# A Discontinuity Preserving Relaxation Scheme for Spectral–Spatial Hyperspectral Image Classification

Jun Li, *Member, IEEE*, Mahdi Khodadadzadeh, *Student Member, IEEE*, Antonio Plaza, *Fellow, IEEE*, Xiuping Jia, *Senior Member, IEEE*, and José M. Bioucas-Dias, *Member, IEEE*

***Abstract*—In remote sensing image processing, relaxation is defined as a method that uses the local relationship among neighboring pixels to correct spectral or spatial distortions. In recent years, relaxation methods have shown great success in classification of remotely sensed data. Relaxation, as a preprocessing step, can reduce noise and improve the class separability in the spectral domain. On the other hand, relaxation (as a postprocessing approach) works on the label image or class probabilities obtained from pixelwise classifiers. In this work, we develop a discontinuity preserving relaxation strategy, which can be used for postprocessing of class probability estimates, as well as preprocessing of the original hyperspectral image. The newly proposed method is an iterative relaxation procedure, which exploits spatial information in such a way that it considers discontinuities existing in the data cube. Our experimental results indicate that the proposed methodology leads to state-of-the-art classification results when combined with probabilistic classifiers for several widely used hyperspectral data sets, even when very limited training samples are available.**

***Index Terms*—Border delineation, hyperspectral image classification, relaxation methods, spectral–spatial relaxation, probabilistic relaxation (PR).**

## I. Introduction

REMOTELY sensed hyperspectral image classification has been a very active area of research in recent years [1]. Although techniques for unsupervised classification and/or clustering have also been used in the literature [2], supervised classification has been more popular in many applications [3]. Still, there are several important challenges when performing supervised hyperspectral image classification [4], such as the unbalance between high dimensionality and limited training samples, or the presence of mixed pixels in the data (which may compromise classification results for coarse spatial resolutions). Another relevant challenge is the need to integrate the spatial and spectral information to take advantage of the complementarities that both sources of information can provide. Such integration can reduce the negative impact of the aforementioned challenges.

According to the principle that, in remote sensing images, neighboring pixels are likely to have the same contextual properties, spectral–spatial techniques can be effectively exploited to improve the classification accurary [2]. For example, in [5], simply adding the mean of neighboring pixel values for each band to the original spectral feature vector of central pixel has shown better classification performance than conventional spectral methods. In [6], authors proposed to extract textural features from the hyperspectral image using efficient image enhancement algorithms and then combine them with spectral information via kernels in a semisupervised graph-based framework for classification. In other approaches, modeling different kinds of the structural information contained in hyperspectral images by using morphological filters and integrating with spectral information have been successfully used for hyperspectral image classification [7]–[9].

The important category of spectral–spatial techniques comprises relaxation methods which are defined as methods that use the local relationship among neighboring pixels to correct spectral or spatial distortions. As preprocessing, spatial smoothing over the hyperspectral data can remove noise and enhance spatial texture information [10]–[12]. For example, in [11], in order to classify land cover mathematical morphology-based noise reduction filter has been used before the maximum-likelihood (ML) classification algorithm. In [10], authors showed that anisotropic diffusion algorithm can reduce the spatial and spectral variability of the image, while preserving the edges of objects, which will improve the classification accuracy of hyperspectral imagery. On the other hand, as a postprocessing method, relaxation-based approaches can be an effective tool to improve classification accuracies [2]. These normally iterative methods are broadly referred to as *continuous relaxation* (CR) or *probabilistic relaxation* (PR) [13]–[16], which incorporate spatial-contextual information into the obtained probabilistic classification results. In other words, after a probabilistic pixelwise classification of the hyperspectral image, the process of PR is applied to exploit the continuity, in probability sense, of neighboring labels. Perhaps the most popular PR strategy is

Manuscript received February 16, 2015; revised June 08, 2015; accepted August 08, 2015. Date of publication September 13, 2015; date of current version February 09, 2016. This work was supported by the Portuguese Science and Technology Foundation under Project UID/EEA/50008/2013 and Project PTDC/EEI-PRO/1470/2012, and in part by the Chinese National Science Foundation under Project 41431178.

J. Li is with the Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: lijun48@mail.sysu.edu.cn).

M. Khodadadzadeh and A. Plaza are with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres E-10003, Spain.

X. Jia is with the Department of Electrical Engineering, School of Engineering and Information Technology, University of New South Wales, Sydney, NSW 2904, Australia.

J. M. Bioucas-Dias is with the Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa, Lisboa 1649-004, Portugal.

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

Digital Object Identifier 10.1109/JSTARS.2015.2470129

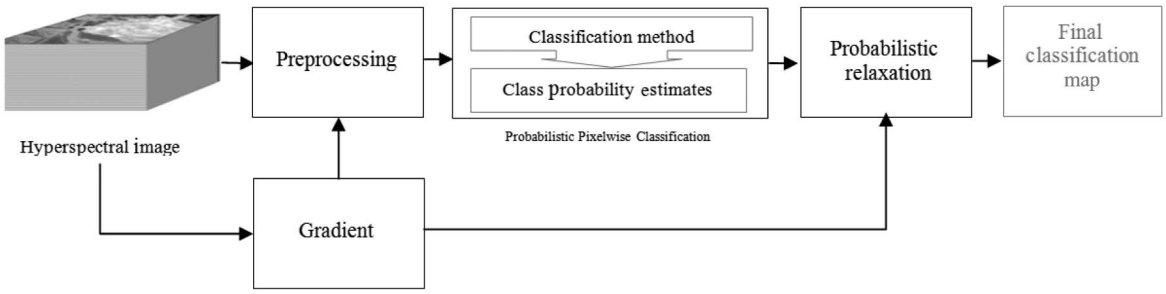


Fig. 1. Flowchart of the proposed method.

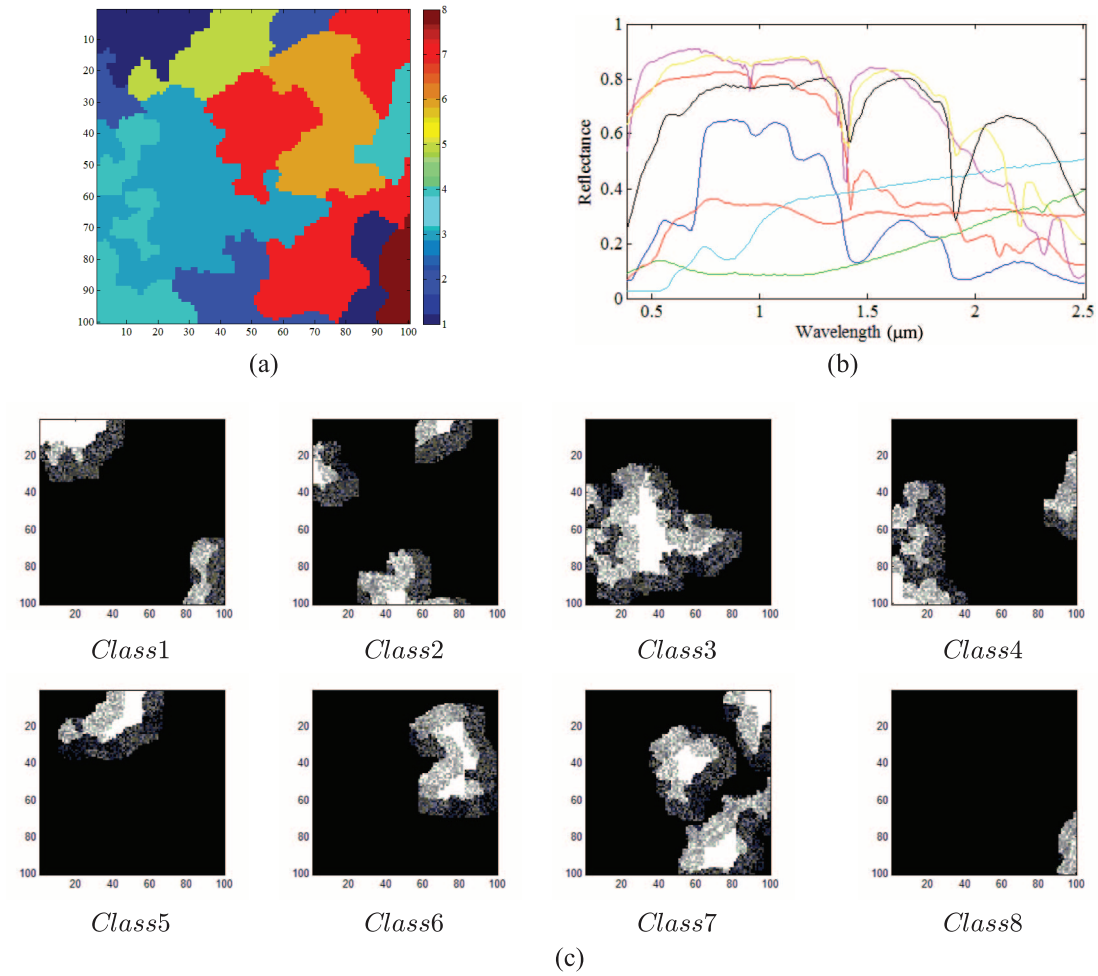


Fig. 2. (a) Classes in a synthetic scene with $n = 100 \times 100$. (b) Spectral signatures of randomly selected materials from the USGS digital library used in the simulation. (c) Fractional abundance distributions considered for generating mixed pixels using a fixed window of size $9 \times 9$ pixels.

