

# Feature Fusion Via Dual-Resolution Compressive Measurement Matrix Analysis For Spectral Image Classification

Juan Marcos Ramirez<sup>a</sup>, José Ignacio Martínez Torre<sup>a</sup>, Henry Arguello<sup>b</sup>

<sup>a</sup>*Computer Science Department, Universidad Rey Juan Carlos, Móstoles, Spain*

<sup>b</sup>*Computer Science Department, Universidad Industrial de Santander, Bucaramanga, Colombia*

---

## Abstract

In the compressive spectral imaging (CSI) framework, different architectures have been proposed to recover high-resolution spectral images from compressive measurements. Since CSI architectures compactly capture the relevant information of the spectral image, various methods that extract classification features from compressive samples have been recently proposed. However, these techniques require a feature extraction procedure that reorders measurements using the information embedded in the coded aperture patterns. In this paper, a method that fuses features directly from dual-resolution compressive measurements is proposed for spectral image classification. More precisely, the fusion method is formulated as an inverse problem that estimates high-spatial-resolution and low-dimensional feature bands from compressive measurements. To this end, the decimation matrices that describe the compressive measurements as degraded versions of the fused features are mathematically modeled using the information embedded in the coded aperture patterns. Furthermore, we include both a sparsity-promoting and a total-variation (TV) regularization terms to the fusion problem in order to consider the correlations between neighbor pixels, and therefore, improve the accuracy of pixel-based classifiers. To solve the fusion problem, we describe an algorithm based on the accelerated variant of the alternating direction method of multipliers (accelerated-ADMM). Additionally, a classification approach that includes the developed fusion method and a multilayer neural network is introduced. Finally, the proposed approach is evaluated on three remote sensing spectral images and a set of compressive measurements captured in the laboratory. Extensive simulations show that the proposed classification approach outperforms other approaches under various performance metrics.

**Keywords:** Compressive spectral imaging, Dual-resolution acquisition systems, Feature fusion, Spectral image classification

---



---

\*Corresponding author

Email address: [juanmarcos.ramirez@urjc.es](mailto:juanmarcos.ramirez@urjc.es) (Juan Marcos Ramirez)

## 1. Introduction

Hyperspectral (HS) images are three-dimensional (3-D) datasets that capture the spectral information of bidimensional (2-D) scenes across a large number of spectral bands. These images have been used in different applications such as precision agriculture, urban planning, and disaster management [1, 2]. In particular, the rich spectral information provided by HS images has been extensively considered to identify materials. However, HS sensors acquire low-spatial-resolution images to ensure measurements with a high signal-to-noise ratio (SNR) [3]. On the other hand, multispectral (MS) images are high-spatial-resolution datasets with poor spectral information. Therefore, to exploit the information embedded in multi-resolution data, image fusion has emerged as a class of techniques that combines the information in HS and MS images for building high-resolution datasets [4].

Notice that the large sizes of both HS and MS images challenge the storing and processing capabilities of the acquisition systems. To overcome this drawback, compressive spectral imaging (CSI) has emerged as an alternative acquisition framework that captures the relevant information of spectral images by sampling a reduced number of measurements [5]. In this regard, the coded aperture snapshot spectral imaging (CASSI) is the most representative architecture whose functioning consists of capturing projections of an encoded and spectrally-dispersed version of the input image onto an imaging detector [6]. Multiple CASSI based architectures have been proposed such as the three-dimensional CASSI (3D-CASSI) [5], the colored CASSI [7], and the snapshot colored compressive spectral imager (SCCSI) [8]. Moreover, various architectures that combine different kinds of measurements have been recently proposed to recover high-resolution spectral images [9, 10, 11].

Feature fusion from spectral images is a research field that focuses on exploiting the spectral and spatial information embedded in spectral images to improve identification and detection tasks. More precisely, this research area attempts to complement the spectral information with spatial features such as local smoothness, shape, and texture with the aim of improving the scene analysis [12]. In this context, feature fusion from CSI compressive measurements has become a challenging task due to the random nature of the coded aperture patterns and the nonlinearity of the encoding operation. In this sense, various feature fusion methods from CSI compressive measurements have been proposed. Specifically, a method that applies a compressive sensing reconstruction algorithm to recover the image PCA bands from CASSI measurements was developed in [13]. Recently, a feature extraction algorithm for spectral image classification from single-pixel measurements has been presented in [14] using a low-rank matrix approximation model. Furthermore, various feature fusion methods from dual-resolution compressive measurements have been recently proposed [15, 16, 17]. These methods exploit the fact that CSI systems can obtain the relevant information of the high-resolution image in a low dimensional space. Additionally, these approaches require a feature extraction stage that reorders measurements using the information embedded in the coded aperture patterns. More

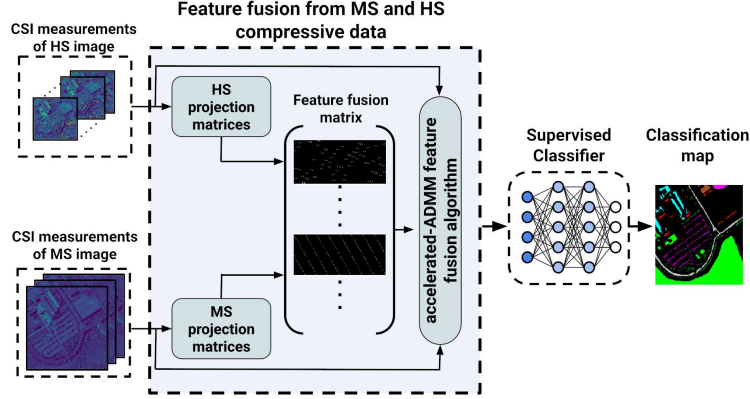


Figure 1: Flowchart of the spectral image classification approach based on the proposed feature fusion method.

precisely, the method reported in [15] stacks an interpolated version of the HS features and a superpixel map yielded by the MS features. Furthermore, the methods proposed in [16] and [17] fuse the extracted features by solving regularized optimization problems. Notice that the main difference with respect to the previous works relies on that the proposed approach obtains the fused features directly from compressive measurements without resorting to a feature extraction stage.

This paper proposes a method that fuses features directly from HS and MS compressive measurements. In particular, the compressive measurements are obtained from a dual-arm optical architecture consisting of two 3D-CASSI branches. The contributions of this work are described as follows.

1. First, we focus on deriving the expressions of the projection matrices that describe the compressive measurements as degraded versions of the fused features. Notice that these expressions incorporate the information embedded in the coded aperture patterns, therefore, a feature extraction procedure that reorders the measurements is not required.
2. From the observational model, the feature fusion is formulated as a regularized least-squares problem. Specifically, a sparsity-inducing term and a total-variation (TV) term are included in the cost function as regularizers. The sparsity-inducing term considers the spatial structures in spectral images while the TV term minimizes noise effects on the classification performance. Furthermore, an algorithm under an accelerated approach of the alternating direction method of multipliers [18] is described to numerically solve the feature fusion problem.
3. Finally, a deep multilayer neural network is used as a supervised pixel-based classifier for labeling high-resolution spectral images.

Figure 1 illustrates the flowchart of the proposed feature fusion method in a spectral image classification framework. The proposed approach is evaluated on three datasets and a set of compressive measurements captured in the laboratory. Extensive simulations show that the proposed approach outperforms other

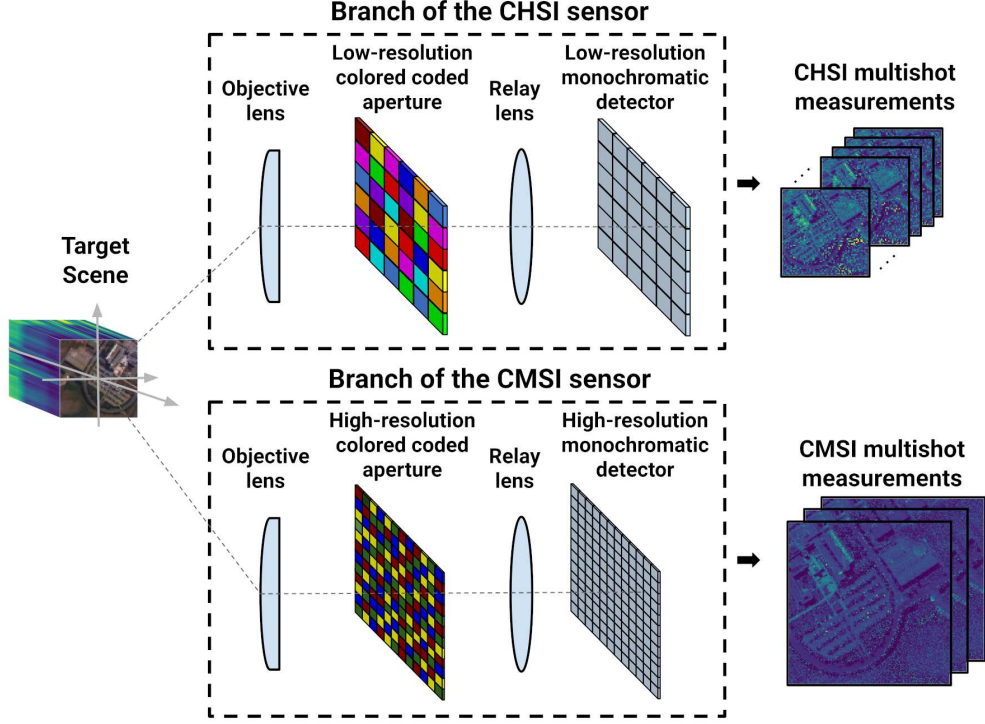


Figure 2: Schematic of the dual-arm optical architecture.

methods that obtain features from dual-resolution compressive measurements. Furthermore, the developed method exhibits a remarkable performance when the projections are contaminated with additive noise.

The paper is organized as follows. Section 2 illustrates the acquisition system and the proposed feature fusion approach is developed in Section 3. To evaluate the performance of the proposed feature fusion approach, the results of extensive simulations are shown in Section 4. Finally, the concluding remarks are synthesized in Section 5.

