Convex Formulation for Multiband Image Classification With Superpixel-Based Spatial Regularization

Yi Liu , *Student Member, IEEE*, Filipe Condessa, *Member, IEEE*, José M. Bioucas-Dias, *Fellow, IEEE*, Jun Li, *Member, IEEE*, Peijun Du , *Senior Member, IEEE*, and Antonio Plaza, *Fellow, IEEE*

Abstract—Superpixels are a powerful device to characterize the spatial–contextual information in remotely sensed hyperspectral image (HSI) interpretation. However, the exploitation of superpixels in classification problems is not straightforward, often leading to unbearable NP-hard discrete integer optimization problems. In this paper, we attack this hurdle by leveraging on a convex relaxation of the original integer optimization problem, which opens the door to include oversegmented superpixel-based regularizers. Specifically, we develop a new method for generating oversegmented superpixels. Then, we introduce a family of convex regularizers in the form of graph total variation, which promotes the same labeling in each superpixel. Vectorial total variation is also included in order to promote piecewise smoothness and align discontinuities along the class boundaries. The solution of the obtained convex optimization problem is computed with the split-augmented Lagrangian shrinkage algorithm. Experiments on HSIs yield classification maps with precise boundaries and inner consistency inside oversegmented superpixels, leading to the state-of-the-art classification accuracies.

Index Terms—Convex relaxation, graph total variation (GTV), oversegmented superpixels, remote sensing image classification, split-augmented Lagrangian shrinkage algorithm (SALSA), vectorial total variation (VTV).

NOMENCLATURE

AVIRIS	Airborne Visible Infrared Imaging Spectrometer.
QB	QuickBird.
ROSIS	Reflective optics system imaging spectrometer.

Fig. 1. Examples of superpixels with multiscale sizes and their performance in image contextual classification.

DPR	Discontinuity preserving relaxation.
GTV	Graph total variation.
MAP	Maximum a posteriori.
MLR	Multinomial logistic regression.
MRF	Markov random field.
LP	Linear program.
PDE	Partial differential equations.
SALSA	Split-augmented Lagrangian shrinkage algorithm.
SuperSALSA	Superpixel-based SALSA.
SVD	Singular value decomposition.
VTV	Vectorial total variation.

I. INTRODUCTION

SUPERVISED classification is an important task for hyperspectral remotely sensed data exploitation, which assigns a set of class labels to each pixel in the scene, given an available training reference [1]. In this context, spatial information has been shown capable of greatly improving the classification performance from the viewpoint of statistical accuracy and mapping effectiveness [2]–[8]. For the example of Fig. 1, the classification map can be remarkably improved with better boundary recall and outliers removed after the inclusion of spatial information that comes from the superpixels of multisize scales. The inclusion of spatial information is often tackled by means of image segmentation. As a discrete problem, segmentation aims to partition an image into multiple segments, which consist of a set of pixels that share some common characteristics (i.e., they belong to the same object or may

have the same surface orientation). As an important source of spatial information, image segments or superpixels play a significant role in remote sensing image analysis [1], machine vision [9], medical imaging [10], and so on. The pixels comprising the same superpixel (especially the oversegmented ones) are generally believed to share highly similar characteristics, such as class labels [11]. At the superpixel level, hyperspectral image (HSI) classification can also be processed much faster than pixelwise techniques [5], [11].

Many techniques and methods have been developed to deal with the image segmentation problem in the spatial domain, such as thresholding [9], clustering, compression/ histogram/ edge-based algorithms [12], [13], region growing [14], integer optimization by graph-cuts (GC) [15], [16], variational methods [17], [18], as well as Bayesian theory-based algorithms such as MRFs [19], [20]. However, image segmentation usually leads to an integer optimization problem that is NP-hard, and thus hard to be solved exactly. This is because the label image is naturally a discrete representation of the original image. Actually, in the context of supervised image segmentation, apart from a few examples, almost all functions associated with a realistic model are nonconvex and even NP-hard [21]. This means that they are hard to solve, and hence, a direct minimization may lead to poor local minima. A popular and well-established paradigm for modeling these problems is function or energy minimization, where the spatial information is tackled with the Potts model in the MRF community or the minimal partition problem in the PDE community. More interesting details regarding binary labeling problems are experimentally surveyed in the work of Klodt *et al.* [22], and the more general case of multilabel problems is reported in the work of Nieuwenhuis *et al.* [21].

In the MRF community, it is often assumed that labels of neighboring pixels follow a Gibbs distribution [19]. In this context, GC algorithms have been developed to model the resulting integer optimization problem. Boykov and Jolly [23] first present the optimal solution of image binary segmentation by GC algorithms that solve max-flow problems. Recent efforts [23]–[26] attack this problem under a discrete optimization framework, by introducing prior regularizers to promote the spatial patterns of the label image and approximate the solutions by GC algorithms. Kohli *et al.* [27] designed a novel model for enforcing label consistency, which is able to combine multiple image segmentations in a principled manner based on the higher order conditional random fields (CRFs). Following this line, an exact solution can be found for the binary case and can only be approximated for the multilabel case. Other techniques also attack the labeling problem through block coordinate descent [28] in a dual objective, such as map LP-relaxation [29], max-sum duality [30], tree reweighted schemes [16], and quadratic pseudo-Boolean optimization [31], [32], among many others.

In the hyperspectral remote sensing image classification literature, the concept of superpixel has been widely explored and utilized. Superpixels, especially oversegmented ones, play a significant role in promoting contextual consistency of hyperspectral classification, which is usually limited by the imbalance between the high spectral dimensionality and limited training samples [5], [6]. The advantages of utilizing oversegmented superpixels include higher homogeneity of pixel characteristics, better definition of trivial (but relevant targets), and sensitive recall of boundaries, as well as lower risks and object function values in convex optimization problems [5], [6], [33]. Using oversegmented image segments, Tarabalka *et al.* [3], [34], [35] developed several postprocessing techniques for improving HSI classification. Fang *et al.* [6] addressed this problem under a sparse model with superpixels. Among these techniques, a straightforward utilization of spatial information usually leads to convergence (optimal or suboptimal solutions), since both the classifier and the spatial information are simultaneously considered. In a more straightforward way, Li *et al.* [4] embedded the GC algorithm associated with spatial information into a novel active learning scheme that iteratively updates the data term to remarkably improve the labeling process. The work by Zhang *et al.* [5] tackled this problem by following a Bayesian framework also called the superpixel-based graphical model. In the work [7], the discontinuity information provided by boundaries has been formulated to reinforce the label consistency for HSI classification. In polarimetric synthetic aperture radar image analysis, Xu *et al.* [33] combine the statistical information with the spatial–contextual information using the stochastic expectation maximization algorithm. However, in spite of its great success, the NP-hardness of the integer optimization problems renders a little flexibility with respect to including superpixels as a spatial prior, mainly due to its discrete nature. In addition, it is generally very difficult to decide which is the best available image superpixelization considering both the variety of image segmentation techniques and the parametric adjustments. Thus, it remains very challenging to naturally exploit superpixels in supervised labeling scenarios for HSI classification. This problem becomes even more complicated when we consider multiple superpixelizations.

A. Contributions

As mentioned before, the Bayesian framework is widely used in order to exploit the spatial–contextual information. Under the Bayesian perspective, spatial or contextual information can both be viewed as priors to the conditional probabilities. However, the MAP segmentation leads to integer optimization problems that are hard to solve due to their discrete nature [36]. In order to deal with this issue, the LP [30], [37] or convex relaxation has been used to relax the discrete labeling problem into a compact domain. It is then much easier to convexly model the prior regularizers over the compact set, which opens the door to the inclusion of different priors resulting from real-world knowledge. The solution of the original problem is then approximated (or even obtained) under a primal-dual scheme [28]–[30], [32] using a linearly relaxed approach in polynomial time.

Based on the convex relaxation program, and in the spirit of [36], [38], [39], this paper introduces a new image labeling mechanism that is extremely flexible with respect to the inclusion of spatial information coming from superpixels, in the form of a spatial regularizer. In this regularizer, each superpixel is formulated as a constraint of GTV that reinforces

the pairwise label consistency in between its comprising pixels. In addition to the GTV associated with superpixels, this paper also utilizes a second spatial regularizer and the VTV, which promotes piecewise smoothness and aligns discontinuities along the edges in the image domain [36], thus improving the boundary recall of the resulting classification. Besides, a new framework based on SVD is designed for the purpose of adapting the superpixelization or superpixel generation method to HSIs, considering their high spectral dimensionality. Here, we specifically generate oversegmented superpixels with the fast simple linear iterative clustering (SLIC) algorithm [40] considering that the oversegmented ones are more likely to be homogeneous.

The main contributions of this paper can be summarized as follows: 1) development of a new strategy to relax the NP-hard integer optimization problem related to image labeling into a compact domain and characterization of oversegmented superpixels as a GTV regularizer under a Bayesian image segmentation perspective; 2) introduction of a VTV as a second spatial regularizer for boundary recalling purposes and development of a new algorithm based on the SALSA [41] method to solve the resulting problem; 3) adaptation of the SLIC algorithm [40] to HSI superpixelization; and 4) provision of experimental evidences, illustrating the potential of the proposed methodology in the context of hyperspectral remote sensing image classification.

B. Related Work

In order to tackle the aforementioned integer optimization problem associated with image segmentation, the work by Marroquin *et al.* [42] extended the Bayesian segmentation framework with a hidden MRF paradigm, which transforms the NP-hard optimization problem into a continuous domain. Under this paradigm, one can include additional information as a prior to the maximum likelihood function, such as MRF [42] or a wavelet-based prior [43]. In the context of convex optimization, some NP-hard problems associated with integer optimization problems, like the shortest path, max-flow, and so on, are often first relaxed and then solved or approximated as LP or semidefinite programming problems. The hidden fields' paradigm can actually be viewed as a statistical interpretation of the relaxation technique, and the linear relaxation program is generally a very close approximation to the unobserved hidden fields, given no prior information over a specific physical process [39]. Condessa *et al.* [38] sidestep the discrete nature of image segmentation by formulating the problem in a Bayesian framework with a hidden set of real-valued random fields. Then, the segmentation via the constrained split-augmented Lagrangian shrinkage algorithm (SegSALSA) is introduced to infer the hidden fields. In turn, the labels can be obtained by marginalized MAP. By this means, prior information, such as structure tensor regularization [44] and VTV [45], [46], is incorporated under a convex scheme.

This paper also has strong connections with the work of Bioucas-Dias *et al.* [36] and Condessa *et al.* [38]. There is, however, a major difference. The methodologies presented in [36] and [38] compute the probabilities of labelings to use

elsewhere, namely, in statistical inference problems. Specifically, in [36] and [38], hidden layers play a relevant role in describing the assumed unobserved variables that are associated with discrete class labels but belong to a continuous domain. The authors thus formulate spatial priors or regularizers based on these hidden fields in order to sidestep the NP-hard discrete optimization problems. Our objective, more in line with [39] and [47], is to use convex relaxation to approximate the original discrete problem into a compact domain in a linear manner. Similar to the work [38], [48], our resulting algorithm is also convex, time-efficient, and highly parallelizable. The remainder of this paper is organized as follows: Section II introduces the problem and the objective function associated with the VTV and GTV regularizers under the MAP framework. Section III describes our newly developed algorithm, which is an instance of the SALSA, to solve the resulted problem in Section II. Section IV presents the experimental evidence of the performance of our proposed method in the context of hyperspectral remote sensing classification. Section V concludes this paper with some remarks and hints at plausible future lines.

II. PROBLEM FORMULATION

The mathematical terms of the image segmentation problem are formulated with the following notations. Let $\mathcal{S} \equiv \{1, 2, \ldots, n\}$ be a set of integers indexing the n pixels of an image and $\mathbf{x} \equiv [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ a matrix of n vectors across d dimensions. Let $\mathbf{y} \equiv (y_1, \ldots, y_n) \in \mathcal{L}^n \equiv \{1, \ldots, K\}^n$ be an image of class labels, termed segmentation or labeling, such that $y_i = k$ if and only if the label of pixel i belongs to class k. Given $\mathbf{x}$, supervised image segmentation aims to find a partition $P \equiv \{R_1, \ldots, R_K\}$ of $\mathcal{S}$, such that the features indexed by a given set R_i, for $i \in \{1, \ldots, K\}$, are *similar* in some sense. Similarly, unsupervised image segmentation may be represented by another partition $\mathcal{N} \equiv \{\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_T\}$ of $\mathcal{S}$, where the set of pixels $\mathcal{V}_t \subset \mathcal{S}$, for $t = 1, \ldots, T$, termed superpixels, comprises n_t pixels, and thus $\sum n_t = n$. We remark that a main goal of this paper is to straightforwardly enforce the spatial coherence provided by the superpixels into the supervised image classification problem.

