Spectral–Spatial Classification of Multispectral Images Using Kernel Feature Space Representation

Sergio Bernabé, *Student Member, IEEE,* Prashanth Reddy Marpu, *Member, IEEE,* Antonio Plaza, *Senior Member, IEEE,* Mauro Dalla Mura, *Member, IEEE,* and Jón Atli Benediktsson, *Fellow, IEEE*

***Abstract*—Over the last few years, several new strategies have been proposed for spectral–spatial classification of remotely sensed image data, for cases when high spatial and spectral resolutions are available. In this letter, we focus on the possibility of performing advanced spectral–spatial classification of remote sensing images with limited spectral resolution (often called multispectral). A new strategy is proposed, where the spectral dimensionality of the multispectral data is first expanded by using nonlinear feature extraction with kernel methods such as kernel principal component analysis. Then, extended multiattribute profiles (EMAPs), built on the expanded set of spectral features, are used to include spatial information. This strategy allows us to first decompose different spectral clusters into different spectral features and further improve the spatial discrimination. The resulting EMAPs are used for classification using advanced classifiers such as support vector machines and random forests. We test our proposed methodology with different multispectral data sets obtaining state-of-the-art classification results.**

***Index Terms*—Extended multiattribute profiles (EMAPs), kernel principal component analysis (KPCA), random forests (RFs), spectral–spatial classification, support vector machines (SVMs).**

I. Introduction

OVER the last few years, many efforts have been directed toward the use of spatial information to refine spectral-based classifiers, assuming high spectral resolution in the remotely sensed data to be processed. For instance, morphological profiles [1] have been widely used for spatial characterization of hyperspectral imagery [2], [3]. Markov random fields (MRFs) have also been employed for spatial characterization in hyperspectral classification, as described in [4] and [5] or as a postprocessing step as in [6]. Despite of the success of these approaches, fewer efforts have been directed toward exploiting spatial and spectral information in remotely sensed data with limited spectral resolution [7]. In this context, attribute profiles (APs) [8], and their extension to multiattribute profiles (EMAPs) [9], have been successfully used for classification of remotely sensed data with more limited spectral resolution [10].

In this letter, we develop a new strategy to perform spectral–spatial classification of images with low spectral resolution (often called multispectral images), which constitute a very important source of remotely sensed data. Our proposed strategy first expands the spectral dimensionality of the data by using nonlinear feature extraction with kernel methods, such as kernel principal component analysis (KPCA) [11]. While feature extraction is often performed with hyperspectral data to reduce dimensionality and select relevant features for classification, it is seldom used with lower spectral resolution data, where the multispectral bands (along with the panchromatic band) are generally used to extract spatial features. Here, we propose to use kernel feature extraction as a mechanism to expand spectral dimensionality and then extract the spatial information based on the extracted spectral features. The motivation behind this is to first decompose the multispectral data into features representing the inherent spectral clusters, which would increase the contrast between different classes after the kernel transformation. This is because this transformation brings the data to a higher dimensional space, in which the data may be easier to separate. Then, EMAPs built on the expanded set of spectral features are used to extract spatial information more optimally prior to classification using advanced techniques such as support vector machines (SVMs) [12] or random forests (RFs) [13]. The proposed strategy is a simple but effective extension of the regular processing chain with EMAPs presented in [9].

The remainder of the letter is organized as follows. Section II describes the proposed methodology, which is based on two main ingredients: KPCA and EMAPs. Section III describes our experimental results. Concluding remarks are given in Section IV.

II. Proposed Methodology

A. Kernel-Based Feature Extraction Using KPCA

PCA has been widely used for feature extraction in remote sensing image analysis [14]. The principal components (PCs) of a stochastic multivariate data are calculated based on a linear transformation, which produces uncorrelated bands of decreasing variance using the covariance matrix as a dispersion

Manuscript received November 23, 2012; revised March 1, 2013; accepted March 28, 2013. Date of publication June 6, 2013; date of current version November 8, 2013. This work was supported in part by the Icelandic Research Fund and the Spanish Ministry of Science and Innovation (CEOS-SPAIN project, reference AYA2011-29334-C02-02).

S. Bernabé and A. Plaza are with the Hyperspectral Computing Laboratory, University of Extremadura, Cáceres E-10003, Spain (e-mail: sergiobernabe@unex.es; aplaza@unex.es).

