Coded Aperture Design for Compressive Spectral Subspace Clustering

Carlos Hinojosa, *Student Member, IEEE*, Jorge Bacca, *Student Member, IEEE*, and Henry Arguello, *Senior Member, IEEE*

***Abstract*—Compressive spectral imaging (CSI) acquires compressed observations of a spectral scene by applying different coding patterns at each spatial location and then performing a spectral-wise integration. Relying on compressive sensing, spectral image reconstruction is achieved by using nonlinear and relatively expensive optimization-based algorithms. In the CSI literature, several works have focused on improving reconstructions quality by properly designing the set of coding patterns. However, signal recovery is not actually necessary in many signal processing applications. For instance, assuming that compressed measurements with similar characteristics lie on the same subspace, unsupervised methods such as subspace clustering can be used to separate them into the same cluster. Since the structure of compressed measurements is defined by the applied codification, it is possible to improve clustering performance. This paper proposes to design a set of coding patterns such that inter-class and intra-class data structure is preserved after the CSI acquisition in order to improve clustering results directly on the compressed domain. To validate the coding pattern design, an algorithm based on sparse subspace clustering (SSC) is proposed to perform clustering on the compressed measurements. The proposed algorithm adds a three-dimensional (3-D) spatial regularizer to the SSC problem exploiting the spatial correlation of spectral images. In general, an overall accuracy up to 83.81% is obtained, when noisy measurements are assumed. In addition, a difference of at most 4% in terms of overall accuracy was observed when comparing the clustering results obtained by the full 3-D data with those achieved using CSI measurements acquired with the proposed coding pattern design.**

***Index Terms*—Compressive spectral imaging, coded aperture design, spectral subspace clustering.**

I. Introduction

SPECTRAL imaging senses two-dimensional (2D) spatial information of a scene across a range of spectral wavelengths. These data sets are regarded as three-dimensional (3D) images, where two of the coordinates correspond to the spatial domain and the third one represents the spectral wavelengths. Traditional sensing techniques construct a spatio-spectral data cube by scanning the scene, either spectrally or spatially in proportion to the desired spatial or spectral resolution, which in turn, increases acquisition times. Exploiting the fact that different objects in the scene reflect, scatter, absorb, and emit electromagnetic energy in distinctive patterns related to their molecular composition, spectral imaging has emerged as a valuable tool for remote sensing applications including agriculture, urban planning, military surveillance, etc [1], [2]. A common task underlying these applications is to detect and classify different materials based on their reflectance spectrum or spectral signatures of the pixels (a.k.a spectral pixels).

Assuming that spectral signatures with similar characteristics lie in the same low-dimensional subspace, subspace clustering theory can be used for modeling the spectral image classification problem [3]. Different methods for subspace clustering, which take into account the multi-subspace structure of the data, have been proposed in the past two decades. Among them, sparse subspace clustering (SSC) constructs a similarity matrix by expressing each data point as a linear combination of all other points in the dataset and then, imposing a sparsity restriction over the coefficients matrix [3]–[6]. Once obtained the sparse coefficient matrix, the segmentation of data points is performed using the spectral clustering (SC) algorithm. In general, spectral image clustering is a very difficult task due to the inherent data complexity and computational cost, which grows in proportion to the dimensions of spectral data sets. In addition, processing such high dimensional data also requires huge computational resources and storage capacities. Therefore, a preprocessing step to reduce the dimension of the spectral imagery is often used in order to perform different image processing techniques[7].

Recently, compressive spectral imaging (CSI) has emerged as a new spectral imaging approach which acquires compressed 2D projections of the entire data cube rather than direct measurements of all voxels [8]. This approach allows to sense and simultaneously reduce the data dimensionality without further processing steps [9], [10]. Therefore, the cost of sensing, storing, transmitting and processing spectral images using this method is significantly reduced. Until now, different CSI sensing methods have been proposed to compress spectral images [9]–[13]. Particularly, in this work the spatial-spectral coded compressive spectral imager (3D-CASSI) is adopted. This CSI sensing

Manuscript received April 16, 2018; revised August 12, 2018 and October 19, 2018; accepted October 22, 2018. Date of publication October 26, 2018; date of current version December 17, 2018. This work was supported by the vicepresidency of research and extension of the Universidad Industrial de Santander under Grant 1872. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Zhouchen Lin. *(Corresponding author: Carlos Hinojosa.)*

The authors are with the Department of System Engineering and Informatics, Universidad Industrial de Santander, Bucaramanga 680002 Colombia (e-mail: carlos.hinojosa@saber.uis.edu.co; jorge.bacca1@correo.uis.edu.co; henarfu@uis.edu.co).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org provided by the authors. The Supplemental Materials contain the majority of demonstration of the theorems listed in the manuscript. This material is 0.194 MB in size..