A. Maximum A-Posteriori Segmentation

We adopt a Bayesian perspective to the segmentation problem as a popular perspective to include prior information regularization information into the image labeling problem. Given the posterior probability $p_{\mathbb{Y}|\mathbb{X}}(\mathbf{y}|\mathbf{x})$, the observation model $p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y})$, and the prior probability $p_{\mathbb{Y}}(\mathbf{y})$ (often an MRF), the MAP segmentation is given by

$$\widehat{\mathbf{y}}_{\text{MAP}} \in \arg \max_{\mathbf{y} \in \mathcal{L}^n} p_{\mathbb{Y}|\mathbb{X}}(\mathbf{y}|\mathbf{x}) = \arg \max_{\mathbf{y} \in \mathcal{L}^n} p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y}) p_{\mathbb{Y}}(\mathbf{y}). \quad (1)$$

Under the conditional independence assumption, we have

$$p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y}) = \prod_{i=1}^{n} p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i) = \prod_{k=1}^{K} \prod_{i \in R_k} p(\mathbf{x}_i) \quad (2)$$

where $p(\mathbf{x}_i) = p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i = k)$. In a supervised scenario, the class probabilities $p_{\mathbb{X}_i|\mathbb{Y}_i}(\cdot|y_i = k)$ for $k \in \mathcal{L}$ are already

known or learned from a training set. Considering (2), we may write

$$\widehat{\mathbf{y}}_{\text{MAP}} \in \arg \min_{\mathbf{y} \in \mathcal{L}^n} - \log \left(p_{\mathbb{X}|\mathbb{Y}}(\mathbf{x}|\mathbf{y}) p_{\mathbb{Y}}(\mathbf{y}) \right)$$

$$= \arg \min_{\mathbf{y} \in \mathcal{L}^n} \sum_{i=1}^{n} D_i(y_i) + \lambda U(\mathbf{y}) \qquad (3)$$

where $D_i(y_i) = - \log p_{\mathbb{X}_i|\mathbb{Y}_i}(\mathbf{x}_i|y_i)$ denotes the log-likelihood probability density (often called a data term in the Bayesian image segmentation scenario), $\lambda U(\mathbf{y}) = - \log p_{\mathbb{Y}}(\mathbf{y})$ corresponds to the prior function, and $\lambda > 0$ is a tunable regularization parameter controlling the power of the spatial prior that is often an MRF. The minimization of (3) is an integer optimization problem over the discrete domain $\mathcal{L}^n$. In the case of Potts [20] prior and $K = 2$, the problem has an exact solution obtained by mapping the problem into of min-cut computation on a suitable graph [49]. However, for $K > 2$, the optimization (3) is proved NP-hard [16], and therefore, usually only approximations may be computed. To further complicate the use of integer formulations, the class of regularizers U that may be used in (3) is quite narrow; for example, it is not a simple task to include the prior coming from superpixels into U.

B. Convex Relaxation

In order to formulate the convex relaxation of (3), first, we replace the domain $\mathcal{L}^n$ in (3) by a more common integer constraint in convex optimization problems. Let $\mathbf{z}_i = [z_{1i}, \ldots, z_{Ki}]^T \in \{0, 1\}^K$ be a "1-of-K" representation of y_i, that is, $(y_i = k) \Leftrightarrow [z_{il} = 0$ for $l \neq k$ and $z_{ik} = 1]$. Using this representation, optimization (3) may be equivalently written as

$$\widehat{\mathbf{z}} \in \arg \min_{\mathbf{z}} \sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i + \lambda \phi(\mathbf{z})$$

$$\text{s.t.: } \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T$$

$$\mathbf{z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \{0, 1\}^{K \times n} \qquad (4)$$

where $\mathbf{q}_i = [D(y_i = 1), \ldots, D(y_i = K)]^T$, for $i \in \{1, \ldots, n\}$ indexing the samples, $\phi(\mathbf{z}) = U(\mathbf{y})$, and $\mathbf{1}_p$ is a p-dimensional column vector of 1s.

As proposed in [39] and [47] and also related to [36] and [38], we relax the optimization (4) by replacing the discrete set $\{0, 1\}$ to the interval $[0, 1]$, obtaining the optimization

$$\widehat{\mathbf{z}} \in \arg \min_{\mathbf{z} \in \mathbb{R}^{K \times n}} \sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i + \lambda \phi(\mathbf{z}),$$

$$\text{s.t.: } \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T$$

$$\mathbf{z} \geq 0 \qquad (5)$$

where $\lambda > 0$ is a tunable regularization parameter. Although optimization (5) is not equivalent to (4), it has been shown to provide very close approximations [39], [47]. The solution $\widehat{\mathbf{z}}$ of (5) can be used to gain information about the solution to the original integer program. Although the solutions $\widehat{z}_{ki}$ yielded by (5) are mostly discrete [21], a few elements, mostly in

the boundary of the classes, may not be in $\{0, 1\}$. In order to recover a complete discrete solution, we compute

$$\widehat{y}_i = \arg \max_{k} \widehat{z}_{ki}, \quad i \in \mathcal{S}, \ k \in \mathcal{L}.$$

The formulation (5) yields excellent results when compared with the original integer formulation, as extensively illustrated in [21]. In addition, a proper tailoring to the function $\phi(\cdot)$ allows to embrace a much larger group of prior information sources that comes from the real-world knowledge of the scene, besides the Gibbs distributions, namely MRFs or CRFs. This is exactly the possibility that we explore in this paper.

C. Spatial Regularizers

In this paper, as already mentioned, we use two spatial regularizers: the VTV and the GTV. Below we provide the details of both.

We adopt the following form of VTV [46]:

$$\phi_{\text{VTV}}(\mathbf{z}) \equiv \lambda_1 \sum_{i \in \mathcal{S}} \eta_i \sqrt{\|[\mathbf{z}\mathbf{D}_h]_{(:,i)}\|^2 + \|[\mathbf{z}\mathbf{D}_v]_{(:,i)}\|^2} \qquad (6)$$

where $\mathbf{D}_h, \mathbf{D}_v \in \mathbb{R}^{n \times n}$ are the matrices acting on the bands of $\mathbf{z}$ and computing horizontal and vertical first-order backward vector differences, respectively, $[\mathbf{z}\mathbf{D}_h]_{(:,i)}$ and $[\mathbf{z}\mathbf{D}_v]_{(:,i)}$ denote the vectors of K horizontal differences (one per latent image) computed at pixel i, η_i is a weight computed beforehand and affecting the magnitude of the discrete gradient at pixel i (this is more sensitive to different directions than only considering horizontal and vertical gradients), $\|\cdot\|$ is the standard Euclidean norm, and $\lambda_1 \geq 0$ is the regularizer parameter controlling the strength of the spatial VTV prior. The regularizer (6) is used in order to promote the piecewise smoothness of $\mathbf{z}$ and also to preserve aligned edges across $\mathbf{z}$ in the image domain. The VTV regularizer is convex, although not strictly, allowing optimization by proximal methods relying on Moreau proximity operators (MPOs) [38], [50].

Given the latent multiband image $\mathbf{z} \in \mathbb{R}^{K \times n}$, the GTV induced by an image segmentation $\mathcal{N} = \{\mathcal{V}_t, t = 1, \ldots, T\}$, where $\mathcal{V}_t$ denotes the tth superpixel, is defined as

$$\phi_{\text{GTV}}(\mathbf{z}) = \sum_{t=1}^{T} \sum_{(m,l) \in \mathcal{V}_t} \frac{\omega_{ml}}{2} \|(\mathbf{z}_m - \mathbf{z}_l)\|^2$$

where $\omega_{ml} \geq 0$ denotes a pairwise weight between nodes m and l. We remark that $\omega_{ml} = 0$ if m, l are not in the same superpixel. Therefore, minimizing ϕ_{GTV} promotes small variations of the latent vectors $\mathbf{z}_i$ inside the superpixels, and thus label consistency [51], [52].

In this paper, we set $\omega_{ml} = (1/n_t)$ if $m, l \in \mathcal{V}_t$. Since $\omega_{ml} = 0$ for m and l in different superpixels, then we may write

$$\phi_{\text{GTV}}(\mathbf{z}) = \sum_{t=1}^{T} \|(\mathbf{A}_t - \mathbf{I})(\mathbf{z}_{|\mathcal{V}_t})^T\|_F^2$$

where $\| \|_F$ is the Frobenious norm, $\mathbf{A}_t$ is the adjacency matrix associated with the superpixel $\mathcal{V}_t$ normalized by $n_t = |\mathcal{V}_t|$, and $\mathbf{z}_{|\mathcal{V}_t}$ is the subset of columns of $\mathbf{z}$ with indexes in $\mathcal{V}_t$.

The GTV regularizer may encompass multiple segmentations of which we have different degrees of confidence. Let $\mathcal{N}_i = \{\mathcal{V}_{i,t}, t = 1, \ldots, T_i\}$, for $i = 1, \ldots, C$, represent C segmentations, where $\mathcal{V}_{i,t}$ denotes the tth superpixel of segmentation i. By summing C GTV terms ϕ_{GTV}^i, one per segmentation, we obtain

$$\phi_{\mathrm{GTV}}(\mathbf{z}) = \sum_{i=1}^{C} \omega_i \phi_{\mathrm{GTV}}^i(\mathbf{z}) \tag{7}$$

where $\omega_i \geq 0$ expresses the confidence degree in the ith segmentation and

$$\phi_{\mathrm{GTV}}^i(\mathbf{z}) = \sum_{t=1}^{T_i} \|(\mathbf{A}_{i,t} - \mathbf{I})(\mathbf{z}_{|\mathcal{V}_{i,t}})^T\|_F^2 \tag{8}$$

where $\mathbf{A}_{i,t}$ is the adjacency matrix associated with the superpixel $\mathcal{V}_{i,t}$ normalized by $n_{i,t} = |\mathcal{V}_{i,t}|$ [52].

By taking the Frobenius norm, we can decouple (8) pixelwisely and superpixelwisely, thus opening the possibility to flexibly weight specific superpixelizations, objects, or classes for more practical purposes. On the other side, note that here we are allowed to combine multiple superpixelizations/segmentations at the same time, which avoids the dilemma of selecting "the best" segmentation.

In this paper, we use the fast superpixel clustering algorithm SLIC [40] to obtain multiple oversegmented spatial partitions.

III. OPTIMIZATION ALGORITHM

Having in mind the data term $-(\log \mathbf{p}_i^T)\mathbf{z}_i$ or $\mathbf{q}_i^T \mathbf{z}_i$ ($\mathbf{p}$ is the probability of vectors that are often known or learned by supervised classifiers), and the VTV and GTV regularizers, the resulting optimization problem turns out to be

$$\widehat{\mathbf{z}} \in \arg \min_{\mathbf{z} \in \mathbb{R}^{K \times n}} \sum_{i=1}^{n} \mathbf{q}_i^T \mathbf{z}_i + \lambda_1 \phi_{\mathrm{VTV}}(\mathbf{z}) + \lambda_2 \phi_{\mathrm{GTV}}(\mathbf{z})$$
$$\text{s.t.: } \mathbf{z} \geq 0, \quad \mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T \tag{9}$$

where the constraint $\mathbf{z} \in [0, 1]^{K \times n}$ is removed, as it is implied by $\mathbf{z} \geq 0$ and $\mathbf{1}_K^T \mathbf{z} = \mathbf{1}_n^T$, and $\lambda_1, \lambda_2 \geq 0$ are The regularization parameters weighting ϕ_{VTV} and ϕ_{GTV}, respectively.

Optimization (9) is convex as the data term and the regularizers are convex and the constraint set is also convex. In addition, it has a solution, as it is a convex problem defined on a compact set.

At this point, we remark that the ability to compute classification maps in a convex optimization framework, compatible with unsupervised segmentations obtained beforehand, was exactly what we originally targeted in this paper.

To solve the optimization (9), we use the SALSA [41], in way similar to that of SegSALSA [36], [38]. A major difference concerns the data term and the use of GTV regularization. In addition, we modify the original SALSA to use matrices as optimization variables instead of vectors.

We start by writing the optimization (9) in the equivalent form

$$\min_{\mathbf{z} \in \mathbb{R}^{K \times n}} \sum_{j=1}^{J} g_j(\mathbf{z} \mathbf{H}_j) \tag{10}$$

where g_j, for $j = 1, \ldots, J$, are convex, proper, and closed functions and $\mathbf{H}_j$ are matrices defined as follows:

$$g_1: \ \mathbb{R}^{K \times n} \to \mathbb{R} \quad \mathbf{H}_1 = \mathbf{I} \quad \boldsymbol{\xi} \ \mapsto \sum_{i=1}^{n} \mathbf{q}_i^T \boldsymbol{\xi}_i$$

$$g_2: \ \mathbb{R}^{K \times n} \to \mathbb{R} \quad \mathbf{H}_2 = \mathbf{I} \quad \boldsymbol{\xi} \ \mapsto \sum_{i=1}^{n} \iota_{\{1\}}(\mathbf{1}_K^T \boldsymbol{\xi}_i)$$

$$g_3: \ \mathbb{R}^{K \times n} \to \mathbb{R} \quad \mathbf{H}_3 = \mathbf{I} \quad \boldsymbol{\xi} \ \mapsto \iota_{\mathbb{R}_+^{K \times n}}(\boldsymbol{\xi})$$

$$g_4: \ \mathbb{R}^{K \times 2n} \to \mathbb{R} \quad \mathbf{H}_4 = [\mathbf{D}_h \ \mathbf{D}_v]$$
$$(\boldsymbol{\xi}^h, \boldsymbol{\xi}^v) \mapsto \lambda_1 \sum_{i=1}^{n} \eta_i \sqrt{\|\boldsymbol{\xi}_i^h\|^2 + \|\boldsymbol{\xi}_i^v\|^2}$$

for $i = 1, \ldots, C$

$$g_{4+i}: \ \mathbb{R}^{K \times n} \to \mathbb{R} \quad \mathbf{H}_{4+i} = \mathbf{I}$$
$$\boldsymbol{\xi} \mapsto \lambda_2 \omega_i \sum_{t=1}^{T_i} \|(\mathbf{A}_{i,t} - \mathbf{I})(\boldsymbol{\xi}_{|\mathcal{V}_{i,t}})^T\|_F^2$$

where the symbols $\iota_{\{1\}}$ in function g_2 and $\iota_{\mathbb{R}_+^{K \times n}}$ in function g_3 represent, respectively, the indicator function in the sets $\{1\}$ and $\mathbb{R}_+^{K \times n}$. Function g_1 is the data term, g_2 is an equivalent form of representing the sum-to-one constraint, g_3 is an equivalent form of representing the nonnegativity constraint, g_4 is the VTV regularizer, and g_5 to g_{4+C} are the GTV regulalizers. Note that $4 + C = J$.

