Fusion of Hyperspectral and LiDAR Remote Sensing Data Using Multiple Feature Learning

Mahdi Khodadadzadeh, *Student Member, IEEE*, Jun Li, *Member, IEEE*, Saurabh Prasad, *Senior Member, IEEE*, and Antonio Plaza, *Fellow, IEEE*

Abstract—Hyperspectral image classification has been an active topic of research. In recent years, it has been found that light detection and ranging (LiDAR) data provide a source of complementary information that can greatly assist in the classification of hyperspectral data, in particular when it is difficult to separate complex classes. This is because, in addition to the spatial and the spectral information provided by hyperspectral data, LiDAR can provide very valuable information about the height of the surveyed area that can help with the discrimination of classes and their separability. In the past, several efforts have been investigated for fusion of hyperspectral and LiDAR data, with some efforts driven by the morphological information that can be derived from both data sources. However, a main challenge for the learning approaches is how to exploit the information coming from multiple features. Specifically, it has been found that simple concatenation or stacking of features such as morphological attribute profiles (APs) may contain redundant information. In addition, a significant increase in the number of features may lead to very high-dimensional input features. This is in contrast with the limited number of training samples often available in remote-sensing applications, which may lead to the Hughes effect. In this work, we develop a new efficient strategy for fusion and classification of hyperspectral and LiDAR data. Our approach has been designed to integrate multiple types of features extracted from these data. An important characteristic of the presented approach is that it does not require any regularization parameters, so that different types of features can be efficiently exploited and integrated in a collaborative and flexible way. Our experimental results, conducted using a hyperspectral image and a LiDAR-derived digital surface model (DSM) collected over the University of Houston campus and the neighboring urban area, indicate that the proposed framework for multiple feature learning provides state-of-the-art classification results.

Index Terms—Digital surface model (DSM), hyperspectral, light detection and ranging (LiDAR), multiple feature learning.

I. Introduction

HYPERSPECTRAL imaging is concerned with the extraction of information from objects or scenes lying on the Earth surface, using hundreds of (narrow) spectral bands typically covering the visible and near infra-red domains [1]. In hyperspectral imaging, also termed imaging spectroscopy [2], the sensor acquires a spectral vector with hundreds or thousands of elements from every pixel in a given scene. The result is the so-called hyperspectral image or hyperspectral data *cube*. It should be noted that hyperspectral images are spectrally smooth and spatially piece-wise smooth; this means that the values in neighboring locations and wavelengths are often highly correlated [3].

Hyperspectral image classification has been a very active area of research in recent years [4]. Given a set of observations (i.e., pixel vectors in a hyperspectral image), the goal of classification is to assign a unique label to each pixel vector so that it is well defined by a given class. The wider availability of hyperspectral data with high spatial resolution has been quite important for classification techniques. However, in some cases, the spatial resolution of the hyperspectral data is not enough to separate complex classes such as those present in urban environments [4]. This aspect, together with the expected (linear or nonlinear) mixing happening at subpixel scales [5], complicates the classification process significantly. In some cases, the data coming from other sources can be used to improve and/or refine the results of classification. A good example is the use of light detection and ranging (LiDAR) data [6], which can provide information about the height of the same surveyed area. LiDAR has been shown to be a very useful source of data for classification purposes [7].

In the literature, many techniques have been developed for fusion of hyperspectral and LiDAR data for classification purposes [8]. In 2013, the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) organized a contest[1] involving two data sets: a hyperspectral image and a LiDAR derived digital surface model (DSM), both at the same spatial resolution (2.5 m), and two parallel competitions were established to devise advanced methods for fusion and classification of hyperspectral and LiDAR data.[2] Many other examples can be found in classification of urban areas [9], but also in classification of complex forest areas [10]. Techniques based on morphological features have been quite successful in the literature. For instance, the methodology in [11] jointly considered the features extracted by morphological attribute profiles (APs) [12] computed on both the hyperspectral and LiDAR data, and then fused the spectral, spatial, and elevation information in a stacked architecture. In [13], it was pointed out that

[1][Online]. Available: http://hyperspectral.ee.uh.edu/?page_id=459
[2][Online]. Available: http://hyperspectral.ee.uh.edu/?page_id=795

the simple concatenation or stacking of features such as morphological APs may contain redundant information. The main challenge in multiple feature learning is that how to adequately exploit the information containing in these features. In addition, a significant increase in the number of features may lead to high dimensionality issues that are in contrast with the limited number of training samples often available in remote-sensing applications [14], which may lead to the Hughes effect. To address these issues, decision fusion techniques have been applied [15].

In this paper, we develop a new strategy for fusing hyperspectral and LiDAR data for classification purposes. The main contribution of our newly proposed approach is its capacity to integrate multiple types of features extracted using spatial and spectral information. For the LiDAR data, the DSM provides rich information in spatial sense, while for the hyperspectral data, the very rich spectral resolution provides detailed spectral signatures that can be very useful for classification purposes. A main characteristic of the presented approach is that it can adaptively exploit information from both spatially and spectrally derived features, thus being able to address practical scenarios in which different sources of information (spatial or spectral) may be useful to separate different types of classes. In order to achieve this goal, the proposed approach has been designed in a way that it exhibits great flexibility to combine different types of features without any regularization parameters, thus taking advantage of the complementarity that the features can provide without any *a priori* restrictions. Our presented approach is thus aimed at exploiting the different properties that both spatial and spectral features can provide. In order to achieve the desired spectral–spatial integration that is normally expected in advanced classification problems, we consider morphological features as an important part of our framework, which also exploits the original spectral information contained in the hyperspectral scene. The integration is achieved by a multiple feature learning approach based on the subspace multinomial logistic regression (MLRsub [16]) classifier.

The remainder of this paper is organized as follows. Section II describes the methodological framework developed in this work for the fusion of hyperspectral and LiDAR data. Section III describes our experimental results, conducted using a hyperspectral image and a LiDAR derived DSM collected over the University of Houston campus and the neighboring urban area. This scene was recently used in the IEEE GRSS contest and, therefore, the results obtained for this scene will be compared with other several approaches already tested with these data. Finally, Section IV concludes this paper with some remarks and hints at plausible future research lines.

II. METHODOLOGICAL FRAMEWORK

In this section, we introduce the proposed approach for the integration of the hyperspectral and LiDAR data. The proposed approach comprises the following main steps. In the first step, we use morphological APs [12] to extract the spatial features in the hyperspectral and LiDAR data. Then, in the second step, we perform classification over all the obtained spatial features and the original spectral features by using the MLRsub classifier [16]. We have selected this classifier as it

provides great flexibility for multiple feature learning based on the logarithmic opinion pool (LOGP) rule [17], while offering great ability for learning ill-posed problems by projecting the data into its class-indexed subspace. In this way, by working in a subspace, the proposed method can adequately handle the unbalance between the increased dimensionality of the data (expanded even more after including the morphological APs) and the very limited availability of training samples in practice. In a final step, we use Markov random fields (MRFs) [18] for spatial regularization to promote spatial smoothness in the final classification result, as indicated in [16]. In the following, we present in detail each step of the proposed approach for classification of hyperspectral and LiDAR data.

A. Feature Extraction by Morphological APs

Let $\mathbf{X}^L \equiv (x_1^L, x_2^L, \ldots, x_n^L)$ be the DSM derived from the LiDAR data, where n is the number of pixels in $\mathbf{X}^L$. Similarly, let us denote the hyperspectral image as $\mathbf{X}^h \equiv (\mathbf{x}_1^h, \mathbf{x}_2^h, \ldots, \mathbf{x}_n^h)$, where $\mathbf{x}_i \in \mathbb{R}^d$, for $i = 1, 2, \ldots, n$, denotes a spectral vector, n is the number of pixels in $\mathbf{X}^h$, and d is the number of spectral bands.