## 2. Dual-arm optical system

In this work, a method that fuses features directly from dual-resolution compressive measurements is proposed for spectral image classification. More precisely, the proposed feature fusion method is developed for a dual-arm optical architecture, where each arm consists of a multi-frame 3D-CASSI sensor. Figure 2 illustrates a schematic of the dual-arm optical architecture that obtains the compressive measurements. As can be seen in this figure, one branch is a 3D-CASSI sensor equipped with imaging lenses, a low-resolution colored coded aperture, and a low-resolution imaging detector. This branch, referred to as compressive hyperspectral imaging (CHSI), projects rich spectral information of the scene onto low-spatial resolution detectors. The second branch, referred to as compressive multispectral imaging (CMSI), is a 3D-CASSI sensor that includes imaging lenses, a high-spatial-resolution colored coded aperture, and a high-resolution camera

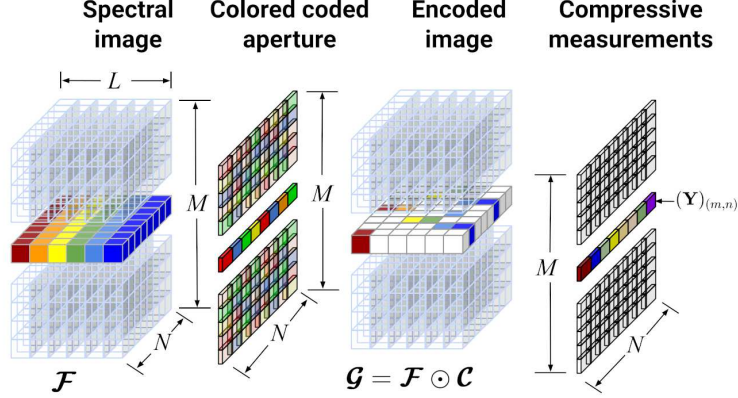


Figure 3: Model of the acquisition process in the 3D-CASSI architecture.

detector. In general, this system obtains high-spatial-resolution snapshots with poor spectral information. Notice that the dual-arm optical systems have been widely used to acquire the high-resolution information of spectral images and spectral videos [19, 20, 21]. More precisely, these systems exploit the high and the low resolution coding masks to acquire dual-resolution compressive measurements, and thus, to capture the relevant information of the image of interest.

### 2.1. Multi-frame 3D-CASSI architecture

Basically, the 3D-CASSI architecture projects an encoded version of the input image onto an imaging detector [5]. Figure 3 illustrates the model of the acquisition process. To describe 3D-CASSI measurements, consider  $\mathcal{F}$  as a discrete version of the input image with a spatial resolution of  $N \times M$  and  $L$  spectral bands. Furthermore, let  $\mathcal{F}_{(m,n,\ell)}$  be an element of the spectral image  $\mathcal{F}$ , where  $(m, n)$  denotes the spatial coordinate and  $\ell$  represents the  $\ell$ -th spectral band. In this context, the input spectral image is encoded by a coded aperture that can be described as an array of optical filters with dimensions  $M \times N$ . Notice that each optical filter modulates the spectral signature at the spatial coordinate  $(m, n)$ . In general, the coded aperture is modeled as a discrete cube  $\mathcal{C}$  with dimensions  $M \times N \times L$  and entries  $\mathcal{C}_{(m,n,\ell)} \in \{0, 1\}$ . In consequence, the encoded image  $\mathcal{G}$  is obtained by applying an element-wise product between the input image and the model of the coded aperture, i.e.  $\mathcal{G} = \mathcal{F} \odot \mathcal{C}$ , where  $\odot$  denotes the element-wise tensor multiplication. Subsequently, the encoded image is integrated along the sensor spectral sensitivity before it is projected onto a detector with dimensions  $M \times N$  [5]. Assuming a perfect registration between each encoded spectral signature and the corresponding detector pixel, the intensity captured at the spatial coordinate  $(m, n)$  can be expressed as

$$(\mathbf{Y})_{(m,n)} = \sum_{\ell=1}^L \mathcal{F}_{(m,n,\ell)} \mathcal{C}_{(m,n,\ell)}, \quad (1)$$

for  $m = 1, \dots, M$  and  $n = 1, \dots, N$ .

Multiple snapshots are required to properly reconstruct the image from compressed measurements, where each snapshot is captured using a different coded aperture pattern. Various algorithms have been developed for designing coded aperture patterns under the restricted isometry property (RIP) constraint [7, 22, 23]. In this work, we assume that a reduced set of optical filters are available with non-overlapping responses covering the entire wavelength spectrum. Furthermore, consider  $\Theta$  a binary matrix with dimensions  $L \times P$  whose columns  $\{\theta_i \in \{0, 1\}^L\}_{i=1}^P$  contain the responses of the available optical filters, where  $P$  is the number of available optical filters. Therefore, the intensity captured by the imaging detector at the spatial location  $(m, n)$  and acquired at the  $k$ -th snapshot  $(\mathbf{Y}^{(k)})_{(m,n)}$  can be described as

$$(\mathbf{Y}^{(k)})_{(m,n)} = \sum_{\ell=1}^L \mathcal{F}_{(m,n,\ell)} (\theta_{\mathcal{S}_{(m,n,k)}})_{\ell} \quad (2)$$

for  $m = 1, \dots, M$ ,  $n = 1, \dots, N$ ,  $k = 1, \dots, K$ , where  $K$  denotes the number of captured snapshots,  $\theta_{\mathcal{S}_{(m,n,k)}}$  represents the filter response used to encode the spectral signature at the coordinate  $(m, n)$  and the  $k$ -th snapshot with  $\mathcal{S}_{(m,n,k)} \in \{1, 2, \dots, P\}$ . To improve the performance of the spectral image recovery methods from multiple snapshots, the coded apertures should be designed such that projections capture the entire information of the spectral image [7]. In this sense, Algorithm 1 has been introduced in [16] to optimally design coded apertures. Specifically, to capture the entire information of the image, each spectral signature should be modulated by all available filters. To this end, for every spatial coordinate  $(m, n)$ , the design algorithm randomly distributes the optical filters across the snapshots. This distribution is included in  $\alpha$  that contains a random permutation of the set  $\{1, \dots, P\}$ , where each component indexes a particular optical filter. In consequence, this algorithm generates a data cube  $\mathcal{S}$  with dimensions  $M \times N \times K$  that contains the filter patterns to capture measurements. The optical filter response used at the spatial location  $(m, n)$  and the snapshot  $k$  is denoted as  $\theta_{\mathcal{S}_{(m,n,k)}}$ . In this work, we consider the information provided by the coded aperture patterns to build the decimation matrices of the proposed feature fusion algorithm.

### 3. Feature fusion method

#### 3.1. High-resolution features

The fused features should exploit both the high-spatial-resolution acquired by the CMSI sensor and the rich spectral information captured by the CHSI system. In this work, each fused feature is modeled as the projections of the spectral signature of the high-resolution image modulated by the set of optical filters used by the CHSI sensor, in other words,

$$(\mathcal{X})_{(m,n,k)} = \sum_{\ell=1}^L \mathcal{F}_{(m,n,\ell)} (\theta_{\mathcal{S}_{(m,n,k)}}^{(\text{hs})})_{\ell} \quad (3)$$

for  $k = 1, \dots, K$ , where  $K$  is the number of captured snapshots, and  $\{\theta_i^{(\text{hs})}\}_{i=1}^{P^{(\text{hs})}}$  are column vectors of the matrix  $\Theta^{(\text{hs})}$  that contains the optical filter responses used to obtain the HS compressive measurements.

---

**Algorithm 1** Design of the colored coded apertures patterns [16]

---

**Input:**  $\Theta \in \{0, 1\}$

**Output:**  $\mathcal{S}_{(m,n,k)}, \theta_{\mathcal{S}_{(m,n,k)}}$  for  $m = 1, \dots, M$ ;  $n = 1, \dots, N$ ; and  $k = 1, \dots, K$

```

1: for  $m = 1$  to  $M$  do
2:   for  $n = 1$  to  $N$  do
3:      $\alpha = \text{randperm}(K)$ 
4:     for  $k = 1$  to  $K$  do
5:        $\mathcal{S}_{(m,n,k)} = \alpha(k)$ 
6:        $\theta_{\mathcal{S}_{(m,n,k)}} = \theta_{\alpha(k)}$ 
7:     end for
8:   end for
9: end for

```

---

On the other hand, the CSI measurements do not offer discriminative properties required by pixel-based classification methods. This is because each CSI projection is acquired using a distinct coded aperture pattern. In previous works, a feature extraction procedure has been proposed before performing the feature fusion method [15, 16, 17]. This feature extraction procedure reorders measurements such that each band corresponds to the spectral image response to a single optical filter. Conversely, in this work, we consider the information embedded in the coded aperture patterns to build the decimation matrices that describe the feature fusion model. Therefore, both feature extraction and feature fusion are included in a unique optimization problem.

### 3.2. Multispectral downsampling model

A CMSI measurement at the spatial coordinate  $(m, n)$  and the snapshot  $w$  can be expressed as

$$(\mathcal{X}_{(\text{ms})})_{(m,n,w)} = \sum_{\ell=1}^L \mathcal{F}_{(m,n,\ell)} \left( \theta_{\mathcal{S}_{(m,n,w)}}^{(\text{ms})} \right)_{\ell}, \quad (4)$$

for  $w = 1, \dots, W$ , where  $W$  represents the number of snapshots captured by the CMSI system, and  $\{\theta_i^{(\text{ms})}\}_{i=1}^{P^{(\text{ms})}}$  are columns of the matrix  $\Theta^{(\text{ms})}$  containing the responses of the optical filters used by the CMSI sensor. Without loss of generality, consider that the response of an optical filter used by a CMSI sensor, at the spatial coordinate  $(m, n)$  and the  $w$ -th snapshot, can be obtained as the summation of  $q$  responses of optical filters used by a CHSI sensor, i.e.

$$\theta_{\mathcal{S}_{(m,n,w)}}^{(\text{ms})} = \sum_{a=1}^q \theta_{\mathcal{S}_{(m,n,(w-1)q+a)}}^{(\text{hs})}. \quad (5)$$

Substituting (5) in (4), we obtain

$$(\mathcal{X}_{(\text{ms})})_{(m,n,w)} = \sum_{a=1}^q \sum_{\ell=1}^L \mathcal{F}_{(m,n,\ell)} \left( \theta_{\mathcal{S}_{(m,n,(w-1)q+a)}}^{(\text{hs})} \right)_{\ell}. \quad (6)$$



Figure 4: (a) An example of the projection matrix  $\mathbf{H}_{(\text{ms})}$  for  $M = 6$ ,  $N = 6$ ,  $P = 8$ , and  $q = 4$ ; (b) an example of the projection matrix  $\mathbf{H}_{(\text{hs})}$  for  $M = 6$ ,  $N = 6$ ,  $P = 8$ , and  $p = 2$ .