P. R. Marpu is with the Masdar Institute of Science and Technology, P.O. Box 54224—Abu Dhabi, UAE (e-mail: prashanthmarpu@ieee.org).

M. Dalla Mura is with GIPSA-Lab, Grenoble Institute of Technology, Grenoble BP 46-38402, France (e-mail: mauro.dalla-mura@gipsa-lab.grenoble-inp.fr).

S. Bernabé and J. A. Benediktsson are with the Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavik 101, Iceland (e-mail: sbg30@hi.is; 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/LGRS.2013.2256336

1545–598X © 2013 IEEE

matrix. The PCs are calculated so as to maximize the variance in every component. PC components are obtained as the linear combination of the original bands with maximum variance subject to the constraint that it is uncorrelated with all the other components. For an image of n spectral bands represented as a random vector $\mathbf{G}$ with zero mean, the covariance matrix can be simply calculated as

$$\Sigma = \left\langle \mathbf{G}\mathbf{G}^T \right\rangle. \tag{1}$$

Normally, only the first few principal components account for most of the variance in the data. This fact is used to reduce the dimensionality of the data by considering only the first few principal components, which account for most of the variance.

In turn, the kernel formulation of PCA [11] is obtained by replacing the inner products in the Gram matrix $\left\langle \mathbf{G}\mathbf{G}^T \right\rangle$ with the kernel functions $K(\mathbf{x}_i, \mathbf{x}_j)$, where $\mathbf{x}_i$ and $\mathbf{x}_j$ represent data observations. The kernel functions represent the inner products of vectors $\Phi(\mathbf{x}_i)$ and $\Phi(\mathbf{x}_j)$, which are the nonlinear mappings of $\mathbf{x}_i$ and $\mathbf{x}_j$ into a higher dimensional feature space. When there is a large number of observations, as is the case for remote sensing images, these observations are generally subsampled to perform KPCA. Even that will lead to a high number of features, so only the top few features that account for most of the variance are considered, as mentioned above. Since the data are projected onto a higher dimensional feature space, the clusters may now be easier to discriminate in the kernel feature space [10]. This is the reason why KPCA is chosen as a preprocessing step to decompose the spectral information in the kernel feature space, so that the spectral clusters will have less variance in the obtained features, and may be easier to separate in the kernel feature space.

B. Spatial Characterization Using EMAPs

EMAPs [9] are an extension of APs [8] obtained using different types of attributes and stacked together. The filtering operation implemented in APs is based on the evaluation of how a generic attribute $\mathcal{A}$, computed for every connected components of a scalar image, compares to a given reference value λ in a binary predicate $\mathcal{P}$ (e.g., $\mathcal{P} := \mathcal{A}(C_i) > \lambda$, with C_i being the ith connected component of the image). If $\mathcal{P}$ holds true 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 ordered threshold values $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$, an AP is obtained by applying a sequence of attribute thinning and attribute thickening operations as indicated in (2), where ϕ_i and γ_i, respectively, denote the thickening and thinning transformations, and $f_j(\mathbf{x}_i)$ denotes a feature extracted from the original observation $\mathbf{x}_i$.

In [2], it was suggested to use several PCs of original data to address this issue. In this way, APs are built on each of the first q PCs. This leads to the following definition of the extended attribute profile (EAP) for the observation $\mathbf{x}_i$

$$\begin{aligned} \mathrm{AP}(f_j(\mathbf{x}_i)) \quad := \quad & \big\{\phi_n(f_j(\mathbf{x}_i)), \ldots, \phi_1(f_j(\mathbf{x}_i)), f_j(\mathbf{x}_i), \\ & \gamma_1(f_j(\mathbf{x}_i)), \ldots, \gamma_n(f_j(\mathbf{x}_i))\big\} \end{aligned} \tag{2}$$

$$\mathrm{EAP}(\mathbf{x}_i) := \big\{\mathrm{AP}(f_1(\mathbf{x}_i)), \mathrm{AP}(f_2(\mathbf{x}_i)), \ldots, \mathrm{AP}(f_q(\mathbf{x}_i))\big\} \tag{3}$$

