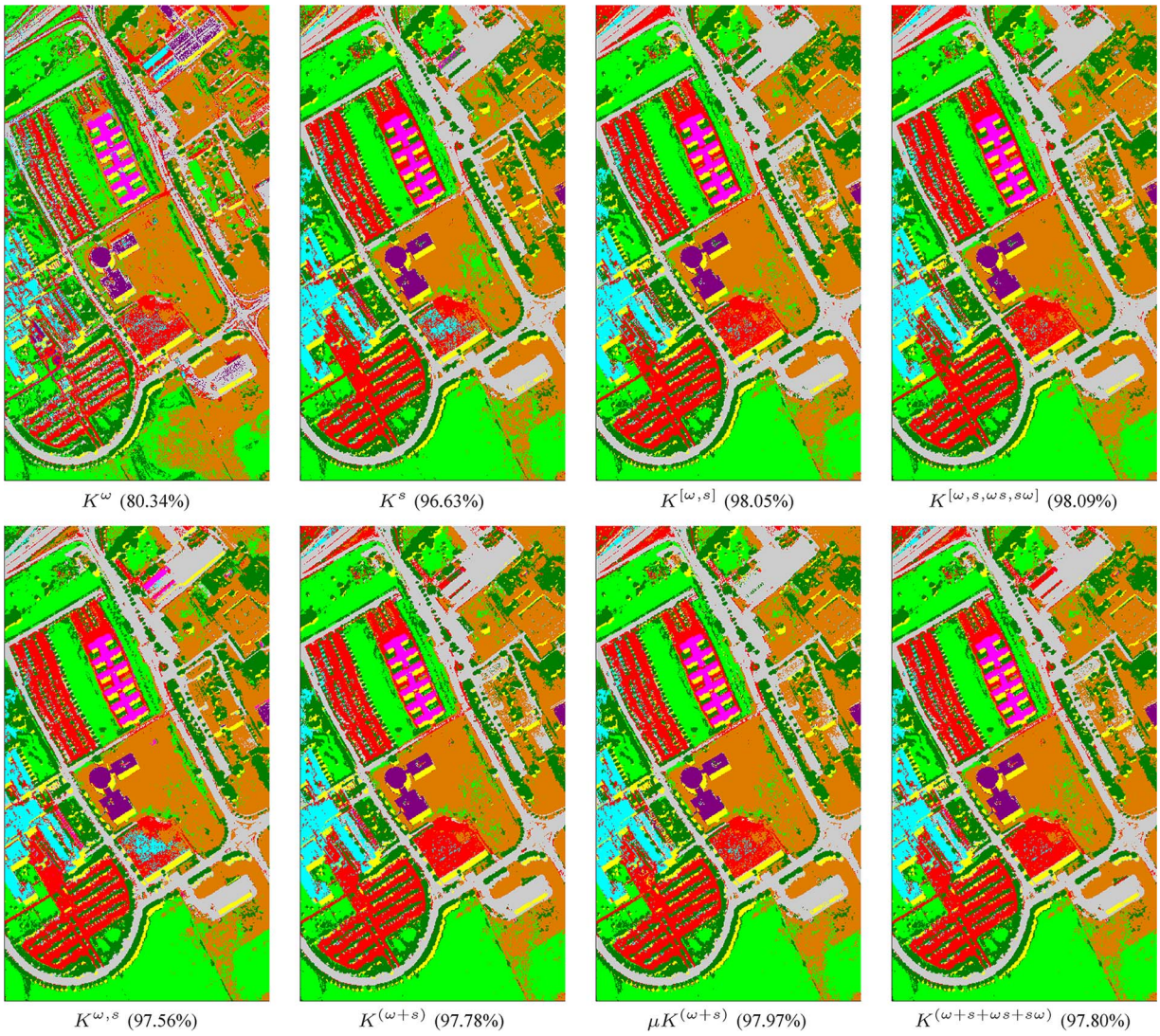

Fig. 7. Classification maps and overall classification accuracies (in the parentheses) obtained for the ROSIS University of Pavia data set.