In order to perform feature extraction from the hyperspectral and LiDAR data, we use mathematical morphology (MM) [19], a widely used approach for modeling the spatial characteristics of the objects in remotely sensed images. Using a structural element (SE) of predefined size and shape, morphological profile (MP) of a gray-scale image can be generated by applying morphological opening or closing operations. For example, by applying an opening operation on the image, the objects smaller than the size of the SE are removed. Hence, MP is built by changing the size of the SE and repeating the opening and closing operations. Advanced MM techniques such as derivative morphological profiles (DMPs) [20] have been successfully used for multispectral image classification by processing the panchromatic band of these instruments. This strategy has also been extended to hyperspectral image classification, by extracting the first few principal components of the data using, for instance, principal component analysis (PCA) [21] and then building so-called extended morphological profiles (EMPs) on the first few PCA components to extract relevant features for classification [22]. Since redundant information is generally present in DMPs and EMPs with high dimensionality, feature extraction and selection techniques have been used to extract the most relevant information prior to classification [23]. Recently, morphological APs [24]–[26] were introduced as an advanced mechanism to obtain a detailed multilevel characterization created by the sequential application of morphological attribute filters that can be used (prior to classification) to model different kinds of the structural information. Similar to generating of MPs, in the procedure of construction of APs, if the measured attribute of an object (e.g., area, volume, standard deviation, etc.) meets a predefined condition λ, then that object is kept unchanged; otherwise, it is merged to the most similar surrounding object. Hence, the merging operation to the adjacent object of a lower and higher gray level are called thinning and thickening, respectively. Fig. 1 shows an example of generating AP based on attribute of area which
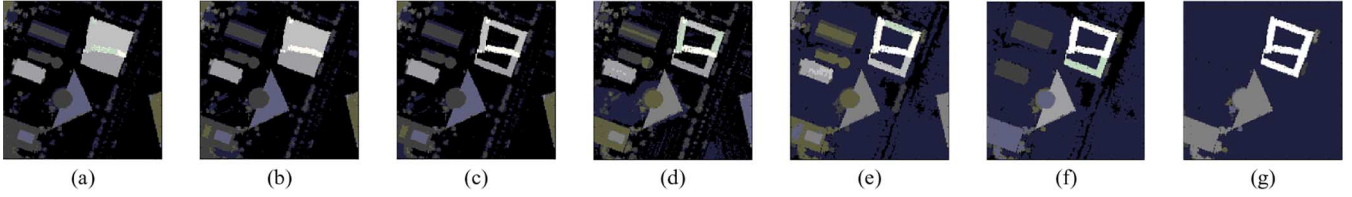
Fig. 1. Example of an AP using attribute of area with conditions $\lambda_1 = 100$, $\lambda_2 = 500$, and $\lambda_3 = 1000$. (a)–(c) AP is composed of thickening images. (d) Original image. (e)–(g) Thinning images.

can provide information about the size and the shape of objects in the image. Extended attribute profiles (EAPs) are typically obtained by generating an AP on each of the first few PCA components (or any other features retained after applying feature selection on the original hyperspectral image), thus building a stacked vector using the AP on each feature, as illustrated in Fig. 2. From the EAP definition, the consideration of multiple attributes leads to the concept of extended multiattribute profile (EMAP) [27] which combines the EAPs by concatenating them in a single vector of features and improves the capability in extracting the spatial characteristics of the structures in the scene, where attribute filtering can be efficiently computed by applying a Max-tree algorithm [28].

In our specific context, the multiple features available $\mathcal{X}$ are defined as follows: $\mathcal{X} \equiv (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3, \tilde{\mathbf{X}}_4)$, where $\tilde{\mathbf{X}}_1 = \mathbf{X}^h$ (i.e., the hyperspectral data); $\tilde{\mathbf{X}}_2 = \mathbf{X}^L$ (i.e., the LiDAR data); $\tilde{\mathbf{X}}_3 = \text{EMAP}(\mathbf{X}^h)$ (i.e., an EMAP built on the first few PCA components of the original hyperspectal data); and $\tilde{\mathbf{X}}_4 = \text{AP}(\mathbf{X}^L)$ (i.e., the AP of the LiDAR data). Notice that, for the LiDAR data, we could only generate its APs as it is a single-channel DSM image.

B. Probabilistic Classification Using Spectral Features

Let $\mathbf{Y} \equiv (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ denote an image of labels, $\mathbf{y}_i \equiv [y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(k)}]^T$, where k is the number of classes. For $c = 1, \ldots, k$, if pixel i belongs to class c, $y_i^{(c)} = 1$, otherwise, $y_i^{(c)} = 0$. With these definitions in place, let $p_m(y_i^{(c)} = 1 | (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)$ be the posterior density associated with feature $\tilde{\mathbf{X}}_m$, for $m = 1, 2, 3, 4$, where $\boldsymbol{\omega}_m$ is the parameter associated with the considered classifier.

In the literature, various probabilistic techniques have been suggested for classification of remote-sensing data. In this paper, we focus on two techniques: the support vector machine (SVM) [29]–[31] and MLRsub [16], [32], which have shown good performance in hyperspectral data classification.

1) SVM Classification Technique: The SVM is originally a binary hard classifier that separates two classes by using a linear hyperplane. For finding the separating hyperplane, the training samples located near the class boundaries (i.e., the support vectors) that maximize the separation between the two classes are found, by using a constrained optimization process [29]. An important advantage of SVM algorithm is the possibility of using a kernel trick in the formulation [33]. In this paper, we use the multiclass probabilistic SVM method [34], [35] implemented in the popular LIBSVM library [36], considering Gaussian radial basis function kernel $K((\tilde{\mathbf{x}}_i)_m, (\tilde{\mathbf{x}}_j)_m) = \exp(-\gamma \|(\tilde{\mathbf{x}}_i)_m - (\tilde{\mathbf{x}}_j)_m\|^2)$.
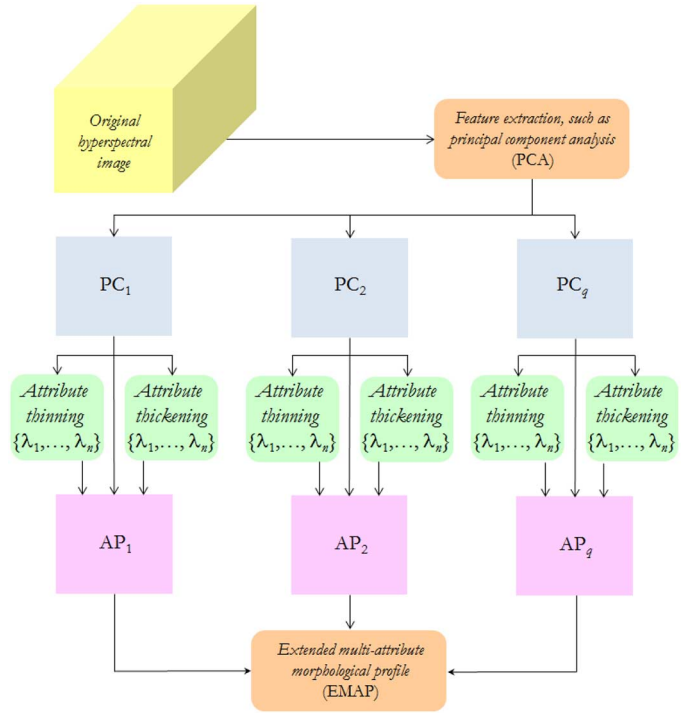

Fig. 2. Graphical illustration of the procedure adopted to construct an EAP from a hyperspectral image. The EMAP is a combination of EAPs obtained with different attributes.