As can be observed in (6), a CMSI measurement can be expressed as the linear expansion of  $q$  components of the fused features, in other words,

$$(\mathcal{X}_{(\text{ms})})_{(m,n,w)} = \sum_{a=1}^q (\mathcal{X})_{(m,n,(w-1)q+a)}, \quad (7)$$

for  $m = 1, \dots, M$ ;  $n = 1, \dots, N$ ; and  $w = 1, \dots, W$ . Indeed, the model for the entire set of measurements can be synthesized in matrix form as follows

$$\mathbf{x}_{(\text{ms})} = \mathbf{H}_{(\text{ms})}\mathbf{x} + \boldsymbol{\eta}_{(\text{ms})} \quad (8)$$

where  $\mathbf{x}_{(\text{ms})} \in \mathbb{R}^{MNW}$  are the CMSI measurements in vector form,  $\mathbf{H}_{(\text{ms})} \in \mathbb{R}^{MNW \times MNK}$  is the projection matrix that includes both downsampling operation and the information of the coded aperture patterns,  $\mathbf{x} \in \mathbb{R}^{MNK}$  is the fused feature vector, and  $\boldsymbol{\eta}_{(\text{ms})} \in \mathbb{R}^{MNW}$  is the noise vector that contaminates the CMSI measurements. In general, the noise perturbations are characterized as random samples that follow a common statistical model. From the observational model (8), each element of the projection matrix is obtained as

$$(\mathbf{H}_{(\text{ms})})_{(u,v)} = \begin{cases} 1, & \text{if } u = m + (n-1)M + (w-1)MN \text{ and } v = m + (n-1)M + \left( \left( \mathcal{S}_{(m,n,w)}^{(\text{ms})} - 1 \right) * q + z \right) MN \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

for  $z = 1, \dots, q$ , where  $\mathcal{S}^{(\text{ms})}$  contains the coded aperture patterns generated to obtain the CMSI measurements. The projection matrix  $\mathbf{H}_{(\text{ms})}$  includes the information of the coded aperture patterns, therefore, the proposed method does not require a feature extraction procedure. Upon a closer examination of the sensing matrix dimensions, it can be observed that the measurement rate with respect to the fused features is reduced to  $\varepsilon_{(\text{ms})} = \frac{MN(K/q)}{MNK} = \frac{1}{q}$ . Figure 4a illustrates an example of the projection matrix structure for  $M = 6$ ,  $N = 6$ ,  $P = 8$ , and  $q = 4$ .

### 3.3. Hyperspectral downsampling model

On the other hand, a low-spatial-resolution measurement acquired by a CHSI system at the spatial coordinate  $(\mu, \nu)$  and snapshot  $k$  can be described as

$$(\mathcal{X}_{(\text{hs})})_{(\mu,\nu,k)} = \sum_{\ell=1}^L \mathcal{F}_{(\mu,\nu,\ell)} \left( \boldsymbol{\theta}_{\mathcal{S}_{(\mu,\nu,k)}}^{(\text{hs})} \right)_{\ell}, \quad (10)$$



with  $\mu = 1, \dots, M/p$  and  $\nu = 1, \dots, N/p$ , where  $p$  is the spatial decimation factor. In this case, each CHSI measurement can be obtained by averaging  $p^2$  entries of the high-resolution features, this is,

$$(\mathbf{x}_{(\text{hs})})_{(\mu, \nu, k)} = \frac{1}{p^2} \sum_{a=0}^{p-1} \sum_{b=0}^{p-1} \sum_{\ell=1}^L \mathcal{F}_{(m+a, n+b, \ell)} \left( \boldsymbol{\theta}_{\mathcal{S}_{(\mu, \nu, k)}}^{(\text{hs})} \right)_{\ell}, \quad (11)$$

therefore, a CHSI sample can be expressed as

$$\left( \mathbf{x}^{(\text{hs})} \right)_{(\mu, \nu, k)} = \frac{1}{p^2} \sum_{a=0}^{p-1} \sum_{b=0}^{p-1} (\mathbf{x})_{(m+a, n+b, k)}. \quad (12)$$

The model for the entire set of CHSI measurements in matrix form can be compactly described as

$$\mathbf{x}_{(\text{hs})} = \mathbf{H}_{(\text{hs})} \mathbf{x} + \boldsymbol{\eta}_{(\text{hs})} \quad (13)$$

where  $\mathbf{x}_{(\text{hs})} \in \mathbb{R}^{(M/p)(N/p)K}$  is the vector that contains the CHSI measurements,  $\mathbf{H}_{(\text{hs})} \in \mathbb{R}^{(M/p)(N/p)K \times MNK}$  is the projection matrix that includes the spatial downsampling,  $\mathbf{x} \in \mathbb{R}^{MNK}$  are the fused features in vector form, and  $\boldsymbol{\eta}_{(\text{hs})} \in \mathbb{R}^{(M/p)(N/p)K}$  is the additive noise vector corrupting the CHSI measurements. Notice that the projection matrix  $\mathbf{H}_{(\text{hs})}$  also includes the information of the coded aperture patterns generated to capture the CHSI samples, therefore, a feature extraction stage is not required. Particularly, each component  $(u, v)$  of the projection matrix can be obtained as

$$(\mathbf{H}_{(\text{hs})})_{(u, v)} = \begin{cases} \frac{1}{p^2}, & \text{if } u = \mu + (\nu - 1)(M/p) + (k - 1)(M/p)(N/p) \text{ and } v = \mu + (\nu - 1)(M/p) + tM + z + d \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

for  $z = 1, \dots, p$  and  $z = 1, \dots, p$ ; where  $d = \lfloor \frac{\mu p}{M} \rfloor + \text{mod}(\mu, (M/p))p + \left( \mathcal{S}_{(\mu, \nu, t)}^{(\text{hs})} - 1 \right) MN$ ,  $\mathcal{S}^{(\text{hs})}$  contains the coded aperture patterns used for acquiring the CHSI measurements. Furthermore,  $\lfloor \cdot \rfloor$  is the floor operator that extracts the integer part of its argument,  $\text{mod}(\mu, (M/p))$  and outputs the reminder after division between  $\mu$  and  $M/p$ . In this case, the measurement rate with respect to the fused features is reduced to  $\varepsilon_{(\text{hs})} = \frac{(M/p)(N/p)K}{MNK} = \frac{1}{p^2}$ . Figure 4b displays an example of the projection matrix structure for  $M = 6$ ,  $N = 6$ ,  $P = 8$ , and  $p = 2$ .

### 3.4. Feature fusion algorithm

Let assume the entries of  $\boldsymbol{\eta}_{(\text{ms})}$  and  $\boldsymbol{\eta}_{(\text{hs})}$  are described as random samples that follow a Gaussian distribution with variance  $\sigma_{(\text{ms})}^2$  and  $\sigma_{(\text{hs})}^2$ , respectively, for  $a = 1, \dots, MNW$  and  $b = 1, \dots, (M/p)(N/p)K$ . Furthermore, without loss of generality, we assume that  $\sigma_{(\text{ms})}^2 = \sigma_{(\text{hs})}^2$ . This leads to the minimization of the sum of squared errors as the optimization approach to describe the fusion problem. Two regularization terms are aggregated to this formulation to improve the quality of the fused features. The feature fusion problem is formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda_1 \|\boldsymbol{\Psi}^\top \mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_{TV} \right\}, \quad (15)$$

where  $\lambda_1$  and  $\lambda_2$  are the regularization parameter that control the trade-off between the sum of squared errors and the regularization terms, with

$$\mathbf{y} = [\mathbf{x}_{ms}^\top, \mathbf{x}_{hs}^\top]^\top, \quad (16)$$

$$\mathbf{H} = [\mathbf{H}_{ms}^\top, \mathbf{H}_{hs}^\top]^\top. \quad (17)$$

As shown in Section 3.1, each band of the high-resolution features can be considered as the response of the input spectral density to a particular optical filter. To preserve the spatial structure of the scene, the sparsity-inducing term  $\lambda_1 \|\Psi^\top \mathbf{x}\|_1$  is included. In other words, we assume that the merged features can be sparsely described in an orthogonal transform  $\Psi$ . This approach has been widely exploited in image denoising, improving thus, the quality of the recovered images. On the other hand, the total-variation (TV) term  $\lambda_2 \|\mathbf{x}\|_{TV}$  exploits the local spatial information present in spectral image features. Specifically, the used TV norm is described as follow

$$\|\mathbf{x}\|_{TV} = \sum_{(i,j,k)} \{|x_{(i,j,k)} - x_{(i+1,j,k)}| + |x_{(i,j,k)} - x_{(i,j+1,k)}| + |x_{(i,j,k)} - x_{(i,j,k+1)}|\}, \quad (18)$$

where  $x_{(i,j,k)}$  is the spectral image voxel at the location  $(i, j, k)$ . The total variation norm is typically rewritten as

$$\|\mathbf{x}\|_{TV} = \|\Phi \mathbf{x}\|_1 \quad (19)$$

where  $\Phi$  is the first-order difference operator. This term induces piece-wise spatial smoothness on the recovered features preserving, in turn, the edges of the scene [24]. It is worth noting that TV norm has been considered in various feature extraction methods to improve the classification accuracy [25, 26].

The feature fusion problem can be also described as

$$\begin{aligned} \hat{\mathbf{x}} = \arg \min_{\mathbf{x}} & \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda_1 \|\boldsymbol{\gamma}_1\|_1 + \lambda_2 \|\boldsymbol{\gamma}_2\|_1 \right\} \\ \text{s.t. } & \Psi^\top \mathbf{x} - \boldsymbol{\gamma}_1 = 0, \quad \Phi \mathbf{x} - \boldsymbol{\gamma}_2 = 0, \end{aligned} \quad (20)$$

with  $\boldsymbol{\gamma}_1$  and  $\boldsymbol{\gamma}_2$  as auxiliary variables in a alternating direction optimization approach. Thus, the augmented Lagrangian for (20) can be expressed as

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda_1 \|\boldsymbol{\gamma}_1\|_1 + \lambda_2 \|\boldsymbol{\gamma}_2\|_1 + \frac{\rho}{2} \|\Psi^\top \mathbf{x} - \boldsymbol{\gamma}_1 - \boldsymbol{\delta}_1\|_2^2 + \frac{\rho}{2} \|\Phi \mathbf{x} - \boldsymbol{\gamma}_2 - \boldsymbol{\delta}_2\|_2^2, \quad (21)$$

where  $\rho > 0$  is a penalty parameter, and  $(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2)$  are Lagrange multiplier vectors. To estimate the fused features, the minimization of the augmented Lagrangian is performed. Algorithm 2 displays the pseudocode to fuse features using the accelerated approach of ADMM [27]. Notice that we aim at minimizing (21) with respect to  $\mathbf{x}$ ,  $\boldsymbol{\gamma}_1$ , and  $\boldsymbol{\gamma}_2$ . As can be seen in steps 4, 7, and 8 of Algorithm 2, the update of each variable is computed as a weighted sum of previous estimates  $(\mathbf{x}^{(j)}, \boldsymbol{\gamma}_1^{(j)}, \boldsymbol{\gamma}_2^{(j)})$  and current approximations  $(\mathbf{x}^{(j+1)}, \boldsymbol{\gamma}_1^{(j+1)}, \boldsymbol{\gamma}_2^{(j+1)})$ . To compute the current estimate  $\mathbf{x}^{(j+1)}$ , we focus on minimizing the augmented Lagrangian

with respect to the target variable. Under this approach, the estimation of  $\mathbf{x}$  at the iteration  $(j + 1)$  is obtained by solving the minimization problem

$$\mathbf{x}^{(j+1)} \leftarrow \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\rho}{2} \|\Psi^\top \mathbf{x} - \boldsymbol{\gamma}_1^{(j)} - \boldsymbol{\delta}_1^{(j)}\|_2^2 + \frac{\rho}{2} \|\Phi \mathbf{x} - \boldsymbol{\gamma}_2^{(j)} - \boldsymbol{\delta}_2^{(j)}\|_2^2 \right\}. \quad (22)$$