TABLE IV

NUMBER OF SAMPLES DOMINATED BY THE SPECTRAL AND THE SPATIAL INFORMATION IN THE PROPOSED GENERALIZED COMPOSITE KERNEL METHOD ($K^{[\omega,s]}$) USING 20% OF THE ORIGINAL SET OF LABELED SAMPLES IN THE ROSIS PAVIA UNIVERSITY SCENE (THE ORIGINAL SET IS FORMED BY A TOTAL OF 3291 SAMPLES, AND WE ARE USING A TOTAL OF 784 SAMPLES FOR TRAINING). ALL THE RESULTS AND ACCURACIES (IN PERCENT) ARE OBTAINED BASED ON TEN MONTE CARLO RUNS PERFORMED BY MERGING/SPLITTING THE TESTING SET. THE OA IN THIS EXPERIMENT IS 95.55%, THE AA IS 96.21%, AND THE κ STATISTIC IS 94.63%

	Number of samples								
Class	Asphalt	Bare soil	Bitumen	Bricks	Gravel	Meadows	Metal sheets	Shadows	Trees
$(\boldsymbol{\nu}^{\omega})^T K^{\omega} \geq (\boldsymbol{\nu}^{s})^T K^{s}$	0	1860	372	3143	799	1478	604	3989	-
$(\boldsymbol{\nu}^{\omega})^T K^{\omega} < (\boldsymbol{\nu}^{s})^T K^{s}$	6508	15995	1527	0	544	4254	724	22	-
Total number of samples	6508	17855	1899	3143	1343	5732	1328	4011	957
Class accuracy	96.60	94.73	85.40	98.17	99.81	95.92	99.80	97.09	98.34

2) Since the proposed framework explicitly includes the spatial information, we could adopt this framework to perform spectral–spatial active learning. In comparison with our previous work in [33] which addresses spectral–spatial active learning by estimating the conditional marginals depending on the whole image, the proposed approach only considers the labeled information for the learning stage, thus reducing computational complexity.
3) The proposed method is computationally expensive in comparison with a single kernel-based approach. In order to address this issue, we are planning on developing computationally efficient implementations of the proposed

TABLE V
STATISTICAL SIGNIFICANCE OF THE MLR CLASSIFICATION RESULTS, OBTAINED USING DIFFERENT KERNELS FOR THE PAVIA UNIVERSITY DATA SET USING 3921 SAMPLES FOR TRAINING, USING MCNEMAR'S TEST

	K^{ω}	K^{s}	$K^{\omega,s}$	$K^{(\omega+s)}$	$\mu K^{(\omega+s)}$	$K^{(\omega+s+\omega s+s\omega)}$	$K^{[\omega,s]}$	$K^{[\omega,s,\omega s,s\omega]}$
K^{ω}	-	-74.76	-82.44	-81.83	-82.83	-81.93	-83.05	-83.20
K^{s}	74.76	-	-11.86	-15.76	-18.74	-15.97	-19.21	-19.57
$K^{\omega,s}$	82.44	11.86	-	-3.50	-6.73	-3.73	-7.81	-8.33
$K^{(\omega+s)}$	81.83	15.76	3.50	-	-4.92	-1.28	-7.33	-8.29
$\mu K^{(\omega+s)}$	82.83	18.74	6.73	4.92	-	4.59	-2.76	-3.64
$K^{(\omega+s+\omega s+s\omega)}$	81.93	15.97	3.73	1.28	-4.59	-	-7.01	-8.00
$K^{[\omega,s]}$	83.05	19.21	7.81	7.33	2.76	7.01	-	-2.08
$K^{[\omega,s,\omega s,s\omega]}$	83.20	19.57	8.33	8.29	3.64	8.00	2.08	-

approaches by resorting to shared memory parallel machines such as multicores or graphical processing units, as the feasibility of parallelizing the EMAP framework in this kind of architectures has already been proven in the literature [35].