The next step consists in replacing (10) with the equivalent optimization

$$\min_{\mathbf{z}, \mathbf{u}} \sum_{j=1}^{J} g_j(\mathbf{u}_j) \quad \text{s.t: } \mathbf{z} \mathbf{G} = \mathbf{u} \tag{11}$$

where $\mathbf{u}_j = \mathbf{z} \mathbf{H}_j$, for $j = 1, \ldots, J$, and

$$\mathbf{G} = [\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_J]$$
$$\mathbf{u} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_J] \tag{12}$$

with $\mathbf{u}_j \in \mathbb{R}^{K \times n}$ for $j \neq 4$ and $\mathbf{u}_4 \in \mathbb{R}^{K \times 2n}$.

By writing the augmented Lagrangian for (11) as proposed in [41], and iteratively optimizing with respect to $\mathbf{z}$ and $\mathbf{u}$ and then updating the scaled Lagrange multipliers, we obtain Algorithm 1, termed SuperSALSA.

Algorithm 1 SuperSALSA

initialization:
choose $(\mathbf{u}_0^j, \mathbf{d}_0^j) \in \mathbb{R}^{K \times n_j}$, $j = 1, \ldots, J$
define $\mathbf{K} = \mathbf{G} \mathbf{G}^T = \sum_{j=1}^{J} \mathbf{H}^j (\mathbf{H}^j)^T$
set $\mu \in \,]0, +\infty[$
for $k = 0, 1, \ldots$ **do**
$\quad \mathbf{z}_{k+1} = \left(\sum_{j=1}^{J} \left(\mathbf{u}_k^j + \mathbf{d}_k^j \right) (\mathbf{H}^j)^T \right) \mathbf{K}^{-1}$
$\quad$ **for** $j = 1$ **to** J **do**
$\qquad \mathbf{u}_{k+1}^j = \mathrm{prox}_{g_j/\mu}(\mathbf{z}_{k+1} \mathbf{H}^j - \mathbf{d}_k^j)$
$\qquad \mathbf{d}_{k+1}^j = \mathbf{d}_k^j - (\mathbf{z}_{k+1} \mathbf{H}^j - \mathbf{u}_{k+1}^j)$
return $\mathbf{z}_{k+1}$

TABLE I
COMPUTATIONAL COMPLEXITY FOR THE MPO OF EACH
RELEVANT TERM IN ALGORITHM 1: SUPERSALSA

Index j	Term	Computational complexity
$j = 1$	Data fit	$O(Kn)$
$j = 2$	Sum-to-one constraint	$O(Kn)$
$j = 3$	Nonnegativity constraint	$O(Kn)$
$j = 4$	VTV	$O(Kn)$
$j > 4$	GTV	$O(CKn)$
Total	Σ	$O(JKn)$

In Algorithm 1, the linear operators $\mathbf{G}$ represent a cyclic convolution. Therefore, the computation of $\mathbf{z}_{k+1}$ can be implemented through cyclic convolution operations, thus diagonalizable in the frequency domain and consequently easily performed using the fast Fourier transform with $O(Kn \ln n)$ complexity. The computation of $\mathbf{u}_{k+1}^j$, for $j = 1, \ldots, J$, is carried by the proximity operator (MPO) [50] of g_j/μ given by

$$\text{prox}_{g_j/\mu}(\mathbf{y}) = \arg\min_{\mathbf{x}} \ g_j(\mathbf{x}) + (\mu/2)\|\mathbf{y} - \mathbf{x}\|^2.$$

Specifically, in this paper, prox_{g_j}, for $j = 1, 2, 3, 4$, correspond to, respectively, to the data fit, the sum-to-one constraint, the nonnegativity constraint, and the VTV. The terms for $j > 4$ correspond to the GTVs. The optimizations underlying the prox operators are given in Appendix B. In addition, we also present the computational complexity for each prox_{g_j} in Table I.

Regarding the stopping criterion, we impose that the primal and dual residuals be smaller than a given threshold, as suggested in [53, Ch. 3.3.2]. We have observed, however, that a fixed number of iterations in the order of 200 provide excellent results.

IV. EXPERIMENTS

In this section, we evaluate the proposed method with remote sensing data sets acquired by different types of sensors, namely, HSIs and multispectral images (MSIs). Before reporting our experimental results, we introduce our newly designed framework in Fig. 2.

The data term, the VTV regularizers, and the GTV regularizers are parameterized, respectively, by the probabilities $\mathbf{q}_n$, $n \in \mathcal{S}$, the weights g_n, $n \in \mathcal{S}$, and the oversegmented superpixelizations $\mathcal{N}_i$, for $i = 1, \ldots, C$. First of all, classifier MLR in our case, whose regressors are learned by the logistic regression via the variable splitting and the augmented Lagrangian (LORSAL) algorithm [54], [55], is used to estimate the class probabilities of the image, in preparation for the data term. Before obtaining the remaining two terms (VTV and GTV), the spectral vectors are projected on a low-dimensional subspace using SVD, for dealing with the high spectral dimensionality of the HSIs. Then, in order to weaken the trivial textural details and emphasize on the edges of the image, anisotropic edge-preserving filtering [56] is then
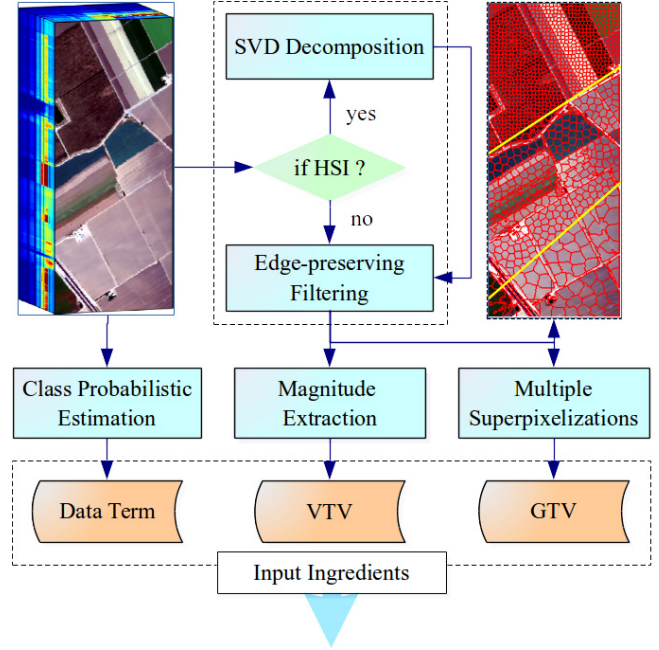

Fig. 2. Experimental framework for our proposed method with respect to hyperspectral and multispectral remote sensing image data sets.

utilized to the main transformed components and to the MSI. Finally, we extract the gradient map with the Sobel operator and generate multiple oversegmented superpixelizations with the fast spatial clustering algorithm SLIC [40], available in the VLFeat toolbox,[1] with varying parameters.

A. Experiments With Hyperspectral Images

In this section, we evaluate our proposed algorithm with the ROSIS Pavia University data set as well as the AVIRIS Salinas data set. The first HSI used in our experiments [see Fig. 3(a)] was collected by ROSIS over the University of Pavia, Italy. The data set consists of 115 spectral bands, covering the wavelength range from 0.43 to 0.86 μm, with the size of 610×340 pixels. The noisy bands had been removed, yielding 103 spectral bands that are actually used in this paper. The ground-truth (GT) image contains nine GT classes, including 42 776 labeled samples. The second HSI used in our experiment is the well-known AVIRIS Salinas[2] data set, collected by the AVIRIS sensor of 224 bands in the wavelength range from 0.38 to 2.50 μm over the Salinas Valley, California. As displayed in Fig. 3(f), it comprises 512 lines by 217 samples across 204 spectral bands after discarding 20 water absorption bands. For the reference collection, a total of 54 129 pixels are available in the labeled GT, including 16 mutually exclusive classes. For both hyperspectral data sets, we also prepared the gradient map [Fig. 3(c) and (h)] and three oversegmented superpixelizations [Fig. 3(d) and (i)], respectively.

Before displaying the experimental results, we first introduce our experimental setup for the analysis of HSIs. The

---

[1]http://www.vlfeat.org/doc/api/slic.html
[2]http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Salinas_scene
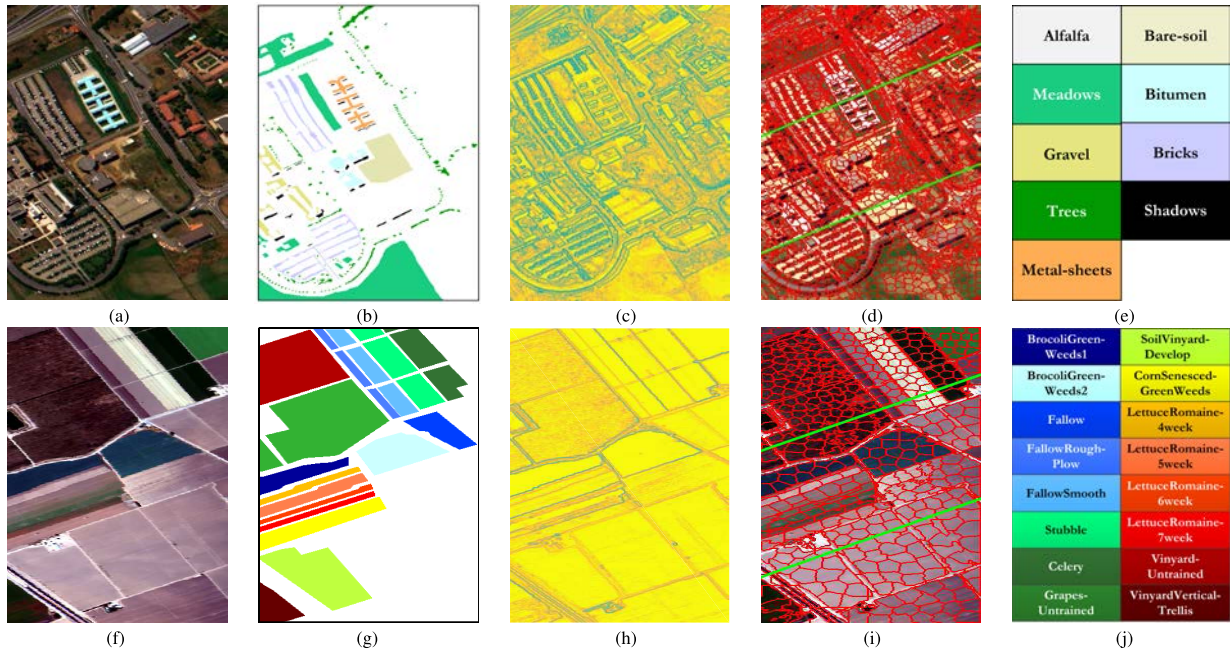
Fig. 3. Experiments with hyperspectral data. First row: ROSIS Pavia University data set. Second row: Results of the AVIRIS Salinas data set. (a) and (f) HSI RGB composite. (b) and (g) GT reference. (c) and (h) Gratitude map. (d) and (i) Multiple oversegmented superpixelizations. (e) and (j) Class legends. The parameter values used for the SLIC algorithm are superpixel size = {10, 13, 16} and shape regularizer $\lambda_{\text{SLIC}} = 0.04$ for the ROSIS Pavia University data and size = {11, 15, 19} and $\lambda_{\text{SLIC}} = 0.01$ for the AVIRIS Salinas data.

class probabilities are estimated by an MLR, where the logistic regressors (assumed to be independent Laplacian random vectors) are learned using the LORSAL algorithm [54], [55]. The MLR classifiers are learned with 15 training samples per class for both the ROSIS Pavia University data and the AVIRIS Salinas data. For simplicity, we use three over-segmented superpixelization representations (with the SLIC algorithm) to construct the GTV. The maximum number of SuperSALSA iterations is set to 200 iterations, since we have systematically observed the convergence from a practical point of view. As for the SVD decomposition step in Fig. 2, we used five and six major components, respectively, for the ROSIS and AVIRIS data sets. Meanwhile, some recently developed state-of-the-art methods, namely, the majority voting (MV) approaches [35], GC [23] algorithm, and discontinuity-preserving relation scheme [7], are also considered in this paper for evaluation and comparison purposes. Also, for statistical purposes of all the relevant methods, 20 and 70 Monte Carlo runs are, respectively, performed for the ROSIS Pavia University data and the AVIRIS Salinas data.