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

Digital Object Identifier 10.1109/JSTSP.2018.2878293

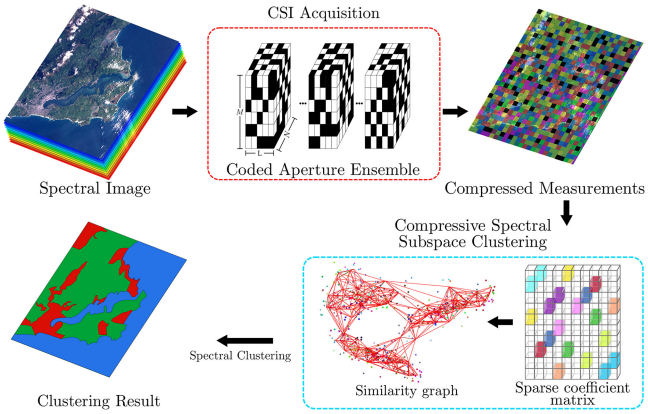

Fig. 1. CSI subspace clustering workflow. Spectral signatures of the scene are encoded by the coded aperture ensemble before the spectral-wise integration. Proper coded aperture design to preserve spectral similarities, allows to perform clustering directly on the compressed measurements.

scheme modulates the spectral data cube in spatial and spectral dimensions using a 3D coded aperture (ensembles of 2D coded apertures) or a coding pattern array. Then, the coded spectral data cube is integrated along the spectral dimension such that each spatial position of the acquired measurements contains the compressed information of a single coded spectral signature [11]. Works in [14]–[16] describe different architectures considered as 3D-CASSI scheme (see [11] for more details).

Because the coded aperture determines how the spectral pixels are encoded, it is possible to define the general structure of the acquired compressed data by properly designing the 3D coded aperture ensemble. Assuming that the compressed measurements with similar characteristics lie on the same subspace, this work focuses on the problem of spectral pixels subspace clustering, directly from compressed measurements, without recovering the original spectral scene. Specifically, our main contribution is to provide a coded aperture design such that the information, and the similarity between the spectral signatures, is approximately preserved after the sensing process, enhancing the spectral pixel clustering directly in the compressed measurements. The proposed design is based on three criteria: considering that surface-emitted spectral signatures are, in general, relatively smooth functions of the wavelength, only neighboring spectral bands are sensed; the cosine of the angle between each pair of spectral pixels should be approximately preserved in the compressed domain; each 3D coded aperture realization should extract new information from the scene. Once obtaining the designed set of coded apertures, 2D projections of the spectral image data are acquired using the 3D-CASSI sensing approach. Finally a sparse subspace clustering-based algorithm is proposed in order to perform the spectral image clustering. Such algorithm incorporates a 3D spatial regularizer in the SSC problem to exploit the spatial correlation of spectral images. The overall CSI subspace clustering workflow is depicted in Fig. 1. Despite recent works have successfully performed subspace clustering on compressed measurements [17], this is the first work that particularly design a binary coded aperture ensemble to improve clustering results directly on CSI measurements without recovering the underlying spectral scene.

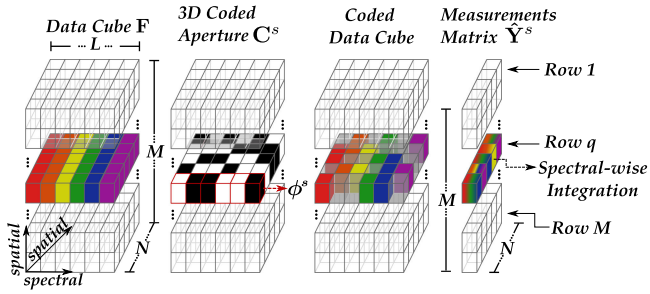

Fig. 2. CSI sensing approach used for compressed measurements acquisition at snapshot s.