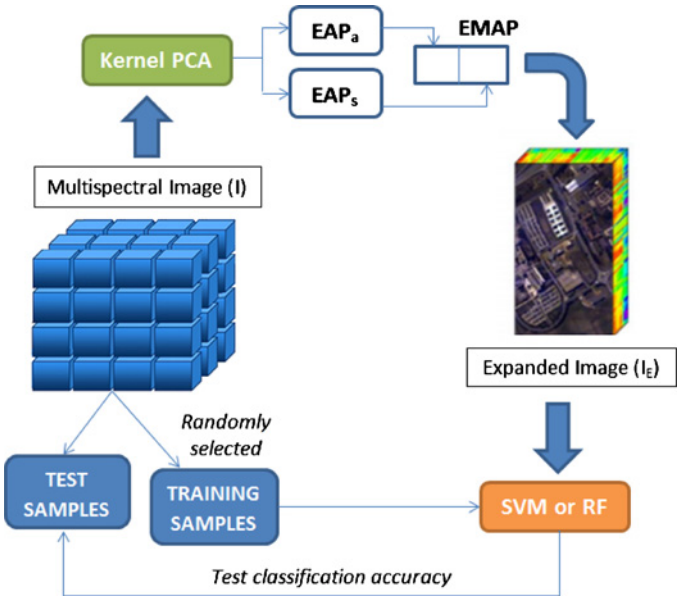

Fig. 1. Proposed framework for multispectral image classification.

where q is the number of retained features. From the EAP definition in (3), the consideration of multiple attributes leads to the concept of EMAP, which improves the capability to extract the spatial characteristics of the structures in the scene. An increase in the dimensionality of the data is also obtained. In this letter, we use only two attributes: area and standard deviation of the pixel values. This specific choice was supported by previous works [10] in which these attributes were proved to effectively extract the spatial characteristics of the regions. However, it is important to note that any measure computed on image regions may be considered as an attribute and can be used depending on the application. We refer to [8] and [9] (and references therein) for additional details on APs and EMAPs, their implementation details and their computational complexity.

C. Multispectral Image Classification Framework

The proposed framework for multispectral image classification is summarized by the flowchart in Fig. 1. First, we use KPCA to extract the features and to expand the dimensionality of the original multispectral image with n bands. In this way, the contrast between different spectral clusters increases, but the variance within the cluster decreases in the corresponding KPCA component representative of the cluster. The second step is to build EMAPs based on the features derived using KPCA, exploiting the spatial information. This strategy will increase the dimensionality of the data. However, it embeds an intelligent combination of spatial and spectral information. Classification is finally performed on the obtained EMAPs by using a nonlinear classifier such as the SVM or RF. It has been observed in [10] and [15] that SVMs are more sensitive than RFs to the Hughes phenomenon [14] when dealing with high dimensional data such as EMAPs, whereas RFs provide consistent and better performance. This is because RFs usually do not consider all the input features in the final model, which implicitly makes them, and thus the final ensemble, more

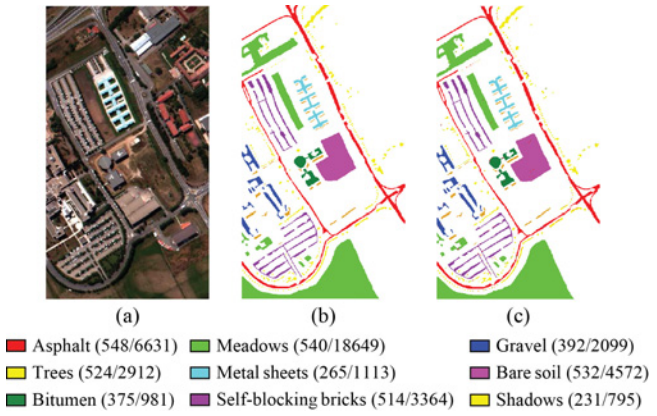

Fig. 2. (a) RGB composite of the hyperspectral ROSIS Pavia scene. (b) Reference map for the ROSIS Pavia University data with nine reference classes. (c) Classification result with EMAP (KPCA) using 5% training and the RF classifier: 98.81% accuracy. The numbers in the parentheses represent the number of (training/test) pixels available for each class.

robust to high dimensional input data. Both SVMs and RFs are used in this letter for comparative purposes.