based on the use of Markov random fields (MRFs) [2], [17]–[20]. Specifically, the MRF has been shown to be a very successful technique for refining the classification results provided by the probabilistic SVM classifier. For instance, spectral–spatial hyperspectral image classification was performed in [21], given an initial SVM classification map and a final MRF-based relaxation process. This work suggests to incorporate a "fuzzy no-edge/edge" function into the MRF-based relaxation procedure for preserving discontinuities. An adaptive MRF approach was proposed, in [22], for hyperspectral image classification. This work introduced a relative homogeneity index for each pixel to determine an appropriate weighting coefficient for the spatial contribution in the MRF-based relaxation procedure. In [23], a novel and rigorous framework was proposed for contextual hyperspectral image classification, which combines SVMs and MRFs in a unique formulation. More recently, integration of the multinomial logistic regression (MLR) and MRF algorithms has shown significant performance in hyperspectral image classification. For instance, in [24], combining MRF-based multilevel logistic (MLL) prior with subspace-based MLR (MLR$sub$) algorithm was proposed for hyperspectral image classification. In another effort to use MRF-based prior, in [25], the hyperspectral classification results were obtained by maximizing the marginal probability of the posterior

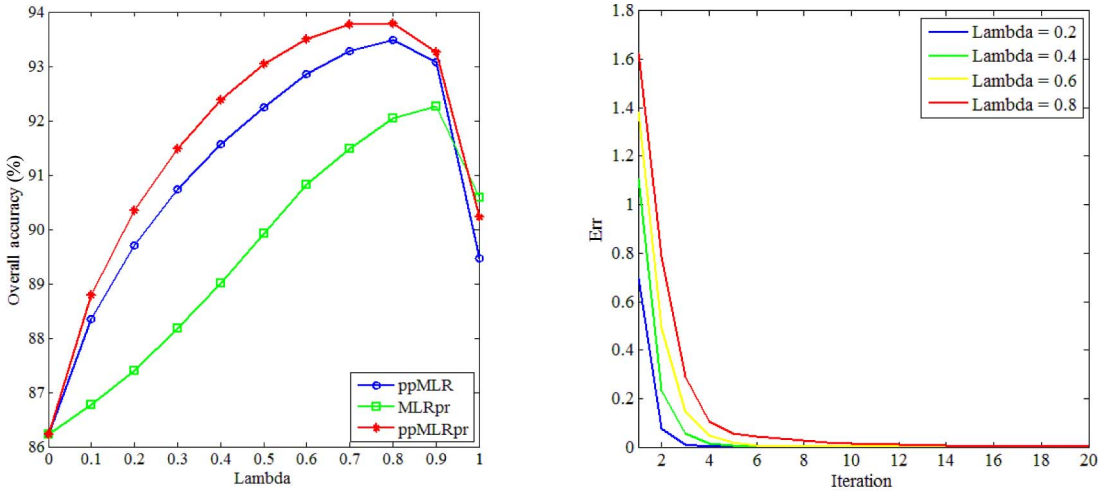


Fig. 3. Impact of parameter $\lambda$.

distribution using the loopy belief propagation method, where the posterior class probability was modeled as an MLR classifier and an MRF. Very recently, combining sparse MLR (SMLR) algorithm [26] with a spatially adaptive total variation (SpATV) regularization was proposed, which showed significant performance [27]. However, one of the first approaches to include spatial–contextual information in probabilistic classification was *probabilistic label relaxation* (PLR), [2], [28], [29]. PLR strategies use the probabilities of neighboring pixels iteratively to update the class probabilities for the center pixel based on a neighborhood function [2]. It has been observed, quite often, the use of spatial information as relaxation, although, on one hand, it clearly improves the classification accuracy in smooth image areas, on the other hand, it degrades the classification performance in the neighborhood of the class boundaries. Fundamentally, this is a consequence of enforcing smoothness across the boundaries. Based on this observation, in this work, we develop a new relaxation strategy for hyperspectral image classification which aims at introducing spatial relaxation while accurately preserving the edges of class boundaries. This edge preserving strategy relies on discontinuity maps estimated from the original image cube. These maps are accurate because they are inferred from the many image bands, usually on the order of hundreds, with aligned discontinuities.

The proposed approach can also be used as a preprocessing step to logically relax the original spectral vectors by considering discontinuities from the data cube. This step is able to reduce noise and improve the class separability while preserving discontinuities by including edge information. However, as a postprocessing, the proposed approach is based on the main principles of PLR-based methods, which can be considered as a form of PR since they iteratively improve the probabilistic output of the considered classifier by naturally imposing spatial consistency in the final classified image. This is important, as some spatial postprocessing strategies tend to generate an undesired blob-like effect in the final classification results. In this regard, our experimental results indicate that the proposed methodology leads to state-of-the-art classification results when compared with other widely used PR-based methods (e.g., PLR and MRF). The probabilistic outputs and the fact that the presented method does not require prior information about the scene are other important features of the proposed approach.

TABLE I
OVERALL (OA) AND AVERAGE (AA) CLASSIFICATION ACCURACIES (%) (AS A FUNCTION OF PARAMETER $\sigma$) NOISE

| Methods | Accuracies | $\sigma$ | | | | |
|---|---|---|---|---|---|---|
| | | 0 | 0.05 | 0.10 | 0.15 | 0.20 |
| MLR | OA | 92.63 | 91.09 | 83.02 | 69.92 | 56.16 |
| | AA | 93.77 | 92.09 | 85.07 | 72.71 | 59.74 |
| ppMLR | OA | 93.08 | 95.50 | 91.50 | 79.19 | 65.74 |
| | AA | 94.09 | 96.05 | 92.46 | 81.17 | 68.49 |
| MLRpr | OA | 95.70 | 95.77 | 87.76 | 72.69 | 57.82 |
| | AA | 96.23 | 96.19 | 89.57 | 75.63 | 61.64 |
| ppMLRpr | OA | 94.57 | 95.90 | 91.80 | 79.62 | 65.90 |
| | AA | 95.43 | 96.44 | 92.76 | 81.76 | 68.94 |

This paper is organized as follows. Section II describes the main stages of the proposed classification framework, including preprocessing, classification, and edge-preserving probability relaxation. Section III presents an experimental validation of the method, conducted using three well-known hyperspectral data sets collected by the airborne visible infrared imaging spectrometer (AVIRIS) over the Indian Pines, Indiana, and Salinas Valley, California, and by the reflective optics spectrographic imaging system (ROSIS) over the city of Pavia, Italy. Section IV concludes the paper with some remarks and hints at plausible future research.

## II. PROPOSED FRAMEWORK

In this section, we first present probabilistic pixelwise classification method, which is applied in this work and then we

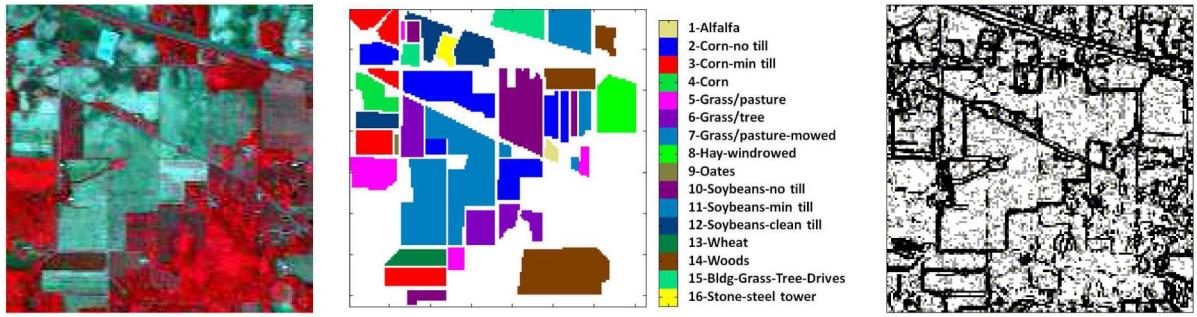


Fig. 4. AVIRIS Indian Pines data set. (a) False color composition. (b) Ground-truth. (c) Discontinuity map.

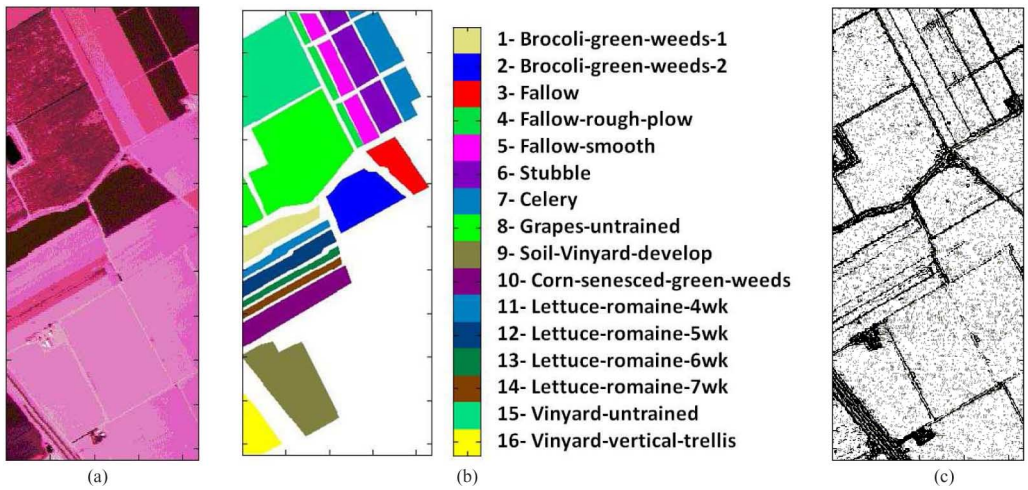


Fig. 5. AVIRIS Salinas data set. (a) False color composition. (b) Ground-truth. (c) Discontinuity map.