In the following sections, the mathematical model of the CSI sensing approach is described. Then, the coding pattern design is formulated as an optimization problem and an algorithm to solve it is also developed. Further, theoretical bounds on the ℓ_2 norm preservation of the spectral pixels after being compressed by the proposed patterns is provided. Simulations are included to analyze the performance of the presented spectral image clustering approach.

II. Compressed Measurements Acquisition

Denote $\boldsymbol{\mathcal{F}}$ as the spatio-spectral input data cube, with $M \times N$ spatial dimensions, L spectral bands and entries denoted as $\mathcal{F}_{m,n,k}$, where m and n index the spatial coordinates, and k determines the k-th spectral band. As shown in Fig. 2, the 3D-CASSI first modulates the voxels of the spectral scene using a 3D coded aperture $\mathbf{C}^s$, whose entries are indexed as $C^s_{m,n,k}$. Then, the coded spectral scene is integrated in the focal plane array (FPA) detector, along the spectral axis. In CSI it is possible to acquire $S \ll L$ measurement shots, each one employing a different coded aperture, such that different measurements of the spectral data cube are acquired each time. Therefore, the output of the sensing process, at the (m,n)-th detector pixel and a specific snapshot s, can be expressed as

$$\hat{Y}^s_{m,n} = \sum_{k=0}^{L-1} \mathcal{F}_{m,n,k} C^s_{m,n,k}. \tag{1}$$

Note that to each spatial location of $\mathbf{C}^s$ is assigned a coding pattern $\boldsymbol{\phi}^s \in \mathbb{R}^L$, with entries $(\boldsymbol{\phi}^s)_k \in \{0,1\}$, that modulates a spectral pixel in that particular position before being integrated at the FPA detector. Further, observe that there is a finite number of coding patterns randomly distributed in $\mathbf{C}^s$.

At each measurement shot, the spectral pixels are encoded differently using distinct coding patterns, hence new information is acquired from the underlying data cube. Denoting P as the number of different coding patterns distributed on $\mathbf{C}^s$, if the number of measurement shots S is greater than P, then some pixels are oversampled, thus redundant information is acquired. On the other hand, if $S < P$ clustering becomes hard since the similarity among two spectral pixels in the same cluster may decrease if they are encoded (projected) differently. Finally, if $S = P$ the 3D coded aperture for each measurement shot can be designed such that a particular pixel is modulated once by a different coding pattern in a particular snapshot, acquiring new information each time.

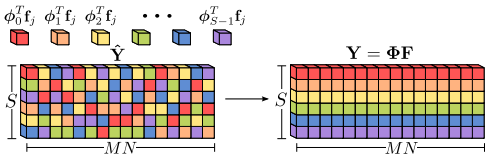

Fig. 3. Rearrangement of the matrix $\hat{\mathbf{Y}}$ such that the s-th row of $\mathbf{Y}$ contains the compressed measurements acquired with the s-th coding pattern $\boldsymbol{\phi}^s$.

The set of compressed measurements from (1) can be arranged in a $S \times MN$ matrix $\hat{\mathbf{Y}} = [[\hat{Y}^0_{0,0}, \hat{Y}^0_{1,0}, \ldots, \hat{Y}^0_{0,1}, \ldots, \hat{Y}^0_{(M-1),(N-1)}]^T, \ldots, [\hat{Y}^{(S-1)}_{0,0}, \ldots, \hat{Y}^{(S-1)}_{(M-1),(N-1)}]^T]^T$, where each column value corresponds to a compressed spectral signature. Note that, each row of $\hat{\mathbf{Y}}$ contains the compressed information (spectral responses) of the pixels acquired in the s-th snapshot. However, column vectors of $\hat{\mathbf{Y}}$ may contain the measurement acquired with a specific coding pattern $\boldsymbol{\phi}^s$ in different rows since, in a particular snapshot, all the spectral pixels are not necessarily encoded by the same coding pattern. Then, the matrix $\hat{\mathbf{Y}}$ is not convenient for SSC as its structure makes difficult to discriminate among compressed measurements. Therefore, using $S = P$, the entries of $\hat{\mathbf{Y}}$ are rearranged to form a new matrix $\mathbf{Y}$, such that each row contains the compressed information acquired with a specific coding pattern $\boldsymbol{\phi}^s$. Note that this rearrangement is possible only when $S = P$ since in this case it can be guaranteed that, at a specific snapshot, one pixel is encoded only once by a different pattern and, at the end of the sensing procedure, all pixels were encoded by the whole set of S coding patterns. Formally, the rearrangement can be expressed as