III. Experimental Results

In this section, we present a quantitative and comparative assessment of the proposed methodology using multispectral data. The main goal is to incorporate spectral and spatial information in an effective way to improve the multispectral image classification results. The combination of feature extraction using KPCA and spatial characterization using EMAPs provides a processing approach that has not been explored in previous contributions focused on multispectral data. Before describing the results obtained in our experiments, we first describe the data sets and experimental setup considered in our experiments. Then, we continue with the discussion of the classification results obtained for the different approaches compared in this letter.

A. Data Set Description

The experimental analysis was carried using three multispectral images, two of them are obtained by extracting the red, green, and blue (RGB) bands from two well known hyperspectral images. The reasons for choosing the RGB images from hyperspectral images are twofold. First, this allows us to use the highly reliable reference data available for these scenes. Second, this also allows us to compare the results obtained with the original hyperspectral data with much higher spectral resolution. The third image considered in experiments is a multispectral image acquired using the IKONOS instrument. This scene is used to validate the presented methodology with real multispectral data.

1) The first image used in experiments was derived from a hyperspectral image acquired using the reflective optics system imaging spectrometer (ROSIS) sensor over the University of Pavia, Italy. For this scene, with 610×340 pixels and spatial resolution of 1.3 m/pixel, we have selected three spectral bands (12, 26, and 51) corresponding to the RGB channels, out of the 103 spectral channels (covering the wavelength range from 0.4 to 0.9 μm available in the original image. Fig. 2(b) shows the reference map available for the scene, which comprises urban features as well as soil and vegetation features.
2) The second image used in experiments was derived from a hyperspectral image acquired using the airborne visible infrared imaging spectrometer (AVIRIS) sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene, with a size of 145×145 pixels and 202 spectral bands in the range from 0.4 to 2.5 μm, was acquired over a mixed agricultural/forest area, early in the growing season. In this scene, we have selected three spectral bands (5, 12, and 24) corresponding to the RGB data channels, to form a multispectral (color) image. The spatial resolution of the scene is 20 m/pixel. Fig. 3(b) shows the reference map available for this scene.[1]
3) The third image used in experiments was derived from a multispectral image acquired by the IKONOS satellite on August 9, 2001, over a urban area of Reykjavik, Iceland. This scene is made up of 975×637 pixels and four spectral bands and the spatial resolution is 1 m/pixel. Fig. 4(b) shows the reference map available for the scene. This data set is used to evaluate the

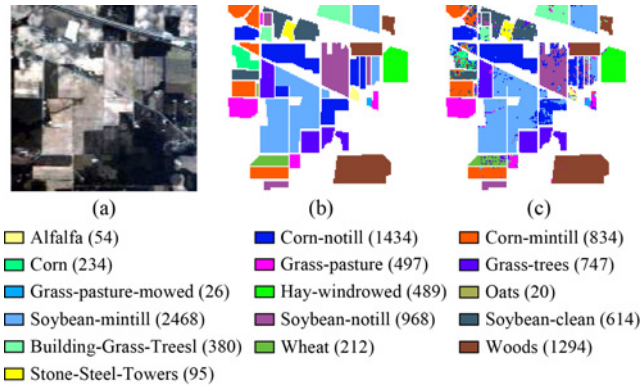

Fig. 3. (a) RGB composite of the hyperspectral AVIRIS Indian Pines scene. (b) Reference map for the AVIRIS Indian Pines data with sixteen reference classes. (c) Classification result with EMAP (KPCA) using 5% training and the RF classifier: 88.74% accuracy.

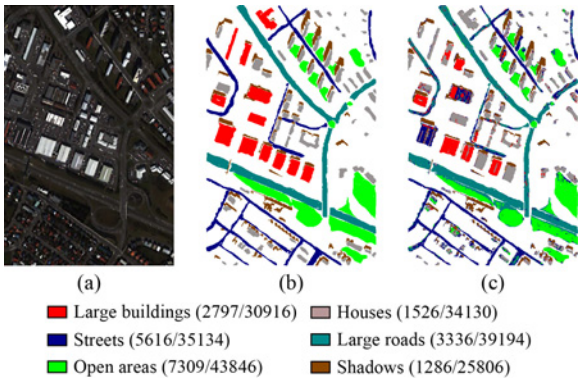

Fig. 4. (a) RGB composite of the Multispectral IKONOS Reykjavik scene. (b) Reference map for the IKONOS Reykjavik data with six reference classes. (c) Classification result with EMAP (KPCA) using the fixed training and the SVM classifier: 72.02% accuracy.

