



# Supervised spatio-spectral classification of fused images using superpixels

KAREN SANCHEZ,<sup>1</sup> CARLOS HINOJOSA,<sup>2</sup> AND HENRY ARGUELLO<sup>2,\*</sup>

<sup>1</sup>Department of Electrical Engineering, Universidad Industrial de Santander, Bucaramanga, 680002, Colombia

<sup>2</sup>Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, 680002, Colombia

\*Corresponding author: henarf@uis.edu.co

Received 17 October 2018; revised 31 December 2018; accepted 8 January 2019; posted 9 January 2019 (Doc. ID 348427); published 19 February 2019

The low spatial resolution of hyperspectral (HS) images generally limits the classification accuracy. Therefore, different multiresolution data fusion techniques have been proposed in the literature. In this paper, a method for supervised classification of spectral images from data fusion measurements is proposed. Specifically, the proposed approach exploits the spatial information of an RGB image by grouping pixels with similar characteristics into superpixels and fuses such features with the spectral information of an HS image. Simulations results on three datasets show that the proposed classification method improves the overall accuracy and reduces the computational complexity compared to the traditional approach that first performs the fusion followed by the classification. © 2019 Optical Society of America

<https://doi.org/10.1364/AO.58.0000B9>

## 1. INTRODUCTION

Spectral imaging (SI) is a technique used to acquire spatial information of a scene in many contiguous electromagnetic frequencies. The acquired image is commonly represented as a 3D data cube, where two dimensions correspond to the spatial information and the third one to the spectral domain. Traditional sensing techniques construct such a data cube by scanning the scene, either spectrally or spatially, in proportion to the desired spatial or spectral resolution. In general, SI is a valuable tool for remote sensing applications including precision agriculture, urban planning, military surveillance, etc. [1–3].

Based on the spectral/spatial resolution, spectral range, and width and contiguity of bands, SI sensors can be categorized as hyperspectral (HS) or multispectral (MS). In particular, HS devices capture hundreds of spectral bands of the scene of interest in a wavelength interval ranging from the visible region (0.4–0.7 μm) to the infrared region (≈2.4 μm). However, despite its high spectral resolution, HS imagery usually suffers from low spatial resolution compared to that obtained with an MS or a typical RGB sensor. Therefore, spectral image fusion has emerged as a challenging processing task in the remote sensing field that consists of the combination of a high-spectral- but low-spatial-resolution image with a low-spectral- but high-spatial-resolution image for obtaining a high-spatial- and high-spectral-resolution image [4]. Based on the used theoretical approach, data fusion algorithms can be categorized as component substitution, spectral unmixing, sparse signal representation, and Bayesian estimation [5–10]. Traditionally, these algorithms fuse the information acquired by traditional spatio-spectral imaging sensors, resulting

in high storage and processing costs [11]. After applying a fusion technique, the spectral signatures of the pixels (a.k.a spectral pixels) are commonly processed to detect and classify different materials within the enhanced spatio-spectral data cube. Particularly, spectral image classification, which aims at assigning each spectral pixel to one class belonging to a set of categories, has drawn broad attention and has led to a variety of methods [12–14]. In the literature, many of the methods have concentrated on exploring the role of the spectral signatures in classification, employing exclusively the spectral signatures for determining the classification maps. Since spectral images can be seen as a set of 2D images acquired at different spectral wavelengths, the spatial information is naturally another source of data that can be considered to improve the classification performance. Then, the introduction of spatial dependency offers the possibility to boost the pixel-wise classification methods [15–18].

Among image processing methods, the superpixel technique is well known to efficiently embed the spatial neighboring information into the classification procedure. This technique is based on the oversegmentation strategy that partitions an image into homogeneous subregions [19,20]. Such subregions are usually irregular spatial units, but perceptually consistent, i.e., all pixel values in a superpixel area are most likely uniform, meaning that superpixel methods are adapted to real scenes [21]. Therefore, combining superpixel partitioning with conventional spectral classification approaches is expected to boost the classification performance, considering not only the spectral behavior of the spectral signatures but the variations of spatial characteristics of the scene of interest.

In this paper, a supervised spatial-spectral classification approach of high-spectral- and high-spatial-resolution fused images is developed, where the superpixel segmentation technique is used to incorporate the spatial information in the fused data. Specifically, we first apply a superpixel algorithm to the rich spatial resolution RGB image in order to group spatial neighboring pixels, with similar characteristics, in segments or superpixels. Then, an optimization problem is formulated to fuse the high spectral information of an HS image with the spatial information given by the superpixels of the MS image. Finally, the pixels of the fused data cube are classified using a supervised classification algorithm. The main contribution of this paper is a methodology that fuses and, simultaneously, incorporates spatial information of the scene for boosting the classification accuracy. This approach, in turn, provides a framework for low computational complexity and accurate classification.

The paper is organized as follows: in Section 2, the matrix model and some notation used through the paper are introduced. The mathematical description and flow chart of the proposed classification methodology is described in Section 3. In order to evaluate the performance of the proposed method, several simulation results are presented in Section 4 with different spectral image datasets. Some concluding remarks are summarized in Section 5.

## 2. MATRIX MODEL

In general, we denote  $\mathcal{F}$  as the spatio-spectral data cube, with  $M \times N$  spatial dimensions,  $L$  spectral bands, and entries denoted as  $\mathcal{F}_{n_1, n_2, k}$ , where  $n_1$  and  $n_2$  index the spatial coordinates, and  $k$  determines the  $k$ -th spectral band. The data cube can be rearranged as a matrix  $\mathbf{F} \in \mathbb{R}^{MN \times L}$  whose rows  $\{\mathbf{f}_j^T\}_{j=0}^{MN}$  are the spectral signatures of the data cube. Specifically, the rearrangement can be expressed as

$$\mathbf{F} = [[\mathcal{F}_{0,0,0}, \mathcal{F}_{1,0,0}, \dots, \mathcal{F}_{0,1,0}, \dots, \mathcal{F}_{(M-1),(N-1),0}]^T, \\ \dots, [\mathcal{F}_{0,0,(L-1)}, \dots, \mathcal{F}_{(M-1),(N-1),(L-1)}]^T]. \quad (1)$$

In this work, the classification maps of a high-spatial- and high-spectral-resolution image  $\mathbf{F}$  are found from a fused image obtained from an RGB and an HS image, denoted as  $\mathbf{F}_m \in \mathbb{R}^{MN \times 3}$  and  $\mathbf{F}_h \in \mathbb{R}^{M_h N_h \times L_h}$ , respectively. Note that  $\mathbf{F}_m$  can be composed by combining three spectral bands representing

the red, green, and blue colors. Further,  $\mathbf{F}_h$  can be seen as a blurred and spatially downsampled instance of  $\mathbf{F}$  with  $N_h = N/p$  and  $M_h = M/p$ , where  $p$  is the downsampling factor along the spatial coordinates. Mathematically, the matrices  $\mathbf{F}_m$  and  $\mathbf{F}_h$  can be expressed as

$$\mathbf{F}_m = \mathbf{F} \mathbf{D}_m, \quad (2)$$

$$\mathbf{F}_h = \mathbf{D}_h \mathbf{F}, \quad (3)$$

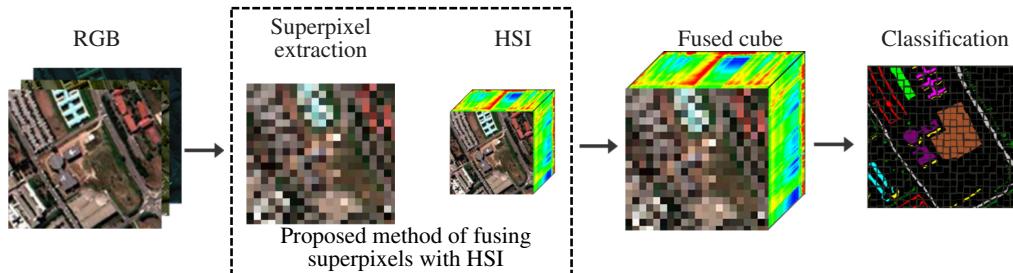
where  $\mathbf{D}_m \in \mathbb{R}^{L \times 3}$  is a sparse selection matrix with only three nonzero values, i.e., a single one per column that indicates the selected spectral bands composing the RGB image and  $\mathbf{D}_h \in \mathbb{R}^{M_h N_h \times M N}$  is a spatial downsampling that includes both the blur operation and the spatial decimation of  $\mathbf{F}$  with downsampling factor  $p$ .