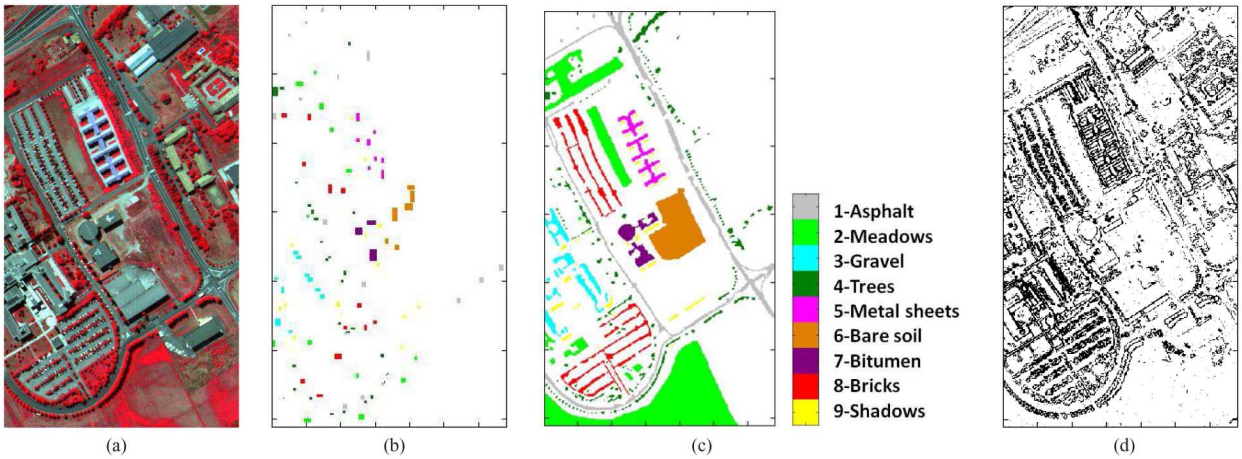


Fig. 6. ROSIS Pavia University data set. (a) False color composition. (b) Training data. (c) Ground-truth. (d) Discontinuity map.

describe the proposed relaxation approach which is used in both preprocessing and postprocessing methods. The flowchart of proposed method has been shown in Fig. 1.

### A. Probabilistic Pixelwise Classification

Let $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ denote the observed data from an input image, where $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T$ denotes a spectral vector associated with an image pixel $i \in S$, $d$ is the number of spectral bands, and $S = \{1, 2, \ldots, n\}$ is the set of integers indexing the $n$ pixels of an image. In probabilistic pixelwise classification, the goal is to make a decision for a pixel $\mathbf{x}_i$ regarding a label assignment $y_i \in \{1, 2, \ldots, K\}$. The decision is made on the basis of the posterior probabilities that the pixel belongs to each of the $K$ classes, i.e., $p(y_i = k|\mathbf{x}_i)$. With these definitions in mind, and adopting the *maximum a posteriori*

TABLE II
OA AND AA CLASSIFICATION ACCURACIES (%) OBTAINED BY DIFFERENT METHODS FOR THE AVIRIS INDIAN PINES DATA SET

| Class | Train/test | Methods | | | | |
|---|---|---|---|---|---|---|
| | | MLR | MLR-MLL | Proposed relaxation methods | | |
| | | | | ppMLR | MLRpr | ppMLRpr |
| Alfalfa | 15 / 54 | 86.67 (4.06) | 92.96 (5.29) | 98.61 (1.33) | 97.31 (0.95) | 97.96 (1.02) |
| Corn-no till | 15 / 1434 | 50.36 (7.87) | 61.98 (9.16) | 82.49 (6.39) | 70.25 (8.89) | 85.54 (6.27) |
| Corn-min till | 15 / 834 | 58.62 (10.93) | 70.85 (13.01) | 86.60 (6.57) | 74.03 (14.26) | 89.90 (7.09) |
| Corn | 15 / 234 | 69.70 (10.81) | 90.49 (8.71) | 94.94 (4.96) | 95.60 (9.13) | 97.44 (4.48) |
| Grass/trees | 15 / 497 | 82.27 (5.68) | 88.16 (5.32) | 91.18 (4.04) | 90.99 (7.57) | 91.64 (5.11) |
| Grass/pasture | 15 / 747 | 88.80 (5.06) | 96.06 (2.68) | 97.48 (1.15) | 99.65 (0.42) | 98.25 (0.75) |
| Grass/pasture-mowed | 15 / 26 | 96.73 (2.87) | 98.65 (2.26) | 99.81 (0.86) | 97.31 (4.69) | 94.23 (22.34) |
| Hay-windrowed | 15 / 489 | 90.26 (4.35) | 97.29 (3.14) | 99.16 (0.71) | 99.64 (1.36) | 99.98 (0.06) |
| Oats | 15 / 20 | 99.75 (1.12) | 100.00 (0.00) | 100.00 (0.00) | 43.50 (28.75) | 92.00 (11.17) |
| Soybeans-notill | 15 / 968 | 53.36 (9.30) | 68.16 (12.31) | 87.57 (6.85) | 74.83 (12.18) | 88.57 (5.76) |
| Soybeans-min till | 15 / 2468 | 49.59 (9.57) | 60.65 (10.74) | 80.68 (5.78) | 68.07 (14.52) | 86.29 (5.72) |
| Soybeans-clean till | 15 / 614 | 58.36 (10.92) | 76.69 (12.36) | 88.88 (5.32) | 84.45 (14.14) | 93.20 (6.16) |
| Wheat | 15 / 212 | 97.81 (2.31) | 99.46 (0.41) | 99.58 (0.34) | 99.98 (0.11) | 99.62 (0.42) |
| Woods | 15 / 1294 | 86.81 (7.63) | 94.11 (7.10) | 93.96 (3.72) | 95.89 (7.05) | 95.24 (4.56) |
| Bldg-grass-tree-drives | 15 / 380 | 44.30 (12.02) | 57.25 (15.18) | 94.99 (6.40) | 72.33 (18.72) | 95.96 (5.84) |
| Stone-steel towers | 15 / 95 | 92.16 (4.29) | 96.58 (4.19) | 98.63 (2.11) | 97.95 (2.01) | 96.79 (1.95) |
| OA | | 64.30 (2.29) | 75.09 (2.86) | 88.36 (1.67) | 80.67 (3.12) | 91.05 (1.87) |
| AA | | 75.35 (1.60) | 84.33 (1.36) | 93.41 (1.03) | 85.11 (2.52) | 93.91 (2.27) |
| $\kappa$ | | 60.03 (2.45) | 72.03 (3.10) | 86.88 (1.86) | 78.22 (3.36) | 89.87 (2.09) |

*probability* (MAP) classification criterion, we can write the discriminant function as follows:

$$y_i = k \quad \text{if} \quad p(y_i = k|\mathbf{x}_i) \geq p(y_i = c|\mathbf{x}_i) \quad \forall c \neq k. \tag{1}$$

Various probabilistic classification techniques have been successfully used for hyperspectral data [24], [30], [31]. In this work, we consider an MLR algorithm. MLR-based techniques exhibit the advantage of modeling directly the posterior class distributions. In this context, the densities $p(y_i|\mathbf{x}_i)$ can be modeled by the MLR, which corresponds to a discriminative model of the discriminative–generative pair for $p(\mathbf{x}_i|y_i)$ (Gaussian) and $p(y_i)$ (multinomial). The MLR model is formally given by [32]

$$p(y_i = k|\mathbf{x}_i, \omega) = \frac{\exp\left(\omega^{(k)}\mathbf{h}(\mathbf{x}_i)\right)}{\sum_{k=1}^{K}\exp\left(\omega^{(k)}\mathbf{h}(\mathbf{x}_i)\right)} \tag{2}$$

where $\mathbf{h}(\mathbf{x}) \equiv [\mathbf{h}_1(\mathbf{x}), \ldots, \mathbf{h}_l(\mathbf{x})]^T$ is a vector of $l$ fixed functions of the input, often termed as features, $\omega^{(k)}$ is the set of logistic regressors for class $k$, and $\omega \equiv [{\omega^{(1)}}^T, \ldots, {\omega^{(K-1)}}^T]^T$.

Recently, [24] proposed to combine the MLR with a subspace projection method called MLR$sub$. 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 is the distance from the class subspace [24]. In [33], a modified version of MLR$sub$ is proposed, which uses the following input function $\mathbf{h}(\mathbf{x}_i)$ in (2) and is given by

$$\mathbf{h}(\mathbf{x}_i) = [\|\mathbf{x}_i\|^2, \|\mathbf{x}_i^{\mathrm{T}}\mathbf{U}^{(1)}\|^2, \ldots, \|\mathbf{x}_i^{\mathrm{T}}\mathbf{U}^{(K)}\|^2]^{\mathrm{T}} \tag{3}$$

where $\mathbf{U}^{(k)} = \{\mathbf{u}_1^{(k)}, \ldots, \mathbf{u}_{r^{(k)}}^{(k)}\}$, $k = 1, 2, \ldots, K$, is a set of $r^{(k)}$-dimensional orthonormal basis vectors for the subspace associated with class $k$ ($r^{(k)} \ll d$).

The fact that hyperspectral vectors tend to live in unions of subspaces underlies the input function (3). In the following, we simply refer to the MLR$sub$ classifier adopted in this work as MLR for simplicity.