Notice that the analytical solution of (22) involve computationally expensive operations, therefore, a rough estimation is obtained by using the first-order approximation of the cost function around  $\mathbf{x}^{(j)}$ , in other words,

$$\begin{aligned} \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{\rho}{2} \|\Psi^\top \mathbf{x} - \boldsymbol{\gamma}_1^{(j)} - \boldsymbol{\delta}_1^{(j)}\|_2^2 + \frac{\rho}{2} \|\Phi \mathbf{x} - \boldsymbol{\gamma}_2^{(j)} - \boldsymbol{\delta}_2^{(j)}\|_2^2 &\approx \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}^{(j)}\|_2^2 + \frac{\rho}{2} \|\Psi^\top \mathbf{x}^{(j)} - \boldsymbol{\gamma}_1^{(j)} - \boldsymbol{\delta}_1^{(j)}\|_2^2 \\ &\quad + \frac{\rho}{2} \|\Phi \mathbf{x}^{(j)} - \boldsymbol{\gamma}_2^{(j)} - \boldsymbol{\delta}_2^{(j)}\|_2^2 + \left\langle \mathbf{x} - \mathbf{x}^{(j)}, \vartheta(\mathbf{x}^{(j)}) \right\rangle \\ &\quad + \frac{\beta}{2} \|\mathbf{x} - \mathbf{x}^{(j)}\|_2^2 \\ &\approx \frac{1}{2} \|\mathbf{y} - \mathbf{H}\mathbf{x}^{(j)}\|_2^2 + \frac{\rho}{2} \|\Psi^\top \mathbf{x}^{(j)} - \boldsymbol{\gamma}_1^{(j)} - \boldsymbol{\delta}_1^{(j)}\|_2^2 \\ &\quad + \frac{\rho}{2} \|\Phi \mathbf{x}^{(j)} - \boldsymbol{\gamma}_2^{(j)} - \boldsymbol{\delta}_2^{(j)}\|_2^2 \\ &\quad + \frac{\beta}{2} \|\mathbf{x} - \mathbf{x}^{(j)} + \frac{1}{\beta} \vartheta(\mathbf{x}^{(j)})\|_2^2, \end{aligned} \quad (23)$$

where  $\vartheta(\mathbf{x}^{(j)})$  is the gradient of the cost function around  $\mathbf{x}^{(j)}$  that is computed as follows

$$\vartheta(\mathbf{x}^{(j)}) = \mathbf{H}^\top (\mathbf{y} - \mathbf{H}\mathbf{x}^{(j)}) + \rho (\mathbf{x}^{(j)} - \Psi^\top (\boldsymbol{\gamma}_1^{(j)} - \boldsymbol{\delta}_1^{(j)})) + \rho (\Phi^\top (\Phi \mathbf{x}^{(j)} - \boldsymbol{\gamma}_2^{(j)} - \boldsymbol{\delta}_2^{(j)})), \quad (24)$$

with  $\beta$  as a step parameter. By substituting (24) in (23), and setting the derivative to zero, the current approximation  $\mathbf{x}^{(j+1)}$  is obtained by

$$\mathbf{x}^{(j+1)} = \mathbf{x}^{(j)} - \left( \frac{1}{\beta} \right) \left( \mathbf{H}^\top (\mathbf{y} - \mathbf{H}\mathbf{x}^{(j)}) + \rho (\mathbf{x}^{(j)} - \Psi^\top (\boldsymbol{\gamma}_1^{(j)} - \boldsymbol{\delta}_1^{(j)})) + \rho (\Phi^\top (\Phi \mathbf{x}^{(j)} - \boldsymbol{\gamma}_2^{(j)} - \boldsymbol{\delta}_2^{(j)})) \right). \quad (25)$$

Additionally, to update the auxiliary variables  $\boldsymbol{\gamma}_1$  and  $\boldsymbol{\gamma}_2$  at the iteration  $(j + 1)$ , the optimization problems should be solved

$$\boldsymbol{\gamma}_1^{(j+1)} \leftarrow \arg \min_{\boldsymbol{\gamma}_1} \left\{ \frac{\rho}{2} \|\Psi^\top \mathbf{x}^{(j+1)} - \boldsymbol{\gamma}_1 - \boldsymbol{\delta}_1^{(j)}\|_2^2 + \lambda_1 \|\boldsymbol{\gamma}_1\|_1 \right\} \quad (26)$$

$$\boldsymbol{\gamma}_2^{(j+1)} \leftarrow \arg \min_{\boldsymbol{\gamma}_2} \left\{ \frac{\rho}{2} \|\Phi \mathbf{x}^{(j+1)} - \boldsymbol{\gamma}_2 - \boldsymbol{\delta}_2^{(j)}\|_2^2 + \lambda_2 \|\boldsymbol{\gamma}_2\|_1 \right\} \quad (27)$$

whose solutions are obtained basically applying soft-thresholding operators, as shown as follows

$$\boldsymbol{\gamma}_1^{(j+1)} = \text{soft} \left( \Psi^\top \mathbf{x}^{(j+1)} + \boldsymbol{\delta}_1^{(j)}, \lambda_1 / \rho \right) \quad (28)$$

$$\boldsymbol{\gamma}_2^{(j+1)} = \text{soft} \left( \Phi \mathbf{x}^{(j+1)} + \boldsymbol{\delta}_2^{(j)}, \lambda_2 / \rho \right) \quad (29)$$

where  $\text{soft}(\gamma, \lambda) = \text{sign}(\gamma) \max(\gamma - \lambda, 0)$ . Then, Lagrange multiplier vectors are updated by computing the operations described in steps 9 and 10 of Algorithm 2. Finally, the computational complexity for computing

$\mathbf{x}$  can be determined as  $\mathcal{O}((MN)^2K)$ , while the soft-thresholding operations exhibit a complexity  $\mathcal{O}(MNK)$ , with  $M \times N \times K$  as the size of the fused features. In consequence, the computational complexity per iteration is obtained as  $\mathcal{O}((MN)^2K)$ . The source code for the proposed feature fusion method can be downloaded from this link: [https://github.com/JuanMarcosRamirez/featurefusion\\_getfund](https://github.com/JuanMarcosRamirez/featurefusion_getfund)

---

**Algorithm 2** Accelerated ADMM for feature fusion from compressive measurements

---

**Input:**  $\mathbf{y}, \Psi, \Phi, \mathbf{H}, \lambda_1, \lambda_2, \rho, \alpha^{(0)}$

$\mathbf{x}^{(0)} \rightarrow \mathbf{0}, \gamma_1^{(0)} \rightarrow \mathbf{0}, \gamma_2^{(0)} \rightarrow \mathbf{0}, \delta_1^{(0)} \rightarrow \mathbf{0}, \delta_2^{(0)} \rightarrow \mathbf{0}$

**Output:**  $\mathbf{x}^{(j+1)}$

```

1: repeat
2:    $\mathbf{x}^{(j)} = (1 - \alpha^{(j+1)})\mathbf{x}_2^{(j)} + \alpha^{(j+1)}\mathbf{x}^{(j)}$ 
3:    $\mathbf{x}^{(j+1)} \leftarrow \arg \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \gamma_1^{(j)}, \gamma_2^{(j)})$  Eq.(25)
4:    $\mathbf{x}_2^{(j+1)} = (1 - \alpha^{(j+1)})\mathbf{x}_2^{(j)} + \alpha^{(j+1)}\mathbf{x}^{(j+1)}$ 
5:    $\gamma_1^{(j+1)} \leftarrow \arg \min_{\gamma_1} \mathcal{L}(\mathbf{x}^{(j)}, \gamma_1, \gamma_2^{(j)})$  Eq.(28)
6:    $\gamma_2^{(j+1)} \leftarrow \arg \min_{\gamma_2} \mathcal{L}(\mathbf{x}^{(j)}, \gamma_1^{(j)}, \gamma_2)$  Eq.(29)
7:    $\gamma_1^{(j+1)} = (1 - \alpha^{(j+1)})\gamma_1^{(j)} + \alpha^{(j+1)}\gamma_1^{(j+1)}$ 
8:    $\gamma_2^{(j+1)} = (1 - \alpha^{(j+1)})\gamma_2^{(j)} + \alpha^{(j+1)}\gamma_2^{(j+1)}$ 
9:    $\delta_1^{(j+1)} = \delta_1^{(j)} + \Psi^\top \mathbf{x}^{(j+1)} - \gamma_1^{(j+1)}$ 
10:   $\delta_2^{(j+1)} = \delta_2^{(j)} + \Phi \mathbf{x}^{(j+1)} - \gamma_2^{(j+1)}$ 
11: until a stopping rule is satisfied

```

---

### 3.5. Supervised classification method

The feature fusion method is included in a spectral image classification approach as shown in Fig. 1. To this end, a multilayer perceptron neural network (MLPNN) is selected as a supervised classification method. More precisely, an MLPNN is a feed-forward neural network whose structure can learn nonlinear classification boundaries [28]. This classifier typically exhibits a deep structure that can handle large data sets and it does not require a huge amount of training data for achieving the desired classification performance compared to that needed by the convolutional neural networks (CNN). Notice that various methods based on deep learning have been recently developed for spectral image classification exhibiting remarkable results [29, 30]. However, these methods require a high-resolution spectral image. Furthermore, these techniques involve complex network structures that consider separately the spectral signature and local spatial information. Indeed, to achieve the desired classification performance, both a huge amount of training data and a large number of learning epochs are needed.