## 3. PROPOSED SPATIO-SPECTRAL CLASSIFICATION APPROACH

The aim of the proposed method is to incorporate the spatial information of the spectral scene in the classification features using an RGB image as side information. The flow chart of the algorithm is depicted in Fig. 1. In general, a rich spatial and spectral fused cube  $\mathbf{F} \in \mathbb{R}^{M_m N_m \times L_h}$  is obtained using the spectral information of an HS image and the superpixels extracted from the RGB image. Then, the fused cube is later used as the input for a supervised pixel-based classifier.

### A. Superpixel Segmentation

In Fig. 1, the first step in the proposed methodology is to segment an RGB image in superpixels. The superpixels technique is based on the oversegmentation strategy that gathers a group of uniform pixels into homogeneous subregions. In this work, we use the simple linear iterative clustering (SLIC) superpixel algorithm [22,23]. SLIC works in the 5D space, where two coordinate components  $(x, y)$  depict the spatial location of the segment and the other three components corresponds to the RGB color channels. Given  $N_{seg}$  desired equally sized superpixels, where the approximation size of each superpixel is  $N^2/N_{seg}$ , the first step of SLIC is to define a cluster center at every grid interval  $S = \sqrt{N^2/N_{seg}} = N/\sqrt{N_{seg}}$ . The algorithm assumes that the pixels associated with a cluster lie in a  $2S \times 2S$  area around the superpixel center on the  $(x, y)$  plane. Therefore, this becomes the search area for the pixels near to each cluster center. To avoid keeping the center on the edge of



**Fig. 1.** Flow chart of the proposed classification framework. An RGB image is segmented into a predefined number of superpixels. The superpixel image and the HS image are used to obtain a high spectral and spatial resolution, which are used as classification features. Finally, a supervised classification algorithm is applied on the fused image.

an object, it is transferred to the lowest gradient position in a  $3 \times 3$  neighborhood. In the next step, for each cluster center, SLIC assigns the best-matching pixels from the search area according to the distance measure,

$$D_c = \sqrt{(R_j - R_{j'})^2 + (G_j - G_{j'})^2 + (B_j - B_{j'})^2}, \quad (4)$$

$$D_p = \sqrt{(x_j - x_{j'})^2 + (y_j - y_{j'})^2}, \quad (5)$$

$$D_t = D_c + \frac{m}{S} D_p, \quad (6)$$

where  $R_j, G_j, B_j$  corresponds to the color of the  $j$ -th pixel, and  $m$  controls the compactness of a superpixel, which is usually chosen as  $m = 10$ . The complexity of SLIC algorithm is  $O(MN)$ , since it needs to compute distances from any point to no more than eight cluster centers and the number of iterations is constant [23]. Notice that the region grows at most 2 times the cluster radius with the SLIC algorithm.

Once the segmentation map of the RGB image has been created using SLIC, the next step in the proposed method is to create the upsampling matrix  $\mathbf{U} \in \mathbb{R}^{N_{\text{seg}} \times M\bar{N}}$ , which collects the superpixel information. First, denote  $\mathbf{p}^e$  as the vector of size  $n_e$  containing the indices of all pixels belonging to the superpixel  $e$ . Then, the nonzero values of the  $e$ -th row of  $\mathbf{U}$ , denoted in vector form as  $(\mathbf{u}_e)^T$ , are determined by the entries of  $\mathbf{p}^e$  and the value of  $n_e$  as follows:

$$(\mathbf{u}_e)_{(\mathbf{p}^e)_l}^T = \frac{1}{n_e}, \quad \text{for } l = 0, \dots, n_e, \quad (7)$$

where  $(\mathbf{u}_e)_{(\mathbf{p}^e)_l}^T$  denotes the position in  $\mathbf{u}_e$  indexed by the  $l$ -th entry of the  $\mathbf{p}^e$  vector.

## B. Problem Formulation

The following step in the proposed method is to perform the fusion operation  $\mathbf{F}_h$  spectral information and the superpixels extracted from the RGB image  $\mathbf{F}_m$ . Using the notation given in Section 2, the proposed fusion operation is formulated as the optimization problem,

$$\min_{\mathbf{F}} \frac{1}{2} \|\mathbf{F}_h - \mathbf{D}_h \mathbf{F}\|_F^2 + \lambda \|\mathbf{F}\|_*, \quad (8)$$

where  $\|\cdot\|_F$  stands for the Frobenius norm and  $\lambda$  is a regularization parameter. Since pixels in  $\mathbf{F}$  that belongs to the same class share a common low-rank pattern, the nuclear norm minimization  $\|\mathbf{F}\|_*$ , in Eq. (8), is used to obtain the lowest rank representation that captures the global structure of the data [24].

As the matrix  $\mathbf{F}$  is obtained from the superpixel information, it is possible to relate the  $\mathbf{U}$  matrix with  $\mathbf{F}$  as

$$\mathbf{F} = \mathbf{U}^T \bar{\mathbf{F}}, \quad (9)$$

where  $\bar{\mathbf{F}} \in \mathbb{R}^{N_{\text{seg}} \times L}$  contains the fused spectral information of each superpixel. Then, assuming that  $\text{rank}(\bar{\mathbf{F}}) \approx \text{rank}(\mathbf{F})$ , the optimization problem in Eq. (8) can be rewritten as

$$\min_{\bar{\mathbf{F}}} \frac{1}{2} \|\mathbf{F}_h - \mathbf{D}_h \mathbf{U}^T \bar{\mathbf{F}}\|_F^2 + \lambda \|\bar{\mathbf{F}}\|_*. \quad (10)$$

## 1. ADMM Algorithm to Solve the Proposed Optimization Problem

The minimization in Eq. (10) can be efficiently solved by the well-known alternating direction method of multipliers (ADMM). First, an auxiliary matrix  $\mathbf{Z}$  with the same size of  $\bar{\mathbf{F}}$  is used to separate the problem. In this way, problem in Eq. (10) becomes

$$\begin{aligned} & \min_{\bar{\mathbf{F}}, \mathbf{Z}} \frac{1}{2} \|\mathbf{F}_h - \mathbf{D}_h \mathbf{U}^T \bar{\mathbf{F}}\|_F^2 + \lambda \|\mathbf{Z}\|_* \\ & \text{s.t. } \mathbf{Z} = \bar{\mathbf{F}}. \end{aligned} \quad (11)$$