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

TABLE I

OVERALL ACCURACY (%) AND STANDARD DEVIATION (AFTER 10 MONTE CARLO RUNS) OBTAINED AFTER APPLYING THE SVM CLASSIFIER TO THE ROSIS PAVIA UNIVERSITY, THE AVIRIS INDIAN PINES, AND THE IKONOS REYKJAVIK DATA SETS USING DIFFERENT TYPES OF FEATURES

Overall Accuracy	**ROSIS Pavia University**			**AVIRIS Indian Pines**			**IKONOS Reykjavik**
Features	**Standard training set**	**Subset of 50 pixels**	**5% Training**	**5% Training**	**10% Training**	**15% Training**	**Standard training set**
Hyperspectral	80.64% ± 0.00	84.09% ± 2.06	93.38% ± 0.21	75.60% ± 1.15	81.95% ± 0.54	84.46% ± 0.45	–
Multispectral	65.81% ± 0.00	62.67% ± 5.62	78.87% ± 0.40	48.33% ± 0.71	49.63% ± 0.29	49.86% ± 0.38	58.73% ± 0.00
EMAP	76.85% ± 0.00	91.87% ± 0.63	96.89% ± 0.23	**79.63% ± 0.73**	**83.71% ± 0.47**	**85.89% ± 0.44**	65.40% ± 0.00
$KPCA_{\sigma=1.0}$	66.13% ± 0.00	65.60% ± 4.26	78.92% ± 0.26	47.97% ± 1.20	49.50% ± 0.46	49.89% ± 0.17	59.31% ± 0.00
$KPCA_{\sigma=1.5}$	66.01% ± 0.00	63.93% ± 4.61	78.62% ± 0.38	48.83% ± 0.38	49.59% ± 0.38	49.95% ± 0.38	61.33% ± 0.00
$KPCA_{\sigma=2.0}$	67.94% ± 0.00	62.86% ± 3.71	78.61% ± 0.32	48.25% ± 0.62	49.25% ± 0.46	49.70% ± 0.30	61.01% ± 0.00
EMAP ($KPCA_{\sigma=1.0}$)	93.41% ± 0.00	**93.22% ± 0.71**	97.18% ± 0.64	62.75% ± 3.72	74.45% ± 0.63	77.32% ± 2.90	70.55% ± 0.00
EMAP ($KPCA_{\sigma=1.5}$)	**94.09% ± 0.00**	90.61% ± 2.38	97.18% ± 0.62	63.86% ± 2.74	73.91% ± 2.16	77.34% ± 3.14	71.88% ± 0.00
EMAP ($KPCA_{\sigma=2.0}$)	91.87% ± 0.00	91.61% ± 1.36	**97.25% ± 0.63**	61.40% ± 3.14	71.62% ± 3.68	77.86% ± 2.30	**72.02% ± 0.00**

TABLE II

OVERALL ACCURACY (%) AND STANDARD DEVIATION (AFTER 10 MONTE CARLO RUNS) OBTAINED AFTER APPLYING THE RF CLASSIFIER TO THE ROSIS PAVIA UNIVERSITY, THE AVIRIS INDIAN PINES, AND THE IKONOS REYKJAVIK DATA SETS USING DIFFERENT TYPES OF FEATURES