A schematic of the supervised classification approach is illustrated in Fig. 5. As can be observed in this figure, the MLPNN consists of an input layer, a set of hidden layers, and an output layer. Specifically, the proposed classification approach firstly extracts the vector from the fused features at the coordinate  $(m, n)$

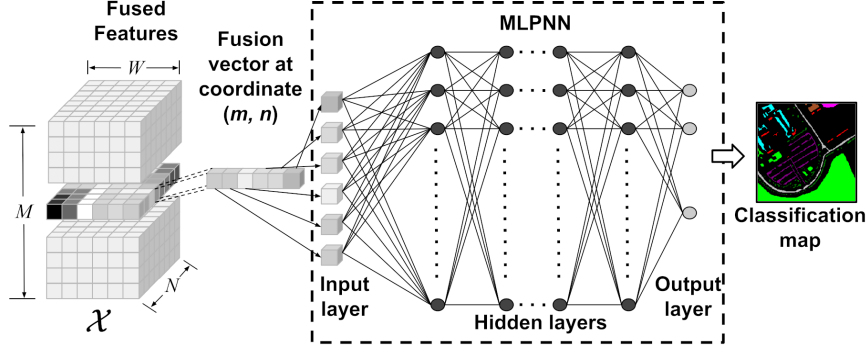


Figure 5: Schematic of the supervised classification approach.

and its elements reduce to the entries of the MLPNN input layer. Notice that the size of the input layer depends on the dimensions of the fused vectors, and consequently, it also depends on the compression ratio. Then, the MLPNN comprises a set of fully-connected hidden layers whose  $j$ -th neuron at the  $k$ -th layer outputs

$$\zeta_j^k = f \left( \sum_{i=1}^{N_n} \Upsilon_{i,j}^{k-1} \zeta_{i,j}^{k-1} + \xi_j^{k-1} \right) \quad (30)$$

where  $\Upsilon_{i,j}^{k-1}$  is the weight that connects the  $i$ -th neuron at the  $k-1$  layer with the  $j$ -th neuron in layer  $k$ ,  $\xi_j^{k-1}$  is the bias term of the  $j$ -th neuron in layer  $k$ , and  $f(\cdot)$  represents a nonlinear activation function. In this work, we select ten hidden layers with ten neurons each, where the number of hidden layers and the number of neurons by layer were determined by implementing a classification performance search. Every neuron is activated with the rectified linear unit (ReLU) function, i.e.  $f(u) = \max(u, 0)$ . The output layer size depends on the number of the classes and the network parameters are updated by using the backpropagation algorithm with the binary cross-entropy loss function [31]. To be more precise, the MLPNN structure attempts to learn the network parameter set  $\Theta = \{\Upsilon, \xi\}$  that basically contains the connection weight matrix  $\Upsilon$  and the bias vector  $\xi$ . Under the multiclass classification framework, let  $\mathbf{z}_n \in \{0, 1\}^{N_c}$  be the  $n$ -th ground truth label vector with  $N_c$  as the number of output classes, and consider  $\mathbf{s}_n$  the corresponding input vector, therefore, the training set can be represented as  $\Gamma = \{\mathbf{z}_n, \mathbf{s}_n\}_{n=1}^{N_t}$  with  $N_t$  as the number of training samples. Hence, the training stage attempts learn the network parameters  $\Theta$  that minimize the loss function given by

$$\mathcal{E}(\Theta) = - \sum_{n=1}^{N_t} \sum_{c=1}^{N_c} (\mathbf{z}_n)_c \log((\mathbf{p}_n(\Theta, \mathbf{s}_n))_c) \quad (31)$$

where  $(\mathbf{z}_n)_c$  is the  $c$ -th element of the ground truth label vector and  $(\mathbf{p}_n(\Theta, \mathbf{s}_n))_c$  is the probability predicted by the network at the  $c$ -th output layer node. Notice that the network parameters are randomly initialized.

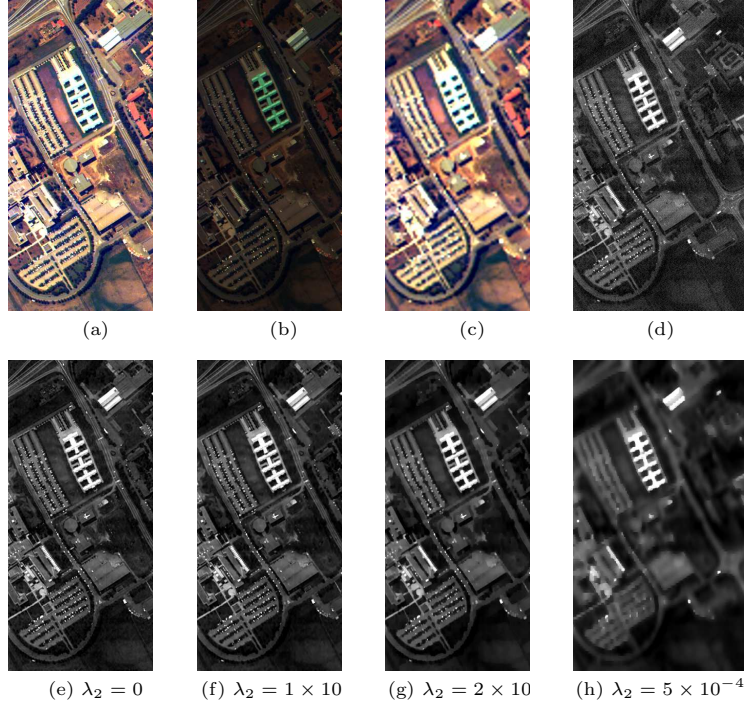


Figure 6: Pavia University spectral image. RGB composite of (a) the original image, (b) the multispectral image, and (c) the hyperspectral image; (d) reference feature band; (e)-(h) feature bands obtained by the proposed fusion method for different values the TV regularization parameter  $\lambda_2$  with of  $\lambda_1 = 0.001 \times \|\mathbf{H}^\top \mathbf{y}\|$ .

## 4. Results and analysis

### 4.1. Pavia University

This spectral image was acquired by the airborne sensor known as Reflective Optics Spectrographic Imaging System (ROSIS-03) over the University of Pavia, Italy [32]. This data set exhibits a high-spatial-resolution (1.3 m per pixel) with dimensions  $610 \times 340$  pixels and 96 spectral bands in the wavelength range from 0.43 to  $0.86 \mu\text{m}$ . Figure 6a displays the RGB composite of the Pavia University image.

Notice that the proposed fusion method is directly applied to compressive measurements of the MS image and the HS image. First, we built the MS and the HS versions of the input spectral image. To this end, the MS image was derived by integrating contiguous spectral bands of the original image with  $q = 4$ . Therefore, the MS image is a high-spatial-resolution data cube with dimensions  $610 \times 340 \times 24$ . The RGB composite of the MS image is shown in Fig 6b. The HS image was obtained by applying spatial downsampling to every spectral band of the original image with  $p = 4$ . The resulting image exhibits dimensions of  $152 \times 85 \times 96$ . Figure 6c shows the RGB composite of the HS image. For comparison purposes, Fig. display a band of the high-resolution features using the model described in (3).

Then, compressive measurements were obtained by simulating the 3D-CASSI dual-arm optical system.

Classes	# Samples		High-resolution spectral image				Proposed feature fusion method			
	Train	Test	SVM-RBF	RF	CNN	MLPNN	SVM-RBF	RF	CNN	MLPNN
Asphalt	661	5944	88.04	<b>91.14</b>	40.81	90.55	95.54	94.55	39.86	<b>95.98</b>
Meadows	1865	16784	<b>98.13</b>	97.60	96.26	96.78	99.41	99.01	95.89	<b>99.48</b>
Gravel	210	1889	65.08	68.68	19.34	<b>72.92</b>	69.18	74.42	20.01	<b>80.22</b>
Trees	304	2736	70.26	88.94	5.54	<b>90.95</b>	95.43	93.35	8.98	<b>95.61</b>
Metal	135	1210	59.73	98.66	43.74	<b>99.32</b>	99.08	99.15	58.41	<b>99.42</b>
Bare Soil	503	4526	56.46	63.07	<b>94.73</b>	87.30	94.28	91.19	96.14	<b>98.07</b>
Bitumen	133	1197	75.22	76.30	13.80	<b>81.62</b>	80.00	70.33	15.47	<b>82.45</b>
Bricks	368	3314	84.82	87.53	12.61	<b>89.04</b>	<b>90.30</b>	87.30	20.31	86.69
Shadows	95	852	92.69	<b>99.77</b>	21.95	99.39	96.85	<b>97.92</b>	19.39	96.58
Overall accuracy (%)			84.87 $\pm$ 0.30	89.05 $\pm$ 0.34	64.21 $\pm$ 1.85	<b>92.11 <math>\pm</math> 0.87</b>	94.73 $\pm$ 0.20	93.87 $\pm$ 0.29	65.46 $\pm$ 1.82	<b>95.98 <math>\pm</math> 0.50</b>
Average accuracy (%)			76.71 $\pm$ 0.68	85.74 $\pm$ 0.57	38.76 $\pm$ 2.74	<b>89.76 <math>\pm</math> 1.19</b>	90.23 $\pm$ 0.40	89.69 $\pm$ 0.41	41.61 $\pm$ 2.06	<b>92.72 <math>\pm</math> 0.79</b>
Kappa statistic			0.790 $\pm$ 0.005	0.852 $\pm$ 0.005	0.491 $\pm$ 0.028	<b>0.895 <math>\pm</math> 0.012</b>	0.930 $\pm$ 0.002	0.918 $\pm$ 0.004	0.516 $\pm$ 0.021	<b>0.945 <math>\pm</math> 0.008</b>

Table 1: Performance of the feature fusion method using different supervised classifiers on the Pavia University data set.

More precisely, the compressive measurements of the MS image comprise 6 high-spatial-resolution snapshots (compression ratio: 25%) while the compressive measurements of the HS image are 24 low-spatial-resolution projections (compression ratio: 25%). Subsequently, the proposed feature fusion method was applied directly to the compressive measurements. Figures 6e-6h display the bands recovered by the proposed feature fusion algorithm for different values of the TV regularization parameter  $\lambda_2$ . For this experiment, the sparsity regularization parameter was set to  $\lambda_1 = 0.001 \times \|\mathbf{H}^\top \mathbf{y}\|_\infty$ . As can be seen in these figures, the recovered bands preserve the spatial structures of the underlying scene. It is relevant to notice that the fused cube is estimated directly from compressive measurements without applying a previous feature extraction stage.

The bands of the fused features exhibit smoother regions as  $\lambda_2$  increases, preserving, in turn, the spatial structure and the edges of the scene. Additionally, image noise is minimized as  $\lambda_2$  increases, and this effect efficiently improve the labeling performance. To evaluate this behavior, the classification maps obtained for different values of the TV parameter are illustrated in Fig. 7. The ground truth map is shown in Fig. 7a. As can be seen in this figure, the classification noise is reduced as  $\lambda_2$  increases.