TABLE III
OA AND AVERAGE (AA) CLASSIFICATION ACCURACIES (%) OBTAINED BY DIFFERENT METHODS FOR THE AVIRIS SALINAS DATA SET

| Class | Train/test | Methods | | | | |
|---|---|---|---|---|---|---|
| | | MLR | MLR-MLL | Proposed relaxation methods | | |
| | | | | ppMLR | MLRpr | ppMLRpr |
| Brocoli-green-weeds-1 | 15 / 2009 | 99.31 (0.55) | 99.93 (0.40) | 99.88 (0.27) | 99.98 (0.13) | 99.90 (0.24) |
| Brocol-green-weeds-2 | 15 / 3726 | 98.42 (1.14) | 99.12 (0.09) | 99.41 (1.18) | 99.79 (0.37) | 96.67 (15.90) |
| Fallow | 15 / 1976 | 91.81 (6.20) | 94.21 (5.43) | 98.85 (2.85) | 95.65 (6.10) | 96.08 (18.27) |
| Fallow-rough-plow | 15 / 1394 | 98.35 (2.99) | 98.70 (2.75) | 99.79 (0.32) | 99.94 (0.13) | 98.90 (5.62) |
| Fallow-smooth | 15 / 2678 | 95.88 (2.70) | 98.71 (1.53) | 97.93 (1.00) | 98.72 (0.22) | 97.42 (6.96) |
| Stubble | 15 / 3959 | 98.62 (1.08) | 99.31 (0.37) | 99.69 (0.36) | 99.82 (0.22) | 98.49 (7.43) |
| Celery | 15 / 3579 | 98.78 (0.62) | 99.22 (0.20) | 99.76 (0.17) | 99.91 (0.01) | 98.25 (9.13) |
| Grapes-untrained | 15 / 11271 | 66.18 (9.09) | 74.39 (24.50) | 81.76 (5.80) | 77.46 (12.22) | 84.68 (6.42) |
| Soil-vinyard-develop | 15 / 6203 | 97.30 (0.90) | 98.58 (1.04) | 98.91 (0.76) | 100.00 (0.01) | 97.46 (9.05) |
| Corn-senesced-green-weeds | 15 / 3278 | 81.72 (5.92) | 85.67 (2.08) | 90.69 (3.24) | 88.77 (8.35) | 92.17 (3.51) |
| Lettuce-romaine-4wk | 15 / 1068 | 93.12 (3.61) | 95.55 (1.89) | 97.18 (2.80) | 99.46 (0.77) | 99.58 (0.52) |
| Lettuce-romaine-5wk | 15 / 1927 | 97.62 (2.98) | 99.15 (0.00) | 98.82 (1.97) | 99.60 (1.28) | 98.94 (1.98) |
| Lettuce-romaine-6wk | 15 / 916 | 98.70 (0.86) | 99.01 (0.80) | 98.73 (1.05) | 98.86 (0.34) | 98.73 (0.64) |
| Lettuce-romaine-7wk | 15 / 1070 | 94.20 (3.01) | 96.81 (3.41) | 97.47 (2.25) | 97.21 (1.77) | 97.13 (2.25) |
| Vinyard-untrained | 15 / 7268 | 63.90 (9.33) | 71.53 (41.58) | 87.32 (9.06) | 78.01 (14.28) | 90.55 (8.65) |
| Vinyard-vertical-trellis | 15 / 1807 | 94.16 (2.89) | 96.32 (1.64) | 96.71 (3.39) | 98.36 (0.78) | 96.93 (3.69) |
| OA | | 85.28 (1.51) | 89.02 (6.54) | 93.30 (1.70) | 91.26 (2.30) | 93.79 (4.46) |
| AA | | 91.76 (0.76) | 94.14 (2.60) | 96.43 (0.81) | 95.72 (1.05) | 96.37 (4.48) |
| $\kappa$ | | 83.67 (1.66) | 87.80 (7.28) | 92.56 (1.89) | 90.29 (2.55) | 93.11 (4.91) |

### *B. Discontinuity Preserving Relaxation*

In this work, we introduce a new relaxation method to logically smooth the classification results or the original hyperspectral image using both spatial and spectral information while preserving the discontinuities extracted from the data cube.

Let $\mathbf{p} = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \in \mathbb{R}^{K \times n}$, $\mathbf{p}_i = [p_i(1), \ldots, p_i(K)]^T$ for $i \in \mathcal{S}$ be the $K$-dimensional multivariate vector of probabilities defined on site $i$. Let $\mathbf{u} = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \in \mathbb{R}^{n \times K}$, for $i \in \mathcal{S}$, $\mathbf{u}_i = [u_i(1), \ldots, u_i(K)]^T$ be the final vectors of probabilities obtained from the relaxation process. In this work, we implement a relaxation scheme that is the solution of the following optimization problem:

$$\min_{\mathbf{u}} \quad (1-\lambda)\|\mathbf{u}-\mathbf{p}\|^2 + \lambda \sum_i \sum_{j \in \partial_i} \varepsilon_j \|\mathbf{u}_j - \mathbf{u}_i\|^2$$
$$\text{s.t.:} \quad \mathbf{u}_i \geq 0, \quad \mathbf{1}^T \mathbf{u}_i = 1 \tag{4}$$

where the constraints are justified by the fact that the vectors $\mathbf{u}_i$ represent probabilities $\mathbf{1}$ is a vector column of $K$ 1s, $\lambda$ $(0 \leq \lambda \leq 1)$ is a weight parameter controlling the relative impact of the both terms in the objective function, $\partial_i$ denotes the eight-neighborhood of pixel $i$ (other types of neighborhood can be applied), and $\varepsilon_j$ is a value in the site $j \in S$ of edge image $\varepsilon$ given by

$$\varepsilon = \exp\left(-\sum_{i=1}^{d} \text{sobel}(\mathbf{X}^{(i)})\right) \tag{5}$$

where sobel() denotes the Sobel filter, which detects the discontinuities in an image and the output at each pixel is 0 or 1. The Sobel filter is applied on each spectral channel in a specific direction and $\mathbf{X}^{(i)}$ denotes the $i$th band of the original data cube $\mathbf{X}$. Note that here, to have a better interpretation of the edges, we considered the average of the results obtained by applying sobel() in two vertical and horizontal directions.

In the proposed relaxation scheme (4), the first term in the objective function measures the data misfit and the second term promotes smooth solutions weighted by the parameter $\varepsilon_j$, which, according to its definition, is large when there are no discontinuities between the neighboring pixels it connects and small when there are discontinuities. The solution of (4) corresponds, therefore, to tradeoff between adjustment to the "noisy" classification, imposed by the first term, and smoothness imposed by the second term. We stress, however,

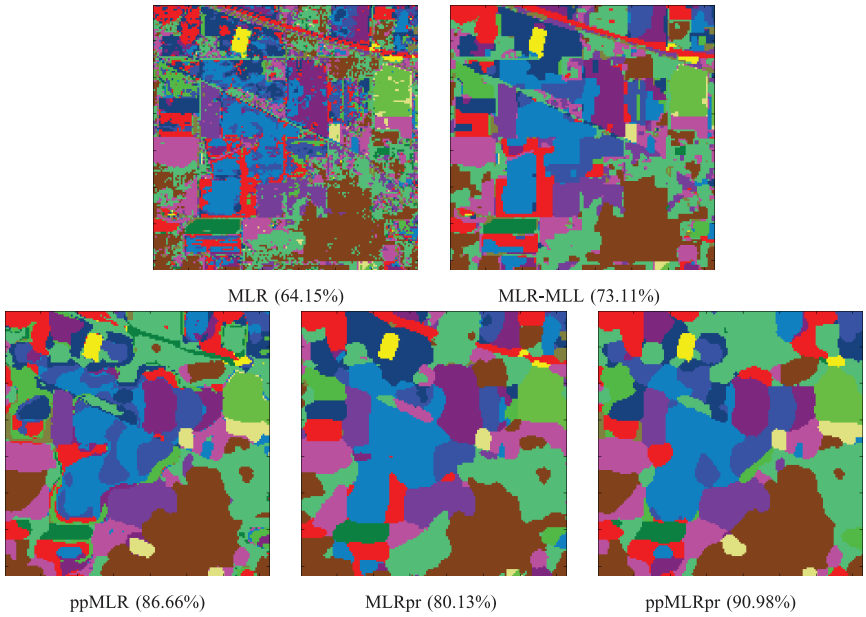


Fig. 7. Classification maps obtained by different methods for the AVIRIS Indian Pines scene (PP refers to preprocessing and the OAs are reported in the parentheses).

that due to the presence of map $\varepsilon$, the smoothness is not applied across the discontinuities.

At this point, we would like to make reference to edge-preserving image restoration methods such as those based on total variation (TV) [16] or based on anisotropic diffusion (AD) [34]. In both cases (i.e., TV and AD), the objective is similar to ours: to apply strong smoothing in areas away from edges and avoid smoothing the edges. However, in our case, we know the edges in advance, which is not the case of those methods. This is a considerable advantage, which results from the availability of many hyperspectral bands.

Problem (4) is strictly convex and, therefore, has a unique solution. Herein, we apply a projected iterative Gauss Seidel scheme, which consists in iteratively minimizing the objective function in (4) with respect to each optimization variable $u_i(k)$ and, after a complete sweep, project on the probabilities at each pixel onto the probability simplex. The obtained algorithm is shown in Algorithm 1, where $iters$ is the number of maximum iterations defined in advance, $Err^{(t+1)} = \frac{\|\mathbf{u}^{t+1}-\mathbf{u}^t\|}{\|\mathbf{u}^t\|}$ is an error parameter and $\tau$ is the error threshold parameter controlling the degree of convergence.

**Algorithm 1.** Discontinuity Preserving Relaxation (PR)

**Input:** $\mathbf{p}$, $\varepsilon$, $\lambda$, *iters*, $Err^{(1)} = \|\mathbf{p}\|$, $\tau$
**Output:** $\mathbf{u}$
$t := 1$
**while** $Err^{(t+1)} - Err^{(t)} \leq \tau$ **or** $t \leq iters$ **do**
  **for** $k := 1$ to $K$ **do**
    $u_i^{(t+1)}(k) = \frac{(1-\lambda)p_i(k)+\lambda\sum_{j\in\partial_i}\varepsilon_j u_j^{(t)}(k)}{(1-\lambda)+\lambda\sum_{j\in\partial_i}\varepsilon_j}$
  **end for**
  $u_i^{(t+1)} = u_i^{(t+1)} / \sum_{k=1}^{K} u_i^{(t+1)}(k)$
  $Err^{(t+1)} = \frac{\|\mathbf{u}^{t+1}-\mathbf{u}^t\|}{\|\mathbf{u}^t\|}$
**end while**

At this point, we would like to call attention to the fact that, apart from the constraints used in (4) linked to the fact that we are estimating probabilities, the ratione used to carry out PR can be used to denoise the original bands of the hyperspectral image $\mathbf{X}$ ensuring the preservation of the discontinuities. The correspondent algorithm, which may be used as a preprocessing step, is shown in Algorithm 2, where $Err_{:k}$ denotes the error parameter for the $k$th band, $\tilde{\mathbf{x}}_{:k}$ is the processed image of the $k$th band, which corresponds to the original $k$th band, i.e., $\mathbf{x}_{:k} = [x_{1k}, \ldots, x_{nk}]$. Finally, we empirically find out that both algorithms converge very fast, say, less than 20 iterations.