Overall Accuracy	**ROSIS Pavia University**			**AVIRIS Indian Pines**			**IKONOS Reykjavik**
Features	**Standard training set**	**Subset of 50 pixels**	**5% Training**	**5% Training**	**10% Training**	**15% Training**	**Standard training set**
Hyperspectral	71.42% ± 0.13	72.63% ± 2.66	87.66% ± 0.28	69.84% ± 1.30	75.38% ± 0.56	77.70% ± 0.51	–
Multispectral	64.87% ± 0.11	61.68% ± 2.62	77.08% ± 0.27	44.27% ± 0.54	46.17% ± 0.43	47.05% ± 0.28	61.38% ± 0.09
EMAP	80.81% ± 0.30	88.61% ± 1.38	96.76% ± 0.23	80.93% ± 1.07	84.78% ± 0.49	86.46% ± 0.37	62.07% ± 0.42
$KPCA_{\sigma=1.0}$	65.26% ± 0.11	65.32% ± 2.83	78.06% ± 0.22	44.62% ± 0.62	45.95% ± 0.52	46.30% ± 0.37	60.91% ± 0.07
$KPCA_{\sigma=1.5}$	65.44% ± 0.18	65.64% ± 2.39	78.48% ± 0.26	44.36% ± 0.62	45.83% ± 0.47	46.24% ± 0.53	61.67% ± 0.09
$KPCA_{\sigma=2.0}$	65.97% ± 0.14	65.50% ± 2.59	78.46% ± 0.24	44.95% ± 0.73	46.45% ± 0.47	46.81% ± 0.36	61.69% ± 0.08
EMAP ($KPCA_{\sigma=1.0}$)	92.45% ± 0.15	94.72% ± 0.88	98.61% ± 0.13	88.65% ± 0.63	**92.67% ± 0.59**	**94.25% ± 0.42**	**67.83% ± 0.33**
EMAP ($KPCA_{\sigma=1.5}$)	**94.23% ± 0.35**	94.92% ± 1.15	**98.81% ± 0.14**	**88.74% ± 0.64**	92.65% ± 0.40	93.96% ± 0.28	67.29% ± 0.62
EMAP ($KPCA_{\sigma=2.0}$)	93.96% ± 0.57	**95.05% ± 1.09**	98.67% ± 0.14	88.25% ± 0.53	92.26% ± 0.61	93.90% ± 0.33	67.76% ± 0.73

proposed methodology using a real multispectral data set with standard broad bands (as opposed to the other two data sets, which are obtained from hyperspectral scenes with narrow spectral bands).

B. Experimental Setup

For the KPCA stage, we used the Gaussian radial basis function (RBF) kernel: $K(\mathbf{x}_i, \mathbf{x}_j) := \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2)$ kernel, which is widely used in hyperspectral image classification. Three different values were chosen for parameter σ in order to study the impact of this parameter on the results. The values chosen were $\sigma = \{1.0, 1.5, 2.0\}$ times the mean value of the mutual distances between 2000 randomly chosen pixels. The corresponding KPCA features are referred to as $KPCA_{\sigma=1.0}$, $KPCA_{\sigma=1.5}$, and $KPCA_{\sigma=2.0}$, respectively.

The EMAPs are built using the area (related to the size of the regions) and standard deviation (which measures the homogeneity of the pixels enclosed by the regions) attributes. The threshold values λ were chosen in the range $\{50, 500\}$ with a stepwise increment of 50 for the area attribute. For the standard deviation, attribute values ranging from 2.5% to 20% of the mean of the feature with a stepwise increment of 2.5% were chosen [15]. These values are selected to accommodate the possible characterization of connected components of the classes of interest.

Then, a supervised classification process was performed using both the SVM classifier (with the RBF kernel) and the RF classifier. The parameters of the SVM were tuned by using fivefold cross-validation, while RF does not require any parameter tuning. This is because every individual decision tree is built to over-fit, which means that no parameters are required to facilitate the pruning of the trees. Classification was performed using three different configurations of the training process in each dataset.

1) For the ROSIS Pavia University data set, we considered: 1) A standard training set widely used in the state-of-the-art (referred to hereinafter as *standard training set*); 2) a subset of the entire reference set in Fig. 2(b), where 50 pixels are randomly sampled for each reference class; and 3) a subset of the entire reference set, where 5% of the pixels in each class are randomly sampled.
2) For the AVIRIS Indian Pines data, we randomly sampled 5%, 10%, and 15% of the reference data in Fig. 3(b).
3) For the IKONOS Reykjavik data, a subset of the entire reference set shown in Fig. 4(b) was considered.

KPCA was used to expand the dimensionality of the multispectral data prior to the construction of the EMAPs. Then, only the top 20 features resulting from KPCA (which in all cases comprise more than 99% of the data variance) were used, and the final obtained EMAPs consisted of 740 features [10 levels of filtering using area attribute resulting in 20 features and eight levels of filtering for standard deviation attribute resulting in 16 features for each KPCA component, so 20 KPCA components × (20 area attribute features + 16 standard deviation attribute features + 1 KPCA component) = 740 total features].