To evaluate the performance of the proposed feature fusion method, Table 1 shows the labeling accuracy results obtained by different supervised classification methods: support vector machines with radial basis function (SVM-RBF) kernel, random forest (RF), the convolutional neural network (CNN) and the multi-layer perceptron neural network (MLPNN). To be more precise, the labeling accuracy is determined for each class, where each value is obtained by averaging ten realizations of the respective experiment. For each trial, a different pattern of the coded aperture is generated and a set of 10% of features are randomly selected as training samples. For testing the corresponding classification method, the remaining 90% of features are used. Furthermore, Table 1 displays the overall accuracy (OA), the average accuracy (AA), and the Kappa statistic ( $\kappa$ ) [33]. For comparison purposes, the accuracy results obtained from the high-resolution image

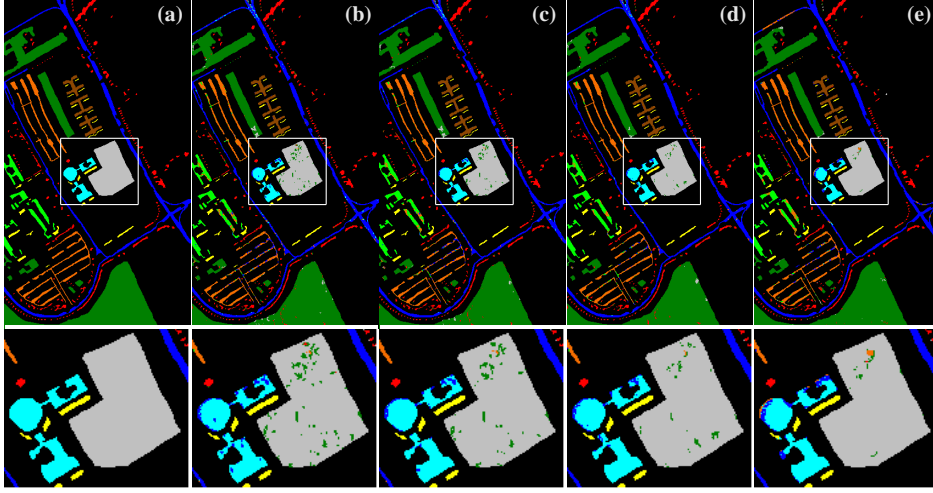


Figure 7: Pavia University spectral image. (a) The classification map of the ground truth; classification maps obtained by proposed approach for different TV regularization parameters (b)  $\lambda_2 = 0$  (c)  $\lambda_1 = 5 \times 10^{-6}$  (d)  $\lambda_1 = 5 \times 10^{-5}$  (e)  $\lambda_1 = 5 \times 10^{-4}$ .

are also included, where spectral signatures are considered classification features. The best accuracy values obtained from either the high-resolution image and the fused features are in bold font. The fused features were estimated using  $\lambda_1 = 0.01$  and  $\lambda_2 = 5 \times 10^{-4}$ .

For this experiment, the parameters of SVM-RBF classifier were set to  $\sigma = 1$  and  $C = 1$ , where *one-against-one* multi-classification rule was implemented [34]. The number of trees of the RF classifier was fixed to 300. The CNN approach reduces the dimensionality of the input data via principal component analysis (PCA). Then, the first three principal components are used as input data of the CNN architecture. For each pixel, a patch of size  $27 \times 27 \times 3$  is obtained with the aim of building the training and test sets. The CNN architecture consists of four convolutional layers of 16, 32, 64, and 128 filters, respectively, with a kernel size of  $5 \times 5$ . Furthermore, each hidden layer includes a max-pooling layer with a size of  $2 \times 2$ . Finally, a fully connected layer is aggregated at the output end. As can be seen in Table 1, the accuracy results obtained from fused features outperform those yielded from the high-resolution image. In addition, the results obtained by the MLPNN classifier are, in general, better than those yielded by the other labeling methods.

#### 4.2. Indian Pines

This data set was captured by an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over an agricultural region in Indiana, USA. This Pines spectral image consists of a dataset with  $145 \times 145$  pixels and 200 spectral bands recording the wavelength range from 0.40 to  $2.50 \mu\text{m}$  [36]. Furthermore, this data set exhibits a spatial-resolution of 20 m per pixel. The RGB composite of this spectral image is shown in Fig. 8a and the ground truth map is displayed in Fig. 8b.



Method	Input	Description
OTVCA [25]	High-resolution spectral image	A feature extraction method for hyperspectral images that solves a TV penalized optimization problem.
SSLRA [26]	High-resolution spectral image	A feature extraction method based on a low-rank model. This approach decomposes the hyperspectral image into smooth and sparse components.
EP [35]	High-resolution spectral image	A feature extraction method based on the technique referred to as extended extinction profiles.
FES [15]	Dual-resolution compressive measurements	Extracts features by reordering the compressive data yielding HS features and MS features. The method stacks an interpolated version of the HS features and a superpixel map of the MS features.
FEF-TR [17]	Dual-resolution compressive measurements	Implements the feature extraction stage [15]. The extracted features are fused by solving an inverse problem penalized by the Tikhonov term that leads a closed-form solution.
FEF-L1TV [16]	Dual-resolution compressive measurements	Implements the feature extraction stage [15]. The extracted features are fused by solving an inverse problem regularized by L1 and TV norms. An ADMM-based algorithm was developed to solve the formulated problem.
Proposed	Dual-resolution compressive measurements	Fuses features directly from compressive measurements by solving an inverse problem regularized by L1 and TV norms. An accelerated version of the ADMM algorithm is developed to solve the problem.

Table 2: General description of the comparison methods.

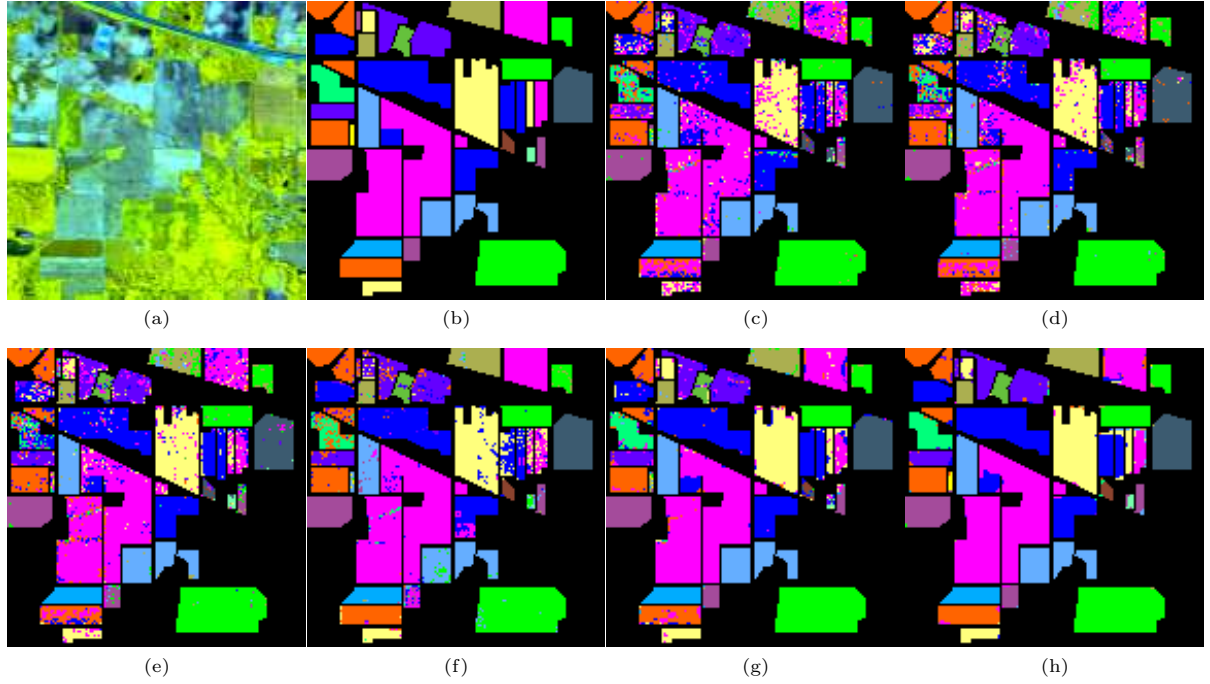


Figure 8: Indian Pines spectral image: (a) the RGB composite of the original image and (b) the classification map of the ground truth. Classification maps obtained by (c) the HR image (HR), OA: 79.40%, (d) the OTVCA method, OA: 79.47%, (e) the SSLRA method, OA: 81.31%, (f) the EP stacking, OA: 88.42%, (g) the FEF-L1TV method, OA: 92.99%, and (h) the proposed approach, OA: 96.54%.

Classes	# Samples		HR [34]	OTVCA [25]	SSLRA [26]	EP stacking [35]	FEF-L1TV [16]	Proposed
	Train	Test						
Alfalfa	9	37	51.08 $\pm$ 14.27	18.54 $\pm$ 9.60	46.08 $\pm$ 12.50	68.35 $\pm$ 8.31	<b>89.19 <math>\pm</math> 8.55</b>	86.22 $\pm$ 8.96
Corn-notill	286	1142	71.18 $\pm$ 3.38	75.87 $\pm$ 1.40	74.59 $\pm$ 2.53	69.57 $\pm$ 1.32	95.58 $\pm$ 1.26	<b>96.29 <math>\pm</math> 1.26</b>
Corn mintill	166	664	54.86 $\pm$ 2.56	55.78 $\pm$ 2.53	57.24 $\pm$ 2.36	87.45 $\pm$ 2.91	95.32 $\pm$ 1.63	<b>96.45 <math>\pm</math> 1.25</b>
Corn	47	190	60.63 $\pm$ 6.24	32.76 $\pm$ 4.78	39.02 $\pm$ 5.42	54.52 $\pm$ 11.10	93.42 $\pm$ 3.49	<b>95.21 <math>\pm</math> 3.27</b>
Grass-pasture	97	386	86.76 $\pm$ 2.51	84.70 $\pm$ 2.80	88.89 $\pm$ 2.19	71.04 $\pm$ 9.31	<b>94.61 <math>\pm</math> 1.50</b>	92.95 $\pm$ 1.64
Grass-trees	146	584	96.23 $\pm$ 0.97	95.55 $\pm$ 1.30	97.68 $\pm$ 0.95	78.16 $\pm$ 3.23	99.26 $\pm$ 0.31	<b>99.50 <math>\pm</math> 0.40</b>
Grass-pasture-mowed	6	22	1.82 $\pm$ 5.75	49.86 $\pm$ 11.97	<b>69.27 <math>\pm</math> 12.43</b>	44.77 $\pm$ 11.78	67.73 $\pm$ 24.19	42.73 $\pm$ 29.00
Hay-windrowed	96	382	<b>99.40 <math>\pm</math> 0.35</b>	98.96 $\pm$ 0.73	99.19 $\pm$ 0.63	99.23 $\pm$ 0.38	98.77 $\pm$ 0.69	98.95 $\pm$ 0.76
Oats	4	16	0.00 $\pm$ 0.00	2.56 $\pm$ 4.36	3.06 $\pm$ 5.65	0.19 $\pm$ 1.39	<b>65.00 <math>\pm</math> 27.83</b>	52.50 $\pm$ 26.55
Soybean-notill	194	778	67.04 $\pm$ 1.90	70.55 $\pm$ 2.53	71.63 $\pm$ 2.18	87.61 $\pm$ 1.93	95.66 $\pm$ 1.45	<b>96.75 <math>\pm</math> 1.06</b>
soybean-mintill	491	1964	87.72 $\pm$ 1.30	86.11 $\pm$ 1.15	87.22 $\pm$ 1.28	95.04 $\pm$ 1.35	96.62 $\pm$ 0.72	<b>97.56 <math>\pm</math> 0.43</b>
Soybean-clean	119	474	69.11 $\pm$ 2.86	65.76 $\pm$ 3.64	70.54 $\pm$ 4.70	72.93 $\pm$ 2.58	92.09 $\pm$ 2.37	<b>94.66 <math>\pm</math> 2.76</b>
Wheat	41	164	93.90 $\pm$ 3.76	96.76 $\pm$ 1.87	95.04 $\pm$ 2.53	94.11 $\pm$ 2.25	<b>97.99 <math>\pm</math> 1.29</b>	97.62 $\pm$ 3.35
Woods	253	1012	97.80 $\pm$ 0.28	97.07 $\pm$ 0.77	98.12 $\pm$ 0.64	92.65 $\pm$ 3.23	<b>99.25 <math>\pm</math> 0.38</b>	99.16 $\pm$ 0.65
Buildings-Drives	77	309	54.95 $\pm$ 2.97	64.68 $\pm$ 4.02	75.06 $\pm$ 3.92	91.16 $\pm$ 3.54	96.28 $\pm$ 2.16	<b>98.61 <math>\pm</math> 1.40</b>
Stone-Steel-Towers	19	74	76.89 $\pm$ 4.05	69.03 $\pm$ 6.68	74.61 $\pm$ 5.90	81.77 $\pm$ 6.81	93.65 $\pm$ 4.23	<b>94.86 <math>\pm</math> 3.48</b>
Overall accuracy (%)			79.66 $\pm$ 0.56	79.58 $\pm$ 0.41	81.38 $\pm$ 0.54	84.80 $\pm$ 0.66	96.27 $\pm$ 0.24	<b>96.91 <math>\pm</math> 0.38</b>
Average accuracy (%)			66.84 $\pm$ 1.18	66.53 $\pm$ 1.25	71.70 $\pm$ 1.37	74.28 $\pm$ 1.36	<b>91.90 <math>\pm</math> 2.23</b>	90.00 $\pm$ 1.93
Kappa Statistic			0.766 $\pm$ 0.007	0.765 $\pm$ 0.005	0.786 $\pm$ 0.006	0.827 $\pm$ 0.008	0.965 $\pm$ 0.003	<b>0.958 <math>\pm</math> 0.004</b>