**Algorithm 2.** Discontinuity Preserving Relaxation (Preprocessing)

**Input:** $\mathbf{X}$, $\varepsilon$, $\lambda$, *iters*, $Err^{(1)} = \|\mathbf{X}\|$, $\tau$
**Output:** $\tilde{\mathbf{X}}$
**for** $k := 1$ to $d$ **do**
  $t := 1$
  $Err_{:k}^{(1)} = Err^{(1)}$
  **while** $Err_{:k}^{(t+1)} - Err_{:k}^{(t)} \leq \tau$, **or** $t \leq iters$ **do**
    $\tilde{x}_{ik}^{(t+1)} = \frac{(1-\lambda)x_{ik}+\lambda\sum_{j\in\partial_i}\varepsilon_j\tilde{x}_{jk}^{(t)}}{(1-\lambda)+\lambda\sum_{j\in\partial_i}\varepsilon_j}$
    $Err_{:k}^{(t+1)} = \frac{\|\tilde{\mathbf{x}}_{:k}^{t+1}-\tilde{\mathbf{x}}_{:k}^{t}\|}{\|\tilde{\mathbf{x}}_{:k}^{t}\|}$
  **end while**
**end for**

## III. Experimental Results and Discussion

In this section, we use both simulated and real hyperspectral data to evaluate the proposed approach. The main goal of using simulated data set is to evaluate the performance of the algorithm in a fully controlled environment, while the experiments with real experiments are intended to provide a quantitative

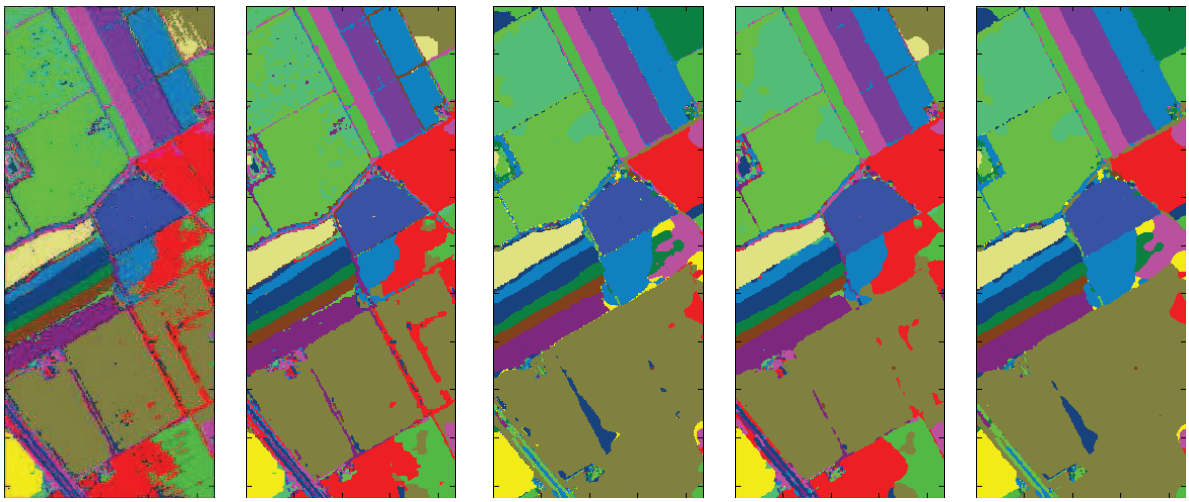


Fig. 8. Classification maps obtained by different methods for the AVIRIS Salinas scene (PP refers to preprocessing and the OAs are reported in the parentheses).

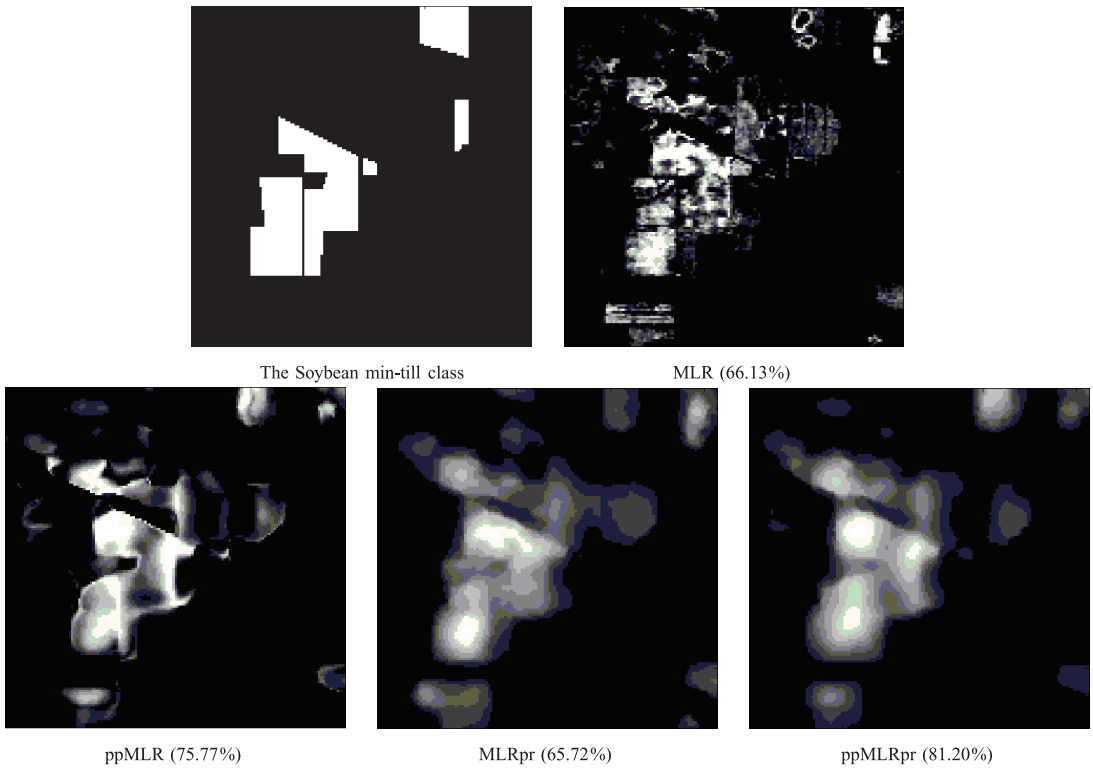


Fig. 9. Probability image of the class *Soybean min-till* resulted from the proposed methods.

evaluation of the method in real-analysis scenarios. For simplicity, in this section, we refer to spatial preprocessing as "pp," while "MLL" and "pr" denote MLL-based [35], [36] and PR-based spatial relaxation, respectively.

### A. Experiments With Simulated Data

In our first experiment, we use a simulated image with eight classes and $100 \times 100$ pixels, in which the spatial distribution is extracted from a real image and the spectral signatures are selected from the U.S. Geological Survey (USGS) digital spectral library.[1] The ground-truth image and the spectral signatures of eight randomly selected mineral signatures allocated to the main classes are shown in Fig. 2. We considered the following linear mixture model for generating a simulated mixed pixel:

$$\mathbf{x}_i = \sum_{j \in \partial_i} \mathbf{m}^{(j)} \gamma_j + \mathbf{n}_i \tag{6}$$

where $\mathbf{m}^{(l)}, l = 1, \ldots, 8$ are spectral signatures obtained randomly from the USGS spectral library, and $\gamma_j$, which follows a

[1] [Online]. Available: http://speclab.cr.usgs.gov/spectral-lib.html

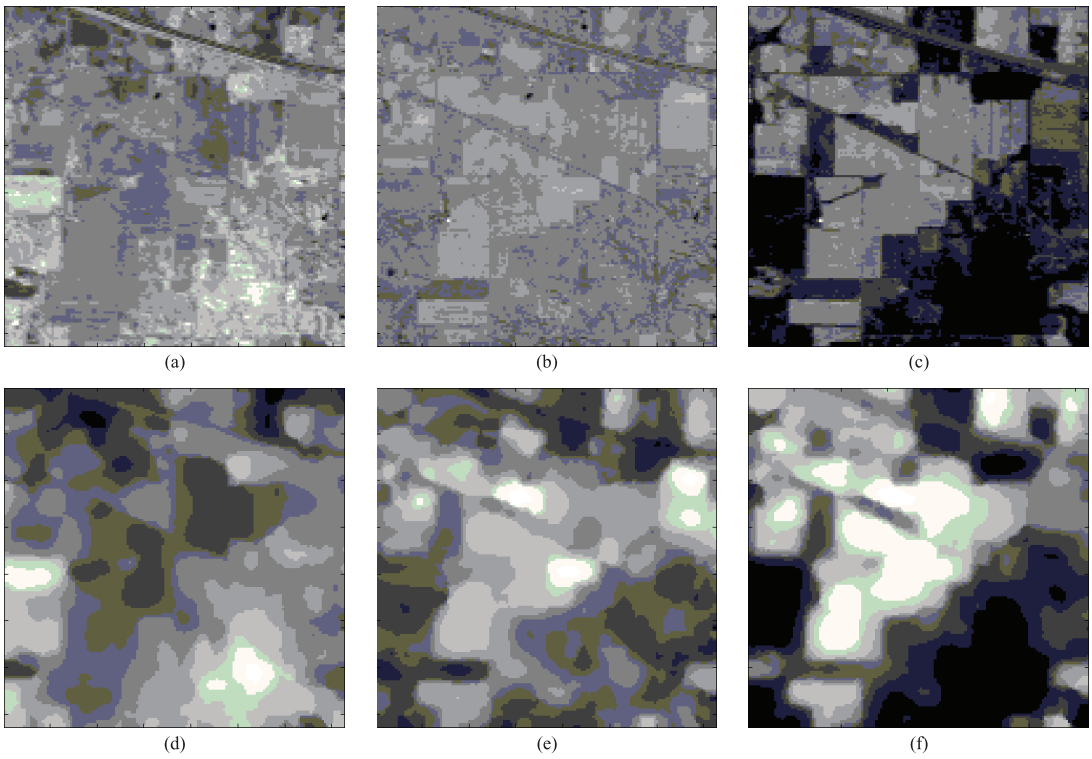


Fig. 10. Bands of numbers 50, 100, and 150 of the original hyperspectal image (a–c) before and (d–f) after preprocessing.