$$Y_{s,j} = \hat{Y}_{s',j} \quad \text{if } \hat{Y}_{s',j} = (\boldsymbol{\phi}^s)^T \mathbf{f}_j \quad \forall s',$$

for $s, s' = 0, \ldots, S-1$, where $\mathbf{f}_j \in \mathbb{R}^L$ denotes the j-th spectral signature for $j = 0, \ldots, MN-1$. This rearrangement, depicted in Fig. 3, preserves the structure of the underlying high dimensional data improving the subspace clustering results. Alternatively, define the matrix of S coding patterns as $\mathbf{\Phi} = [\boldsymbol{\phi}^0, \boldsymbol{\phi}^1, \ldots \boldsymbol{\phi}^{S-1}]^T$ then, the problem of acquiring and rearranging the measurements $\hat{\mathbf{Y}}$ can be succinctly expressed as the random projection

$$\mathbf{Y} = \mathbf{\Phi F}, \tag{2}$$

where $\mathbf{F} = [[\mathcal{F}_{0,0,0}, \mathcal{F}_{1,0,0}, \ldots, \mathcal{F}_{0,1,0}, \ldots, \mathcal{F}_{(M-1),(N-1),0}]^T, \ldots, [\mathcal{F}_{0,0,(L-1)}, \ldots, \mathcal{F}_{(M-1),(N-1),(L-1)}]^T]^T$ is a $L \times MN$ matrix whose columns are the spectral signatures $\mathbf{f}_j$ of the data cube and $\mathbf{\Phi}$ can be viewed as the projection matrix.

A typical procedure after the compressed measurements acquisition is signal recovery, which is achieved using nonlinear and relatively expensive optimization-based or iterative algorithms [18], [19]. CSI signal recovery procedure is explained in [8]. The aim of this work is to avoid the computational cost of recovering all the data cube by performing all the spectral image clustering directly on the compressed measurements using a set of designed coded apertures. Since the coding patterns $\boldsymbol{\phi}^s$ determine the structure of $\mathbf{C}^s$, designing the set of 3D coded apertures is equivalent to the coding pattern design.

III. Coding Pattern Design for CSI Subspace Clustering

Based on the concept of affinity [20], which characterizes the similarity between two subspaces, it has been theoretically proven and numerically verified that several dominant subspace clustering algorithms could successfully perform clustering on the compressed data [21]–[23]. These works employ the random projection method[24], using Gaussian random matrices, in order to acquire the dimensionality-reduced or compressed data. Such random projection matrices preserve the structure of high dimensional data, hence it is possible to directly learn from the low dimensional data [25]. In addition, recent works in [26], [27] have theoretically proved that, with high probability, the Euclidean distance and principal angles between two subspaces remain almost unchanged after Gaussian random projections, which allows to perform sparsity related signal processing tasks directly on compressed measurements. However, in CSI, the measurements are typically acquired using sensing or projection matrices whose entries are not Gaussian. In addition, although work in [23] provides theoretical results for other compression matrices which are not necessarily Gaussian, these do not take into account the structure of hyperspectral data. In this section, an coding pattern design for CSI, which takes into account the hyperspectral data structure, is presented. The proposed design considers the way in which the spectral scene information is encapsulated in the CSI measurements. Further, such design not only allows to learn from the compressed data but also preserves its ℓ_2 norm and the spectral signature similarities among vectors on the compressed domain. Additionally, an algorithm to generate the coding matrix $\mathbf{\Phi}$ is developed in this section.

A. Coding Pattern Design

Recent works in CSI have focused on properly designing the coding patterns in order to better reconstruct the underlying spectral scene [12], [28]. These designs use the restricted isometry property (RIP) as the main optimization criterion. On the other hand, because the aim of this work is to perform classification on the compressed measurements, the design of the coding patterns must preserve the similarity among the spectral signatures. In order to design the coding patterns matrix $\mathbf{\Phi}$, the following three design criteria are considered.