The augmented Lagrangian associated with the optimization in Eq. (11) can be written as

$$\begin{aligned} \mathcal{L}(\bar{\mathbf{F}}, \mathbf{Z}, \mathbf{G}) = & \frac{1}{2} \|\mathbf{F}_h - \mathbf{D}_h \mathbf{U}^T \bar{\mathbf{F}}\|_F^2 + \lambda \|\mathbf{Z}\|_* \\ & + \frac{\rho}{2} \|\bar{\mathbf{F}} - \mathbf{Z} + \mathbf{G}\|_F^2, \end{aligned} \quad (12)$$

where  $\mathbf{G}$  is the scaled dual variable and  $\rho > 0$  is the weighting of the augmented Lagrangian term [25]. The ADMM solution of Eq. (11) is summarized in Algorithm 1 that consists of minimizing  $\bar{\mathbf{F}}$ ,  $\mathbf{Z}$ , and  $\mathbf{G}$ , alternatively. Specifically, the minimization of the three matrices consists of the following updates:

1) *Updating  $\bar{\mathbf{F}}$ :*  $\bar{\mathbf{F}}^{(t)}$  is obtained by minimizing  $\mathcal{L}$  with respect to  $\bar{\mathbf{F}}$ , while  $(\mathbf{Z}, \mathbf{G})$  are fixed. We calculate the derivative of  $\mathcal{L}$  with respect to  $\bar{\mathbf{F}}$ , and set it to zero to obtain  $\bar{\mathbf{F}}$  as follows:

$$\bar{\mathbf{F}}^{(t)} = \frac{\mathbf{U}^T \mathbf{D}_h^T \mathbf{F}_h + \rho \mathbf{I}(\mathbf{Z} - \mathbf{G})}{\rho \mathbf{I} + \mathbf{U}^T \mathbf{D}_h^T \mathbf{D}_h \mathbf{U}}. \quad (13)$$

2) *Updating  $\mathbf{Z}$ :* the update of  $\mathbf{Z}$  is given by the singular value thresholding operator [26]. Specifically, we compute the SVD of the symmetric matrix,

$$\bar{\mathbf{F}} + \mathbf{G} = \hat{\mathbf{U}} \Sigma \hat{\mathbf{U}}^*, \quad (14)$$

and apply the soft-thresholding operator to the resulting singular values,

$$S_{1/\rho}(\Sigma) = \text{diag}\{(\sigma_i - 1/\rho)_+\}, \quad (15)$$

where  $(a)_+ := \max\{a, 0\}$ . Then, the update of  $\mathbf{Z}$  is defined as

$$\mathbf{Z}^{(t)} = \hat{\mathbf{U}} S_{1/\rho}(\Sigma) \hat{\mathbf{U}}^*. \quad (16)$$

After obtaining the estimation of  $\bar{\mathbf{F}}$ , the rich spatio-spectral image can be obtained using Eq. (9).

## Algorithm 1. ADMM Subiterations to Estimate $\bar{\mathbf{F}}$

**Input:**  $\mathbf{F}_h, \lambda, \rho, \mathbf{D}_h, \mathbf{U}$  MAXITER.

*Initialization:*

1:  $\mathbf{Z}^{(0)} = \mathbf{0}, \mathbf{G}^0 = \mathbf{0}$

*Main iterations*

2: **for**  $t = 1$  to MAXITER **do**

3:      $\bar{\mathbf{F}}^{(t)} \leftarrow \arg \min_{\bar{\mathbf{F}}} \mathcal{L}(\bar{\mathbf{F}}, \mathbf{Z}^{(t-1)}, \mathbf{G}^{(t-1)})$

4:      $\mathbf{Z}^{(t)} \leftarrow \arg \min_{\mathbf{Z}} \mathcal{L}(\bar{\mathbf{F}}^{(t)}, \mathbf{Z}, \mathbf{G}^{(t-1)})$

5:      $\mathbf{G}^{(t)} \leftarrow \mathbf{G}^{(t-1)} + \bar{\mathbf{F}} - \mathbf{Z}$

6: **end for**

## C. SUPERVISED CLASSIFICATION

Given the spectral superpixels  $\bar{\mathbf{F}}$  and the respective high-spectral- and high-spatial-resolution image  $\mathbf{F}$ , obtained with

Eq. (9), the multiclass supervised classification method aims at determining the class label that better fits the corresponding  $j_s$ -th pixel, for  $j_s = 0, \dots, N_{\text{seg}} - 1$ . Notice that with the proposed method, it is only necessary to determine the class label of the  $N_{\text{seg}}$  spectral superpixels (rows of  $\bar{\mathbf{F}}$ ). Then, the pixels in  $\mathbf{F}$  acquire the same label of the superpixel, to which they belong. As the number of superpixels  $N_{\text{seg}} \ll MN$ , the proposed approach boosts the computational performance, as the next section shows. In this paper, the classification of the spectral scene is achieved using two well-known algorithms: the support vector machine (SVM) and the  $k$ -nearest neighbor (KNN). In addition, we denote  $\Omega = \{\omega_1, \dots, \omega_c\}$  as the set of  $c$  class labels and  $\Theta = \{\theta_1, \dots, \theta_v\}$  as the set of  $v$  integers indexing the training samples.

## 1. SVM

Let us first consider the binary supervised classification problem, where each feature vector of the training set is associated with a binary output  $\beta_\theta \in \{-1, 1\}$  for  $\Theta = \{\theta_1, \dots, \theta_v\}$ . In this context, the SVM focuses on determining a hyperplane that optimally separates the samples belonging to different classes. This problem is equivalent to minimizing the norm of a weight vector  $\mathbf{h}$  that is normal to the decision hyperplane [27,28]. Specifically, this classification method solves the following optimization problem:

$$\begin{aligned} & \min_{\mathbf{h}, b_0, \mathcal{X}_\Theta} \left\{ \frac{1}{2} \|\mathbf{h}\|_2^2 + \eta \sum_{\Theta} \mathcal{X}_\Theta \right\} \\ & \text{s.t. } \beta_\theta (\gamma(\mathbf{f}_\Theta^T)^T \mathbf{h} + b_0) \geq 1 - \mathcal{X}_\Theta \\ & \quad \mathcal{X}_\Theta \geq 0, \end{aligned} \quad (17)$$

for all  $\Theta = \{\theta_1, \dots, \theta_v\}$ , where  $\mathbf{f}_\Theta^T$  are the  $\Theta$  rows of  $\mathbf{F}$  used as training data,  $\mathcal{X}_\Theta$  is the set of slack variables that considers the nonseparability between sets belonging to different classes,  $\eta$  is the regularization parameter that controls the influence of the misclassified samples,  $\gamma$  is a nonlinear function that maps the feature vector to the kernel space, and  $b_0$  is the offset of the decision hyperplane with respect to the origin of the coordinate system [29].

As mentioned before, SVMs are basically binary classifiers. However, a multiclass strategy for the spectral image classification is required. To overcome this drawback, the one-against-one multiclass strategy is adopted, where two classes are separately analyzed by implementing a binary classifier, ignoring, at the same time, the remaining classes [30]. Finally, using the training sample set  $\{\mathbf{f}_\Theta^T\}_{\Theta=\theta_i}^{\theta_v}$ , the problem of classifying each test pixel consists of finding the class  $\Omega$  whose training pixel  $\Theta$  is the nearest to the test pixel in the Euclidean distance,

$$\text{Class}(\bar{\mathbf{f}}_{j_s}^T) = \min_{\Omega=\Omega_1, \dots, \Omega_c} \|\bar{\mathbf{f}}_{j_s}^T - (\mathbf{f}_\Theta^T)_{(\Omega)}\|_2^2, \quad (18)$$

where  $\bar{\mathbf{f}}_{j_s}^T$  is the  $j_s$ -th row of  $\bar{\mathbf{F}}$  and  $(\mathbf{f}_\Theta^T)_{(\Omega)}$  denotes the training pixels (rows of  $\mathbf{F}$ ) belonging to the class  $\Omega$ . Notice that the SVM training is performed using the rows of  $\mathbf{F}$  as training data, and the classification is performed on the superpixels (rows of  $\bar{\mathbf{F}}$ ).