Table 3: Performance of various spectral image classification methods on the Indian Pines data set.

For comparative purposes, Figs 8c-8h display the classification maps yielded by different types of features extraction and fusion methods. In particular, these methods obtain the classification features from the original spectral image. Specifically, we classify the spectral image from the spectral signatures of the high-resolution image (HR) [34], and the features obtained by the orthogonal total variation component analysis (OTVCA) [25], the sparse-smooth low-rank analysis (SSLRA) [26], and the extinction profile (EP) method [35]. Additionally, Fig. 8g includes the labeling map obtained by the feature extraction and fusion technique with the L1-TV regularization (FEF-L1TV) [16] and the classification map yielded by the proposed classification approach is illustrated in Fig. 8h. Table 2 include a general description of the involved methods. For these figures, simulations selected 20% of the classification attributes as training samples and 80% of the remaining features to test the various labeling approaches. Furthermore, we use the same training and test sets to evaluate the various approaches. As can be seen, the proposed approach is not affected by the classification noise compared with respect to the other approaches, yielding homogeneous labeling regions. Notice that the labeling map obtained by the proposed classification method provides the best OA value.

Table 3 shows the accuracy results obtained by various classification techniques on the labeling of the Indian Pines dataset. The accuracy values are obtained by averaging ten realizations of the corresponding experiment. Note that the best accuracy results are in bold font. Furthermore, the overall accuracy (OA), the average accuracy (AA), and the Kappa Statistic ( $\kappa$ ) are included in the last three rows of Table 3. As can be observed in this table, the classification approach that contains the proposed feature fusion method exhibits, in general, the best performance.

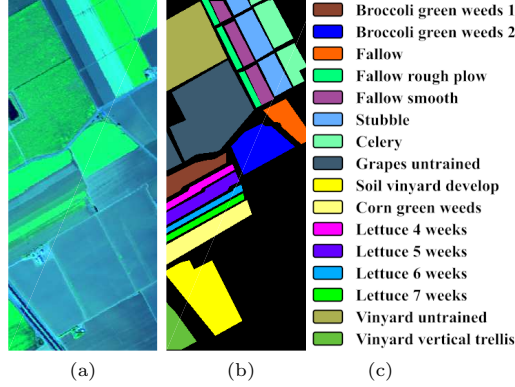


Figure 9: Salinas Valley spectral image: (a) the RGB composite of the original image and (b) the classification map of the ground truth.

#### 4.3. Salinas Valley

The Salinas Valley spectral image was captured by an AVIRIS sensor over the Valley of Salinas, California, USA [37]. This spectral image has dimensions  $512 \times 217 \times 192$  in the wavelength interval from 0.24 to 2.40  $\mu m$ . The RGB composite of this image is shown in Fig. 9a. Furthermore, Figs 9b and 9c display the ground truth labeling map for 16 different classes, where each class corresponds to a particular kind of crop.

Figure 10a shows the overall accuracy curves as the compression ratio increases for different fusion methods that obtain features from dual-resolution compressive measurements. We compare the performance of the proposed method with respect to those yielded by the feature extraction and stacking method (FES) [15], the feature extraction and fusion technique that uses the Tikhonov regularization (FEF-TR) [17], and the feature extraction and fusion method that exploits the L1-TV regularization (FEF-L1TV) [16]. In this work, the information embedded in coded aperture patterns is included in the projection matrices (9) and (14) leading to an algorithm that fuses spectral image features directly from compressive measurements.

As can be observed in Fig. 10a, the proposed classification approach exhibit a competitive performance with respect the other methods. Furthermore, Fig. 10b displays the OA curves versus the training rate for the various feature fusion approaches. For this experiment, the projections were captured by fixing the compression ratio at 25%. As shown in this figure, the proposed approach outperforms the other methods at low rates of training samples.

Additionally, Fig. 10c shows the OA curves versus the signal-to-noise ratio (SNR) exhibited by compressive measurements for the various fusion approaches. More precisely, projections are corrupted with additive noise obeying to a zero-mean Gaussian model. For these curves, five realizations are averaged. Notice that the proposed approach exhibits a superior performance compared to the remaining methods, with at least 4% of accuracy gain. In practical scenarios, detectors are affected by signal-dependent noise, therefore, we test the proposed approach when projections are corrupted by Poisson noise. In this regard,

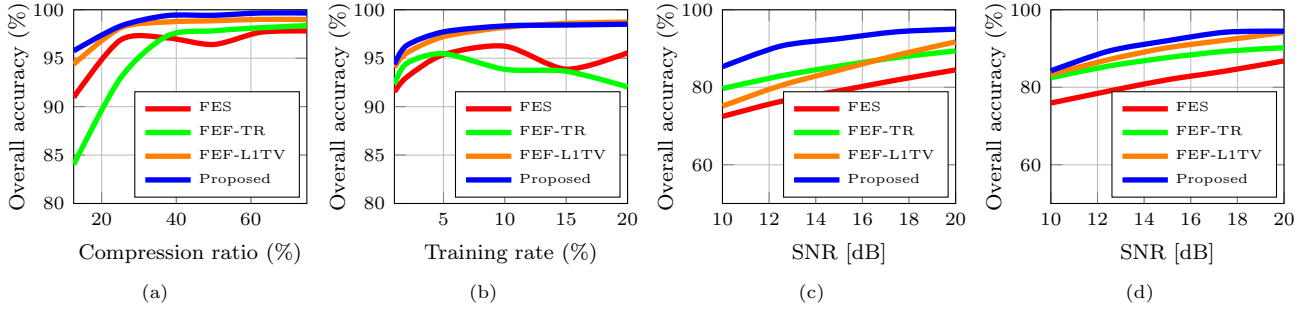


Figure 10: Performance on the Salinas Valley data set of the various classification methods: (a) OA versus the compression ratio and (b) OA versus the rate of training samples. OA on the Salinas Valley spectral image yielded by the various classification approaches as the SNR increases for (c) Gaussian noise and (d) Poisson noise.

Fig. 10d displays the OA curves versus the Poisson noise level. As can be seen in this figure, the proposed fusion strategy outperforms the other techniques for the interval under test.

To show the performance gain of the proposed fusion method with respect to other approaches that fuse features from dual-resolution compressive measurements, Fig. 11a displays the OA versus the compression ratio for the various fusion techniques using a particular supervised classifier. More precisely, we compare the proposed method with respect to FEF-TR and FEF-L1TV methods [23, 16]. To this end, we use a support vector machine with a polynomial kernel as a supervised classifier. Notice that every value of this curve is obtained by averaging of ten realizations of the corresponding experiment, and for each trial, a different coded aperture pattern is generated. Furthermore, for each realization of this experiment, we selected the same 10 % of the features as training samples and the remaining 90 % as test samples. As can be seen in Fig. 11a, the proposed feature fusion approach outperforms the other methods, exhibiting a remarkable accuracy gain for high compression rates. Finally, Fig. 11b shows the computation time spent by the various feature fusion methods as the compression rate increases. These values were obtained using a desktop architecture with an Intel Core i7 CPU, 3.00 GHz, 64-GB RAM, and Ubuntu 18.04 operating system. As can be seen in this figure, the proposed fusion approach exhibits about ten times lower computation times than those spent by the FEF-L1TV method. Additionally, the proposed feature fusion approach shows the best trade-off between labeling accuracy and computation time.

#### 4.4. Real measurements

Finally, the performance of the proposed classification approach from real compressive measurements is evaluated. To this end, we use data captured by a 3D-CASSI optical setup. More precisely, this data set was captured in the optics laboratory of the High-Dimensional Signal Processing (HDSP) group at the Universidad Industrial de Santander, Colombia. Furthermore, this dataset comprises 32 multispectral snapshots with dimensions  $256 \times 256$  and 8 hyperspectral snapshots with dimensions  $128 \times 128$ . Notice that these measurements were acquired to evaluate the performance of the FEF-TR method [17]. Figure 12a

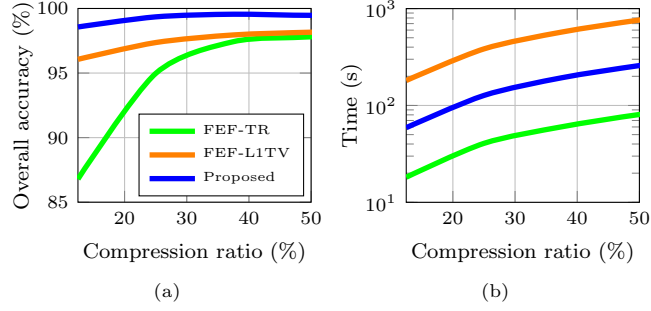


Figure 11: Salinas Valley data set: (a) OA versus the compression ratio yielded by the different compressive fusion approaches using the same supervised classifier and (b) Computation time versus the compression ratio yielded by the different compressive fusion approaches.