1) Sensing Scheme: The entries of the matrix $\mathbf{\Phi}$ can be chosen from a Bernoulli distribution $(\mathbf{\Phi})_{s,k} \sim Be(p)$. Therefore, the entries of the s-th coding pattern can be expressed as

$$(\boldsymbol{\phi}^s)_k = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } q, \end{cases} \tag{3}$$

for $k = 0, 1, \ldots, L-1$, where $q = 1 - p$. A projection matrix with this structure simply carries out a random sampling on the data vectors, across all the spectral bands, before performing element-wise addition. Considering that surface-emitted spectral signatures are, in general, relatively smooth functions of wavelength [29], acquiring the information of different sets of adjacent spectral bands at each snapshot will preserve the original signal structure. Therefore, the intuition is to perform the sampling of neighboring spectral bands instead of randomly

sampling all the spectral data vectors, which could add outliers to the measurements.

For each coding pattern $\boldsymbol{\phi}^s$, two cutoff wavelengths $\lambda_1^s, \lambda_2^s \in \{0, 1, \ldots, L-1\}$ are selected, such that $\lambda_1^s < \lambda_2^s$ and $\lambda_2^s - \lambda_1^s + 1 = \Delta$, where Δ is defined as the coding pattern bandwidth. Then, the band-structured matrix can be expressed as

$$(\boldsymbol{\phi}^s)_k = \begin{cases} \varphi_k^s, \iff \lambda_1^s \leq k \leq \lambda_2^s \\ 0, \quad \text{otherwise}, \end{cases} \tag{4}$$

where $\boldsymbol{\varphi}^s \in \{0,1\}^L$ is a L-long vector whose nonzero elements, within the region delimited by λ_1^s and λ_2^s, represent the sampled spectral bands of the pixels. Equation (4) can be alternatively written as

$$(\boldsymbol{\phi}^s)_k = \delta_{\lfloor \lambda_1^s / k \rfloor} \delta_{\lfloor k / \lambda_2^s \rfloor} \varphi_k^s, \tag{5}$$

where δ_x is the Kronecker delta function, i.e., δ_x is equal to 1 only when $x = 0$.

2) Preserving Similarities: The success of subspace clustering on the compressed measurements depends fundamentally on how the coding matrix $\boldsymbol{\Phi}$ affects the mutual similarities of the spectral signatures. A usual measure of similarity among two vectors is the cosine of the angle between them. Then, assuming that the vectors have unit length, the similarity between two compressed measurements $\mathbf{y}_j = \boldsymbol{\Phi}\mathbf{f}_j$, $\mathbf{y}_{j'} = \boldsymbol{\Phi}\mathbf{f}_{j'}$ is defined as

$$\text{sim}\,(\mathbf{y}_j, \mathbf{y}_{j'}) = \mathbf{y}_j^T \mathbf{y}_{j'} = \mathbf{f}_j^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{f}_{j'} \quad j \neq j', \tag{6}$$

where $\mathbf{y}_j \in \mathbb{R}^S$ and $\mathbf{f}_j \in \mathbb{R}^L$ correspond to the j-th column of the matrices $\mathbf{Y}$ and $\mathbf{F}$, respectively. If the columns of $\boldsymbol{\Phi}$ are normalized, it is possible to decompose the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ as

$$\boldsymbol{\Phi}^T \boldsymbol{\Phi} = \mathbf{I} + \boldsymbol{\Theta}, \tag{7}$$

where

$$\Theta_{kk'} = (\boldsymbol{\Phi}_k)^T \boldsymbol{\Phi}_{\mathbf{k}'} \quad \mathbf{k} \neq \mathbf{k}', \tag{8}$$

$\boldsymbol{\Phi}_{\mathbf{k}}$ denotes the k-th column of $\boldsymbol{\Phi}$, and $\Theta_{kk} = 0$. Observe that the matrix $\boldsymbol{\Theta}$ collects all the entries outside the diagonal of $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$. Therefore, if $\Theta_{jj'} = 0\ \forall j, j'$, the matrix $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ would be equal to $\mathbf{I}$ and the similarities of the spectral signatures would be exactly preserved in the compressed measurements. However, because the matrix $\boldsymbol{\Phi}$ has more columns than rows, all the entries of $\boldsymbol{\Theta}$ could be mostly small but not equal zero [30]. Considering that a linear mapping such as that in (2) can cause significant distortions in the compressed measurements if $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is not approximately $\mathbf{I}$, the proposed coded aperture design should minimize the entries of $\boldsymbol{\Theta}$.