The obtained results are displayed in Tables II and III. Several observations can be made from these results. First, all the segmentation results obtain remarkable improvements compared with the fundamental MLR classifier, particularly after the inclusion of spatial information. Meanwhile, the proposed algorithm also outperforms the compared methods with respect to the overall accuracy for both the HSI data sets. For illustrative purposes, Fig. 4 also displays the corresponding classification maps that are obtained from one of the Monte Carlo runs. Remarkably, our obtained segmentation maps show stronger pixel consistency while keeping more

precise contours for the land objects, which are exactly what this paper explores by formulating the VTV and superpixel-related GTV regularizers. Specifically, when compared with MV and GC methods, our proposed method (with VTV) preserves better the edge/boundary information, since MV does not consider the boundary discontinuity and the GC does not consider the oblique (only the vertical and horizontal) discontinuities. After analyzing the results obtained for both the ROSIS and AVIRIS data sets, we can observe that the boundary recall of SuperSALSA generally surpasses that of the MV and GC methods and achieves the same or higher accuracy levels than those achieved by the DPR method. Also, it can be seen that large values of λ_1 for VTV lead to higher accuracies for boundary recall, according to Fig. 4, while very large values result in lower boundary recall accuracies due to the loss of trivial details. On the other hand, the map acquired by the DPR method misses some small-scale details due to the fact that the DPR method relies greatly on the quality of edge extraction, which is a challenging task in HSI processing. Bearing these observations in mind, our proposed method turns out to be the state of the art. Besides, according to (8) and the experimental results in Figs. 3(d) and (i) and 4 (third and sixth rows), we can infer that the spatially regularized classification results will almost rigidly follow the edges of superpixels if higher values of λ_2 are used. Therefore, we conclude that the edge-preserving filter on the main transformed components is necessary and significant in the task of improving the segmentation results (before the application of the SLIC algorithm) and also the final classification results. Regarding the proposed SuperSALSA algorithm itself, highly discrete classification results are also obtained, as expected by (5) and indicated

TABLE II

ACCURACY STATISTICS [%] OF THE ROSIS PAVIA UNIVERSITY DATA SET OBTAINED BY THE PRESENTED CLASSIFICATION FRAMEWORK IMPLEMENTED USING THE MLR CLASSIFIER IN COMPARISON WITH THE STATE-OF-THE-ART METHODS, MV, GC, AND DPR. IN PARTICULAR, WE SET THE PARAMETER OF THE PROPOSED METHOD $\lambda_1 = 5$, $\lambda_2 \times c_i = 2$, FOR $i = \{1, \ldots, 3\}$. THE AVERAGES AND CORRESPONDING STANDARD DEVIATIONS ARE CALCULATED UNDER 20 MONTE CARLO RUNS. IN ALL CASES, ONLY 15 RANDOMLY SELECTED TRAINING SAMPLES PER CLASS HAVE BEEN USED. CORRESPONDING TIME CONSUMPTION STATISTICS ARE ALSO LISTED IN THE FINAL LINE FOR EACH SPECTRAL–SPATIAL CLASSIFICATION METHOD

Class	MLR	MVs	Graphcut	DPR	SuperSALSA	(discrete rate /%)
Alfalfa	76.85 ± 3.75	87.31 ± 2.92	91.12 ± 3.46	92.33 ± 2.90	**92.48 ± 4.99**	93.95 ± 4.56
Meadows	88.16 ± 3.47	96.51 ± 2.84	96.48 ± 2.54	96.42 ± 2.37	**97.74 ± 2.42**	97.74 ± 0.74
Gravel	72.58 ± 6.56	84.59 ± 9.70	85.39 ± 7.37	84.16 ± 7.78	**92.27 ± 9.16**	70.84 ± 7.48
Trees	84.75 ± 5.20	79.63 ± 6.67	**87.41 ± 4.27**	86.80 ± 4.13	81.96 ± 7.67	72.98 ± 3.16
Metal sheets	99.80 ± 0.16	98.46 ± 1.32	99.95 ± 0.05	**100.00 ± 0.00**	99.93 ± 0.02	99.85 ± 0.02
Bare soil	72.64 ± 5.99	86.71 ± 7.54	86.68 ± 6.49	88.32 ± 6.78	**93.84 ± 6.20**	89.96 ± 1.49
Bitumen	84.14 ± 5.73	98.07 ± 2.82	95.46 ± 5.44	97.24 ± 4.52	**98.56 ± 2.26**	91.88 ± 0.44
Bricks	76.04 ± 5.03	86.42 ± 6.68	85.20 ± 5.88	**90.39 ± 4.11**	87.85 ± 10.22	63.01 ± 3.91
Shadows	99.50 ± 0.40	99.63 ± 0.51	99.59 ± 0.44	**99.89 ± 0.04**	99.11 ± 0.36	84.37 ± 0.47
Overall accuracy	83.02 ± 1.54	91.45 ± 1.43	92.48 ± 1.48	93.24 ± 1.14	**94.34 ± 1.77**	(overall discrete
Average accuracy	83.83 ± 1.24	90.82 ± 1.55	91.92 ± 1.48	92.84 ± 1.40	**93.75 ± 2.06**	rate)
κ statistic	77.80 ± 1.88	88.63 ± 1.88	90.02 ± 1.94	91.03 ± 1.51	**92.48 ± 2.35**	92.15 ± 4.52
Time/s	*	4.18 ± 0.21	8.44 ± 1.25	25.30 ± 0.47	176.98 ± 8.74	

TABLE III

ACCURACY STATISTICS [%] OF THE AVIRIS SALINAS DATA SET OBTAINED BY THE PRESENTED CLASSIFICATION FRAMEWORK IMPLEMENTED USING THE MLR CLASSIFIER IN COMPARISON WITH THE STATE-OF-THE-ART METHODS, MV, GC, AND DPR. IN PARTICULAR, WE SET THE PARAMETER OF THE PROPOSED METHOD $\lambda_1 = 5$, $\lambda_2 \times c_j = 5$, FOR $j = \{1, \ldots, 3\}$. THE AVERAGES AND CORRESPONDING STANDARD DEVIATIONS ARE CALCULATED UNDER 70 MONTE CARLO RUNS. IN ALL CASES, ONLY 15 RANDOMLY SELECTED TRAINING SAMPLES PER CLASS HAVE BEEN USED. CORRESPONDING TIME CONSUMPTION STATISTICS ARE ALSO LISTED IN THE FINAL LINE FOR EACH SPECTRAL–SPATIAL CLASSIFICATION METHOD

Class	MLR	MVs	Graphcut	DPR	SuperSALSA	(discrete rate /%)
C1	99.12 ± 0.49	**100 ± 0.00**	99.97 ± 0.12	**100 ± 0.00**	**100 ± 0.00**	98.86 ± 0.33
C2	99.86 ± 0.41	99.92 ± 0.02	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	99.92 ± 0.22
C3	93.33 ± 4.42	95.28 ± 8.97	**99.96 ± 0.10**	99.27 ± 2.08	99.84 ± 0.15	78.85 ± 0.23
C4	99.66 ± 0.15	98.98 ± 0.00	99.75 ± 0.09	**100 ± 0.00**	99.85 ± 0.00	95.55 ± 0.73
C5	97.41 ± 0.81	86.34 ± 5.49	98.68 ± 0.32	**98.80 ± 0.34**	94.18 ± 4.73	80.88 ± 2.89
C6	99.22 ± 0.23	99.92 ± 0.00	99.73 ± 0.17	**99.97 ± 0.01**	99.92 ± 0.00	99.32 ± 0.63
C7	99.88 ± 0.06	99.29 ± 0.08	99.91 ± 0.03	**100 ± 0.00**	99.92 ± 0.01	99.53 ± 1.07
C8	67.57 ± 9.79	81.02 ± 13.41	78.88 ± 13.12	78.70 ± 11.84	**82.51 ± 18.94**	73.57 ± 9.15
C9	98.45 ± 0.73	98.66 ± 1.02	98.93 ± 0.82	**99.55 ± 0.55**	99.44 ± 0.58	95.65 ± 0.61
C10	85.74 ± 4.43	90.51 ± 4.58	91.45 ± 3.57	93.20 ± 3.15	**93.62 ± 2.03**	86.08 ± 2.30
C11	91.68 ± 2.89	96.81 ± 0.07	96.42 ± 2.77	98.74 ± 0.72	96.23 ± 1.32	91.85 ± 1.12
C12	99.61 ± 0.57	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	**100 ± 0.00**	100.00 ± 0.51
C13	89.14 ± 4.81	71.20 ± 10.64	**95.16 ± 4.00**	93.31 ± 4.37	90.58 ± 8.34	64.08 ± 3.08
C14	92.29 ± 0.92	90.60 ± 6.34	97.95 ± 1.49	**99.70 ± 0.15**	97.27 ± 0.52	82.52 ± 0.29
C15	66.14 ± 8.10	78.94 ± 16.07	79.12 ± 13.37	78.52 ± 11.13	**84.34 ± 15.33**	64.93 ± 19.89
C16	96.11 ± 2.62	98.12 ± 0.04	**99.32 ± 0.74**	98.90 ± 0.91	98.43 ± 0.84	97.12 ± 0.90
Overall accuracy	86.50 ± 1.43	90.75 ± 2.17	91.82 ± 1.96	91.93 ± 1.65	**93.14 ± 3.19**	(overall discrete
Average accuracy	92.20 ± 0.48	92.85 ± 0.83	95.95 ± 0.65	**96.17 ± 0.54**	96.01 ± 1.03	rate)
κ statistic	85.02 ± 1.56	89.72 ± 2.41	90.92 ± 2.16	91.03 ± 1.81	**92.39 ± 3.50**	85.27 ± 2.42
Time/s	*	1.15 ± 1.55	6.03 ± 1.55	24.89 ± 0.56	159.26 ± 7.25	

by [21]. It should be noted that the discrete rate varies a lot by class. Possible reasons for this are: 1) the presence of rich spatial textures in some of the classes leads to higher hardness of label consistency reinforcement and boundary realignment and 2) it is generally difficult to classify certain classes whose most likely probabilistic estimations are not prominent enough, and thus are hard to become discrete using the SuperSALSA algorithm.
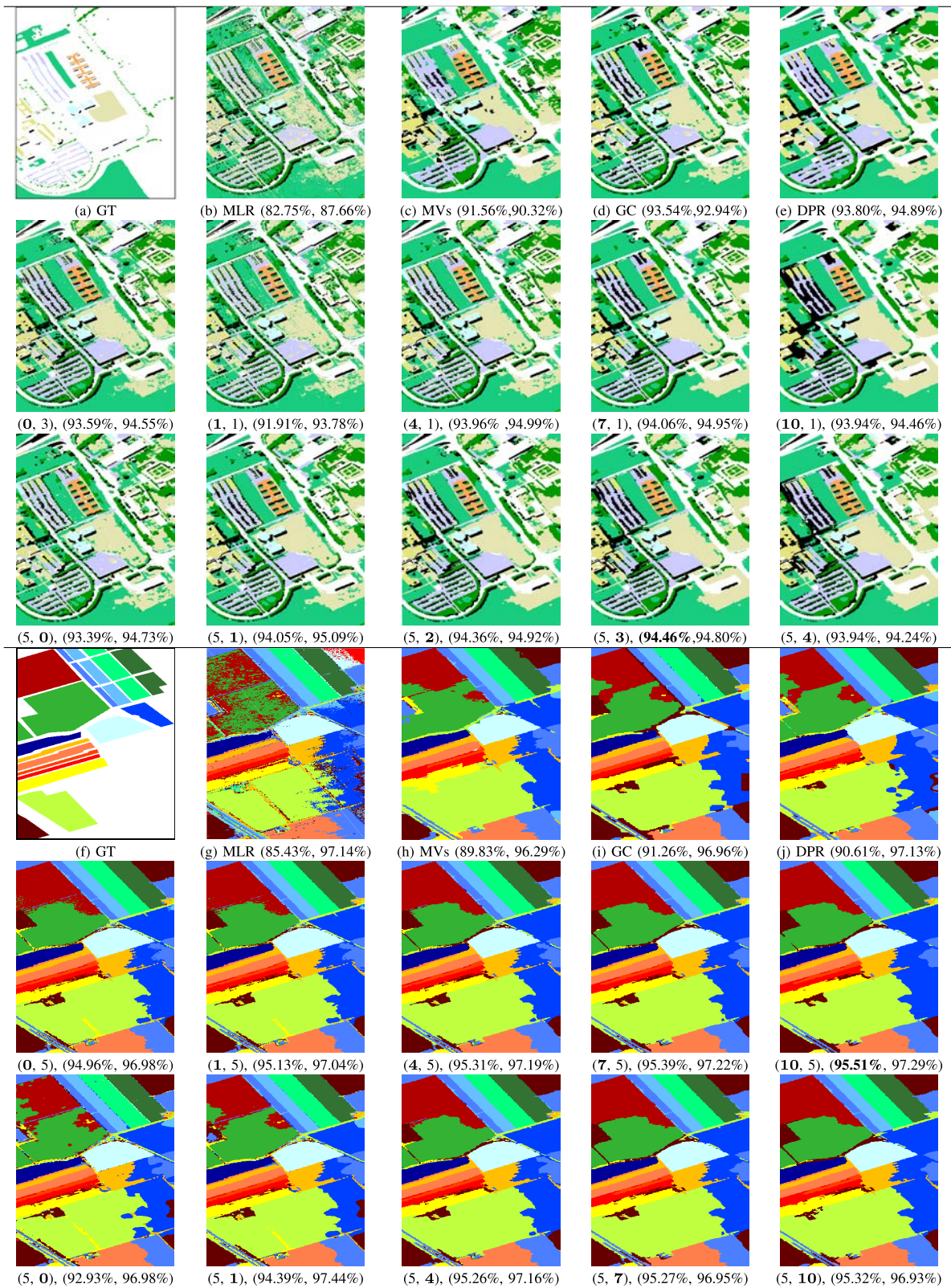
Fig. 4.   Classification maps of two HSI data sets. First row: ROSIS Pavia University data set. Fourth row: AVIRIS Salinas data set. The results of the ROSIS data set by the proposed SuperSALSA algorithm are displayed in the second and third rows along with different values of parameter λ_1, $\lambda_2 \times c_i$, for $i = \{1, \ldots, 3\}$, followed by the overall accuracy and overall accuracy of edges with a radius of 5 pixels. In the fourth to sixth rows are the corresponding results of the AVIRIS Salinas data set. (a) and (f) GT. (b) and (g) MLR classification. (c) and (h) MV. (d) and (i) GC. (e) and (j) DPR.