4) Last but not the least, we are planning to extend our proposed generalized composite kernel framework to include MKL, as described in [20]. Here, the use of multiple kernels may introduce the advantage of removing convex constraints. It is expected that the proposed framework may provide more flexibility for MKL, where the simple MKL method in [20] can be formulated as a special case of generalized MKL. We are also planning on exploring the impact of reducing the dimensionality of the original data set prior to applying the proposed approach.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe for making the Airborne Visible/Infrared Imaging Spectrometer Indian Pines hyperspectral data set available to the community and Prof. P. Gamba for providing the Reflective Optics Spectrographic Imaging System data over Pavia, Italy, along with the training and test sets. The authors would also like to thank the Associate Editor and the two anonymous reviewers for their outstanding comments and suggestions, which greatly helped to improve the technical quality and presentation of this paper.

REFERENCES

[1] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.

[2] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

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

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

[5] B. Scholkopf and A. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization and Beyond*. Cambridge, MA: MIT Press, 2002.

[6] 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.

[7] M. Fauvel, J. Benediktsson, J. Chanussot, and J. 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.

[8] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.

[9] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Inst. Superior Técnico, Lisbon, Portugal, 2009, Tech. Rep..

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

[11] J. Li, J. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

[12] J. S. Borges, J. M. Bioucas-Dias, and A. R. S. Marcal, "Bayesian hyperspectral image segmentation with discriminative class learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 2151–2164, Jun. 2011.

[13] J. M. Bioucas-Dias and J. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

[14] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[15] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote-sensing data over urban areas," *EURASIP J. Adv. Signal Process.*, vol. 2009, pp. 1–14, 2009.

[16] W. Li, S. Prasad, J. E. Fowled, 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.

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

[18] F. Bach, G. Lanckriet, and M. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *Proc. 21st Int. Conf. Mach. Learn.*, 2004, pp. 1–6.

[19] G. Lanckriet, N. Cristianini, P. Bartlett, L. Ghaoui, and M. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Dec. 2004.

[20] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, Nov. 2008.

[21] D. Tuia, G. Matasci, G. Camps-Valls, and M. Kanevski, "Learning relevant image features with multiple kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.
[22] C. Wang, D. You, Y. Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012.
[23] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[24] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
[25] 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.
[26] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams, "Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS)," *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
[27] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Statist. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992.
[28] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Jul. 2010.
[29] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin, Germany: Springer-Verlag, 2003.
[30] E. J. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Comput. Vis. Image Underst.*, vol. 64, no. 3, pp. 377–389, Nov. 1996.
[31] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer-Verlag, 2006.
[32] P. R. Marpu, M. Pedergnana, M. D. Mura, S. Peetersd, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral data using extended attribute profiles based on supervised and unsupervised feature extraction techniques," *Int. J. Image Data Fusion*, vol. 3, no. 3, pp. 269–298, Sep. 2012.
[33] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, 2012, to be published.
[34] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.
[35] M. H. F. Wilkinson, H. Gao, W. H. Hesselink, J.-E. Jonker, and A. Meijster, "Concurrent computation of attribute filters on shared memory parallel machines," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1800–1813, Oct. 2008.

Jun Li received the B.S. degree in geographic information systems from Hunan Normal University, Changsha, China, in 2004, the M.E. degree in remote sensing from Peking University, Beijing, China, in 2007, and the Ph.D. degree in electrical engineering from the Instituto de Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotécnica e de Computadores and the Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). She has also been actively involved in the Hyperspectral Imaging Network, a Marie Curie Research Training Network involving 15 partners in 12 countries and intended to foster research, training, and cooperation on hyperspectral imaging at the European level. Since 2011, she has been a Postdoctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. She has been a Reviewer of several journals, including *Pattern Recognition*, *Optical Engineering*, *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li is also 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 and has also been a Reviewer of the IEEE TRANSACTIONS ON IMAGE PROCESSING and the IEEE TRANSACTIONS ON FUZZY SYSTEMS. She was recognized as an outstanding Ph.D. student outside China by the Chinese Government in 2011.

Prashanth Reddy Marpu (M'05) received the M.Sc. degree from the Technical University of Denmark, Lyngby, Denmark, in 2006 and the Ph.D. degree from the Freiberg University of Mining and Technology, Freiberg, Germany, in 2009.