random distribution with $0 \leq \gamma_j \leq 1$ and $\sum_{j \in \partial_i} \gamma_j = 1$, determines the abundance of the signatures which contribute to the mixture model. Note that, here, the maximum abundance value of $\gamma_j$ is assigned to the objective class according to the ground-truth image. $\partial_i$ is a neighborhood with a specific size around the central pixel $i$ over the considered ground-truth image. $\partial_i$ determines a set of class labels to contribute in the mixture. So that the pixels near the borders of the regions are generated by mixtures of different class labels and the pixels far from the borders are considered pure.

In our simulations, we set the size of the neighborhood to $9 \times 9$ pixels. For illustrative purposes, Fig. 2(c) shows an example of the abundance maps associated with the eight classes of the simulation image. In each pixel of the scene, the fractional abundances vary from 0% (black color) to 100% (white color) and sum to unity. Note that, using the suggested procedure, signature abundance is not constant over class regions and the pixels closer to the discontinuities are more heavily mixed, as expected in real scenarios. Zero-mean Gaussian noise with covariance $\sigma^2\mathbf{I}$, i.e., $\mathbf{n}_i \sim \mathcal{N}(0, \sigma^2\mathbf{I})$, is finally added to the generated synthetic image. For each class, we randomly chose 10 samples (in total 80 samples) from the ground-truth image in Fig. 2(a) for training purposes.

We have conducted different experiments with the simulated hyperspectral image described earlier. These experiments have been carefully designed to analyze several relevant aspects of our proposed method in a fully controlled environment. All of the results reported in this paper with the simulated data sets were obtained after 30 Monte Carlo runs in which we randomly select eight different materials and also randomly select different training sets.

### B. Impact of Parameter $\lambda$

In our first experiment, we analyze the impact of the tunable parameter $\lambda$ intended to control the relative impact of both the terms in the proposed relaxation scheme. It should be noted that, if $\lambda = 0$, only the first term is considered and the method remains as the original MLR algorithm. If $\lambda = 1$, only the smoothing term is used. Fig. 3(a) plots the obtained OA results as a function of $\lambda$, with $\sigma = 0.1$ and the maximum number of iterations as 20. From Fig. 3(a), we can conclude that the relaxation performance indeed depends on the setting of $\lambda$. However, even with $0.7 \leq \lambda \leq 0.9$, the proposed relaxation method leads to significant classification results for the considered problem. Fig. 3(b) shows convergence of the proposed PR method with different values of $\lambda$ parameter. As can be observed, the proposed approach converged very fast, i.e., less than 20 iterations, for all cases with different value of $\lambda$. Hence, in this paper, we set the parameter $\lambda = 0.9$ and the maximum number of iterations as 20 for the remaining simulated experiments.

### C. Impact of Noise

In the other experiment with simulated data, we evaluate the impact of noise on the proposed relaxation method. Table I shows the classification results obtained by the proposed approach using different values of noise standard deviation $\sigma$. Several conclusions can be obtained from Table I. First and foremost, it is remarkable that the proposed approach, which carefully uses the local relationship among neighboring pixels, has improved the performance of MLR-based classification

TABLE IV
OA AND AA CLASSIFICATION ACCURACIES (%) OBTAINED BY DIFFERENT METHODS FOR THE ROSIS PAVIA UNIVERSITY SCENE

| Class | Train/test | Methods | | | | |
|---|---|---|---|---|---|---|
| | | MLR | MLR-MLL | Proposed relaxation methods | | |
| | | | | ppMLR | MLRpr | ppMLRpr |
| Asphalt | 15 / 6631 | 61.32 (5.93) | 75.38 (7.74) | 76.81 (6.95) | 89.03 (6.09) | 78.14 (8.38) |
| Meadows | 15 / 18649 | 66.96 (9.98) | 79.36 (12.78) | 90.87 (5.14) | 80.74 (12.94) | 90.36 (6.37) |
| Gravel | 15 / 2099 | 57.06 (10.19) | 57.39 (17.56) | 79.47 (7.42) | 71.59 (18.08) | 81.55 (9.32) |
| Trees | 15 / 3064 | 89.64 (9.79) | 91.72 (12.68) | 86.11 (8.44) | 84.68 (15.58) | 83.49 (8.49) |
| Metal sheets | 15 / 1345 | 96.02 (5.15) | 98.35 (4.05) | 97.17 (3.11) | 97.99 (3.55) | 97.73 (4.44) |
| Bare soil | 15 / 5029 | 45.45 (8.64) | 42.48 (18.93) | 68.79 (12.56) | 59.58 (17.55) | 70.61 (14.96) |
| Bitumen | 15 / 1330 | 78.92 (9.50) | 89.55 (13.16) | 81.48 (13.29) | 96.42 (11.89) | 80.09 (12.51) |
| Bricks | 15 / 3682 | 70.67 (8.09) | 90.10 (5.13) | 87.65 (5.50) | 92.25 (6.35) | 88.88 (4.33) |
| Shadows | 15 / 947 | 98.50 (2.50) | 99.50 (1.26) | 91.38 (6.36) | 99.90 (0.11) | 92.41 (5.13) |
| OA | | 67.00 (3.55) | 76.50 (5.35) | 84.83 (2.70) | 81.81 (5.25) | 85.05 (2.87) |
| AA | | 73.84 (1.95) | 80.42 (3.49) | 84.41 (3.00) | 85.80 (4.12) | 84.80 (2.69) |
| $\kappa$ | | 58.23 (3.72) | 69.46 (6.27) | 80.06 (3.44) | 76.40 (6.34) | 80.36 (3.62) |

accuracy. Clearly, the performance of the proposed relaxation method decreases as $\sigma$ increases. When the noise is low, using the proposed method as PR shows better performance than preprocessing, however, in high noise images, relaxation method as preprocessing shows significant improvements. Note that the results obtained using both preprocessing and PR, i.e., ppMLRpr, are always superior.

### D. Real Experiments

Three different hyperspectral images were used for the experiments. These data sets have different characteristics and contexts (two agricultural areas and an urban area, with different spectral and spatial resolutions).

1) The first one is the well-known AVIRIS Indian Pines scene [see Fig. 4(a)], collected over Northwestern Indiana in June 1992 [3]. The scene is available online[2] and contains $145 \times 145$ pixels and 220 spectral bands between 0.4 and 2.5 μm. The spatial resolution of the scene is 20 m/pixel. A total of 20 spectral bands were removed prior to experiments due to noise and water absorption in those channels. The ground-truth image displayed in Fig. 4(b) contains 10 366 samples and 16 mutually exclusive classes having from 20 to 2468 samples. These data are widely used as a benchmark for testing the accuracy of hyperspectral data classification algorithms, mainly because it constitutes a challenging classification problem due to the presence of mixed pixels in available classes, and also because of the unbalanced number of available labeled pixels per class.
2) The second image considered in experiments is the AVIRIS Salinas image, collected over the Valley of Salinas, Southern California, USA, in 1998. It contains $217 \times 512$ pixels and 204 spectral bands and is characterized by 3.7 m/pixel spatial resolution. Fig. 5(a) shows the ground-truth map with 16 mutually exclusive classes. Due to the spectral similarity of most classes, this data set also represents a very challenging classification problem.
3) The third image used in our experiments was collected by the ROSIS instrument. These data were acquired over the urban area of the University of Pavia, Pavia, Italy. The flight was operated by the Deutschen Zentrum for Luftund Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Commission. The image size in pixels is $610 \times 340$, with very high spatial resolution of 1.3 m/pixel. The number of data channels in the acquired image is 103 (with spectral range from 0.43 to 0.86 μm). Fig. 6(a) shows a false color composite of the image, while Fig. 6(c) shows nine ground-truth classes of interest, which comprise urban features, as well as soil and vegetation features. In the original data set, out of the available ground-truth pixels, 3921 were used for training [see Fig. 6(b)] and 42 776 samples were used for testing.

Moreover, for the three considered hyperspectral images, the discontinuities maps were generated using (5), which have been shown in Figs. 4(c), 5(c), and 6(d), respectively.