The regularization parameters play a significant role in adjusting the performance of the whole machinery. In order to illustrate the effect of the two spatial regularizers, we display additional segmentation maps obtained by using different values of the regularizer parameters λ_1 and $\lambda_2 \times c_i$, for $i = \{1, \ldots, 3\}$. For simplicity, we set the same parametric
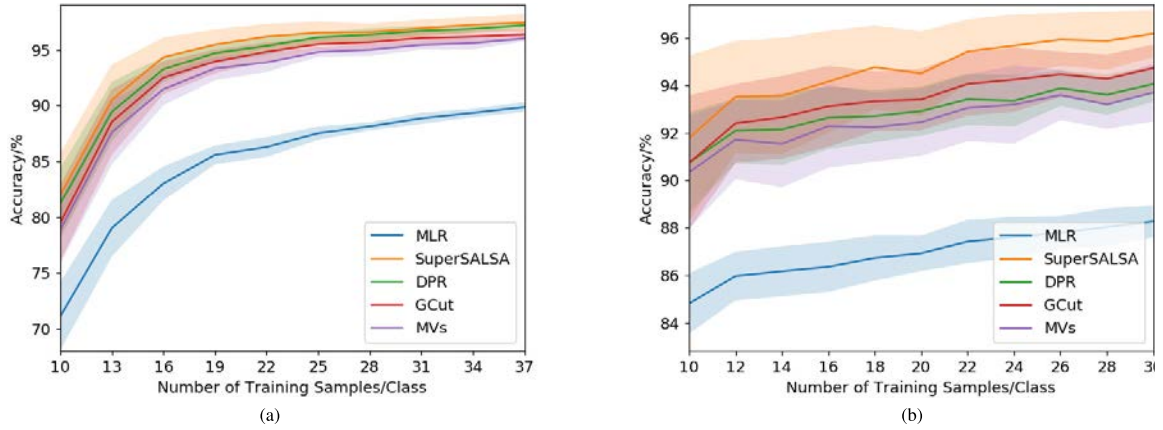
Fig. 5. Obtained OA lines for the two hyperspectral data of scenes. (a) ROSIS Pavia University. (b) AVIRIS Salinas. Solid lines: average of 20 and 70 Monte Carlo runs for the ROSIS Pavia University data and the AVIRIS Salinas data, respectively. Colored area around the lines: standard deviation around the mean.
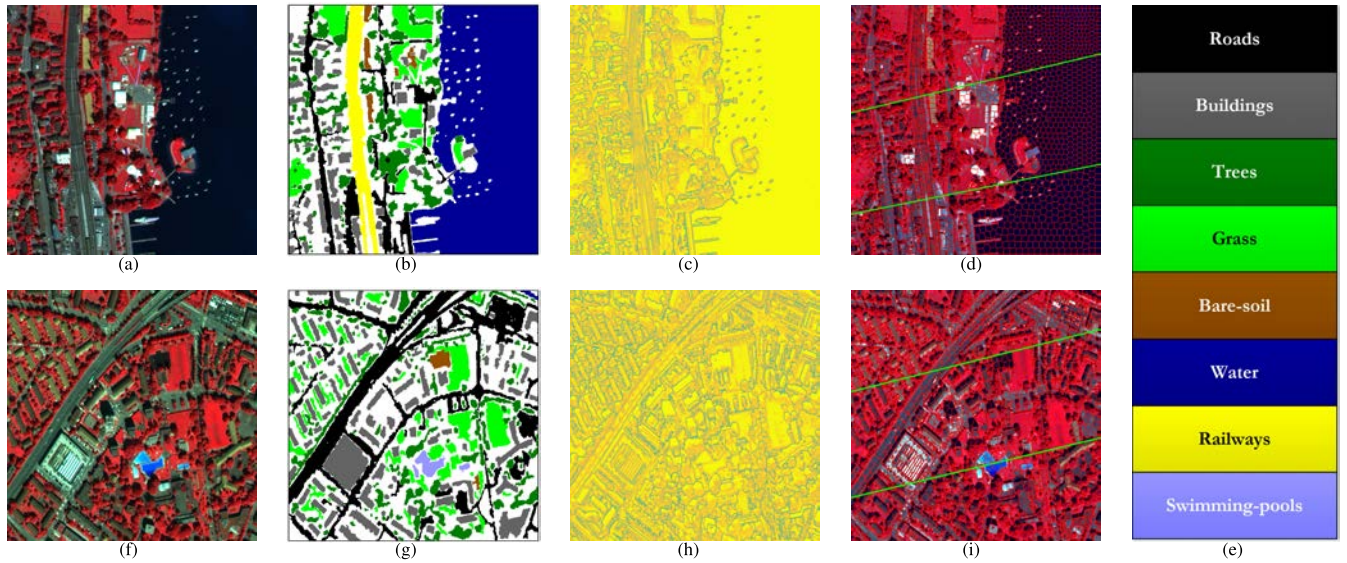

Fig. 6. MSI data sets. First row: QB Zurich3 data set. Second row: QB Zurich6 data set. (a) and (f) MSI RGB composite. (b) and (g) GT reference. (c) and (h) Gratitude map. (d) and (i) Multiple oversegmented superpixelizations. (e) Class legends. Parameter values used for the SLIC algorithm are superpixel size = $\{15, 19, 23\}$ and shape regularizer $\lambda_{\text{SLIC}} = 0.04$ for the Zurich3 data, and size = $\{15, 19, 23\}$ and $\lambda_{\text{SLIC}} = 0.05$ for the Zurich6 data.

values for three oversegmented superpixelizations. The obtained results of both the HSI data sets are shown in Fig. 4, displaying a clue also on how to tune the values of the parameters by hand. Specifically, this means that only the superpixel-based regularizer (GTV) is utilized when setting $\lambda_1 = 0$. And, likewise, when $\lambda_2 \times c_i = 0$, only the VTV regularizer is considered. It is obviously observed that different scales of contours of the land objects can be controlled by tuning the parameters. For both the regularizers, small values of the parameters lead to more details of the land objects, especially the small ones while greater values tend to keep the main contours of the objects. As for the VTV and GTV regularizers, when both of them are considered, the performance improves greatly in comparison with the individual use of one regularizer, which is consistent with what we have anticipated, i.e., involving the VTV to both promote piecewise smoothness and align the discontinuities along the boundaries while incorporating GTV to reinforce the label consistency

over the oversegmented superpixels. Besides, the processing times listed in Tables II and III show that the SuperSALSA algorithm is computationally more expensive than its state-of-the-art competitors, a situation that can be improved using high-performance computing architectures [57], [58].

In addition, the overall accuracies have also been obtained regarding different sizes of training samples for both the ROSIS Pavia University data and the AVIRIS Salinas Scene data. As shown in Fig. 5, similar observations to Tables II and III can also be made. First of all, the involvement of spatial information achieves remarkable promotes from the MLR by roughly 10% and 7% of the overall accuracy, respectively, for the ROSIS Pavia University and the AVIRIS Salinas data. Also, it can be observed that the proposed SuperSALSA algorithm advances the other state of the arts with respect to different sizes of training samples.

In addition, there are differences between the ROSIS and AVIRIS data sets. Note that a different number of

TABLE IV

Accuracy Statistics [%] of the Zurich3 Data Set Obtained by the Presented Classification Framework Implemented Using the MLR Classifier in Comparison With the State-of-the-Art Methods, MV, GC, and DPR. In Particular, We Set the Parameter of the Proposed Method $\lambda_1 = 4$, $\lambda_2 \times c_i = 2$, for $i = \{1, \ldots, 3\}$. The Averages and Corresponding Standard Deviations Are Calculated Under 20 Monte Carlo Runs. In All Cases, 200 Randomly Selected Training Samples per Class Have Been Used. Corresponding Time Consumption Statistics Are Also Listed in the Final Line for Each Spectral–Spatial Classification Method

Class	MLR	MVs	Graphcut	DPR	SuperSALSA	(discrete rate /%)
Roads	77.54 ± 1.36	$\mathbf{84.43 \pm 1.24}$	83.11 ± 1.29	79.01 ± 1.36	70.59 ± 4.78	38.97 ± 5.51
Buildings	52.54 ± 3.51	50.03 ± 5.69	57.15 ± 4.75	60.39 ± 4.17	$\mathbf{72.50 \pm 5.46}$	43.41 ± 6.85
Trees	84.96 ± 0.90	91.20 ± 0.81	88.83 ± 1.09	90.51 ± 1.10	$\mathbf{94.75 \pm 1.02}$	50.59 ± 0.50
Grass	91.06 ± 1.19	94.67 ± 0.62	93.03 ± 1.17	93.09 ± 1.14	$\mathbf{95.43 \pm 1.12}$	21.00 ± 1.41
Bare-Soil	95.02 ± 0.76	95.22 ± 0.96	95.70 ± 0.57	$\mathbf{95.80 \pm 0.80}$	92.87 ± 0.75	49.93 ± 0.80
Water	99.56 ± 0.08	99.77 ± 0.01	99.61 ± 0.08	$\mathbf{99.84 \pm 0.06}$	99.78 ± 0.03	15.87 ± 0.30
Railways	56.60 ± 1.91	69.77 ± 2.29	73.31 ± 2.65	78.79 ± 1.89	$\mathbf{94.32 \pm 1.28}$	65.31 ± 0.77
Overall accuracy	87.87 ± 0.19	90.39 ± 0.48	90.62 ± 0.36	91.16 ± 0.34	$\mathbf{92.98 \pm 0.50}$	(overall discrete
Average accuracy	79.61 ± 0.23	83.58 ± 0.78	84.39 ± 0.60	85.35 ± 0.55	$\mathbf{88.61 \pm 0.80}$	rate)
κ statistic	81.99 ± 0.27	85.71 ± 0.71	86.08 ± 0.53	86.87 ± 0.50	$\mathbf{89.58 \pm 0.74}$	40.73 ± 2.27
Time/s	$*$	20.86 ± 0.59	13.81 ± 5.84	136.02 ± 6.96	1568.59 ± 37.70	

TABLE V

Accuracy Statistics [%] of the QB Zurich6 Data Set Obtained by the Presented Classification Framework Implemented Using the MLR Classifier in Comparison With the State-of-the-Art Methods, MV, GC, and DPR. In Particular, We Set the Parameter of the Proposed Method $\lambda_1 = 4$, $\lambda_2 \times c_i = 2$, for $i = \{1, \ldots, 3\}$. The Averages and Corresponding Standard Deviations Are Calculated Under 20 Monte Carlo Runs. In All Cases, 200 Randomly Selected Training Samples Per Class Have Been Used. Corresponding Time Consumption Statistics Are Also Listed in the Final Line for Each Spectral–Spatial Classification Method

Class	MLR	MVs	Graphcut	DPR	SuperSALSA	(discrete rate /%)
Roads	73.28 ± 2.01	82.03 ± 1.28	82.03 ± 1.89	83.12 ± 1.43	$\mathbf{84.20 \pm 1.77}$	29.91 ± 2.07
Buildings	64.92 ± 2.61	63.10 ± 4.91	65.19 ± 4.10	66.99 ± 2.96	$\mathbf{67.44 \pm 4.15}$	48.93 ± 1.83
Trees	76.60 ± 1.67	84.87 ± 2.23	83.53 ± 1.31	86.17 ± 1.52	$\mathbf{93.03 \pm 1.91}$	66.10 ± 1.44
Grass	88.14 ± 1.22	$\mathbf{92.10 \pm 1.01}$	90.47 ± 1.18	90.66 ± 1.37	91.38 ± 2.24	60.50 ± 2.87
Bare-Soil	72.49 ± 2.08	93.89 ± 3.68	83.42 ± 1.04	83.85 ± 1.96	$\mathbf{94.61 \pm 3.68}$	35.20 ± 4.18
Water	84.61 ± 1.28	$\mathbf{100 \pm 0.00}$	98.10 ± 2.02	93.49 ± 1.80	99.95 ± 0.15	24.97 ± 0.23
Swimming-Pools	92.78 ± 0.83	$\mathbf{97.55 \pm 0.29}$	93.39 ± 0.75	94.91 ± 0.92	96.22 ± 3.40	70.93 ± 3.36
Overall accuracy	74.01 ± 0.54	78.98 ± 1.25	78.93 ± 0.89	80.36 ± 0.71	$\mathbf{82.36 \pm 0.91}$	(overall discrete
Average accuracy	78.97 ± 0.31	87.65 ± 0.93	85.16 ± 0.51	85.60 ± 0.46	$\mathbf{89.55 \pm 0.63}$	rate)
κ statistic	65.91 ± 0.62	72.10 ± 1.63	72.05 ± 1.18	73.82 ± 0.92	$\mathbf{76.32 \pm 1.21}$	40.16 ± 2.09
Time/s	$*$	17.78 ± 0.70	12.74 ± 2.17	122.12 ± 5.83	1318.21 ± 15.67	

Monte Carlo runs are chosen for the ROSIS Pavia University data set (20 runs) and the AVIRIS Salinas data set (70 runs). The standard deviation of the OA around its mean is larger for AVIRIS Salinas than for ROSIS Pavia (whatever the number of training samples per class). Its value naturally decreases with the use of a larger number of training samples per class. Overall, this shows a larger variability in the results obtained (Fig. 5), depending on the data set considered (and thus the data content).