3) Information Acquisition: In order to better discriminate among the classes, new information from the underlying spectral scene should be acquired in each measurement shot, hence the coding patterns should be as different as possible, and the matrix $\boldsymbol{\Phi}$ should be full rank. Note that if all coding patterns (rows of $\boldsymbol{\Phi}$) are different, i.e. a spectral band is only sensed once by a unique coding pattern, the values outside the diagonal of $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ should be zero and the diagonal values will be constants related to the number of nonzero elements in each coding pattern [12]. Since it is desired that all coding patterns have the same bandwidth Δ, $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ will approximate to the matrix $\Delta\mathbf{I}$. Then, the matrix $\boldsymbol{\Phi}\boldsymbol{\Phi}^T$ can be decomposed as

$$\boldsymbol{\Phi}\boldsymbol{\Phi}^T = \mathbf{I} + \boldsymbol{\Lambda}, \tag{9}$$

where

$$\Lambda_{ss'} = \boldsymbol{\phi}^s (\boldsymbol{\phi}^{s'})^T \quad s \neq s', \tag{10}$$

and $\Lambda_{ss} = \Delta - 1$. Therefore, the minimization of the entries of $\boldsymbol{\Lambda}$ should be considered in the coded aperture design.

B. Optimization Algorithm for Coding Patterns Design

Taking into account the previous considerations, the proposed coding pattern design can be succinctly expressed as the following optimization problem

$$\begin{aligned} \underset{\{\boldsymbol{\Phi}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\varphi}^s\}}{\arg\min} \quad & f(\boldsymbol{\Phi}) = \|\boldsymbol{\Phi}^{\mathbf{T}}\boldsymbol{\Phi} - \mathbf{I}\|_F^2 + \|\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathbf{T}} - \mathbf{I}\|_F^2 \\ \text{subject to} \quad & (\boldsymbol{\phi}^s)_k = \delta_{\lfloor \lambda_1^s / k \rfloor} \delta_{\lfloor k / \lambda_2^s \rfloor} \varphi_k^s, \\ & \lambda_2^s = \lambda_1^s + \Delta - 1, \\ & \text{Rank}\,(\boldsymbol{\Phi}) = S, \end{aligned} \tag{11}$$

for $s = 0, \ldots, S-1$, $k = 0, \ldots, L$ and where $\boldsymbol{\lambda}_1 = [\lambda_1^0, \ldots, \lambda_1^{S-1}]$, $\boldsymbol{\lambda}_2 = [\lambda_2^0, \ldots, \lambda_2^{S-1}]$. This optimization problem can be efficiently solved with the procedure summarized in Algorithm 1. Specifically, lines 2 to 4 generate the first coding pattern $\boldsymbol{\phi}^0$, which has a band structure with a predefined bandwidth Δ. Then, lines 6–9 are intended to minimize the number of times a spectral band is sensed. Specifically, the algorithm counts how many spectral bands have been sensed in a certain bandwidth and then λ_1^s is chosen uniformly at random from the set of banded sections with less information A, complying with the criteria of subsection III-A3. Finally, we choose the position in which the inner products are minimized. This is attained by minimizing the elements outside the diagonal, in detail, lines 13–17 count the inner products in the neighborhood. Then, l is sampled uniformly at random from the set that minimizes the inner product (B) to assign a 1 in the coding pattern position l. As observed in Fig. 4, a random design of $\boldsymbol{\Phi}$ entries may lead to oversample a subset of spectral bands (green line) while leaving some spectral bands unsampled (red line).

Note that, Algorithm 1 is a greedy algorithm, since we make locally choices at each stage with the intent of minimizing the objective function. In Theorem III.1, we provide theoretical guarantees that the final result of Algorithm 1 tends to a stationary point.