C. Analysis and Discussion of Results

Table I shows the overall accuracy (in percentage) and the standard deviations (each reported value of accuracy is an average of 10 Monte Carlo runs) obtained by the SVM classifier applied to different feature extraction methods for

the two considered scenes. Here, we compare the SVM classification results using the multispectral (or RGB) scenes, the features resulting from dimensionality expansion of the multispectral/RGB scenes using KPCA (with different values for parameter σ in the RBF kernel), and the features resulting from applying EMAPs to the first 20 features resulting from KPCA dimensionality expansion of the multispectral data. Similarly, Table II shows the classification results obtained by the RF classifier. These results can be compared to those obtained after applying EMAPs to the original hyperspectral data in [10], or after applying a composite kernel-based approach in [16].

An interesting observation from Tables I and II is that the EMAP built on the multispectral/RGB data generally provides good results in terms of classification accuracy (even higher than with the original hyperspectral data in the ROSIS Pavia and AVIRIS Indian Pines scenes). This indicates the importance of including spatial information in the analysis. In turn, as expected the classification accuracies obtained using the multispectral/RGB data alone are always sensibly lower. On the other hand, from Tables I and II it can also be seen that, when the KPCA is applied to the multispectral/RGB data, the obtained classification accuracies are not significantly improved. However, the combination of EMAPs and KPCA produces results that are generally better than those obtained by the EMAPs alone (with the exception of the SVM classifier in the AVIRIS Indian Pines data). This indicates that the extraction of spatial information using EMAPs is better suited when the contrast between spectral clusters is high. The combination of EMAPs and KPCA produced better results (particularly for the RF classifier).

Our experiments generally confirm the observations in [10] that SVMs may be more sensitive to the Hughes phenomenon than RFs. While both the SVM and RF classifiers provide similar results for the ROSIS Pavia University data and for the IKONOS Reykjavik data, Table I shows that the SVM provides inferior results in the case of AVIRIS Indian Pines with limited training samples (particularly when the high dimensional combination of EMAPs built on KPCA, resulting in 740 features, is used). In this case, the performance gradually increases as the size of the training set increases from 5% to 15%. However, this effect was not observed in the results of the RF classifier for the AVIRIS Indian Pines dataset, as shown in Table II.

IV. Conclusion and Future Research

In this letter, we developed a new methodology for spectral–spatial classification of remotely sensed multispectral images with limited spectral resolution. Our proposed method first expanded the dimensionality of multispectral data using kernel feature extraction, and then included spatial information by means of morphological characterization of the expanded set of features. Our experimental results indicated that the proposed approach is attractive for advanced classification of data sets with limited spectral resolution. Here, three-band RGB images were derived from the available hyperspectral data sets to create multispectral data sets. The classification accuracies obtained on such data sets were superior to those provided by EMAPs (built on the multispectral data), and even to those obtained using the full hyperspectral information with hundreds of spectral bands. This revealed the importance of spatial information and indicated that the combination of KPCA and EMAPs provides a simple yet powerful strategy to perform spectral–spatial classification of data sets with limited spectral resolution (in this letter, we considered both multispectral scenes and also RGB scenes derived from real hyperspectral data sets). In our future work, we plan to use higher level strategies to derive spatial features such as object-based image analysis and knowledge-based methods.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] B. Zhang, S. Li, X. Jia, L. Gao, and M. Peng, "Adaptive Markov random field approach for classification of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 973–977, Sep. 2011.

[6] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[7] D. Tuia, F. Pacifici, M. Kanevski, and W. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.

[8] M. Dalla 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.

[9] M. Dalla 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, 2010.

[10] P. R. Marpu, M. Pedergnana, M. Dalla Mura, S. Peeters, 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, 2012.

[11] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, no. 5, pp. 1299–1319, 1998.

[12] B. Schölkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. The MIT Press, 2002.

[13] L. Breiman, "Random forests," *Mach. Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[14] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. Berlin, Germany: Springer, 2006.

[15] P. R. Marpu, M. Pedergnana, M. Dalla Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral–spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, Mar. 2013.

[16] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.