B. Experiments With High Spatial Resolution Images (VHR)

In this section, we evaluate our proposed framework with two remotely sensed MSIs. The "Zurich Summer v1.0" data set[3] is a collection of 20 chips (crops), taken from a QB acqui-

[3]https://sites.google.com/site/michelevolpiresearch/data/zurich-dataset

sition of the city of Zürich (Switzerland) in August 2002. The QB images are composed by four channels (NIR-R-G-B) and were pansharpened to the PAN resolution of 0.62 meters/pixel. In this collection, eight different urban and periurban classes were manually annotated: Roads, Buildings, Trees, Grass, Bare Soil, Water, Railways, and Swimming pools [59]. The cumulative number of class samples is highly unbalanced to reflect real-world situations. And the purpose of distributing data sets is to encourage reproducibility of experiments. In this paper, we employ the third (926×943 pixels) and sixth (812×984 pixels) images, as an example to test our proposed method. As illustrated in the framework (see Fig. 2), the QB MSI images and their corresponding prepared ingredients are shown in Fig. 6. We extracted the gradient maps for both used image and transformed the images into three oversegmented superpixel representations by the SLIC
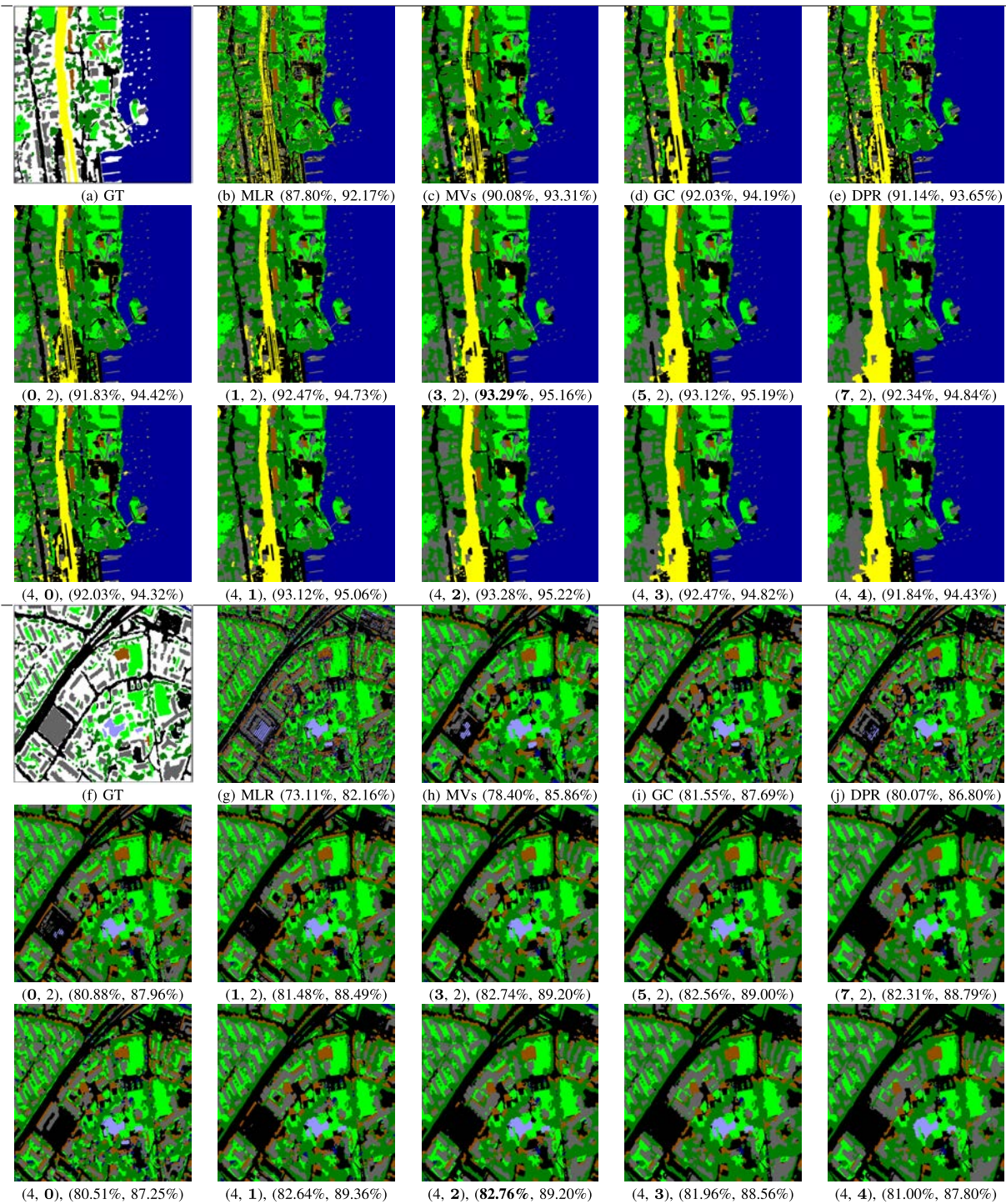
Fig. 7. Classification maps of two QB Zurich v1.0 MSI data sets. First row: Zurich3 data set. Fourth row: Zurich6 data set. The results of the Zurich3 data set by the proposed SuperSALSA algorithm are displayed in the second and third rows along with different values of parameter λ_1, $\lambda_2 \times c_i$, for $i = \{1, \ldots, 3\}$, followed by the overall accuracy and overall accuracy of edges with a radius of 5 pixels. In the fourth to sixth rows are the corresponding results of the Zurich6 data set. (a) and (f) GT. (b) and (g) MLR classification. (c) and (h) MV. (d) and (i) GC. (e) and (j) DPR.

method (by the VLFeat toolbox).[4] The result of our proposed method is obtained after 200 iterations with convergence, while 20 Monte Carlo runs are employed for all the considered methods.

To start with, the accuracy statistics of 20 Monte Carlo runs are shown in Tables IV and V with the mean plus/minus the

standard deviation reported in the tables. From both the tables, we can see that the accuracies are remarkably increased after the spatial information is incorporated by different methods. In comparison with the state-of-the-art methods, namely, MV, GC, and DPR, our proposed method achieves the highest values in terms of overall accuracy, κ statistic, and average accuracy, which is consistent with our results with HSIs. For illustrative purposes, we also display the classification

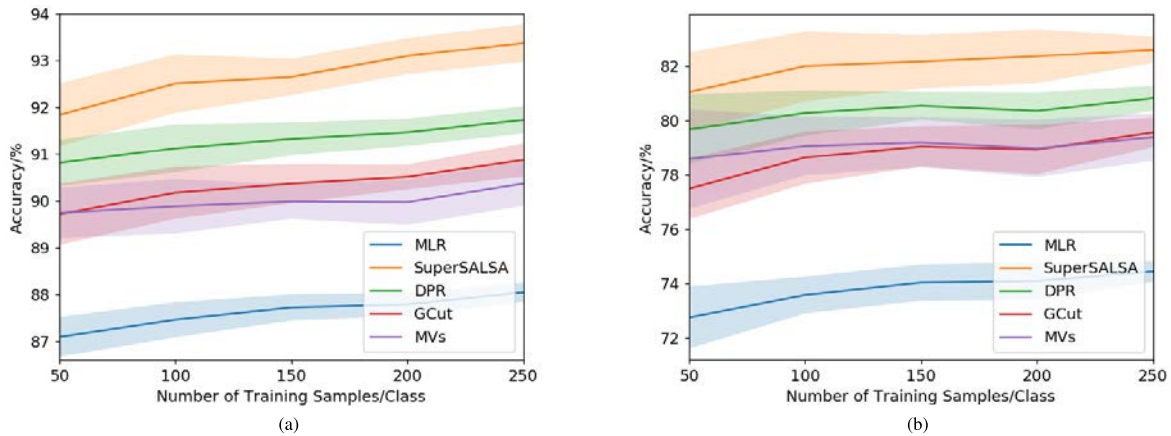[4]http://www.vlfeat.org/doc/api/slic.html

Fig. 8. Obtained OA lines for the two hyperspectral data of (a) QB Zurich 3 and (b) Zurich 6 Salinas data. Solid lines: average of 20 Monte Carlo runs for both the data sets. Colored area around the lines: standard deviation around the mean.

maps that are obtained with the methods in Fig. 7. First of all, it is remarkable that our proposed method obtains classification maps in which both the label consistency and boundary discontinuity are well refined, in comparison with the state-of-the-art methods. Also, the hand-tuned parameters of the VTV and GTV spatial regularizers provide an intuitive description of their effects on the classification performance of our proposed machinery (i.e., small values of the parameters preserve more trivial details of the land objects, while greater values keep the basic contours or strong boundaries of large-scale land objects). This is quite consistent with what have been observed with the experiments on HSIs in Section IV-A and, meanwhile, is what we originally explore in this paper. In turn, the discrete-rate variations for specific classes (possibly induced by the different characteristics of certain classes, such as their spatial textures) are more significant than in the experiments with HSIs. In fact, the discrete rates of the QB Zurich data sets are relatively low as compared with those of the ROSIS Pavia University and AVIRIS Salinas data sets. This is possibly due to the presence of rich spatial details in the QB data sets that introduce additional difficulties in order to converge to discrete results while making consistency and realigning the pixel labels using the SuperSALSA algorithm. Apart from its general performance on the whole image, the proposed SuperSALSA algorithm generally outperforms the state-of-the-art competitors in terms of boundary recall accuracies. This indicates that, spatially, the boundary definition of our proposed method outperforms its competitors in this paper by considering both the VTV and GTV regularizers. Specifically, the increase in λ_1 values for the VTV regularizer leads to a higher boundary accuracy, while very large values result in lower accuracies. These observations are consistent with the experimental results of the hyperspectral data sets (see Fig. 4). Finally, similar to the case of using HSIs, our proposed SuperSALSA algorithm consumes more time with the QB Zurich data sets. This can be addressed by resorting to high-performance computing architectures, since our method is amenable to parallelization.

Fig. 8 also displays the classification overall accuracies regarding different sizes of training samples in order to further assess the proposed SuperSALSA algorithm. It can

be observed that involving spatial information, especially the regularization by superpixels, is capable of dramatically improving the classification accuracy performances. On the other hand, the overall accuracies have been more promoted by proportion in this case when compared with the case of hyperspectral data sets (see Fig. 5). This is reasonable in the sense that the classification tasks for images of very high spatial resolution are often obstructed by the imbalance between the limited spectral bands and rich spatial variations of land objects. Furthermore, a comparison of Fig. 5 with Fig. 8 indicates that the standard deviation decreases as the training size increases, and different variations can also be observed with different data sets, but these are less apparent when using multispectral data sets.

V. CONCLUSION AND FUTURE LINES

This paper proposed a new method, SuperSALSA, which provides convex formulation to exploit the spatial information coming from multiple oversegmented superpixelizations into supervised HSI segmentation. With this method, we sidestep the NP-hardness of the original discrete integer optimization problem by using a relaxation technique based on linear programming, and successfully express the oversegmented superpixels in the form of a graphical spatial regularizer across the relaxed hidden field. In addition, we formulate a convex optimization problem and approximate the solution for the original NP-hard image labeling problem. Specifically, we design a framework to substantiate and validate the method. The experimental results, obtained with remotely sensed HSI and MSIs, demonstrate that our proposed method achieves the state-of-the-art performance when compared with other methods, such as MV, GC, and DPR techniques. Our proposed approach can also be viewed as a general framework for solving a range of similar problems, such as change detection, regression, and so on. Also the proposed approach is highly parallel and pixelwise-decoupled, and thus it can be straight-forwardly implemented in parallel using high-performance computing architectures. Our future perspectives will focus on exploring and evaluating the potential of the proposed approach when dealing with remote sensing data coming from multiple sources.

APPENDIX

We recall that the convex functions g_j, for $j = 1, \ldots, 4$, correspond to the data term, the sum-to-one constraint, the nonnegativity constraints, and the VTV regularizer, respectively, and for $j = 4, \ldots, J$, correspond to a set of closed, proper, and convex functions $g_j(\mathbf{z}\mathbf{H}_j)$ associated with the GTV. Since C is the number of graphs/superpixelizations, the number $J = C + 4$ denotes the total number of terms in the objective function (10).

We define the convex functions g_j, for $j = 1, 2, 3$, as follows:

$$g_1(\zeta) \equiv \sum_{i \in \mathcal{S}} \mathbf{q}_i^T \zeta_i$$
$$g_2(\zeta) \equiv \iota_+(\zeta)$$
$$g_3(\zeta) \equiv \iota_1(\zeta) \tag{13}$$

where $\mathbf{q}_i \equiv -\log(\mathbf{p}_i) \in \mathbb{R}^{K \times 1}$ is to be understood componentwise, and ζ_i and ζ_i are dummy variables whose dimensions depend on functions g_j for $j = 1, \ldots, 3$. $\iota_+(\cdot)$ is the indicator function defined on the set $\zeta \in \mathbb{R}_+^{K \times n}$ with: $\iota_+(\zeta) = 0$ if $\zeta \in \mathbb{R}_+^{K \times n}$ and $\iota_+(\zeta) = +\infty$, otherwise. Likewise, $\iota_1(\zeta)$ is the indicator in set $\{\mathbf{1}_n\}$, with $\iota_1 = 0$ if and only if $\zeta \in \{\mathbf{1}_n\}$, and $+\infty$ otherwise.

A. SALSA Formulation

Adapting the formulations (11) and (12), we apply the C-SALSA methodology [41]. We denote the scaled Lagrange multipliers associated with the constraints $\mathbf{Gz} = \mathbf{u}$ as $\mathbf{d} = [\mathbf{d}^1, \ldots, \mathbf{d}^J]$ and thus have the following C-SALSA-based formulation for (11):

$$\mathbf{z}_{k+1} = \arg\min_{\mathbf{z}} \|\mathbf{z}\mathbf{G} - \mathbf{u} - \mathbf{d}\|_F^2 \tag{14}$$

$$\mathbf{u}_{k+1}^j = \arg\min_{\mathbf{u}^j} g_j(\mathbf{u}^j) + \frac{\mu}{2} \|\mathbf{z}^{k+1}\mathbf{H}^j - \mathbf{u}^j - \mathbf{d}_k^j\|_F^2 \tag{15}$$

$$\mathbf{d}_{k+1}^j = \mathbf{d}_k^j - \mathbf{z}_{k+1}\mathbf{H}^j - \mathbf{u}_{k+1}^j \tag{16}$$

where $\mu > 0$ is the weight of the augmented Lagrangian term. We remark that if the optimization 10 has a solution, the sequence $\{\mathbf{z}_k\}$ converges to it; otherwise, at least one of the sequences $\{\mathbf{u}_k\}$ or $\{\mathbf{d}_k\}$ diverges [41].

B. Moreau Proximity Operators

The optimization subproblems associated with (15) can be solved through proximal methods, by computing the associated MPOs [50] of each of the convex functions. We first present the closed-form expressions of these operators for the data-fit term, and sum-to-one and nonnegativity constraints.