## 2. KNN

Following the same previous introduced notation, denote  $\mathcal{N}_{j_s}$  as the set containing the  $k_n$  nearest neighbors of the  $j_s$  pixel  $\bar{\mathbf{f}}_{j_s}^T$

with labels denoted as  $\{\ell_1, \ell_2, \dots, \ell_{k_n}\}$ . Then, the KNN classifier finds the  $k_n$  nearest neighbors of the  $\bar{\mathbf{f}}_{j_s}^T$  point in the training data and assigns such testing point to the most frequently occurring class of its  $k_n$  neighbors. The KNN performs the classification with the following majority voting rule:

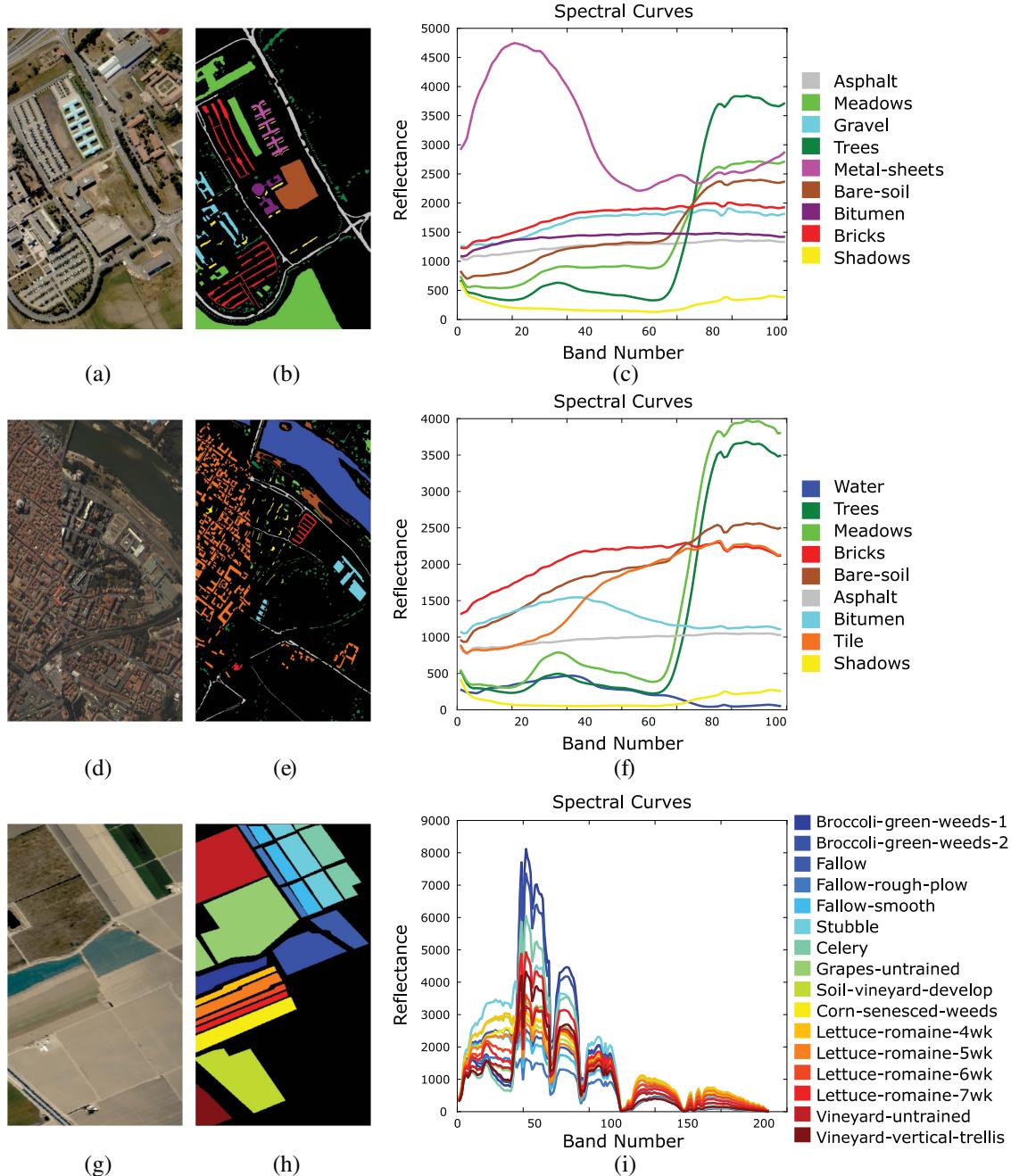
$$\text{Class}(\bar{\mathbf{f}}_{j_s}^T) = \arg \max_{\Omega=\Omega_1, \dots, \Omega_c} \sum_{i=1}^{k_n} \delta(\ell_i, \Omega), \quad (19)$$

where  $\delta(\cdot, \cdot)$  is the Kronecker delta function, which is equal to 1 if all its arguments are equal, and 0 otherwise.

## 4. SIMULATIONS AND RESULTS

In this section, the performance of the proposed spectral image classification method is evaluated. In particular, we test the classification framework on three datasets, the Pavia University, the Pavia Center, and the Salinas Valley. The results are compared against other approaches that fuse high-spatial- and high-spectral-resolution images, followed by the classification. Pavia University is an image acquired over an urban area surrounding the University of Pavia, northern Italy, by the Reflective Optics System Imaging Spectrometer (ROSIS-03) airborne sensor. The size of the dataset is  $610 \times 340$  pixels and 103 spectral bands, with a high spatial resolution of 1.3 meters per pixel and a spectral coverage ranging from 0.43 to 0.84  $\mu\text{m}$ . Figures 2(a)–2(c) show the color composite of Pavia University; the corresponding ground truth image, which differentiates nine land-cover classes: asphalt, meadows, gravel, trees, metal sheets, bare soil, bitumen, bricks, and shadows; and the characteristic spectral signatures of each class, respectively. Pavia Center is an image also acquired with the ROSIS-03 sensor with size of  $1096 \times 715$  pixels and 102 spectral bands, and it retains the same spatial and spectral resolution characteristics as the Pavia University dataset. The ground truth of Pavia Center contains nine classes: water, trees, asphalt, self-blocking bricks, bitumen, tiles, shadows, meadows, and bare soil. A false-color images of the Pavia Center, the ground truth, and its spectral signatures are shown in Figs. 2(d)–2(f). The third spectral dataset with which the proposed method was tested is the Salinas image. This scene was collected in 1998 by the airborne visible/infrared imaging spectrometer (AVIRIS) on Salinas Valley, California, USA. The size of the dataset is  $512 \times 217$  pixels and 204 spectral bands in the range of 0.24 to 2.40  $\mu\text{m}$ . The Salinas ground truth contains 16 land-cover classes: broccoli green weeds 1, broccoli green weeds 2, fallow, fallow rough plow, fallow smooth, stubble, celery, grapes untrained, soil vineyard develop, corn senesced green weeds, lettuce romaine 4 wk, lettuce romaine 5 wk, lettuce romaine 6 wk, lettuce romaine 7 wk, vineyard untrained, and vineyard vertical trellis. Figures 2(g)–2(i) shows a false-color image of Salinas Valley, its ground truth, and the spectral signatures of each class, respectively.