2) MLRsub Classification Technique: Using the notations in this paper, the MLR classifier is given by

$$p_m(y_i^{(c)} = 1 | (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m) = \frac{\exp\left(\boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^{k} \exp\left(\boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)} \quad (1)$$

where $\mathbf{h}((\tilde{\mathbf{x}}_i)_m) \equiv [h_1((\tilde{\mathbf{x}}_i)_m), \ldots, h_l((\tilde{\mathbf{x}}_i)_m)]^T$ is a vector of l fixed functions of the input data, often termed as features, $\boldsymbol{\omega}_m^{(c)}$ is the set of logistic regressors for class c, and $\boldsymbol{\omega}_m \equiv [\boldsymbol{\omega}_m^{(1)^T}, \ldots, \boldsymbol{\omega}_m^{(k)^T}]^T$. The MLRsub combines the classic MLR formulation with a subspace projection method to cope with highly mixed hyperspectral data using limited training samples. The idea of applying subspace projection methods to improve classification relies on the basic assumption that the samples within each class can approximately lie in a lower dimensional subspace. Thus, each class may be represented by a subspace spanned by a set of basis vectors, while the classification criterion for a new input sample would be the distance from the class subspace [16]. In [32], a modified version of MLRsub is
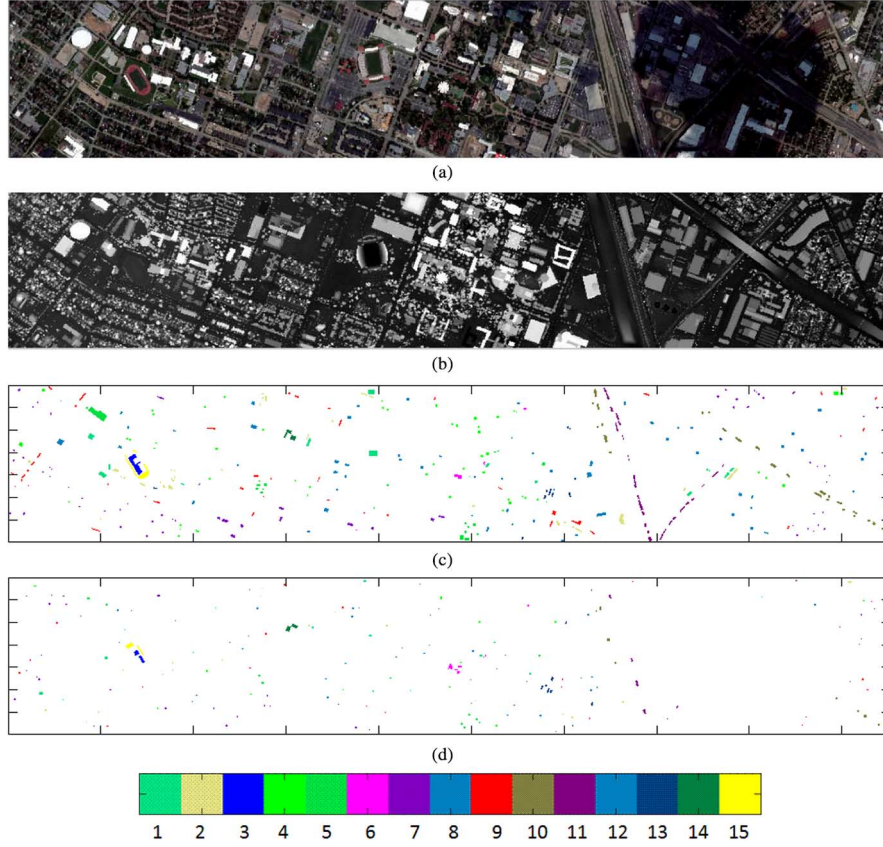
Fig. 3. (a) False color composition of the hyperspectral Houston data. (b) LiDAR-derived DSM for the Houston data. (c) Ground truth available for the considered hyperspectral/LiDAR Houston data. (d) Training set used in our experiments. The class labels are as follows: 1, Grass healthy; 2, Grass stressed; 3, Grass synthetic; 4, Tree; 5, Soil; 6, Water; 7, Residential; 8, Commercial; 9, Road; 10, Highway; 11, Railway; 12, Parking lot; 13, Parking lot; 14, Tennis court; and 15, Running track.

proposed for handling both linear and nonlinear mixtures happening at subpixel level in the data, as it is also common to have nonlinear mixtures in real data set. In this work, we also take nonlinearity into account so that the input function $\mathbf{h}(\tilde{\mathbf{x}}_i)$ in (1) is given by

$$\mathbf{h}((\tilde{\mathbf{x}}_i)_m) = [\|(\tilde{\mathbf{x}}_i)_m\|^2, \|(\tilde{\mathbf{x}}_i)_m^{\mathrm{T}}\mathbf{U}_m^{(1)}\|^2, \dots, \|(\tilde{\mathbf{x}}_i)_m^{\mathrm{T}}\mathbf{U}_m^{(k)}\|^2]^{\mathrm{T}} \quad (2)$$

where $\mathbf{U}_m^{(c)} = \{(\mathbf{u}_1^{(c)})_m, \dots, (\mathbf{u}_{r^{(c)}}^{(c)})m\}$ is a set of $r^{(c)}$-dimensional orthonormal basis vectors for the subspace associated with class c ($r^{(c)} \ll d_m$, and d_m is the dimensionality of feature $\tilde{\mathbf{X}}_m$.

C. Probabilistic Classification Using Multiple Spectral and Spatial Features

According to the LOGP rule [17], which is a decision fusion scheme that is commonly applied to combine information from multiple features, for any pixel $i = 1, \dots, n$, we have

$$p_{\mathrm{LOGP}}(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_1, \dots, (\tilde{\mathbf{x}}_i)_4, \boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_4, \alpha_1, \dots, \alpha_4)$$
$$= \frac{\prod_{m=1}^4 p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)^{\alpha_m}}{\sum_{c=1}^k \prod_{m=1}^4 p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)^{\alpha_m}} \quad (3)$$

where $\{\alpha_m|0 \le \alpha_m \le 1, \sum_{m=1}^4 \alpha_m = 1\}$ is a tunable parameter which controls the impact of each feature vector on the final decision probability. Notice that, for the multiple feature learning problem in (3), we have two different types of parameters: 1) the logistic regressors $\boldsymbol{\omega}_m$ and 2) the weight parameter α_m. These parameters are respectively associated with the classifier and with the features. This leads to difficulties from the viewpoint of both their optimization and the overall computational cost. Learning the weight parameter α_m is generally the most difficult task in the LOGP framework.

In order to relax the difficulty of learning the two different types of parameters, in this work, we propose to use the MLRsub classifier in [16] and [32] to model the posterior density $p_m(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_m)$. This provides the possibility of learning the regressors and the weight parameters in combined fashion, so that we only need to learn one type of parameter. This greatly lightens the computational cost while relaxing the optimization problem.

Under the present setup, by embedding the MLRsub model in (1) into the LOGP framework for multiple feature learning in (3), we can now obtain

$$p_{\mathrm{LOGP}}(y_i^{(c)} = 1|(\tilde{\mathbf{x}}_i)_1, \dots, (\tilde{\mathbf{x}}_i)_m, \boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_4, \alpha_1, \dots, \alpha_4)$$
$$= \frac{\exp\left(\sum_{m=1}^4 \alpha_m \boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^k \exp\left(\sum_{m=1}^4 \alpha_m \boldsymbol{\omega}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}. \quad (4)$$

TABLE I
INFORMATION CLASSES AND TRAINING-TEST SAMPLES

Class		Samples	
Number	Name	Train	Test
1	Healthy grass	198	1053
2	Stressed grass	190	1064
3	Synthetic grass	192	505
4	Trees	188	1056
5	Soil	186	1056
6	Water	182	143
7	Residential	196	1072
8	Commercial	191	1053
9	Road	193	1059
10	Highway	191	1036
11	Railway	181	1054
12	Parking lot 1	192	1041
13	Parking lot 2	184	285
14	Tennis court	181	247
15	Running track	187	473
Total		2832	12 197