He is currently a Researcher with the Masdar Institute of Science and Technology, Masdar, United Arab Emirates. He was a Marie Curie Fellow in the framework of the Hyperspectral Imaging Network project with the University of Pavia, Pavia, Italy, during 2009–2010 and was a Postdoctoral Researcher with the University of Iceland, Reykjavik, Iceland, during 2010–2011.

Antonio Plaza (M'05–SM'07) was born in Cáceres, Spain, in 1975. He received the M.S. and Ph.D. degrees in computer engineering from the University of Extremadura, Cáceres.

He has been a Visiting Researcher with the Remote Sensing Signal and Image Processing Laboratory, University of Maryland, Baltimore; with the Applied Information Sciences Branch, Goddard Space Flight Center, National Aeronautics and Space Administration (NASA), Greenbelt, MD; with the Airborne Visible/Infrared Imaging Spectrometer Data Facility, NASA Jet Propulsion Laboratory, Pasadena, CA; with the Telecommunications and Remote Sensing Laboratory, University of Pavia, Pavia, Italy; and with the Grenoble Images Parole Signal Automatique, Grenoble, France. He is currently an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, where he is the Head of the Hyperspectral Computing Laboratory. He was the Coordinator of the Hyperspectral Imaging Network, a European project designed to build an interdisciplinary research community focused on hyperspectral imaging activities. He is the author or coauthor of more than 350 publications on remotely sensed hyperspectral imaging, including more than 90 journal citation report papers (45 since January 2011), around 20 book chapters, and over 230 conference proceeding papers. He has been a Guest Editor of seven special issues on scientific journals on the topic of remotely sensed hyperspectral imaging. His research interests include remotely sensed hyperspectral imaging, pattern recognition, signal and image processing, and efficient implementation of large-scale scientific problems on parallel and distributed computer architectures.

Dr. Plaza has been the Chair for the IEEE Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Whispers 2011). He has also been serving as the Chair for the SPIE Conference on Satellite Data Compression, Communications, and Processing, since 2009, and for the SPIE Europe Conference on High-Performance Computing in Remote Sensing, since 2011. He was a recipient of the Best Reviewer Award of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2009 and the Best Reviewer Award of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) in 2010, a journal for which he has served as Associate Editor since 2007 and for which he has reviewed more than 260 manuscripts. He is currently serving as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) and as the President of the Spanish Chapter of IEEE GRSS. He has been appointed Editor of the TGRS, with a three-year term starting in January 2013.

José M. Bioucas-Dias (S'87–M'95) received the E.E., M.Sc., Ph.D., and "Agregado" degrees in electrical and computer engineering from Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, Universidade Técnica de Lisboa, where he was an Assistant Professor from 1995 to 2007 and has been an Associate Professor since 2007. Since 1993, he has also been a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing.

Dr. Bioucas-Dias was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–2000), and he is an Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS). He was a Guest Editor of TGRS for the Special Issue on Spectral Unmixing of Remotely Sensed Data, of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING for the Special Issue on Hyperspectral Image and Signal Processing, and of the IEEE Signal Processing Magazine for the Special Issue on Signal and Image Processing in Hyperspectral Remote Sensing. He was the General Cochair of the Third IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (Whispers 2011) and has been a member of program/technical committees of several international conferences.

Jón Atli Benediktsson (S'84–M'90–S'M99–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, in 1987 and 1990, respectively.

He is currently a Pro Rector for Academic Affairs 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, and he has published extensively in those fields.

Prof. Benediktsson was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS AdCom since 2000. He was an Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has served as an Associate Editor of TGRS since 1999 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003. He was a recipient of the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing, the Icelandic Research Council's Outstanding Young Researcher Award in 1997, the IEEE Third Millennium Medal in 2000, the yearly research award from the Engineering Research Institute of the University of Iceland in 2006, and the Outstanding Service Award from the GRSS in 2007. He was a corecipient of the University of Iceland's Technology Innovation Award in 2004 and of the 2012 TGRS Best Paper Award. He is a Fellow of SPIE and a member of Societas Scinetiarum Islandica and Tau Beta Pi.