Following the matrix model described in Section 2, the HS images ( $\mathbf{F}_h$ ) are obtained with a decimation ratio across the spatial coordinates of  $p = 4$ . Therefore, the Pavia University image was resized to a high-resolution spectral image of  $152 \times 85 \times 103$ ; the Pavia Center image was spatially reduced to  $274 \times 178$  pixels and 102 spectral bands; and the Salinas Valley image to  $128 \times 54 \times 204$ . On the other hand, with a selection matrix

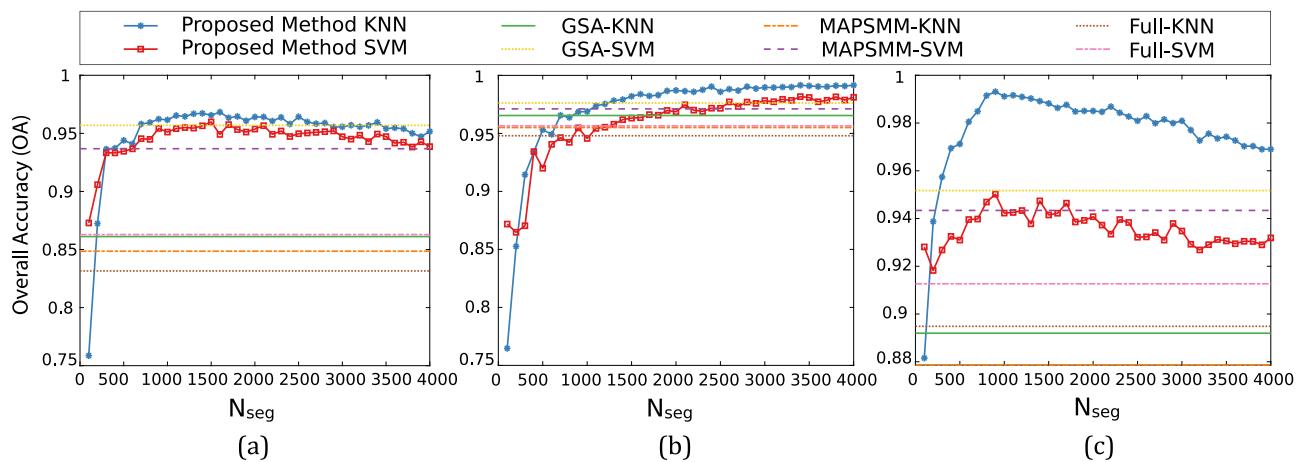


**Fig. 2.** False-color images, ground truth maps, and spectral signatures of each land-cover class for (a)–(c) Pavia University, (d)–(f) Pavia Center, and (g)–(i) Salinas Valley, respectively.

applied to the original datasets, a high-resolution spatial RGB ( $F_m$ ) image of each dataset was obtained. The size of the  $F_m$  images for Pavia University, Pavia Center, and Salinas are  $610 \times 340 \times 3$ ,  $1096 \times 715 \times 3$ , and  $512 \times 217 \times 3$ , respectively.

The experiments described below were performed by randomly selecting 5% of the labeled referenced data as training samples. Note that, in the presented results, we refer to “full-data” when classifying the complete high spatial and spectral image, i.e., no fusion was performed. The aim of adding the full-data results is to evidence the impact of incorporating spatial information in the classification performance. In addition,

we compare the proposed approach with performing the fusion procedure with two state-of-the-art algorithms, before classifying the obtained high spatio-spectral pixels. Specifically, “GSA” refers to the results achieved classifying a high-spectral- and high-spatial-resolution image obtained using the state-of-the-art fusion method GS adaptive (GSA) algorithm [31]. Similarly, “MAP-SMM” corresponds to the results achieved classifying a high-spectral- and high-spatial-resolution image reconstructed with the MAP-SMM algorithm [32]. In all the simulations, the datasets were classified with the SVM and KNN algorithms.



**Fig. 3.** Overall accuracy of classification varying the number of segments ( $N_{\text{seg}}$ ) in the proposed method when classifying (a) Pavia University, (b) Pavia Center, and (c) Salinas Valley. The obtained OA when using the full-data, and the GSA and the MAP-SSM fusion methods are provided for comparison purposes.

#### A. Impact of the Number of Superpixels and Training Samples in the Classification Accuracy

In the first experiment, the number of superpixels to cluster the image was varied to analyze the impact in the overall accuracy (OA). In order to identify an adequate number of segments that provides the highest classification accuracy for each dataset, the value of  $N_{\text{seg}}$  was varied from 100 to 4000 in steps of 100. Figure 3 shows the overall accuracy results of the proposed method with both classifiers, SVM and KNN, by varying the number of segments  $N_{\text{seg}}$  in each dataset.

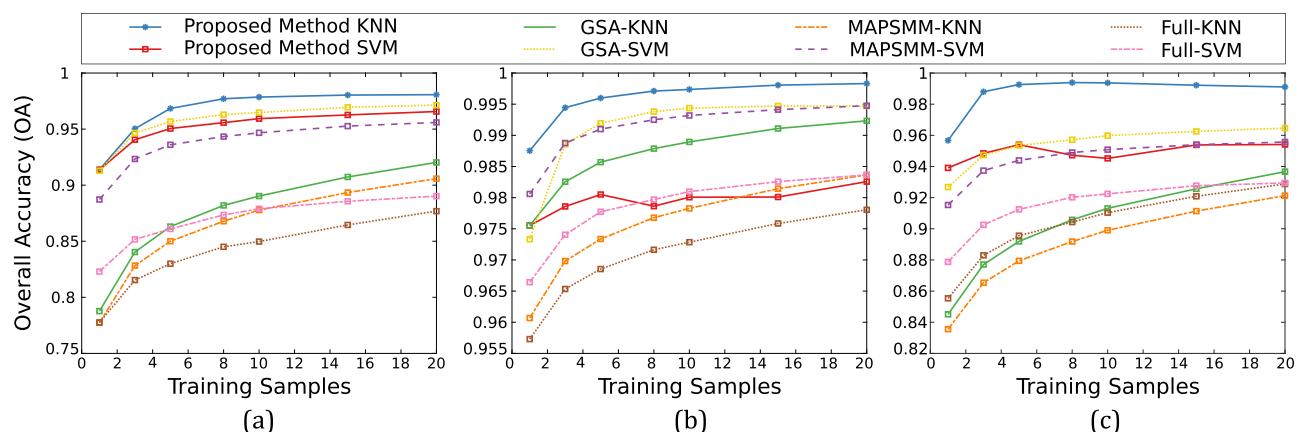
Figure 3(a) presents the obtained results for the Pavia University. Note that, as the number of desired segments ( $N_{\text{seg}}$ ) increases, the overall accuracy increases exponentially until reaching  $N_{\text{seg}} = 1600$ , where the OA begins to decrease slowly, as depicted with the red and blue lines. Furthermore, note that the proposed method with KNN outperforms the results obtained by classifying the full-data and data obtained by the GSA and MAP-SSM fusion methods when  $N_{\text{seg}} \geq 700$  and  $N_{\text{seg}} \leq 2800$ .

The obtained results for Pavia Center are shown in Fig. 3(b). As observed, when the number of superpixels increases, the

performance of the classification improves exponentially, until  $N_{\text{seg}} \approx 3400$  where the overall accuracy begins to tend to  $\approx 99.61\%$ . As observed, in this particular image, all the methods perform well; however, with an adequate selected number of segments, the proposed method with KNN outperforms the others.

Figure 3(c) shows the results obtained for the Salinas image. In this experiment, using the KNN classifier, the classification accuracy increases exponentially when the number of superpixels increases. However, when  $N_{\text{seg}} = 900$ , the OA begins to decrease. In this experiment, it is also important to note that the performance of the KNN classifier was significantly better than SVM. In addition, the results obtained with the proposed method and the KNN classifier outperform all the other methods when  $N_{\text{seg}} \geq 300$ .