TABLE II
COORDINATES

North	West	South	East
29°43′29.20″	95°22′10.60″	29°43′4.09″	95°19′12.91″

Notice that, in (4), we still have two different types of parameters α_m and ω_m. In general, to learn the joint density (3) or (4), learning of the weight parameter α_m is essential. In [17], several strategies were provided to learn the weight parameters. However, as shown in [17] and also as it was mentioned before, it is very time consuming and difficult to search for the optimal value of α_m. However, the MLR*sub* has the potential to overcome these difficulties as it provides a different look to the problem by associating the weight with the regressors instead of the features, i.e., by letting

$$\tilde{\boldsymbol{\omega}}_m^{(c)} = \alpha_m \boldsymbol{\omega}_m^{(c)} \qquad (5)$$

the regressors $\boldsymbol{\omega}_m$ and weight parameters α_m are now combined into a new set of regressors $\tilde{\boldsymbol{\omega}}_m$ associated with the MLR*sub* classifier. In this way, we avoid the weight parameters for the features by introducing them into the regressors. This means that, under this transformation, the weight parameters for the features will be learnt together with the classifier. This provides important advantages from the viewpoint of learning: by transferring the weight parameter used for the features (observations) to the weight parameter used for the classifier, the

proposed approach joins two different problems (parameter and classifier learning) into a single one, which represents a main contribution of this work from the viewpoint of simplifying and making more natural the process of learning from multiple features (i.e., those derived by the hyperspectral and the LiDAR data). By introducing the transformation (5), problem (4) changes to

$$p_{\text{LOGP}}(y_i^{(c)} = 1 | (\tilde{\mathbf{x}}_i)_1, \ldots, (\tilde{\mathbf{x}}_i)_m, \tilde{\boldsymbol{\omega}}_1, \ldots, \tilde{\boldsymbol{\omega}}_4)$$
$$= \frac{\exp\left(\sum_{m=1}^4 \tilde{\boldsymbol{\omega}}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}{\sum_{c=1}^k \exp\left(\sum_{m=1}^4 \tilde{\boldsymbol{\omega}}_m^{(c)} \mathbf{h}((\tilde{\mathbf{x}}_i)_m)\right)}. \qquad (6)$$

Notice that problem (6) is able to handle multiple features with the advantage of having a similar structure as the original MLR*sub* method. Therefore, problem (6) can be solved using the same approach presented in [16], which provides a detailed optimization framework for learning the logistic regressors. As the goal of this work is to adequately exploit the information coming from hyperspectral and LiDAR data, the proposed framework in (6) provides a great flexility in handling features from these two different sources of information, thus being able to naturally integrate spatial and spectral features without the need for *ad hoc* weights or regularization parameters. Although the proposed framework has the ability to integrate multiple types of features, in this work, we constrain ourselves to the set $\mathcal{X} \equiv (\tilde{\mathbf{X}}_1, \tilde{\mathbf{X}}_2, \tilde{\mathbf{X}}_3, \tilde{\mathbf{X}}_4)$ with $\tilde{\mathbf{X}}_1 = \mathbf{X}^h$, $\tilde{\mathbf{X}}_2 = \mathbf{X}^L$, $\tilde{\mathbf{X}}_3 = \text{EMAP}(\mathbf{X}^h)$, and $\tilde{\mathbf{X}}_4 = \text{AP}(\mathbf{X}^L)$ as described in the previous subsection, for the validation of our approach (although additional features can be included in future developments of the method).

III. EXPERIMENTAL RESULTS

In our experiments, we have considered two data sets: a hyperspectral image and a LiDAR-derived DSM, both at the same spatial resolution (2.5 m). The hyperspectral imagery consists of 144 spectral bands in the 380–1050-nm region and has been calibrated to at-sensor spectral radiance units, $\text{SRU} = \mu\text{W}/(\text{cm}^2 \text{ sr nm})$. The corresponding coregistered DSM consists of elevation in meters above sea level (per the Geoid 2012A model). The data were acquired by the National Science Foundation (NSF)-funded Center for Airborne Laser Mapping (NCALM) over the University of Houston campus and the neighboring urban area. The LiDAR data were acquired on June 22, 2012, between the time 14:37:55 and 15:38:10 UTC. The average height of the sensor above ground was 2000 ft. The hyperspectral data were acquired on June 23, 2012 between the time 17:37:10 and 17:39:50 UTC. The average height of the sensor above ground was 5500 ft. For illustrative purposes, Fig. 3(a) shows a false color composition of the hyperspectral data. Fig. 3(b) shows the LiDAR derived DSM. Fig. 3(c) shows the ground truth available for the Houston data, which comprise 15 mutually exclusive classes. Finally, Fig. 3(d) shows the training set used in our experiments. Table I details the classes and the number of available training and test samples. The hyperspectral/LiDAR data and the ground truth are available online

TABLE III
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACIES (%) OBTAINED BY THE SVM AND THE SVM-MRF
FOR DIFFERENT TYPES OF COMBINED FEATURES

#	Class	SVM			SVM-MRF		
		$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)
1	Healthy grass	82.43	42.55	82.15	**83.00**	35.71	80.82
2	Stressed grass	82.05	41.54	83.08	83.83	50.56	**84.02**
3	Synthetic grass	99.80	88.51	**100.00**	**100.00**	88.12	**100.00**
4	Trees	**92.80**	72.16	83.71	93.37	70.17	84.94
5	Soil	98.48	70.36	99.81	**100.00**	75.57	**100.00**
6	Water	95.10	66.43	93.71	**95.80**	66.43	79.02
7	Residential	75.47	71.74	75.00	**81.16**	67.63	77.61
8	Commercial	46.91	**89.08**	48.24	45.87	85.75	49.76
9	Road	77.53	59.02	77.05	**86.02**	66.01	81.78
10	Highway	60.04	63.90	47.49	**65.64**	64.00	48.55
11	Railway	81.02	99.91	91.08	83.21	**100.00**	91.37
12	Parking lot 1	85.49	71.09	79.35	**98.46**	74.83	87.42
13	Parking lot 2	75.09	69.47	72.63	**84.21**	72.28	76.49
14	Tennis court	**100.00**	99.19	**100.00**	**100.00**	**100.00**	**100.00**
15	Running track	98.31	98.94	**100.00**	99.15	**100.00**	**100.00**
	OA	80.49	70.80	79.19	**84.05**	71.67	80.77
	AA	83.37	73.59	82.22	**86.65**	74.47	82.79
	κ	78.98	68.41	77.41	**82.83**	69.34	79.12

The best results for each class are outlined in bold typeface.

for public use, which allows for detailed intercomparisons among different algorithms.

Tables III and IV show the classification results obtained by the SVM and MLRsub classifiers, respectively, and the different individual features considered in this work: the original hyperspectral image (i.e., $\mathbf{X}^h$), the AP built on the LiDAR DSM data [i.e., AP($\mathbf{X}^L$)], and the EMAP built on the hyperspectral data [i.e., EMAP($\mathbf{X}^h$)]. In order to build the EMAP($\mathbf{X}^h$), we fixed the number of principal components to be retained to the number of components that contain more than 98% of the total variance of the original hyperspectral data, where the obtained components are scaled to the range [0, 1000] and converted to integer in order to build the attribute filters. Specifically, 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 are 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% are chosen [11]. Finally, to build the AP($\mathbf{X}^L$), we also used

the area and standard deviation attributes, with the same configuration indicated above, but this time applied to a single component given by the LiDAR DSM data. Concerning the probabilistic SVM and MLRsub classifiers, we optimized the related parameters.

In Tables III and IV, we display the classification results obtained for individual features, with and without the MRF-based postprocessing. As shown by Tables III and IV, the original spectral information contained in the hyperspectral image seems to be the most useful type of feature for classification purposes, while the AP built on the LiDAR DSM is not discriminative enough to separate many of the classes. This is expected, since the LiDAR DSM alone provides information about height that is not expected to be able to discriminate between the different urban classes in the considered data. Interestingly, the EMAP built on the original hyperspectral data is also not as discriminative as the original spectral information in this example. This also comes at no surprise, since the EMAP is mainly based on the spatial features of the objects, but the classes in the considered problem are all difficult to discriminate based on spatial properties only; hence, the use of spectral properties

TABLE IV
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACIES (%) OBTAINED BY THE MLRsub AND THE MLRsub-MRF
FOR DIFFERENT TYPES OF COMBINED FEATURES

#	Class	MLRsub			MLRsub-MRF		
		$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)
1	Grass healthy	82.53	60.97	80.53	**83.10**	57.83	80.72
2	Grass stressed	81.02	12.12	77.16	**84.21**	12.69	79.42
3	Grass synthetic	99.41	88.12	**100**	**100**	78.81	**100**
4	Tree	97.44	50.47	65.25	**100**	45.45	66.57
5	Soil	96.12	22.35	96.40	**99.53**	30.59	99.15
6	Water	94.41	67.83	78.32	**95.80**	66.43	79.02
7	Residential	73.51	76.77	58.68	81.53	**83.77**	52.71
8	Commercial	62.68	**80.72**	40.46	57.55	77.68	41.03
9	Road	71.39	24.27	75.83	**86.87**	30.88	83.00
10	Highway	86.87	56.85	48.17	**88.51**	59.75	46.81
11	Railway	67.65	93.26	94.97	79.89	88.24	**100**
12	Parking lot 1	59.85	61.38	81.65	72.72	53.41	**97.79**
13	Parking lot 2	57.19	67.02	66.32	**84.21**	69.47	82.46
14	Tennis court	97.57	99.60	96.76	**100**	**100**	**100**
15	Running track	95.77	89.64	97.46	98.52	**100**	98.73
	OA	79.60	58.08	74.53	**85.18**	58.26	77.44
	AA	81.56	63.43	77.20	**87.50**	63.67	80.49
	Kappa	77.86	54.67	72.35	**83.92**	54.76	75.52

The best results for each class are outlined in bold typeface.

is quite important. Ideally, the separability of the classes could improve if we considered multiple features simultaneously.