Theorem III.1: Assuming that $\boldsymbol{\Phi}^0 = \mathbf{1}$, where $\mathbf{1}$ is an all-ones matrix, the sequence $\{\boldsymbol{\Phi}^{i_t}\}_{i_t \in \mathbb{N}}$ generated by Algorithm 1 converge to a stationary point. Further, this stationary point satisfies the three equality constraints in (11).

Proof: The proof can be found in Appendix A (see Supplementary Material).

C. Theoretical Results

In the previous section, the optimization algorithm for coding pattern design, proposed in (11), seeks at improving the

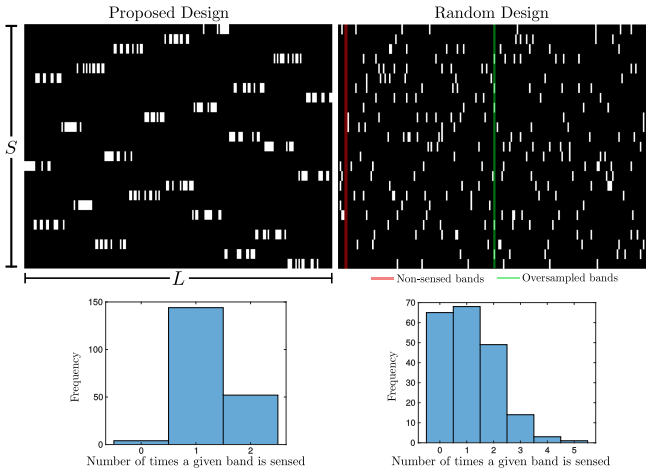

Fig. 4. Examples of coding patterns generated by the proposed (left) and random (right) design, respectively.

Algorithm 1: Coding Patterns Design.

Input: Number of bands L, number of shots S, bandwidth $\Delta > 0$.

Initialization :

$\boldsymbol{\Phi}^0 \leftarrow \mathbf{1}_{S,L}$

Randomly select λ_1^0, λ_2^0 such that $\lambda_2^0 > \lambda_1^0$ with $\lambda_2^0 - \lambda_1^0 + 1 = \Delta$

Select $\boldsymbol{\varphi}^s \in \mathbb{R}^L$ such that $\varphi_k^s \sim Be(\frac{1}{2})$

$\left(\boldsymbol{\phi}^0\right)_k \leftarrow \delta_{\lfloor \lambda_1^0/k \rfloor}\delta_{\lfloor k/\lambda_2^0 \rfloor}\varphi_k^0$ ▷ Banded Structure

for $i_t \leftarrow 1$ to $S-1$ **do**

  for $i \leftarrow 0$ to $(L-\Delta)$ **do**

    $\mathbf{u}_i \leftarrow \sum_{s'=0}^{i_t}\sum_{k=i}^{i+\Delta-1}(\boldsymbol{\phi}^{s'})_k$ ▷ Count sensed spectral bands

  end for

  $A \leftarrow \arg\min_i \mathbf{u}_i$ ▷ $\boldsymbol{\Lambda}$ minimizer set

  $\lambda_1^{i_t} \sim \mathcal{U}[A]$ ▷ $\lambda_1^{i_t}$ is sampled uniformly at random from set A

  $\lambda_2^{i_t} = \lambda_1^{i_t} + \Delta - 1$

  $\tilde{\ell} = 0$

  for $i \leftarrow \lambda_1^{i_t}$ to $\lambda_2^{i_t}$ **do**

    $\varphi_i^{i_t} \leftarrow 0$ ▷ zero values in the bandwidth

    $\mathbf{b}_{\tilde{\ell}} \leftarrow \sum_{s'=0}^{i_t}\prod_{k=(i-1)}^{i}(\boldsymbol{\phi}^{s'})_k$ ▷ Counter of inner products

    $\tilde{\ell} = \tilde{\ell} + 1$

  end for

  for $\tilde{i} \leftarrow 0$ to $\lfloor \frac{1}{2}\Delta \rfloor$ **do**

    $B \leftarrow \arg\min_{\tilde{\ell}} \mathbf{b}_{\tilde{\ell}}$ ▷ $\boldsymbol{\Theta}$ minimizer set