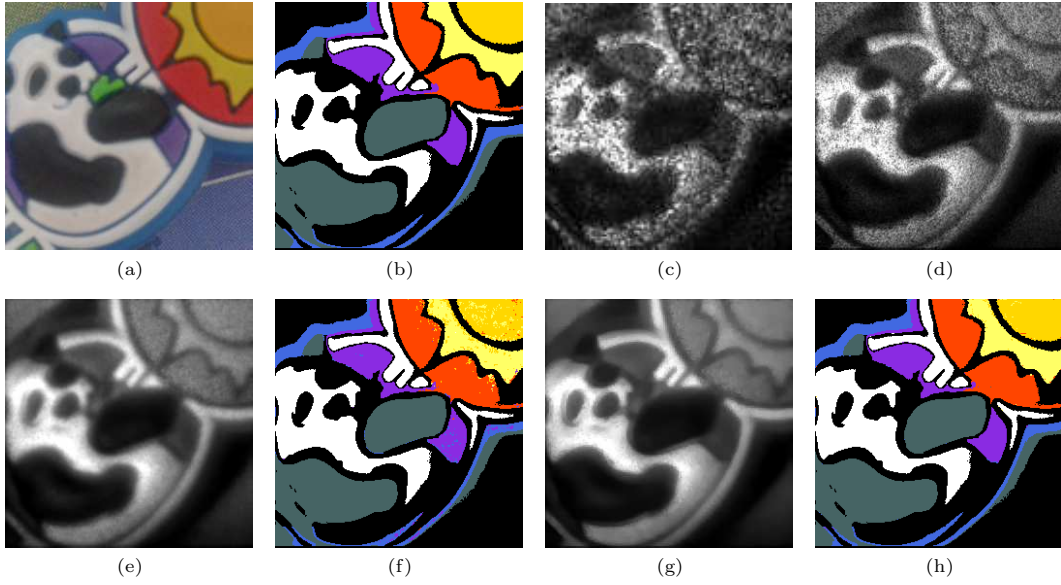


Figure 12: Real CSI measurements. (a) the RGB image of the scene and (b) the classification map of the ground truth; compressive projections captured by (c) a CHSI sensor and (d) a CMSI sensor; (e) and (f) a feature band and the labeling map obtained by the FEF-TR method [17], OA: 96.03; (g) and (h) a feature band and the classification map obtained by the proposed approach, OA: 98.88 %.

shows the RGB image of the scene and the ground truth map is displayed in Fig. 12b. Additionally, Figs 12d and 12c illustrate the camera snapshots captured by both the hyperspectral CSI sensor the multispectral CSI system, respectively.

For comparison purposes, Figs 12e and 12f illustrate a feature band and the classification map obtained by the FEF-TR approach [17]. A feature band and the labeling map estimated by the proposed classification approach are shown in Figs 12g and 12h, respectively. Note that the proposed approach reduces the undesirable artifacts due to the acquisition noise. In addition, the proposed labeling approach remarkably minimize the classification noise providing the best overall accuracy value.

## 5. Conclusions

In this paper, a method that fuses features directly from compressive data has been developed for spectral image classification. More precisely, this approach avoids the feature extraction stage by including the information of the coded aperture patterns to obtain projection matrices. The discrete mathematical model that characterizes the compressive measurements as degraded versions of the high-resolution features was presented. Subsequently, the feature fusion was formulated as a least-squares problem regularized by two penalty terms: a sparsity-promoting term and a total variation (TV) term. Additionally, an accelerated ADMM-based algorithm has been included to numerically solve the formulated problem. The proposed feature fusion method was incorporated into an approach that classifies spectral images from compressive data. The performance of the proposed classification technique was evaluated on four spectral image data sets under different criteria. Notice that the proposed approach was successfully tested on real compressive measurements. The proposed classification approach exhibited a superior performance with respect to other state-of-the-art methods that obtain features from compressive measurements.

## 6. Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754382, GOT ENERGY TALENT. The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed herein lies entirely with the authors.

## References

- [1] G. Camps-Valls, D. Tuia, L. Gómez-Chova, S. Jiménez, J. Malo, Remote sensing image processing, Synthesis Lectures on Image, Video, and Multimedia Processing 5 (2011) 1–192.
- [2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, J. Chanussot, Hyperspectral remote sensing data analysis and future challenges, IEEE Geoscience and Remote Sensing Magazine 1 (2013) 6–36.

- [3] N. Yokoya, C. Grohnfeldt, J. Chanussot, Hyperspectral and multispectral data fusion: A comparative review of the recent literature, *IEEE Geoscience and Remote Sensing Magazine* 5 (2017) 29–56.
- [4] N. Yokoya, T. Yairi, A. Iwasaki, Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion, *IEEE Transactions on Geoscience and Remote Sensing* 50 (2012) 528–537.
- [5] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, D. J. Brady, Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world, *IEEE Signal Processing Magazine* 33 (2016) 95–108.
- [6] A. Wagadarikar, R. John, R. Willett, D. Brady, Single disperser design for coded aperture snapshot spectral imaging, *Applied Optics* 47 (2008) B44–B51.
- [7] H. Arguello, G. R. Arce, Colored coded aperture design by concentration of measure in compressive spectral imaging, *IEEE Transactions on Image Processing* 23 (2014) 1896–1908.
- [8] C. V. Correa, C. Hinojosa, G. R. Arce, H. Arguello, Multiple snapshot colored compressive spectral imager, *Optical Engineering* 56 (2016) 041309.
- [9] H. Rueda, H. Arguello, G. R. Arce, Dual-arm vis/nir compressive spectral imager, in: 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 2572–2576.
- [10] A. Jerez, H. Garcia, H. Arguello, Single pixel spectral image fusion with side information from a grayscale sensor, in: 2018 IEEE 1st Colombian Conference on Applications in Computational Intelligence (ColCACI), 2018, pp. 1–6.
- [11] J. Hauser, M. A. Golub, A. Averbuch, M. Nathan, V. A. Zheludev, M. Kagan, Dual-camera snapshot spectral imaging with a pupil-domain optical diffuser and compressed sensing algorithms, *Applied Optics* 59 (2020) 1058–1070.
- [12] M. Imani, H. Ghassemian, An overview on spectral and spatial information fusion for hyperspectral image classification: Current trends and challenges, *Information fusion* 59 (2020) 59–83.
- [13] A. Ramirez, H. Arguello, G. R. Arce, B. M. Sadler, Spectral image classification from optimal coded-aperture compressive measurements, *IEEE Transactions on Geoscience and Remote Sensing* 52 (2014) 3299–3309.
- [14] H. Vargas, H. Arguello, A low-rank model for compressive spectral image classification, *IEEE Transactions on Geoscience and Remote Sensing* 57 (2019) 9888–9899.
- [15] C. Hinojosa, J. M. Ramirez, H. Arguello, Spectral-spatial classification from multi-sensor compressive measurements using superpixels, in: 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 3143–3147.
- [16] J. M. Ramirez, H. Arguello, Multiresolution compressive feature fusion for spectral image classification, *IEEE Transactions on Geoscience and Remote Sensing* 57 (2019) 9900–9911.
- [17] J. M. Ramirez, H. Arguello, Spectral image classification from multi-sensor compressive measurements, *IEEE Transactions on Geoscience and Remote Sensing* 58 (2020) 626–636.
- [18] Y. Ouyang, Y. Chen, G. Lan, E. Pasiliao Jr, An accelerated linearized alternating direction method of multipliers, *SIAM Journal on Imaging Sciences* 8 (2015) 644–681.
- [19] A. C. Sankaranarayanan, C. Studer, R. G. Baraniuk, Cs-muvi: Video compressive sensing for spatial-multiplexing cameras, in: 2012 IEEE International Conference on Computational Photography (ICCP), 2012, pp. 1–10.
- [20] H. Rueda, H. Arguello, G. R. Arce, Dual-arm vis/nir compressive spectral imager, in: 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 2572–2576.
- [21] X. Lin, G. Wetzstein, Y. Liu, Q. Dai, Dual-coded compressive hyperspectral imaging, *Optics letters* 39 (2014) 2044–2047.
- [22] A. Parada-Mayorga, G. R. Arce, Colored coded aperture design in compressive spectral imaging via minimum coherence, *IEEE Transactions on Computational Imaging* 3 (2017) 202–216.
- [23] C. Hinojosa, J. Bacca, H. Arguello, Coded aperture design for compressive spectral subspace clustering, *IEEE Journal of Selected Topics in Signal Processing* 12 (2018) 1589–1600.
- [24] M.-D. Iordache, J. M. Bioucas-Dias, A. Plaza, Total variation spatial regularization for sparse hyperspectral unmixing, *IEEE Transactions on Geoscience and Remote Sensing* 50 (2012) 4484–4502.

- [25] B. Rasti, M. O. Ulfarsson, J. R. Sveinsson, Hyperspectral feature extraction using total variation component analysis, *IEEE Transactions on Geoscience and Remote Sensing* 54 (2016) 6976–6985.
- [26] B. Rasti, P. Ghamisi, M. O. Ulfarsson, Hyperspectral feature extraction using sparse and smooth low-rank analysis, *Remote Sensing* 11 (2019) 121.
- [27] Y. Ouyang, Y. Chen, G. Lan, E. Pasillao Jr, An accelerated linearized alternating direction method of multipliers, *SIAM Journal on Imaging Sciences* 8 (2015) 644–681.
- [28] I. Goodfellow, Y. Bengio, A. Courville, *Deep learning*, MIT press, 2016.
- [29] X. Mei, E. Pan, Y. Ma, X. Dai, J. Huang, F. Fan, Q. Du, H. Zheng, J. Ma, Spectral-spatial attention networks for hyperspectral image classification, *Remote Sensing* 11 (2019) 963.
- [30] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, J. A. Benediktsson, Deep learning for hyperspectral image classification: An overview, *IEEE Transactions on Geoscience and Remote Sensing* 57 (2019) 6690–6709.
- [31] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O’Reilly Media, 2019.
- [32] Grupo de Inteligencia Computacional, Hyper Remote Sensing Scenes, [http://www.ehu.eus/ccwintco/index.php/Hyperspectral\\_Remote\\_Sen](http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sen) 2008. [Online; accessed 27-January-2020].
- [33] C. M. Bishop, *Machine learning and pattern recognition*, Information science and statistics. Springer, Heidelberg (2006).
- [34] F. Melgani, L. Bruzzone, Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions on Geoscience and Remote Sensing* 42 (2004) 1778–1790.
- [35] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, X. X. Zhu, Hyperspectral data classification using extended extinction profiles, *IEEE Geoscience and Remote Sensing Letters* 13 (2016) 1641–1645.
- [36] M. F. Baumgardner, L. L. Biehl, D. A. Landgrebe, 220 Band AVIRIS Hyperspectral Image Data Set: June 12, 1992 Indian Pines Test Site 3, <https://purrr.purdue.edu/publications/1947/1>, 2019. [Online; accessed 24-February-2020].
- [37] Jet Propulsion Laboratory, NASA, 2006-2019 AVIRIS Data Portal, <https://aviris.jpl.nasa.gov/dataportal/>, 2019. [Online; accessed 19-February-2020].