1) Moreau Proximity Operator for g_1: The MPO for the data-fit term g_1 is

$$\psi_{g_1/\mu}(\mathbf{v}) = \arg\min_{\zeta} \left(\sum_{i \in \mathcal{S}} \mathbf{q}_i^T \zeta_i \right) + \frac{\mu}{2} \|\zeta - \mathbf{v}\|_F^2$$

where $\mathbf{v} \equiv [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbb{R}^{K \times n}$, and $\zeta \equiv [\zeta_1, \ldots, \zeta_n] \in \mathbb{R}^{K \times n}$. This optimization is decoupled (pixelwise) with respect to ζ_i for $i \in \mathcal{S}$, meaning that

$$\psi_{g_1/\mu}(\mathbf{v}) = (\psi_{g_1/\mu}(\mathbf{v}_1), \ldots, \psi_{g_1/\mu}(\mathbf{v}_n))$$

with

$$\psi_{g_1/\mu}(\mathbf{v}_i) = \arg\min_{\zeta_i} \mathbf{q}_i^T \zeta_i + \frac{\mu}{2} \|\zeta_i - \mathbf{v}_i\|_F^2$$

whose solution is given by

$$\psi_{g_1/\mu}(\mathbf{v}_i) = \mathbf{v}_i + \mathbf{q}_i/\mu.$$

This operator has a complexity of $O(Kn)$, the number of classes across the number of pixels.

2) Moreau Proximity Operator for g_2: The MPO for the sum-to-one constraint g_2 is

$$\psi_{g_2/\mu}(\mathbf{v}) = \arg\min_{\zeta} \iota_1(\zeta) + \frac{\mu}{2} \|\zeta - \mathbf{v}\|_F^2$$
$$= \left(\mathbf{I} - \frac{\mathbf{1}_K \mathbf{1}_K^T}{K} \right) \mathbf{v} + \frac{\mathbf{1}_K \mathbf{1}_n^T}{K} \tag{17}$$

where $\mathbf{v}, \zeta \in \mathbb{R}^{K \times n}$. This operator has a complexity of $O(Kn)$, the number of classes $\times$ the number of pixels.

3) Moreau Proximity Operator for g_3: The MPO for the nonnegativity constraint f_3 is

$$\psi_{g_3/\mu}(\mathbf{v}) = \arg\min_{\zeta} \iota_+(\zeta) + \frac{\mu}{2} \|\zeta - \mathbf{v}\|_F^2 = \max\{\mathbf{0}, \mathbf{v}\}$$

where $\mathbf{v}, \zeta \in \mathbb{R}^{K \times n}$. This operator has a complexity of $O(Kn)$, the number of classes $\times$ the number of pixels.

4) Moreau Proximity Operator for g_4: The inclusion of the VTV prior in (9) introduces the term

$$\lambda_1 \sum_{i \in \mathcal{S}} \eta_i \sqrt{\|[\mathbf{z}\mathbf{D}_h](:,i)\|^2 + \|[\mathbf{D}_v](:,i)\mathbf{z})_i\|^2}.$$

We define the linear operator $\mathbf{H}_4 : \mathbb{R}^{K \times n} \to \mathbb{R}^{2K \times n}$ as

$$\mathbf{H}_4 = [\mathbf{D}_h \ \mathbf{D}_v] \tag{18}$$

where $\mathbf{D}_h$ and $\mathbf{D}_v$ correspond to the circular horizontal and vertical difference operators previously defined. The corresponding convex function g_4 is defined as

$$g_4(\zeta) = \lambda_1 \sum_{i \in \mathcal{S}} \eta_i \sqrt{\|\zeta_i^h\|^2 + \|\zeta_i^v\|^2} \tag{19}$$

where $\zeta = [\zeta^h \ \zeta^v] \in \mathbb{R}^{2K \times n}$, and ζ^h and ζ^v belong to the range of the horizontal and vertical difference operators $\mathbf{D}_h$ and $\mathbf{D}_v$, respectively.

The MPO for the VTV prior is thus

$$\psi_{g/\mu}(\mathbf{v}) = \arg\min_{\zeta} \left(\sum_{i \in \mathcal{S}} \eta_i \sqrt{\|\zeta_i^h\|^2 + \|\zeta_i^v\|^2} \right) + \frac{\mu}{2\lambda_1} \|\zeta - \mathbf{v}\|_F^2$$

where $\mathbf{v}, \zeta \in \mathbb{R}^{K \times 2n}$ and $\zeta^h, \zeta^v \in \mathbb{R}^{K \times n}$. This optimization is pixelwise-decoupled and the solution of each subproblem is the vector soft thresholding operator [50]

$$\psi_{g_4/\mu}(\mathbf{v}_i) = \max\{0, \|\mathbf{v}_i\| - \lambda_1 \eta_i/\mu\} \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|} \tag{20}$$

which has complexity $O(Kn)$.

5) Optimization of Graph Total Variation: As introduced in (10), the GTV constraints correspond to g_j, for $j = 5, 6, \ldots, J$, associated with C superpixelizations. Also having (8) in mind, the MPO of the GTV (superpixelization) is

$$
\psi_{g_j/\mu}(\boldsymbol{v}) = \arg\min_{\boldsymbol{\zeta}} \lambda_2 \omega_j \sum_{t=1}^{T_j} \|(\mathbf{A}_{j,t} - \mathbf{I})(\boldsymbol{\zeta}_{|\mathcal{V}_{j,t}})^T\|_F^2
$$
$$
+ \frac{\mu}{2} \|\boldsymbol{\zeta}_{|\mathcal{V}_{j,t}} - \boldsymbol{v}_{|\mathcal{V}_{j,t}}\|_F^2
$$
$$
= \arg\min_{\boldsymbol{\zeta}} \lambda_2 \omega_j \|(\mathbf{A}_j - \mathbf{I})\boldsymbol{\zeta}^T\|_F^2 + \frac{\mu}{2} \|\boldsymbol{\zeta} - \boldsymbol{v}\|_F^2
$$
(21)

where $\boldsymbol{v}, \boldsymbol{\zeta} \in \mathbb{R}^{K \times n}$, with j, for $j \in \{1, \ldots, C\}$, as the jth segmentation and t, for $t \in \{1, \ldots, T_j\}$, as the tth superpixel. The MPO is thus

$$
\psi_{g_j/\mu}(\boldsymbol{v}) = \frac{\mu}{2\lambda_2 \omega_j} \left((\mathbf{A}_j - \mathbf{I})^T (\mathbf{A}_j - \mathbf{I}) + \frac{\mu}{2\lambda_2 \omega_j} \mathbf{I} \right)^{-1} \boldsymbol{v}.
$$
(22)

Note that the solution of (21) can be obtained in a superpixelwise manner with decoupling (21) by each superpixel $\mathcal{V}_{j,t}$. Hence, we represent the coefficient matrix with $\mathbf{F}$ and there comes

$$
\mathbf{F}_j^{-1} = \frac{\mu}{2\lambda_2 \omega_j} \left((\mathbf{A}_j - \mathbf{I})^T (\mathbf{A}_j - \mathbf{I}) + \frac{\mu}{2\lambda_2 \omega_j} \mathbf{I} \right)^{-1}
$$
$$
= \begin{bmatrix} \mathbf{F}_{j,1}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{j,2}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{F}_{j,t}^{-1} \end{bmatrix}
$$

whereby $\mathbf{F}_j^{-1}$ is a diagonal block matrix comprised of $\mathbf{F}_{j,t}^{-1}$, for $i \in \mathcal{V}_j$, with

$$
\mathbf{F}_{j,t}^{-1} = \frac{\mu}{2\lambda_2 \omega_j} \left((\mathbf{A}_{j,t} - \mathbf{I})^T (\mathbf{A}_{j,t} - \mathbf{I}) + \frac{\mu}{2\lambda_2 \omega_j} \mathbf{I} \right)^{-1}.
$$

Then, trivially,

$$
[\mathbf{A}_{j,t} - \mathbf{I}]
$$
$$
= \begin{bmatrix} \frac{1}{n_t} - 1 & \frac{1}{n_t} & \cdots & \frac{1}{n_t} \\ \frac{1}{n_t} & \frac{1}{n_t} - 1 & \cdots & \frac{1}{n_t} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{n_t} & \frac{1}{n_t} & \cdots & \frac{1}{n_t} - 1 \end{bmatrix}
$$
$$
[(\mathbf{A}_{j,t} - \mathbf{I})^T (\mathbf{A}_{j,t} - \mathbf{I})]
$$
$$
= \begin{bmatrix} \frac{n_t - 1}{n_t} & -\frac{1}{n_t} & \cdots & -\frac{1}{n_t} \\ -\frac{1}{n_t} & \frac{n_t - 1}{n_t} & \cdots & -\frac{1}{n_t} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{n_t} & -\frac{1}{n_t} & \cdots & \frac{n_t - 1}{n_t} \end{bmatrix}
$$
$$
= -\frac{1}{n_t} \mathbf{E} + \mathbf{I}.
$$

where $\mathbf{A}_{j,t}$ are actually the diagonal blocks of $\mathbf{A}_j$, which corresponds to a fully connected subgraph of the partition graph with n_t nodes ($n_t = n_{j,t}$ to keep notation light) corresponding to a partition element with n_t pixels, and $\mathbf{E} =$ denotes an $n_t \times n_t$ matrix of 1s. Note that $\mathbf{A}_{j,t}$ is normalized by the pixel number of its corresponding superpixel and that $\mathbf{E} = \mathbf{v}\mathbf{v}^T$, with $\mathbf{v}$ being a column vector of 1s. Defining $\theta_j = (\mu/2\lambda_2 \omega_j)$, we have

$$
\mathbf{F}_{j,t}^{-1} = \theta_j ((\mathbf{A}_{j,t} - \mathbf{I})^T (\mathbf{A}_{j,t} - \mathbf{I}) + \theta_j \mathbf{I})^{-1}
$$
$$
= \theta_j \left(\left(-\frac{1}{n_t} \mathbf{E} + \mathbf{I} \right) + \theta_j \mathbf{I} \right)^{-1}
$$
$$
= -\theta n_t ((-1 - \theta_j) n_t \mathbf{I} + \mathbf{v}\mathbf{v}^T)^{-1}.
$$

Using the matrix inversion lemma and noting that

$$
1 + \mathbf{v}^T ((-1 - \theta_j) n_t \mathbf{I})^{-1} \mathbf{v} = \frac{\theta_j}{1 + \theta_j}
$$

we may write

$$
\mathbf{F}_{j,t}^{-1} = \left(\frac{\theta_j}{1 + \theta_j} \mathbf{I}_{|\mathcal{V}_{j,t}} + \frac{1}{(1 + \theta_j) n_t} \mathbf{E}_{|\mathcal{V}_{j,t}} \right)
$$
$$
= \left(\frac{\mu}{\mu + 2\lambda_2 \omega_j} \mathbf{I}_{|\mathcal{V}_{j,t}} + \frac{2\lambda_2 \omega_j}{(\mu + 2\lambda_2 \omega_j) n_t} \mathbf{E}_{|\mathcal{V}_{j,t}} \right). \quad (23)
$$

Hence, the MPO of the GTV is

$$
\psi_{g_j/\mu}(\boldsymbol{v}) = \left(\frac{\mu}{\mu + 2\lambda_2 \omega_j} \mathbf{I} + \frac{2\lambda_2 \omega_j}{(\mu + 2\lambda_2 \omega_j) n_t} \mathbf{E} \right) \boldsymbol{v} \quad (24)
$$

which can be pixelwise-decoupled as

$$
\psi_{g_j/\mu}(\boldsymbol{v}_i) = \frac{\mu}{\mu + 2\lambda_2 \omega_j} \boldsymbol{v}_i + \frac{2\lambda_2 \omega_j}{(\mu + 2\lambda_2 \omega_j) n_t} \sum_{k=1}^{n_t} \boldsymbol{v}_k \quad (25)
$$

where $i, k \in \mathcal{S} \equiv \{1, \ldots, n\}$ index the pixels that belong to the same tth partition as the ith, kth pixel. To be specific, the first term of (25) corresponds to the value on the ith node itself, the value of the ith pixel, and the second term corresponds to the mean of $\boldsymbol{v}$ on the fully connected subgraph that the ith node belongs to, such that this operator has a complexity of $O(Kn)$. Thus, the computational complexity of all the C GTVs is $O(CKn)$. To summarize, the computational complexity in total of the proposed SuperSALSA algorithm is $O(JKn)$ per iteration.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. 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] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, 2012.

[3] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[4] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[5] G. Zhang, X. Jia, and J. Hu, "Superpixel-based graphical model for remote sensing image mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5861–5871, Nov. 2015.

[6] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.

[7] J. Li, M. Khodadadzadeh, A. Plaza, X. Jia, and J. M. Bioucas-Dias, "A discontinuity preserving relaxation scheme for spectral–spatial hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 625–639, Feb. 2016.

[8] X. Xu, X. Tong, A. Plaza, Y. Zhong, H. Xie, and L. Zhang, "Joint sparse sub-pixel mapping model with endmember variability for remotely sensed imagery," *Remote Sens.*, vol. 9, no. 1, p. 15, 2016.

[9] R. Jain, R. Kasturi, and B. Schunck, *Machine Vision*, vol. 5. New York, NY, USA: McGraw-Hill, 1995.

[10] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.

[11] J. Cheng *et al.*, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE Trans. Med. Imag.*, vol. 32, no. 6, pp. 1019–1032, Jun. 2013.

[12] S. Thayammal and D. Selvathi, "A review on segmentation based image compression techniques," *J. Eng. Sci. Technol. Rev.*, vol. 6, no. 3, pp. 134–140, 2013.

[13] Y. J. Zhang, "A review of recent evaluation methods for image segmentation," in *Proc. 6th Int. Symp. Signal Process. Appl.*, vol. 1. 2001, pp. 148–151.