In the second experiment, the number of samples used in the training stage was varied to analyze the impact in the OA. As observed in Fig. 4, all results obtained from the KNN classifier provide low-accuracy classification results in comparison with SVM, except for the proposed method, which exhibits good



**Fig. 4.** Overall accuracy of classification varying the number of training samples in the proposed method when classifying (a) Pavia University, (b) Pavia Center, and (c) Salinas Valley. The obtained OA when using the full-data, and the GSA and the MAP-SSM fusion methods are provided for comparison purposes.

performance for the tested number of training samples. This is expected since the KNN classifier takes more advantage of the spatial contextual information given by superpixels.

In general, for the proposed method with KNN, increasing the number of training samples will also improve the overall accuracy. However, when the number of training samples is greater than 8, the overall accuracy does not vary notably. In addition, from Fig. 4(a), it can be noted that, for 1% of training samples, the proposed method with both KNN and SVM classifiers provides the same results as those given by GSA-SVM.

## B. Classification Maps and Quantitative Results

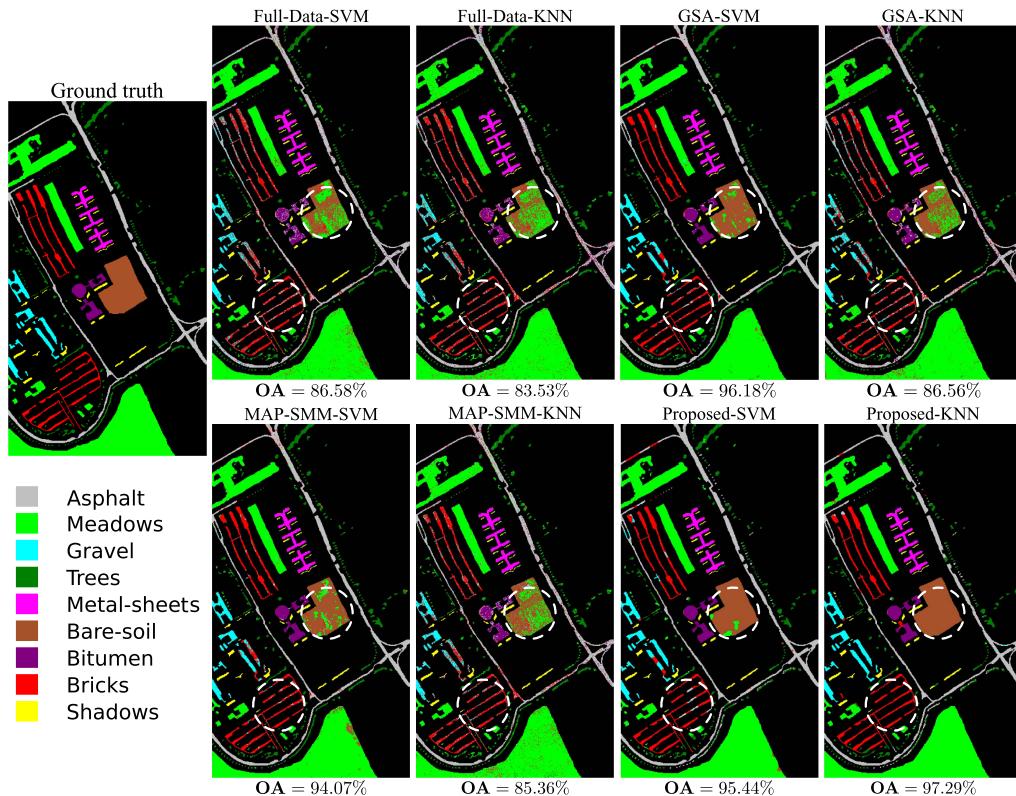
To further validate the performance of the proposed framework, classification maps and quantitative results from full-data, GSA, MAP-SMM, and the proposed approach are presented in Figs. 5–7 and Tables 1–3. In particular, Table 1 shows the numerical results for each in the nine land-cover classes in Pavia University. The presented results of average accuracy (AA), OA, the Kappa coefficient, and the running time are the average of 25 realizations of each experiment. In the table, the optimal values are shown in bold font, and the second-best results are underlined. All the numerical results, except the Kappa coefficients and the running time, are given in percentage.

From Fig. 5 and Table 1, it can be clearly seen that the proposed classification approach provides the best accuracy results in comparison with classifying the pixels of the full and the fused spectral data cubes. The OA values for the different classification approaches are shown in each subfigure of Fig. 5. For this dataset, the best accuracy classification results are provided by the proposed framework using the KNN classifier.

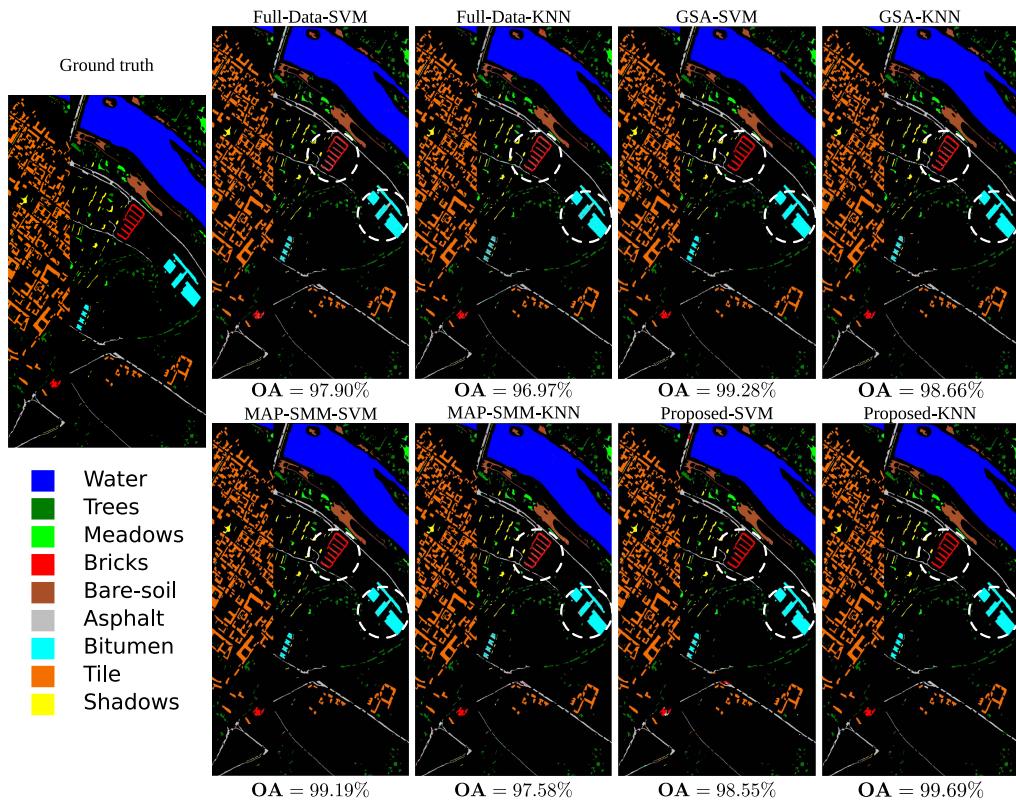
In addition, using the proposed method with the KNN classifier provides the best classification time, as noticed in Table 1. All execution times shown in this paper include the time of the fusion stage.

Figure 6 and Table 2 show the visual and numerical results of the Pavia Center dataset. Notice that when the KNN classifier is used with the proposed approach, the best overall accuracy and Kappa coefficient are obtained. In addition, this method takes the shortest execution time. On the other hand, note that although when using the SVM classifier with the proposed method the classification results are lower than using KNN, these are the second-best results in terms of overall accuracy, Kappa, and time.