*1) Experimental Setup:* Before describing our results, it is important to first report the parameters and main considerations in our experiments. For the experiments with the AVIRIS Indian Pines and Salinas images, the training samples were randomly

[2][Online]. Available: https://engineering.purdue.edu/~biehl/MultiSpec/

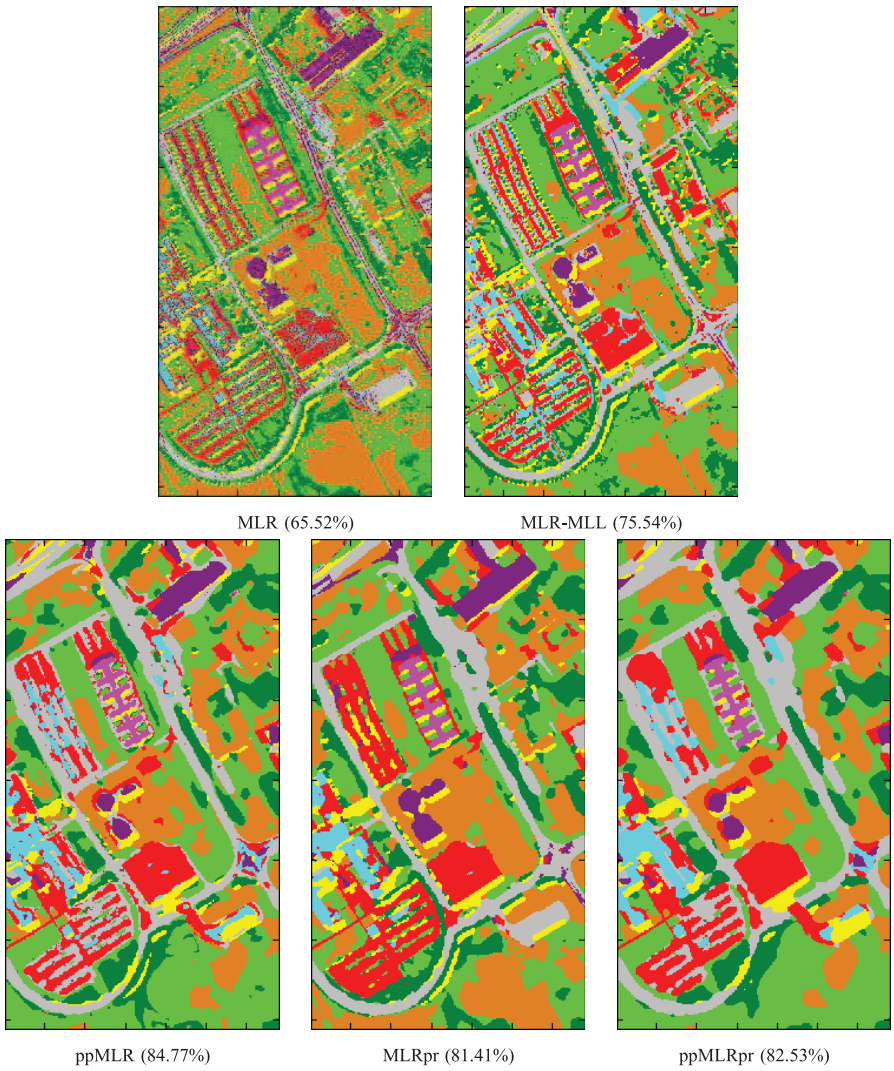


Fig. 11. Classification maps obtained by different methods for the ROSIS Pavia University scene (PP refers to preprocessing and the OAs are reported in the parentheses).

selected from the available ground truth and the remaining samples are used for validation. However, for the ROSIS Pavia University image, small subsets of the original training samples were used. Note that we constructed very small training sets by selecting only 15 labeled samples per class. Obtaining a good performance of the classifier in the presence of very limited training samples is very important as labeled training data are often difficult and expensive to be collected in practice. Concerning the $\lambda$ parameter of the proposed relaxation methods, we considered $\lambda = 0.9$. For the stopping, the maximum number of iterations in all experiments was set to 20. These settings, although suboptimal, lead to very good classification performance. Note that, in all the experiments, the results reported correspond to the average of the results obtained after 20 Monte Carlo runs.

*2) Experiments for AVIRIS Images:* Tables II and III report the obtained classification accuracies for the AVIRIS Indian Pines and Salinas images, respectively. The metrics reported are the individual classification accuracies, as well as the OA, AA, and $\kappa$ statistic. These tables provide the results for each step of the proposed spectral–spatial relaxation method. Moreover, the results have been compared with the recently proposed spectral–spatial classification method MLR$sub$MLL [24]. From the results reported in Tables II and III, we can conclude that our proposed method exhibits state of the art. For instance, Table II reveals that the proposed relaxation method, i.e., ppMLRpr obtained an OA of 91.05% for the AVIRIS Indian Pines image, which contrasts with the OA of 64.30% obtained by the MLR-based classifier. Compared to MLR-MLL, the OA achieved by the presented method improved by about 16% the OA obtained by this method. For the AVIRIS Salinas image, we obtained comparable results.

A more detailed investigation of individual class accuracies is important to assess quantitatively the impact of the proposed method on class separability. As indicated in Tables II and III, the improvement is quite significant for the sets of similar class labels. For example, the classification accuracies obtained by the MLR method with preprocessing for the classes *Corn-no till*, *Corn-min till*, and *Corn* in the AVIRIS Indian Pines scene were 82.49%, 86.60%, and 94.94%, respectively, which are 32.13%, 27.98%, and 25.24% higher than those obtained by the MLR algorithm. It is also remarkable

TABLE V
STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION ACCURACIES

| Data | $Z$ (classification method 1/classification method 2) | | | |
|---|---|---|---|---|
| | ppMLRpr/MLR | ppMLRpr/MLR-MLL | ppMLRpr/ppMLR | ppMLRpr/MLRpr |
| Indian Pines | 48.44 | 36.86 | 12.90 | 28.20 |
| Salinas | 56.61 | 37.64 | 20.06 | 22.88 |
| Pavia University | 64.63 | 34.93 | 13.63 | 21.05 |

TABLE VI
OA AND AA CLASSIFICATION ACCURACIES AFTER COMPARING THE PROPOSED ALGORITHM WITH SOME IMPORTANT RELAXATION BASED ALGORITHMS FOR INDIAN PINES (IP), SALINAS (S), AND PAVIA UNIVERSITY (PU) DATA SETS

| Data | Acc. | Relaxation-based classification methods | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | SVM-MLL | LORSAL-MLL | LORSAL-SegSALSA | SMLR-SpTV | ppMLR | MLRpr | ppMLRpr |
| IP | OA | 78.39 (5.36) | 74.95 (2.93) | 76.35 (2.91) | 77.12 (1.73) | 88.36 (1.67) | 80.67 (3.12) | 91.05 (1.87) |
| | AA | 87.81 (2.29) | 85.56 (0.99) | 87.12 (0.83) | 86.92 (0.99) | 93.41 (1.03) | 85.11 (2.52) | 93.91 (2.27) |
| S | OA | 88.94 (5.82) | 89.78 (1.34) | 90.10 (1.28) | 93.22 (1.78) | 93.30 (1.70) | 91.26 (2.30) | 93.79 (4.46) |
| | AA | 94.61 (2.19) | 94.26 (0.58) | 94.48 (0.56) | 91.11 (1.37) | 96.43 (0.81) | 95.72 (1.05) | 96.37 (4.48) |
| PU | OA | 78.32 (6.65) | 73.08 (5.53) | 76.09 (5.74) | 78.43 (6.62) | 84.83 (2.70) | 81.81 (5.25) | 85.05 (2.87) |
| | AA | 84.03 (3.81) | 81.09 (2.26) | 83.95 (2.25) | 87.05 (3.12) | 84.41 (3.00) | 85.80 (4.12) | 84.80 (2.69) |

that the accuracies for these classes increased in 3.05%, 3.30%, and 2.50%, respectively, when the proposed MLRpr method with preprocessing was used. The same conclusion can be obtained after comparing the individual class accuracies for the sets of *Grass/trees*, *Grass/pasture*, *Grass/pasture-mowed* and *Soybeans-no till*, *Soybeans-min till*, *Soybeans-clean till*. For the AVIRIS Salinas image, it is also possible to consider other sets of similar classes and obtain the same conclusion. For instance, pixelwise classifier MLR obtained low accuracies for class *Vinyard-untrained*, i.e., 63.90%. However, after applying the preprocessing method the accuracy for this class was increased by 23.42%. This improvement is significant because, e.g., the MLR+MLL method obtained 71.53% accuracy for this class, which is just 7.63% higher than MLR result. It is also noticeable that the accuracy obtained by the proposed method ppMLRpr for the class *Vinyard-untrained* is 90.55%, which is 26.65% higher than the result obtained by the MLR algorithm.

For illustrative purposes, Figs. 7 and 8 show the obtained classification maps for the AVIRIS Indian Pines and Salinas data sets. Each of the maps corresponds to one of the 30 Monte Carlo experiments, which were averaged to produce the results reported in Tables II and III. From Figs. 7 and 8, it can be seen that using spatial information (both at the preprocessing and postprocessing level) can lead to more homogeneous regions in classification maps, when compared to the pixelwise classification maps. Most importantly, the proposed method exhibits very good performance in the task of delineating the borders of classes of interest.

Figs. 9 and 10 illustrate the performance of the proposed relaxation method in detailed. For example, Fig. 9 shows the changes of probabilities of class *Soybean min-till* for all the pixels. We can conclude that our proposed method preserve discontinuities during relaxation process. Similarly for preprocessing (Fig. 10), the proposed method obviously smoothed the original hyperspectral image while it considered edge information.

### E. Experiments for the ROSIS Pavia University Image

Table IV details the classification results obtained for the ROSIS Pavia University scene. Several conclusions can be obtained from this table. First and foremost, it is remarkable that the proposed relaxation approach exhibited very good performance using very limited number of training samples. For instance, our proposed method obtained an OA of 85.05%, which is 18.05% higher than the one obtained by the MLR algorithm, whereas the MLR-MLL obtained an OA of 76.50%, which is 9.5% higher than the result obtained by the MLR algorithm. For illustrative purposes, Fig. 11 shows the obtained classification maps for the Pavia University data set.

### F. Evaluation of the Statistical Significance

The McNemar's test [37] is a widely used technique in the remote sensing community for evaluating the statistical significance of the difference in accuracy between two classification methods. 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 ($Z < 0$). Table V reveals the differences in classification accuracies between the proposed ppMLRpr method and the other classification methods used in Tables II–IV are statistically significant. The significant differences in

accuracy between the case of using the ppMLRpr method and MLR method were the most for the Pavia University data and the least for the Indian Pines data. Furthermore, for all the the data sets, the differences in accuracy between the ppMLRpr method and ppMLR method were the least significant.