Additionally, if we compare the results obtained by SVM with the results obtained by the MLRsub algorithm, we may conclude that SVM shows better performance for classification of the different individual features. However, the performance of the two classifiers for classification of the original hyperspectral image is comparable. Even the MLRsub-MRF has shown slightly better results in comparison with SVM-MRF for classification of individual spectral features. This is expected, since the MLRsub algorithm was originally developed for the classification of hyperspectral images based on the assumption that the hyperspectral features lie in a lower dimensional subspace. The other reason is that SVM is originally a hard classifier and the estimated class probabilities used in the MRF-based relaxation procedure are often not reliable. At this point, it is also important to reiterate that the main purpose of this paper is developing a robust technique for improving the classification of hyperspectral images using other complementary features, and the MLRsub method shows good potential for this purpose.

On the other hand, Table V shows the classification results obtained by the proposed framework for multiple feature

learning when different types of features are considered for the classification. Here, we also reported the results obtained by the MLRsub and the MLRsub-MRF with spatial post-processing. The cases considered are as follows: the original hyperspectral image plus the APs built on the LiDAR DSM [i.e., $\mathbf{X}^h + \text{AP}(\mathbf{X}^L)$]; the original hyperspectral image plus the EMAPs built on the same image [i.e., $\mathbf{X}^h + \text{EMAP}(\mathbf{X}^h)$]; the APs built of the LiDAR DSM plus the EMAPs built on the original hyperspectral image [i.e., $\text{AP}(\mathbf{X}^L) + \text{EMAP}(\mathbf{X}^h)$]; and all the features available [i.e., $\mathbf{X}^h + \text{AP}(\mathbf{X}^L) + \text{EMAP}(\mathbf{X}^h)$]. As shown in Table V, the combination of multiple features always increased the classification results with regard to the single-feature cases reported in Table IV. In particular, the classification results improved significantly when the three considered features $\mathbf{X}^h$, $\text{AP}(\mathbf{X}^L)$, and $\text{EMAP}(\mathbf{X}^h)$ were used simultaneously. Interestingly, the combination of $\text{AP}(\mathbf{X}^L)$ and $\text{EMAP}(\mathbf{X}^h)$ without using the original spectral information in $\mathbf{X}^h$ also provided good classification results, which is interesting since the AP and EMAP are spatially guided features. However, the combination of $\mathbf{X}^h$ and $\text{EMAP}(\mathbf{X}^h)$ resulted in the lowest classification accuracies reported in Table V. This observation is also interesting, since both $\mathbf{X}^h$

TABLE V

CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACIES (%) OBTAINED BY THE MLRsub AND THE MLRsub-MRF
FOR MULTIPLE FEATURE COMBINATIONS

#	Class	MLRsub				MLRsub-MRF			
		$\mathbf{X}^h$ +AP($\mathbf{X}^L$)	$\mathbf{X}^h$ +EMAP($\mathbf{X}^h$)	AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$+AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$ +AP($\mathbf{X}^L$)	$\mathbf{X}^h$ +EMAP($\mathbf{X}^h$)	AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$+AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)
1	Grass healthy	83.00	80.91	80.63	82.91	83.00	80.91	80.82	**83.10**
2	Grass stressed	77.44	81.48	74.34	81.48	79.79	**84.30**	74.25	83.08
3	Grass synthetic	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
4	Tree	97.92	86.55	86.17	95.83	**99.62**	88.73	87.69	95.93
5	Soil	98.20	98.58	96.69	99.05	99.72	**100**	98.20	**100**
6	Water	91.61	95.10	79.02	91.61	**95.80**	**95.80**	79.02	**95.80**
7	Residential	86.29	76.77	88.15	87.59	82.56	81.72	**90.67**	84.79
8	Commercial	88.03	64.10	77.97	84.14	**88.32**	57.45	77.30	83.57
9	Road	85.65	82.34	93.77	91.78	92.92	88.86	**97.92**	96.51
10	Highway	69.11	92.28	71.43	86.20	69.69	**99.90**	73.26	86.78
11	Railway	98.10	74.95	98.86	98.58	99.53	82.92	**99.91**	99.72
12	Parking lot 1	85.88	91.64	89.53	92.32	94.91	**100**	96.73	99.81
13	Parking lot 2	70.53	70.53	76.49	76.84	75.79	**86.32**	78.25	85.26
14	Tennis court	99.60	98.79	**100**	99.60	**100**	**100**	**100**	**100**
15	Running track	98.73	98.10	98.73	98.73	98.94	98.73	98.94	98.73
	OA	87.91	84.40	86.86	90.65	89.85	87.86	88.56	**92.05**
	AA	88.67	86.14	87.45	91.11	90.71	89.71	88.86	**92.87**
	Kappa	86.87	83.10	85.74	89.85	88.99	86.85	87.58	**91.37**

The best results for each class are outlined in bold typeface.

TABLE VI

STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION ACCURACIES

Method	Value of z calculated by the McNemar's test					
	$\{\mathbf{X}^h$+AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$)$\}/$					
	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\{\mathbf{X}^h$+AP($\mathbf{X}^L$)$\}$	$\{\mathbf{X}^h$+EMAP($\mathbf{X}^h$)$\}$	$\{$AP($\mathbf{X}^L$)+EMAP($\mathbf{X}^h$)$\}$
MLR	30.78	59.70	42.42	13.61	22.32	16.42
MLR-MRF	24.87	61.73	41.23	11.51	20.11	15.42

and EMAP($\mathbf{X}^h$) are derived from the original hyperspectral image without including any information about the LiDAR DSM. In turn, the inclusion of the LiDAR-based AP($\mathbf{X}^L$) always resulted in an increase in the classification accuracies reported which indicates that, under the proposed framework, the LiDAR information represents a source of complementary information that can improve the analysis of the hyperspectral data alone.