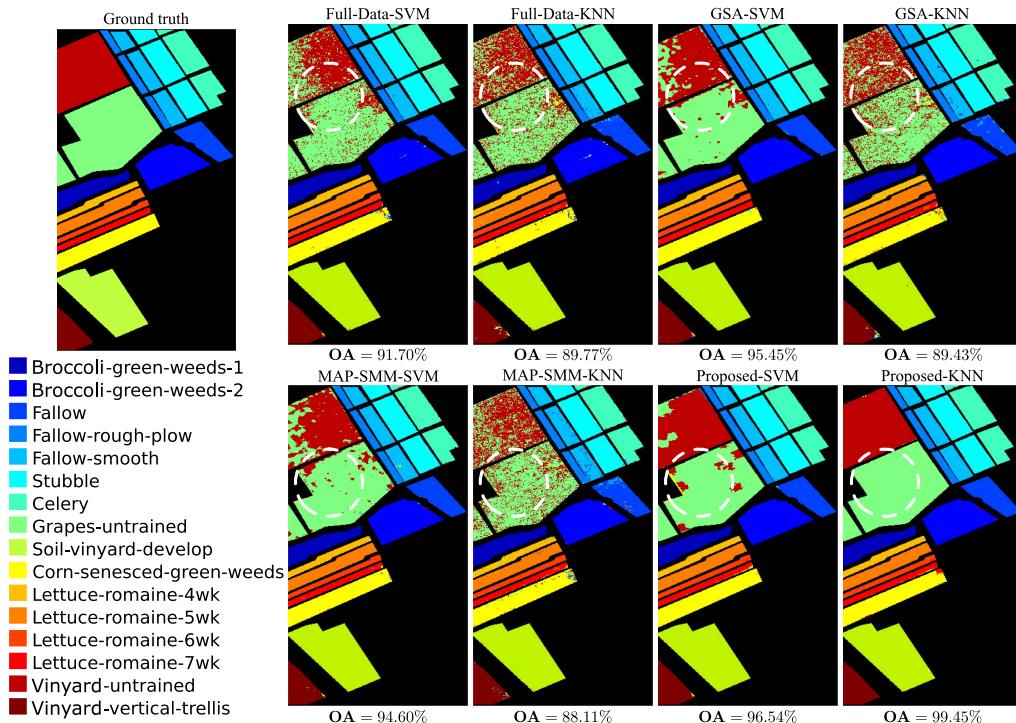
The performance of the classification of the Salinas Valley dataset is shown in Fig. 7, and the quantitative results are presented in Table 3. This spectral image exhibits the greatest increment in classification performance when using the proposed method, compared to the other approaches. By applying the classifier directly on the original full data, OA classification results of 91.29% and 89.47% were obtained, in 33.424 and 17.903 s, respectively, whereas the proposed method using the KNN classifier achieves an OA of 99.21% in 1.414 s. In addition, note that the second-best classification results are obtained using the GSA fusion approach and the SVM classifier. However, the second-best classification time was obtained by the proposed method and the SVM classifier. It is important to note that the good classification results obtained by the GSA and the MAP-SMM are greatly due to the denoising process performed in those methods.



**Fig. 5.** Visual classification results on Pavia University image. Figure shows the ground truth and the results for full-data, GSA, MAP-SMM, and the proposed approach using both SVM and KNN classifiers.



**Fig. 6.** Visual classification results on Pavia Center image. Figure shows the ground truth and the results for full-data, GSA, MAP-SMM, and the proposed approach using both SVM and KNN classifiers.



**Fig. 7.** Visual classification results on Salinas Valley image. Figure shows the ground truth and the results for full-data, GSA, MAP-SMM, and the proposed approach using both SVM and KNN classifiers.

**Table 1. Quantitative Results of the Different Classification Methods for the Pavia University Image**

Class	FULL		GSA		MAP-SMM		PROPOSED	
	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Asphalt	86.41	76.48	<b>97.04</b>	81.00	94.94	82.40	96.10	<u>96.28</u>
Meadows	95.75	94.74	98.89	95.47	97.79	94.28	<u>99.04</u>	<b>99.60</b>
Gravel	69.18	62.34	87.34	66.72	82.93	70.83	<u>90.40</u>	<b>96.61</b>
Trees	91.51	83.74	<b>94.07</b>	85.96	<u>91.91</u>	74.66	74.39	77.81
Metal sheets	99.40	99.43	<u>99.87</u>	<b>99.92</b>	99.00	98.90	99.16	99.27
Bare soil	61.19	55.81	87.59	70.90	82.13	66.00	<u>92.46</u>	<b>99.44</b>
Bitumen	72.41	79.02	94.73	76.27	95.08	80.56	<b>98.70</b>	<u>98.06</u>
Bricks	76.53	77.89	93.03	76.58	91.60	79.32	<u>93.34</u>	<b>93.39</b>
Shadows	95.81	94.91	<u>97.56</u>	<b>98.89</b>	95.73	94.02	95.80	96.60
AA	83.13	80.48	<u>94.46</u>	83.52	92.34	82.33	93.27	<b>95.23</b>
OA	86.24	83.01	<u>95.69</u>	86.11	93.68	84.85	95.00	<b>96.66</b>
Kappa	0.8164	0.7731	<u>0.9429</u>	0.8158	0.9162	0.7985	0.9337	<b>0.9559</b>
Time	16.7052	23.9618	16.5859	25.1008	27.0535	38.1522	<u>10.8399</u>	<b>6.8039</b>

**Table 2. Quantitative Results of the Different Classification Methods for the Pavia Center Image**

Class	FULL		GSA		MAP-SMM		PROPOSED	
	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Water	99.97	99.97	99.98	<u>99.99</u>	99.98	99.98	99.96	<b>100</b>
Trees	93.75	89.98	<u>97.54</u>	94.25	95.15	93.41	97.27	<b>98.56</b>
Meadows	86.27	87.58	<u>95.95</u>	92.02	87.16	87.27	93.07	<b>97.37</b>
Bricks	77.20	74.01	92.45	89.97	91.01	82.88	<b>99.48</b>	<u>97.55</u>
Bare soil	95.94	94.87	97.23	97.50	97.34	92.92	<b>99.38</b>	<u>98.85</u>
Asphalt	91.69	86.24	<u>97.40</u>	94.34	95.39	90.30	97.13	<b>99.29</b>
Bitumen	92.54	90.46	97.59	93.89	87.43	92.02	<b>99.07</b>	<u>98.56</u>
Tile	99.63	99.15	99.67	<u>99.85</u>	99.58	99.20	99.57	<b>99.93</b>
Shadows	96.70	94.32	98.32	<b>99.55</b>	90.97	98.29	90.68	<u>99.10</u>
AA	92.63	90.73	<u>97.35</u>	95.71	93.78	92.92	97.29	<b>98.80</b>
OA	97.74	96.84	99.11	98.54	98.00	97.51	<u>99.13</u>	<b>99.62</b>
Kappa	0.9680	0.9552	0.9874	0.9794	0.9716	0.9647	<u>0.9877</u>	<b>0.9946</b>
Time	35.5553	180.1488	54.4082	359.6640	58.2847	353.3343	<u>10.9937</u>	<b>7.3261</b>