[14] M. M. S. J. Preetha, L. P. Suresh, and M. J. Bosco, "Image segmentation using seeded region growing," in *Proc. Int. Conf. Comput., Electron. Elect. Technol. (ICCEET)*, Mar. 2012, pp. 576–583.

[15] Y. Boykov, V. S. Lee, H. Rusinek, and R. Bansal, "Segmentation of dynamic N-D data sets via graph cuts using Markov models," in *Medical Image Computing and Computer-Assisted Intervention*. Dordrecht, The Netherlands: Springer, 2001, pp. 1058–1066.

[16] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.

[17] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 266–277, Feb. 2001.

[18] C. Li, C.-Y. Kao, J. C. Gore, and Z. Ding, "Minimization of region-scalable fitting energy for image segmentation," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1940–1949, Oct. 2008.

[19] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.

[20] H. D. Li, M. Kallergi, L. P. Clarke, V. K. Jain, and R. A. Clark, "Markov random field for tumor detection in digital mammography," *IEEE Trans. Med. Imag.*, vol. 14, no. 3, pp. 565–576, Sep. 1995.

[21] C. Nieuwenhuis, E. Töppe, and D. Cremers, "A survey and comparison of discrete and continuous multi-label optimization approaches for the Potts model," *Int. J. Comput. Vis.*, vol. 104, no. 3, pp. 223–240, 2013.

[22] M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers, "An experimental comparison of discrete and continuous shape optimization methods," in *Computer Vision*. Marseille, France: Springer, 2008, pp. 332–345.

[23] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," in *Proc. 8th IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1. Jul. 2001, pp. 105–112.

[24] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1124–1137, Sep. 2004.

[25] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient N-D image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, 2006.

[26] V. Jeleň, J. Janáček, Z. Tomori, "Mobility tracking by interactive graph-cut segmentation with Bi-elliptical shape prior," in *Proc. IEEE 8th Int. Symp. Appl. Mach. Intell. Inform. (SAMI)*, Jan. 2010, pp. 225–230.

[27] P. Kohli, L. Ladický, and P. H. S. Torr, "Robust higher order potentials for enforcing label consistency," *Int. J. Comput. Vis.*, vol. 82, no. 3, pp. 302–324, 2009.

[28] H. Wang and D. Koller, "Subproblem-tree calibration: A unified approach to max-product message passing," in *Proc. ICML*, 2013, pp. 190–198.

[29] A. Globerson and T.-S. Jaakkola, "Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2008, pp. 553–560.

[30] T. Werner, "A linear programming approach to max-sum problem: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1165–1179, Jul. 2007.

[31] V. Kolmogorov and C. Rother, "Minimizing nonsubmodular functions with graph cuts-a review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1274–1279, Jul. 2007.

[32] C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer, "Optimizing binary MRFs via extended roof duality," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.

[33] Q. Xu, Q. Chen, S. Yang, and X. Liu, "Superpixel-based classification using *K* distribution and spatial context for polarimetric SAR images," *Remote Sens.*, vol. 8, no. 8, p. 619, 2016.

[34] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[35] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, 2010.

[36] J. M. Bioucas-Dias, F. Condessa, and J. Kovačević, "Alternating direction optimization for image segmentation using hidden Markov measure field models," *Proc. SPIE*, vol. 9019, p. 90190P, Feb. 2014.

[37] V. Kolmogorov and M. Wainwright. (2012). "On the optimality of tree-reweighted max-product message-passing." [Online]. Available: https://arxiv.org/abs/1207.1395

[38] F. Condessa, J. M. Bioucas-Dias, and J. Kovačević, "Supervised hyperspectral image segmentation: A convex formulation using hidden fields," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Lausanne, Switzerland, Jun. 2014. [Online]. Available: http://repository.cmu.edu/cgi/viewcontent.cgi?article=1313&context=ece

[39] T. Pock, A. Chambolle, D. Cremers, and H. Bischof, "A convex relaxation approach for computing minimal partitions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 810–817.

[40] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.

[41] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.

[42] J. L. Marroquin, E. A. Santana, and S. Botello, "Hidden Markov measure field models for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 11, pp. 1380–1387, Nov. 2003.

[43] M. A. T. Figueiredo, "Bayesian image segmentation using wavelet-based priors," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Jun. 2005, pp. 437–443.

[44] S. Lefkimmiatis, A. Roussos, M. Unser, and P. Maragos, "Convex generalizations of total variation based on the structure tensor with applications to inverse problems," in *Proc. Int. Conf. Scale Space Variat. Methods Comput. Vis.*, Leibnitz, Austria, 2013, pp. 48–60.

[45] X. Bresson and T. F. Chan, "Fast dual minimization of the vectorial total variation norm and applications to color image processing," *Inverse Problems Imag.*, vol. 2, no. 4, pp. 455–484, 2008.

[46] B. Goldluecke, E. Strekalovskiy, and D. Cremers, "The natural vectorial total variation which arises from geometric measure theory," *SIAM J. Imag. Sci.*, vol. 5, no. 2, pp. 537–563, 2012.

[47] C. Zach, D. Gallup, J. Frahm, and M. Niethammer, "Fast global labeling for real-time stereo using multiple plane sweeps," in *VMV*. Konstanz, Germany: Aka GmbH, 2008, pp. 243–252.

[48] F. Condessa, J. M. Bioucas-Dias, and J. Kovačević, "SegSALSA-STR: A convex formulation to supervised hyperspectral image segmentation using hidden fields and structure tensor regularization," in *Proc. IEEE GRSS Workshop Hyperspectral Image Signal Process., Evol. Remote Sens. (WHISPERS)*, Jun. 2015, pp. 1–4.

[49] D. M. Greig, B. T. Porteous, and A. H. Seheult, "Exact maximum a posteriori estimation for binary images," *J. Roy. Statist. Soc. Ser. B (Methodol.)*, vol. 51, no. 2, pp. 271–279, 1989.

[50] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011, pp. 185–212.

[51] S. Chen *et al.*, "Signal inpainting on graphs via total variation minimization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2014, pp. 8267–8271.

[52] F. Condessa, "Robust image classification with context and rejection," Ph.D. dissertation, Inst. Superior Técnico, Univ. Lisboa, Lisbon, Portugal, 2016.

[53] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[54] J. M. Bioucas-Dias and M. Figueeredo, "Logistic regression via variable splitting and augmented Lagrangian tools," Inst. Superior Técnico, Univ. Lisboa, Lisbon, Portugal, 2009.

[55] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semi-supervised hyperspectral image classification based on a Markov random field and sparse multinomial logistic regression," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, vol. 3. Jul. 2009, p. III-817.

[56] 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.

[57] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao, and Z. Wei, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2270–2278, Jun. 2016.

[58] Z. Wu *et al.*, "GPU parallel implementation of spatially adaptive hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, to be published. [Online]. Available: http://www.umbc.edu/rssipl/people/aplaza/Papers/Journals/2017.JSTARS.Adaptive.pdf

[59] M. Volpi and V. Ferrari, "Semantic segmentation of urban scenes by learning local class interactions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 1–9.

José M. Bioucas-Dias (S'87–M'95–SM'15–F'17) received the E.E., M.Sc., Ph.D., and Habilitation degrees from the Instituto Superior Técnico (IST), Universidade Técnica de Lisboa (now Universidade de Lisboa), 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 is currently an Associate Professor and teaches inverse problems in imaging and electric communications. He is also a Senior Researcher with the Pattern and Image Analysis Group, Instituto de Telecomunicações, Lisbon, which is a private nonprofit research institution. His research interests include inverse problems, signal and image processing, pattern recognition, optimization, and remote sensing. He has authored or co-authored over 300 scientific publications, including over 90 journal papers (78 of which published in IEEE journals) and 210 peer-reviewed international conference papers and book chapters.

Dr. Bioucas-Dias was the General Co-Chair of the 3rd IEEE Geoscience and Remote Sensing Society Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing in 2011 and has been a member of program/technical committees of several international conferences. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1997 to 2000 and the IEEE TRANSACTIONS ON IMAGE PROCESSING from 2010 to 2014. He was a Guest Editor of six special issues of IEEE journals, such as the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (two times), the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING, the *IEEE Geoscience and Remote Sensing Magazine*. He is currently a Senior Area Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is included in Thomson Reuters' Highly Cited Researchers 2015 list.

Yi Liu (S'13) received the Ph.D. degree from the University of Extremadura, Badajoz, Spain, in 2017. During his Ph.D. research, he was involved in developing spectral partitioning and superpixel-based spatial partitioning techniques and algorithms.

Since 2017, he has been a Post-Doctoral Researcher with the Norwegian University of Science and Technology, where he is currently involved in medical hyperspectral image processing and analysis using inversing problems, pattern recognition, and machine learning techniques. His research interests include image labeling problems for hyperspectral remote sensing images, and hyperspectral image processing and analysis.

Dr. Liu was a recipient of the 2017 top Second GRSS Student Paper Prize at the IEEE Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA.

Filipe Condessa (S'13–M'16) received the Ph.D. degree in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 2016, and the Ph.D. degree from the Universidade de Lisboa, Lisbon, Portugal, in 2016.

He is currently a Post-Doctoral Research Associate in electrical and computer engineering with Carnegie Mellon University, where he develops machine learning algorithms for analysis and prediction of spatiotemporal evolution of social behavior and socioeconomical indicators. His research interests include signal processing, robust machine learning, inverse problems, and social network analysis and prediction.

Dr. Condessa was a recipient of the 2015 Mikio Takagi Prize for Best Student Paper at the IEEE Geoscience and Remote Sensing Symposium, Milan, Italy.

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, Instituto de Telecomunicações, IST, Universidade Técnica de Lisboa, in the framework of the European Doctorate for Signal Processing (SIGNAL). Since 2011, she has been a Post-Doctoral Researcher with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain. She is currently a Professor with Sun Yat-sen University, Guangzhou, China. 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 co-operation on hyperspectral imaging at the European level. Her research interests include hyperspectral image classification and segmentation, spectral unmixing, signal processing, and remote sensing.

Dr. Li 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*, the *Journal of Applied Remote Sensing*, and *Inverse Problems and Imaging*. She is currently an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Peijun Du (M'07–SM'12) received the Ph.D. degree in geomatics engineering from the China University of Mining and Technology, Xuzhou, China, in 2001.

He was a Post-Doctoral Fellow with Shanghai Jiao Tong University, Shanghai, China, from 2002 to 2004, and a Senior Visiting Scholar with the University of Nottingham, Nottingham, U.K., and the Grenoble Images Speech Signals and Automatics Laboratory, Grenoble Institute of Technology, Grenoble, France. He is currently a Professor with the Remote Sensing, Department of Geographic Information Sciences, Nanjing University, Nanjing, China, and the Deputy Director with the Key Laboratory for Satellite Mapping Technology and Applications, National Administration of Surveying, Mapping, and Geoinformation, Beijing, China. His research interests include remote sensing image processing and pattern recognition, hyperspectral remote sensing, and applications of geospatial information technologies.

Dr. Du has been an Associate Editor of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2009. He was a Guest Editor of three special issues of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING. He served as the Co-Chair of the Technical Committee of IEEE GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion Over Urban Areas (URBAN) 2009, International Association for Pattern Recognition-Pattern Analysis in Remote Sensing 2012, and International Workshop on Earth Observation and Remote Sensing Applications (EORSA) 2014, the Co-Chair of the Local Organizing Committee of International Joint Urban Remote Sensing Event 2009, the Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote Sensing (WHISPERS) 2012, and EORSA 2012, and a member of Scientific Committee or Technical Committee of other international conferences, such as Spatial Accuracy 2008, Asian Conference on Remote Sensing 2009, WHISPERS (2010–2016), URBAN (in 2011, 2013, and 2015, respectively), MultiTemp (in 2011, 2013, and 2015, respectively), ISDIF 2011, and the International Society For Optics and Photonics European Conference on Image and Signal Processing for Remote Sensing (2012–2016).

Antonio Plaza (M'05–SM'07–F'15) was born in Caceres, Spain, in 1975.

He is currently 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. He has been the advisor of over 30 M.Sc. dissertations and 12 Ph.D. dissertations. He was the Coordinator of the Hyperspectral Imaging Network, a European project with total funding of 2.8 million Euro. He has authored over 400 publications, including 130 JCR journal papers (82 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* (CRC Press/Taylor & Francis) (the first book on this topic in the published literature) and guest edited eight special issues on hyperspectral remote sensing for different journals. His research interests include remotely sensed hyperspectral image analysis and efficient implementations of large-scale scientific problems on high-performance computing architectures.

Dr. Plaza served as the Director of Education Activities for the IEEE Geoscience and Remote Sensing Society (GRSS) from 2011 to 2012. He was a member of the Editorial Board of the IEEE GEOSCIENCE AND REMOTE SENSING NEWSLETTER from 2011 to 2012 and the *IEEE Geoscience and Remote Sensing Magazine* in 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 has been serving as the President of the Spanish Chapter of IEEE GRSS since 2012. He has served as a Proposal Evaluator for the European Commission (Marie Curie Actions, Engineering Panel), the European Space Agency, the Belgium Science Policy, the Israel Science Foundation, and the Spanish Ministry of Science and Innovation. He has participated in the Tenure Track Selection Committee of different Universities in Italy, Spain, and Australia. He has reviewed over 500 articles for over 50 different journals. He is currently serving as the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He is also an Associate Editor of the IEEE ACCESS and the *Journal of Real-Time Image Processing*. He was a recipient of the Best Ph.D. Dissertation Award at the University of Extremadura in 2002. He was a 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 from 2007 to 2012. He has co-authored the 2011 Best Student Paper at the IEEE International Conference on Space Technology, and a recipient of the 2008 Best Paper award at the IEEE Symposium on Signal Processing and Information Technology. He was a recipient of the most highly cited paper in the *Journal of Parallel and Distributed Computing* from 2005 to 2010 and the 2013 Best Paper Award of the JSTARS journal.