In order to evaluate the statistical significance of the difference in accuracy between two classifications, the McNemar's test has been widely used in the remote-sensing community [37]. In this test, a value of $|Z| > 1.96$ indicates that there is a significant difference in accuracy between two classification results. The sign of Z is also a criterion to indicate whether the first classifier compared is more accurate than the second one ($Z > 0$) or vice versa ($Z < 0$). As it can be seen from Table VI, the differences in classification accuracies between the case of using all considered features and the other combinations are statistically significant. Moreover, in a different scenario, in order to better evaluate the statistical significance of the
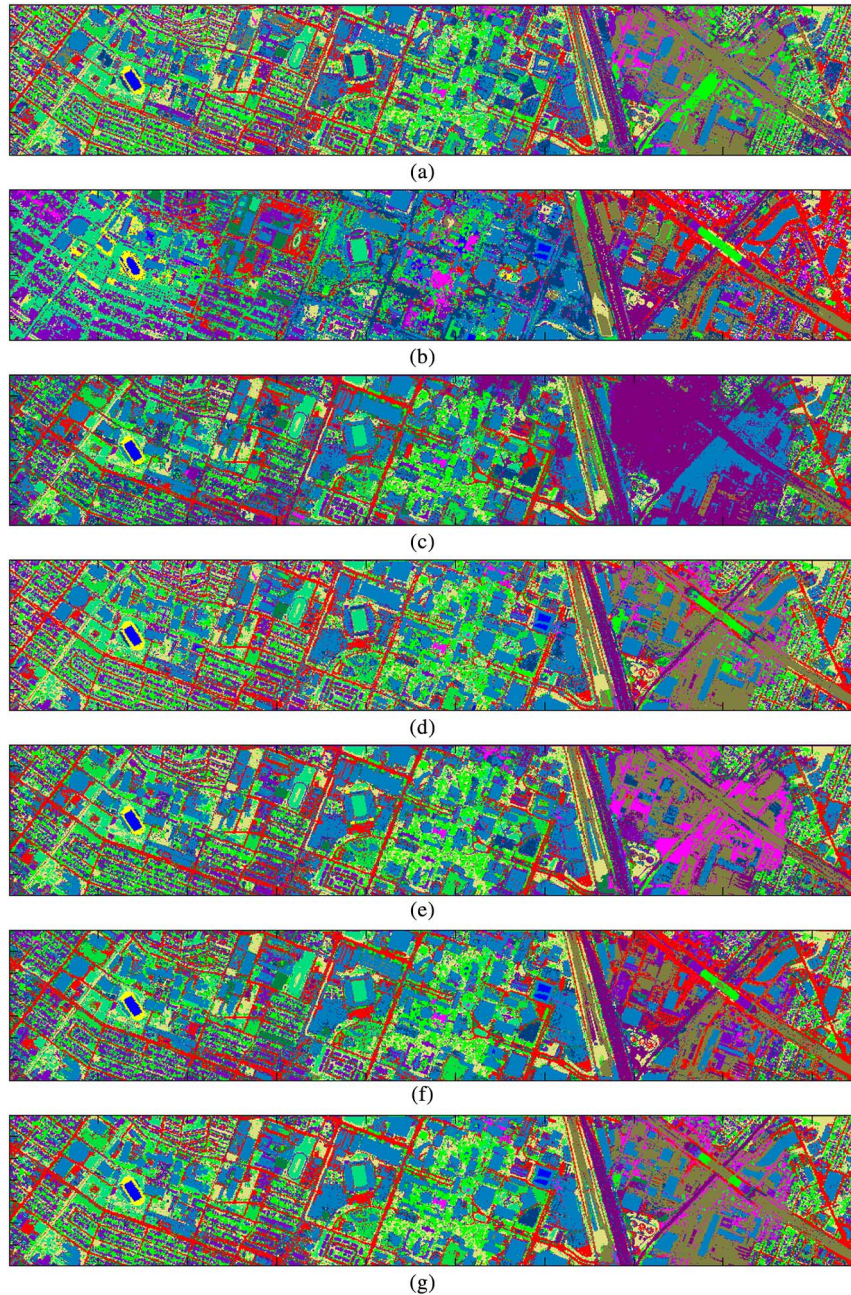
Fig. 4. Classification maps obtained by the MLRsub for different features. (a) MLRsub classification (79.60%) using the hyperspectral image: $\mathbf{X}^h$. (b) MLRsub classification (58.08%) using APs built on the LiDAR DSM: AP($\mathbf{X}^L$). (c) MLRsub classification (74.53%) using EMAPs built on the hyperspectral image: EMAP($\mathbf{X}^h$). (d) MLRsub classification (87.91%) using the hyperspectral image plus APs built on the LiDAR DSM: $\mathbf{X}^h$ + AP($\mathbf{X}^L$). (e) MLRsub classification (84.40%) using the hyperspectral image plus EMAPs built on the hyperspectral image: $\mathbf{X}^h$ + EMAP($\mathbf{X}^h$). (f) MLRsub classification (86.86%) using APs built on the LiDAR DSM plus EMAPs built on the hyperspectral image: AP($\mathbf{X}^L$) + EMAP($\mathbf{X}^h$). (g) MLRsub classification (90.65%) using all the available features: $\mathbf{X}^h$ + AP($\mathbf{X}^L$) + EMAP($\mathbf{X}^h$).

differences, the McNemar's test was used for each object class and Bonferroni correction [38] was applied for McNemar test values to control the family-wise type-I error rate. We considered each P value to indicate a significant difference if P was less than or equal to 0.001. The significant differences in accuracy between the case of using all considered features and EMAP($\mathbf{X}^h$) were the most for the MLR method and the least for the MLR-MRF method. Furthermore, differences in accuracy between the case of using all considered features and

$\mathbf{X}^h$ were the most significant for the MLR-MRF method and the least significant for the MLR method.

At this point, it is important to reiterate that the proposed framework automatically integrates the different sources of information (spatial and spectral) automatically, and without the need to set a parameter to control their relative weight in the final classification result. This is a very important characteristic, as it allows for the integration of multiple features in a very natural way. In fact, additional features such as
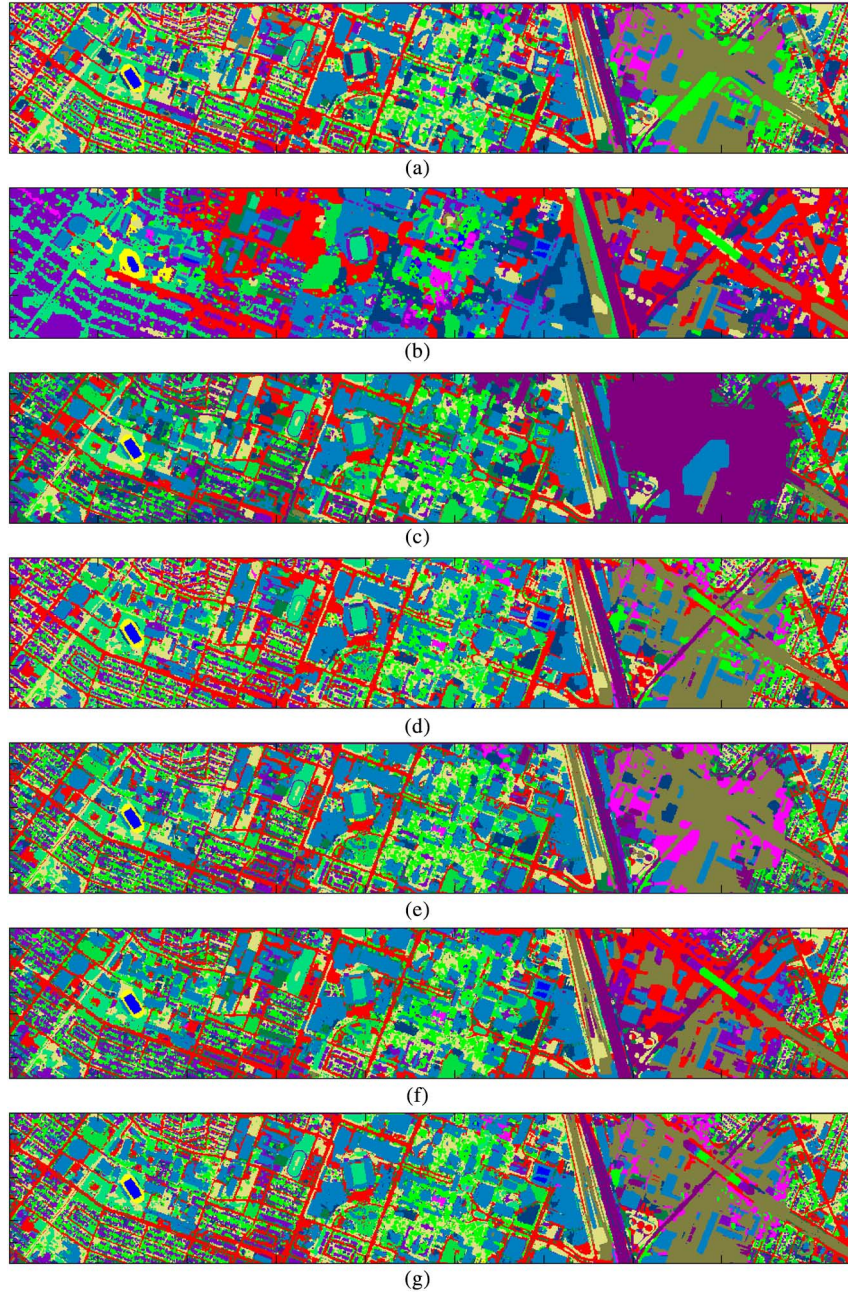
Fig. 5. Classification maps obtained by the MLRsub-MRF for different features. (a) MLRsub-MRF classification (85.18%) using the hyperspectral image: $\mathbf{X}^h$. (b) MLRsub-MRF classification (58.26%) using APs built on the LiDAR DSM: $AP(\mathbf{X}^L)$. (c) MLRsub-MRF classification (77.44%) using EMAPs built on the hyperspectral image: $EMAP(\mathbf{X}^h)$. (d) MLRsub-MRF classification (89.85%) using the hyperspectral image plus APs built on the LiDAR DSM: $\mathbf{X}^h + AP(\mathbf{X}^L)$. (e) MLR$sub$-MRF classification (87.86%) using the hyperspectral image plus EMAPs built on the hyperspectral image: $\mathbf{X}^h + EMAP(\mathbf{X}^h)$. (f) MLR$sub$-MRF classification (88.56%) using APs built on the LiDAR DSM plus EMAPs built on the hyperspectral image: $AP(\mathbf{X}^L) + EMAP(\mathbf{X}^h)$. (g) MLR$sub$-MRF classification (92.05%) using all the available features: $\mathbf{X}^h + AP(\mathbf{X}^L) + EMAP(\mathbf{X}^h)$.