### G. Comparison With Other Algorithms

Using the same data sets in Tables II–IV, Table VI provides a comparison of the proposed relaxation algorithm with other successful methods such as probabilistic SVM [38], [39], and LORSAL [40] with multilevel logistic spatial prior (denoted as SVM-MLL and LORSAL-MLL, respectively), LORSAL with segmentation via the constraint split augmented lagrangian shrinkage (SegSALSA) algorithm [15], and SMLR classifier [26] following with SpATV relaxation algorithm [27]. Note that, for all the tested methods, we carefully optimized the related parameters. From Table VI, we can conclude that ppMLRpr obtained significant results in terms of OA and AA in comparison with those obtained by the other tested algorithms. Moreover, if we compare the results obtained by MLRpr with the results obtained by the other tested algorithms in view of postprocessing, we can conclude that the proposed PR method exhibits better performance.

## IV. Conclusion

In this work, we have developed a new methodology for spectral–spatial classification of remotely sensed hyperspectral scenes. The inclusion of both spectral and spatial information is an important aspect, as it has been shown that the joint exploitation of the information in both domains can significantly improve the final classification results. The main features of our proposed approach can be summarized as follows. First, it provides spatially homogeneous regions after probabilistic classification, thus exploiting the intrinsic correlation which exists between neighboring pixels to improve the final classification results. Second, it specifically models the pixels at the borders of the regions to provide a better delineation of the classified objects. In other words, our proposed approach is able to provide accurate spectral–spatial classification while preserving the edges and the boundaries between classes, which is quite important as the inclusion of spatial regularizers tends to blur the class boundaries and provide nonsmooth delineations. Our experimental results, conducted using a variety of (simulated and real) hyperspectral scenes and spectral–spatial classification strategies, indicate that the proposed approach provides state-of-the-art classification results. In particular, the proposed method provides high classification accuracies when very limited training samples are used, and also provides accurate delineation of classes at the borders.

## V. Acknowledgment

The authors would like to thank the Editors and the Anonymous Reviewers for their detailed and highly constructive comments, which greatly helped us to improve the technical quality and presentation of our paper.

## References

[1] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

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

[3] D. A. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2003.

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

[5] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. CalpeMaravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[6] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 297–301, Apr. 2009.

[7] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 10, pp. 3747–3762, Oct. 2010.

[8] 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.

[9] 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.

[10] Y. Wang, R. Niu, and X. Yu, "Anisotropic diffusion for hyperspectral imagery enhancement," *IEEE Sens. J.*, vol. 10, no. 3, pp. 469–477, Mar. 2010.

[11] I. Yildirim, O. K. Ersoy, and B. Yazgan, "Improvement of classification accuracy in remote sensing using morphological filter," *Adv. Space Res.*, vol. 36, no. 5, pp. 1003–1006, 2005.

[12] P.-F. Hsieh and D. Landgrebe, "Classification of high dimensional data," Ph.D. dissertation, School Elect. Comput. Eng., Purdue Univ., West Lafayette, IN, USA, 1998.

[13] J. Richards, D. Landgrebe, and P. Swain, "Pixel labeling by supervised probabilistic relaxation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-3, no. 2, pp. 188–191, Mar. 1981.

[14] L. Pelkowitz, "A continuous relaxation labeling algorithm for Markov random fields," *IEEE Trans. Syst. Man Cybern.*, vol. 20, no. 3, pp. 709–715, May/Jun. 1990.

[15] J. Bioucas-Dias, F. Condessa, and J. Kovacevic, "Alternating direction optimization for image segmentation using hidden Markov measure field models," in *Proc. IS&T/SPIE Electron. Imag.*, 2014, p. 90 190P.

[16] L. Sun, Z. Wu, J. Liu, and Z. Wei, "Supervised hyperspectral image classification using sparse logistic regression and spatial-TV regularization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 1019–1022.

[17] R. Dubes, A. Jain, S. Nadabar, and C. Chen, "MRF model-based algorithms for image segmentation," in *Proc. 10th Int. Conf. Pattern Recog.*, 1990, vol. 1, pp. 808–814.

[18] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, 3rd ed. Berlin, Germany: Springer-Verlag, 2009.

[19] X. Jia and J. Richards, "Managing the spectral–spatial mix in context classification using Markov random fields," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 311–314, Apr. 2008.

[20] X. Huang, Q. Lu, L. Zhang, and A. Plaza, "New postprocessing methods for remote sensing image classification: A systematic study," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7140–7159, Mar. 2014.

[21] 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.

[22] 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.

[23] G. Moser and S. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.

[24] 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.
[25] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral-spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.
[26] B. Krishnapuram, L. Carin, M. Figueiredo, and A. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 6, pp. 957–968, Jun. 2005.
[27] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral–spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.
[28] B. R. Cuenca, J. A. Malpica, and M. C. Alonso, "A spatial contextual postclassification method for preserving linear objects in multispectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 395–406, Jan. 2013.
[29] A. Rosenfeld, R. Hummel, and S. Zucker, "Scene labelling by relaxation algorithms," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-6, no. 6, pp. 420–433, Jun. 1976.
[30] 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.
[31] 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.
[32] D. Bohning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, 1992.
[33] M. Khodadadzadeh, J. Li, A. Plaza, and J. Bioucas-Dias, "A subspace based multinomial logistic regression for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2105–2109, Dec. 2014.
[34] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
[35] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
[36] S. Bagon. (2006, Dec.). *MATLAB Wrapper for Graph Cut* [Online]. Available: http://www.wisdom.weizmann.ac.il/~bagon/matlab.html
[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] J. Platt, "Probabilities for support vector machines," in *Advances in Large Margin Classifiers*. Cambridge, MA, USA: MIT Press, 2000, pp. 61–74.
[39] 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.
[40] J. Bioucas-Dias and M. Figueiredo, "Logistic regression via variable splitting and augmented lagrangian tools," Instituto Superior Tecnico, Lisbon, Portugal, Tech. Rep., 2009.

**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 Telecomunicações, Instituto Superior Técnico (IST), Universidade Técnica de Lisboa, Lisbon, Portugal, in 2011.

From 2007 to 2011, she was a Marie Curie Research Fellow with the Departamento de Engenharia Electrotcnica 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 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*.

**Mahdi Khodadadzadeh** (S'10) received the B.S. degree in electrical engineering from Sadjad University of Technology, Mashhad, Iran, in 2008, and the M.Sc. degree in electrical engineering and communications from Tarbiat Modares University, Tehran, Iran, in 2011, and the Ph.D. degree in technology of computers and communications from Escuela Politecnica, University of Extremadura, Caceres, Spain, in 2015.

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 has been a Manuscript Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCES AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING AND IEEE GEOSCIENCES AND REMOTE SENSING LETTERS.

**Antonio Plaza** (M'05–SM'07–F'15) was born in Caceres, Spain, in 1975. 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), one of the most productive research groups working on remotely sensed hyperspectral data processing worldwide. He has been the Advisor of 12 Ph.D. dissertations and more than 30 M.Sc. dissertations. He was the Coordinator of the Hyperspectral Imaging Network. He has authored more than 500 publications, including 152 journal papers (more than 100 in IEEE journals), 22 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 (CRC Press/Taylor & Francis) and guest edited 9 special issues on hyperspectral remote sensing for different journals. He has reviewed more than 500 papers for over 50 different journals. His research interests include hyperspectral data processing and parallel computing of remote sensing data.

Dr. Plaza served as an Associate Editor from 2007 to 2012 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 served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012, and is currently serving as a 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. He was the recipient of the recognition of the Best Reviewers of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS (in 2009), the recognition of Best Reviewers of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (in 2010), 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 also 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.

**Xiuping Jia** (M'93–SM'03) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, NSW, Australia, in 1996.

Since 1988, she has been with the School of Information Technology and Electrical Engineering, University of New South Wales, Canberra Campus, Canberra, BC, Australia, where she is currently a Senior Lecturer. She is also a Guest Professor with Harbin Engineering University, Harbin, China, and an Adjunct Researcher with China National Engineering Rsearch Center for Informaiton Technology in Agriculture, Beijing, China. She is the coauthor of the remote sensing textbook titled *Remote Sensing Digital Image Analysis* (Springer-Verlag, 1999, 3rd ed and 2006, 4th ed.). Her research interests include remote sensing and imaging spectrometry.

Dr. Jia is an Editor of the ANNALS OF GIS and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

**José M. Bioucas-Dias** (S'87–M'95) received the E.E., M.Sc., Ph.D., and "Agregado" degrees from the Instituto Superior Técnico (IST), Technical University of Lisbon (TULisbon, now University of Lisbon), Lisbon, Portugal, in 1985, 1991, 1995, and 2007, respectively, all in electrical and computer engineering.

Since 1995, he has been with the Department of Electrical and Computer Engineering, IST, where he was an Assistant Professor from 1995 to 2007 and an Associate Professor since 2007. Since 1993, he has been also a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações. He has authored or coauthored more than 250 scientific publications including more than 70 journal papers (48 of which published in IEEE journals) and 180 peer-reviewed international conference papers and book chapters. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing.

Dr. Bioucas-Dias was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS (1997–2000) and the IEEE TRANSACTIONS ON IMAGE PROCESSING and he is an an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Guest Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING for the Special Issue on *Spectral Unmixing of Remotely Sensed Data*, of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING for the Special Issue on *Hyperspectral Image and Signal Processing*, of the IEEE SIGNAL PROCESSING MAGAZINE for the Special Issue on *Signal and Image Processing in Hyperspectral Remote Sensing*, of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING for the *Advances in Hyperspectral Data Processing and Analysis*, and of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE for the Special Issue on *Advances in Machine Learning for Remote Sensing and Geosciences*. He was the General Co-Chair of the 3rd IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS'2011) and has been a member of program/technical committees of several international conferences.