**Table 3. Quantitative Results of the Different Classification Methods for the Salinas Valley Image**

Class	FULL		GSA		MAP-SMM		PROPOSED	
	SVM	KNN	SVM	KNN	SVM	KNN	SVM	KNN
Broccoli-green-1	98.83	99.00	99.48	98.13	99.47	99.74	<b>100</b>	<u>99.86</u>
Broccoli-green-2	99.23	99.10	<b>99.91</b>	99.78	99.33	99.12	<u>99.80</u>	99.60
Fallow	99.24	97.07	99.76	95.07	<b>99.91</b>	93.85	99.78	<u>99.78</u>
Fallow-rough-plow	98.86	99.52	98.09	98.39	98.58	98.87	<b>99.65</b>	<u>99.57</u>
Fallow-smooth	<b>99.35</b>	96.59	98.98	95.63	98.19	93.10	98.90	<u>98.95</u>
Stubble	99.83	99.70	<b>99.99</b>	99.86	99.86	99.50	<u>99.92</u>	99.79
Celery	99.47	99.41	99.48	98.96	99.08	98.61	<b>99.75</b>	<u>99.60</u>
Grapes-untrained	83.21	76.73	<u>92.60</u>	77.52	90.71	73.89	91.51	<b>99.72</b>
Soil-vineyard-develop	99.50	98.77	<b>99.99</b>	99.46	99.82	98.96	<u>99.94</u>	99.80
Corn-senesced-green-weeds	96.29	93.34	<u>98.32</u>	91.73	97.71	90.53	98.28	<b>98.67</b>
Lettuce-romaine-4wk	97.80	<u>98.25</u>	97.88	89.79	97.45	88.56	97.95	<b>98.93</b>
Lettuce-romaine-5wk	99.83	99.65	<b>99.97</b>	98.42	<u>99.90</u>	98.65	98.72	99.49
Lettuce-romaine-6wk	<u>98.01</u>	97.74	<b>98.67</b>	97.93	97.38	96.59	96.56	92.35
Lettuce-romaine-7wk	96.63	94.51	<b>98.90</b>	94.92	<u>96.80</u>	93.27	95.63	90.88
Vineyard-untrained	66.45	67.07	78.71	66.90	76.90	65.61	<u>78.98</u>	<b>99.28</b>
Vineyard-vertical-trellis	98.76	97.63	97.90	96.38	98.86	95.99	<b>99.02</b>	98.88
AA	95.71	94.63	<u>97.42</u>	93.68	96.87	92.80	97.15	<b>98.45</b>
OA	91.29	89.47	<u>95.17</u>	89.20	94.34	87.86	94.93	<u>99.21</u>
Kappa	0.9030	0.8828	<u>0.9462</u>	0.8797	0.9369	0.8649	0.9435	<b>0.9912</b>
Time	33.4255	17.9035	34.4387	18.8947	35.4682	18.4208	<u>5.4616</u>	<b>1.4140</b>

## 5. CONCLUSION

A supervised classification approach from remote-sensing fused images has been developed. This approach fuses the spatial information of an RGB image, extracted using superpixels, with the spectral information of an HS image. This fusion procedure is effectively formulated as an optimization problem, which was solved using ADMM. The classification, using a supervised method, of the fused image boosts the classification performance, and therefore, it reduces the number of pixels that need to be classified, since the image is grouped in superpixels. Therefore, the overall classification time decreases in comparison with other tested approaches. In all performed experiments, it can be noticed that adding spatial contextual information via superpixel segmentation significantly improves the classification results. Although the classification from the fused images, obtained with GSA and MAP-SMM methods, is good, the classification time is significantly greater than using the proposed method. In general, the framework developed in this paper provides a faster method to directly fuse and classify HS images instead of first reconstructing the high-spatial- and high-spectral-resolution image and then performing classification.

**Funding.** Universidad Industrial de Santander Construcción de una celda de alta presión acoplada a un sistema de medición espectral que permite determinar la solubilidad de moléculas de origen natural en CO<sub>2</sub> supercrítico a partir de medidas de absorbanzia en el rango espectral UV-Vis (VIE project 1872).

**Acknowledgment.** The work of K. Sanchez was supported by the Colciencias Scholarship (Convocatoria 771).

## REFERENCES

- J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.* **1**(2), 6–36 (2013).
- G. Martín and J. M. Bioucas-Dias, "Hyperspectral blind reconstruction from random spectral projections," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **9**, 2390–2399 (2016).
- P. S. Thenkabail and J. G. Lyon, *Hyperspectral Remote Sensing of Vegetation* (CRC Press, 2016).
- N. Yokoya and A. Iwasaki, "Hyperspectral and multispectral data fusion mission on hyperspectral imager suite (HISUI)," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (IEEE, 2013), pp. 4086–4089.
- N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.* **50**, 528–537 (2012).
- N. Akhtar, F. Shafait, and A. S. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *European Conference on Computer Vision (ECCV)* (2014).
- Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.* **53**, 3658–3668 (2015).
- Q. Wei, N. Dobigeon, and J. Tourneret, "Fast fusion of multi-band images based on solving a Sylvester equation," *IEEE Trans. Image Process.* **24**, 4109–4121 (2015).
- M. Simões, J. M. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *CoRR* abs/1411.4005 (2014).
- Q. Wei, J. Bioucas-Dias, N. Dobigeon, J. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.* **54**, 7236–7249 (2016).
- N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: a comparative review of the recent literature," *IEEE Trans. Geosci. Remote Sens. Mag.* **5**(2), 29–56 (2017).
- M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE* **101**, 652–675 (2013).
- P. Ghamisi, N. Yokoya, J. Li, W. Liao, S. Liu, J. Plaza, B. Rasti, and A. Plaza, "Advances in hyperspectral image and signal processing: a comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.* **5**(4), 37–78 (2017).
- P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: a review," *IEEE Geosci. Remote Sens. Mag.* **5**(1), 8–32 (2017).
- J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.* **50**, 809–823 (2012).
- J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* **6**, 2462–2471 (2013).
- G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.* **31**(1), 45–54 (2014).
- R. Hang, Q. Liu, H. Song, and Y. Sun, "Matrix-based discriminant subspace ensemble for hyperspectral image spatial-spectral feature fusion," *IEEE Trans. Geosci. Remote Sens.* **54**, 783–794 (2016).
- Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1356–1363.
- D. Stutz, A. Hermans, and B. Leibe, "Superpixels: an evaluation of the state-of-the-art," *Comput. Vis. Image Understanding* **166**, 1–27 (2018).
- X. Tian, L. Jiao, L. Yi, K. Guo, and X. Zhang, "The image segmentation based on optimized spatial feature of superpixel," *J. Visual Commun. Image Represent.* **26**, 146–160 (2015).
- X. Zhang, S. E. Chew, Z. Xu, and N. D. Cahill, "SLIC superpixels for efficient graph-based dimensionality reduction of hyperspectral imagery," *Proc. SPIE* **9472**, 947209 (2015).
- 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.* **34**, 2274–2282 (2012).
- Z. He, L. Liu, R. Deng, and Y. Shen, "Low-rank group inspired dictionary learning for hyperspectral image classification," *Signal Process.* **120**, 209–221 (2016).
- N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Optim.* **1**, 127–239 (2014).
- F. Lin, M. R. Jovanovic, and T. T. Georgiou, "An ADMM algorithm for matrix completion of partially known state covariances," in *IEEE 52nd Annual Conference on Decision and Control (CDC)* (IEEE, 2013), pp. 1684–1689.
- G. Mountakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: a review," *ISPRS J. Photogr. Remote Sens.* **66**, 247–259 (2011).
- J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.* **51**, 4816–4829 (2013).
- B. Schlkopf, A. J. Smola, and F. Bach, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT, 2018).
- B. Gu and V. S. Sheng, "A robust regularization path algorithm for v-support vector classification," *IEEE Trans. Neural Netw. Learn. Syst.* **28**, 1241–1248 (2017).
- L. Loncan, L. B. Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J.-Y. Tourneret, M. A. Viganjones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: a review," *arXiv:1504.04531* (2015).
- Y. Zhang, "Spatial resolution enhancement of hyperspectral image based on the combination of spectral mixing model and observation model," *Proc. SPIE* **9244**, 924405 (2014).