texture, border-related features, etc. could be integrated in the proposed framework to enhance the obtained classification results. As shown by our experiments, the classification results were obtained using a relatively low number of training samples, and the fact that we increased the number of features did not decrease but rather increase the classification accuracies for the same number of training samples. This is due to the subspace-based nature of the MLRsub and MLRsub-MRF classifiers used to obtain the final classification maps. For illustrative purposes, Figs. 4 and 5, respectively, show

some of the classification maps obtained by using the MLRsub and the MLRsub-MRF for the considered Houston hyperspectral/LiDAR data. Effective classification results can be observed in the final maps reported in these figures.

Moreover, using the described data set, we have conducted another experiment to more carefully analyze the validation set and the effectiveness of the selected features in the proposed method. In this experiment, we have randomly selected 10% of the available labeled samples of each class from the validation set for training purposes. It is important to mention that in the

TABLE VII
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACIES (%) OBTAINED BY THE MLR*sub* FOR DIFFERENT SELECTION OF FEATURE COMBINATIONS, USING 10% OF VALIDATION SAMPLES FOR EACH CLASS

#	Class	Samples		MLR*sub*						
		Train	Test	$\mathbf{X}^h$	AP($\mathbf{X}^L$)	EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$ +AP($\mathbf{X}^L$)	$\mathbf{X}^h$ +EMAP($\mathbf{X}^h$)	AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)	$\mathbf{X}^h$+AP($\mathbf{X}^L$) +EMAP($\mathbf{X}^h$)
1	Healthy grass	105	1053	95.95 (2.09)	38.73 (6.51)	83.16 (3.75)	94.90 (3.01)	**96.22 (2.59)**	87.59 (2.72)	95.23 (3.62)
2	Stressed grass	106	1064	**97.03 (1.37)**	26.21 (4.78)	89.51 (4.07)	96.34 (1.31)	96.85 (1.36)	91.92 (3.65)	96.66 (1.40)
3	Synthetic grass	50	505	99.93 (0.11)	86.66 (7.92)	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**
4	Trees	105	1056	94.13 (2.19)	82.24 (2.87)	88.72 (2.14)	98.91 (0.55)	94.25 (2.46)	98.45 (1.07)	**99.42 (0.57)**
5	Soil	105	1056	98.00 (0.71)	61.49 (2.80)	97.53 (1.34)	98.32 (0.79)	99.23 (0.35)	98.38 (0.94)	**99.51 (0.31)**
6	Water	14	143	93.54 (3.01)	69.72 (5.19)	77.55 (4.89)	92.73 (6.08)	**93.99 (5.34)**	80.58 (7.04)	93.45 (7.30)
7	Residential	107	1072	83.21 (3.35)	82.62 (2.81)	62.22 (5.52)	96.48 (1.61)	89.42 (3.81)	96.45 (1.51)	**98.37 (1.19)**
8	Commercial	105	1053	67.20 (7.34)	90.63 (2.81)	50.05 (3.05)	95.19 (2.08)	76.92 (6.37)	93.75 (2.92)	**95.48 (2.55)**
9	Road	105	1059	72.51 (3.08)	59.35 (7.53)	78.75 (2.47)	89.12 (2.31)	87.07 (2.31)	93.50 (1.62)	**94.65 (1.57)**
10	Highway	103	1036	86.82 (2.22)	65.16 (5.42)	90.00 (3.75)	93.99 (2.55)	96.57 (2.67)	97.70 (0.76)	**99.07 (0.59)**
11	Railway	105	1054	77.75 (2.88)	90.95 (2.64)	90.28 (3.29)	96.39 (0.86)	87.91 (2.88)	**97.44 (0.81)**	97.33 (0.81)
12	Parking lot 1	104	1041	76.96 (4.22)	57.86 (10.11)	85.76 (5.36)	86.42 (4.43)	91.52 (3.58)	88.21 (4.39)	**92.36 (2.90)**
13	Parking lot 2	28	285	48.97 (6.41)	58.48 (7.90)	73.81 (4.59)	73.10 (4.34)	75.57 (4.32)	83.92 (3.84)	**84.30 (3.72)**
14	Tennis court	24	247	96.92 (1.81)	97.91 (0.82)	92.47 (3.28)	99.55 (0.33)	98.18 (1.20)	98.92 (0.50)	**99.69 (0.27)**
15	Running track	47	473	98.09 (0.84)	91.16 (5.01)	97.76 (1.09)	**99.34 (0.58)**	98.91 (0.66)	99.07 (0.60)	99.27 (0.57)
	OA			85.60 (1.17)	67.93 (1.54)	82.94 (0.79)	94.60 (0.51)	92.01 (1.08)	94.45 (0.46)	**96.77 (0.34)**
	AA			85.80 (1.03)	70.61 (1.41)	83.84 (0.67)	94.05 (0.63)	92.17 (0.99)	93.73 (0.50)	**96.32 (0.49)**
	κ			84.38 (1.27)	65.31 (1.65)	81.49 (0.85)	94.14 (0.55)	91.33 (1.17)	93.97 (0.49)	**96.49 (0.37)**

The best results for each class are outlined in bold typeface.

original training set, no training samples were selected from the right part of the image where a large cloud shadow is present. However, in the validation set, there are a significant number of samples from this part. Table VII details the average of the results for this experiment obtained after 30 Monte Carlo runs. Several conclusions can be obtained from Table VII. First and foremost, it is remarkable that the proposed multiple feature learning method showed better performances in comparison with using single kind of features, which means that our proposed method efficiently exploits the information contained in both data sources. More importantly, the performance improvements reported for the proposed method using all kinds of suggested features are quite significant. For example, for the classes *Parking lot 2* corresponded to parked vehicles and *Highway* in the cloud-covered region, we can see a significant improvement in the obtained classification result.

IV. CONCLUSION AND FUTURE LINES

In this paper, we have developed a new efficient strategy for fusion and classification of hyperspectral and LiDAR data. Our approach effectively integrates multiple types of features extracted from these data without the need for any regularization or weight parameters, so that different types of features can be efficiently exploited and integrated in a collaborative and flexible way. In this work, we have considered several types of spatial and spectral features derived from the original hyperspectral image and from the LiDAR-derived DSM, including the full original spectral information and different types of MPs calculated for the hyperspectral and the LiDAR data. Our experimental results, conducted using a hyperspectral image and a LiDAR-derived DSM collected over the University of Houston campus and the neighboring urban area, indicate that the information provided by LiDAR can effectively complement the spectral and the spatial information that can be extracted from the original hyperspectral data, providing an increase in the classification accuracies when the LiDAR data are used as a complementary source of information with regard to the original hyperspectral data. Although our analysis in this work has been constrained to a reduced number and type of features, the proposed framework is completely open and flexible in its capacity to integrate additional types of (spatial and spectral) features. As a result, future work will be directed toward the inclusion of additional types of features such as texture and border-oriented features. Although our experiments have been reported for a data set that has been widely used in the recent literature as it was distributed as part of the IEEE GRSS Data Fusion contest in 2013, in the future, we will also conduct further experiments using additional hyperspectral and LiDAR image pairs and perform comparisons of our approach with regard to other techniques recently presented for fusing hyperspectral and LiDAR data.

ACKNOWLEDGMENT

REFERENCES

[1] J. Bioucas-Dias *et al.*, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, pp. 1147–1153, 1985.

[3] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[4] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. 110–122, 2009.

[5] J. Bioucas-Dias *et al.*, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[6] H. Lee and N. Younan, "DTM extraction of LiDAR returns via adaptive processing," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2063–2069, Sep. 2003.

[7] J. Jung, E. Pasolli, S. Prasad, J. Tilton, and M. Crawford, "A framework for land cover classification using discrete return LiDAR data: Adopting pseudo-waveform and hierarchical segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 491–502, Feb. 2014.

[8] H. Wang, C. Glennie, and S. Prasad, "Voxelization of full waveform LiDAR data for fusion with hyperspectral imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2013, pp. 3407–3410.

[9] D. Lemp and U. Weidner, "Segment-based characterization of roof surfaces using hyperspectral and laser scanning data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2005, vol. 7, pp. 4942–4945.

[10] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[11] M. Pedergnana, P. Marpu, M. Dalla Mura, J. Benediktsson, and L. Bruzzone, "Classification of remote sensing optical and LiDAR data using extended attribute profiles," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 856–865, Nov. 2012.

[12] M. Dalla Mura, J. 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.

[13] W. Liao, R. Bellens, A. Pizurica, S. Gautama, and W. Philips, "Graph-based feature fusion of hyperspectral and LiDAR remote sensing data using morphological features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2013, pp. 4942–4945.

[14] S. Prasad and L. M. Bruce, "Overcoming the small sample size problem in hyperspectral classification and detection tasks," in *Proc. IEEE Int. Conf. Geosci. Remote Sens. (IGARSS)*, 2008, vol. 5, pp. 381–384.

[15] G. Licciardi *et al.*, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.

[16] J. Li, J. 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.

[17] J. Benediktsson, J. Sveinsson, and P. Swain, "Hybrid consensus theoretic classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 4, pp. 833–843, Jul. 1997.

[18] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 2nd ed. Berlin, Germany: Springer-Verlag, 2001.

[19] P. Soille, *Morphological Image Analysis: Principles and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.

[20] M. Pesaresi and J. 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.

[21] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer, 2006.

[22] J. Benediktsson, J. Palmason, and J. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[23] J. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

[24] M. Dalla Mura, J. Atli 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.

[25] P. Ghamisi, J. Benediktsson, and J. Sveinsson, "Automatic spectral-spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.

[26] B. Song *et al.*, "Remotely sensed image classification using sparse representations of morphological attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 5122–5136, Aug. 2014.

[27] M. Dalla Mura, J. Atli 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.

[28] P. Salembier, A. Oliveras, and L. Garrido, "Antiextensive connected operators for image and sequence processing," *IEEE Trans. Image Process.*, vol. 7, no. 4, pp. 555–570, Apr. 1998.

[29] V. Vapnik and A. Chervonenkis, "The necessary and sufficient conditions for consistency in the empirical risk minimization method," *Pattern Recognit. Image Anal.*, vol. 1, no. 3, pp. 283–305, 1991.

[30] C. Huang, L. S. Davis, and J. R. Townshend, "An assessment of support vector machines for land cover classification," *Int. J. Remote Sens.*, vol. 23, no. 4, pp. 725–749, Feb. 2002.

[31] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[32] M. Khodadadzadeh, J. Li, A. Plaza, and J. M. Bioucas-Dias, "A subspace-based multinomial logistic regression for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2015–2109, Dec. 2014.

[33] 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.

[34] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multiclass classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, pp. 975–1005, Dec. 2004.

[35] C.-J. Lin, H.-T. Lin, and R. C. Weng, "A note on platt's probabilistic outputs for support vector machines," *Mach. Learn.*, vol. 68, no. 3, pp. 267–276, 2007.

[36] C. Chang and C. Lin. (2009). *LIBSVM: A Library for Support Vector Machines* [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[37] 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, May 2004.

[38] C. E. Bonferroni, "Teoria statistica delle classi e calcolo delle probabilità," *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze*, vol. 8, pp. 3–62, 1936.

Mahdi Khodadadzadeh (S'10) received the B.Sc. degree in electrical engineering from the Sadjad Institute of Higher Education, Mashhad, Iran, in 2008, and the M.Sc. degree in electrical engineering from Tarbiat Modares University, Tehran, Iran, in 2011. He is currently pursuing the Ph.D. degree at the Hyperspectral Computing Laboratory (HyperComp), Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain.

His research interests include remote sensing, pattern recognition, and signal and image processing, with particular emphasis on spectral and spatial techniques for hyperspectral image classification.

Mr. Khodadadzadeh is a manuscript Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Jun Li (M'13) 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 Telecomunicaes, 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. Currently, she is a Professor with the School of Geography and Planning, Sun Yat-Sen University, Guangzhou, China. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li is an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. She has been a Reviewer of several journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Pattern Recognition*, *Optical Engineering*, *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*.

Saurabh Prasad (S'05–M'09–SM'14) received the B.S. degree in electrical engineering from Jamia Millia Islamia, New Delhi, India, in 2003, the M.S. degree in electrical engineering from Old Dominion University, Norfolk, VA, USA, in 2005, and the Ph.D. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2008.

He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, University of Houston (UH), Houston, TX, USA, where he leads a Research Group on Geospatial Image Analysis. His research interests include statistical pattern recognition, adaptive signal processing, and kernel methods for medical imaging, optical, and synthetic aperture radar remote sensing.

Dr. Prasad is an active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the Elsevier *Pattern Recognition Letters*. He was awarded the Geosystems Research Institute's Graduate Research Assistant of the Year Award in May 2007, and the Office-of-Research Outstanding Graduate Student Research Award in April 2008 at Mississippi State University. He was the recipient of the Best Student Paper Award at IEEE International Geoscience and Remote Sensing Symposium 2008 held in Boston, MA, USA, in July 2008 and the State Pride Faculty Award at Mississippi State University, in October 2010, for his academic and research contributions, and the NASA New Investigator (Early Career) Award in 2014.

Antonio Plaza (M'05–SM'07–F'15) received the Computer Engineer degree in 1997, the M.Sc. degree in 1999, and the Ph.D. degree in 2002, all in computer engineering.

He is an Associate Professor (with accreditation for Full Professor) with the Department of Technology of Computers and Communications, University of Extremadura, Badajoz, Spain, where he is the Head of the Hyperspectral Computing Laboratory (HyperComp). He has been the Advisor of 12 Ph.D. dissertations and more than 30 M.Sc. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with total funding of 2.8 million Euro. He has authored more than 400 publications, including 140 journal papers (90 in IEEE journals), 20 book chapters, and over 240 peer-reviewed conference proceeding papers (94 in IEEE conferences). He has edited a book on *High-Performance Computing in Remote Sensing* for CRC Press/Taylor & Francis and guest edited eight special issues on hyperspectral remote sensing for different journals. He has reviewed more than 500 articles for over 50 different journals. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza was the recipient of the recognition of Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009) and the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), a journal for which he served as Associate Editor in 2007–2012. He is also an Associate Editor for IEEE ACCESS, and was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER (2011–2012) and the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE (2013). He was also a member of the steering committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He was the recipient of the 2013 Best Paper Award of the JSTARS journal, and the most highly cited paper (2005–2010) in the *Journal of Parallel and Distributed Computing*. He was the recipient of the Best Paper Awards at the IEEE International Conference on Space Technology and the IEEE Symposium on Signal Processing and Information Technology, and the Best Ph.D. Dissertation Award at the University of Extremadura, a recognition also received by five of his Ph.D. students. He served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) in 2011–2012, and is currently serving as President of the Spanish Chapter of IEEE GRSS (since November 2012). He has served as a Proposal Evaluator for the European Commission, the National Science Foundation, the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING journal.