    $l \sim \mathcal{U}[B]$ ▷ l is sampled uniformly at random from set B

    $\varphi^{i_t}_{\lambda_1^{i_t}+l-1} \leftarrow 1$

    $\mathbf{b}_l \leftarrow \infty$

  end for

  $\left(\boldsymbol{\phi}^{i_t}\right)_k \leftarrow \delta_{\lfloor \lambda_1^{i_t}/k \rfloor}\delta_{\lfloor k/\lambda_2^{i_t} \rfloor}\varphi_k^{i_t}$

end for

Output: $\boldsymbol{\Phi} = [\boldsymbol{\phi}^0, \boldsymbol{\phi}^1, \ldots \boldsymbol{\phi}^{S-1}]^T$

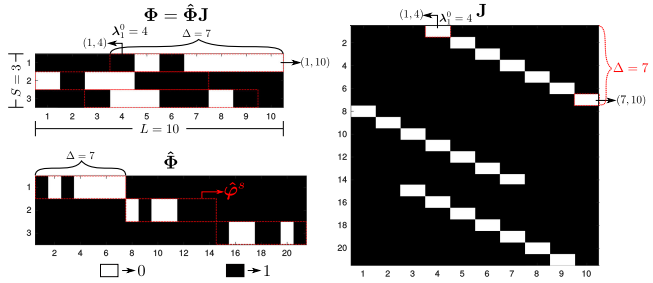

Fig. 5. Example of $\boldsymbol{\Phi}$, $\hat{\boldsymbol{\Phi}}$ and $\mathbf{J}$ structure, using $S = 3$, $L = 10$ and $\Delta = 7$. Note that the first one-valued diagonal in $\mathbf{J}$ will place the first random vector, of length $\Delta = 7$, within columns 4–10 of $\boldsymbol{\Phi}$.

$\boldsymbol{\Phi}$ matrix orthogonality by minimizing $f(\boldsymbol{\Phi})$. Then, it is expected that, with high probability, the main angles between the projected subspaces are preserved, i.e., the separability among subspaces is preserved. In this section, it is shown that if the SSC ℓ_1 minimization program (described in the next section, see (15)) recovers a subspace-sparse solution for the spectral pixels $\mathbf{F}$, then it will also recover a subspace-sparse solution for the compressed pixels $\mathbf{Y}$.

First observe that the matrix $\boldsymbol{\Phi}$ can be decomposed as the product of two matrices as $\boldsymbol{\Phi} = \hat{\boldsymbol{\Phi}}\mathbf{J}$, where $\hat{\boldsymbol{\Phi}} \in \mathbb{R}^{S \times S\Delta}$ is a block rectangular matrix with vectors $\hat{\boldsymbol{\varphi}}^s \in \{0,1\}^\Delta$ placed in diagonal form, where $\hat{\boldsymbol{\varphi}}^s$ correspond to the values in the select bandwidth of $\boldsymbol{\varphi}^s$ for each s. Mathematically, the entries of the s-th row of $\hat{\boldsymbol{\Phi}}$ can be written as

$$\left(\hat{\boldsymbol{\Phi}}^s\right)_{\hat{r}} = \begin{cases} (\hat{\boldsymbol{\varphi}}^s)_{\hat{r} \bmod \Delta}, & \text{if } s = \lfloor \hat{r}/\Delta \rfloor \\ 0, & \text{otherwise,} \end{cases} \tag{12}$$

for $\hat{r} = 0, \ldots, S\Delta - 1$. Moreover, the matrix $\mathbf{J}$ has one-valued diagonals where each diagonal chooses the position of a specific vector $\hat{\boldsymbol{\varphi}}^s$ within the s-th row of $\boldsymbol{\Phi}$. Taking this into account, the entries of the $\hat{r}$-th row of $\mathbf{J}$ can be written as

$$(\mathbf{J}^{\hat{r}})_k = \begin{cases} 1, & \text{if } k = \lambda_1^{\lfloor \hat{r}/\Delta \rfloor} + \hat{r} \bmod \Delta \\ 0, & \text{otherwise,} \end{cases} \tag{13}$$

for $k = 0, \ldots, L-1$. Note that the position of one-valued diagonals in $\mathbf{J}$ is determined by the selected cutoff wavelengths $\lambda_1^{\lfloor \hat{r}/\Delta \rfloor}$ obtained using Algorithm 1. In particular Fig. 5 illustrates the structure of the matrices $\boldsymbol{\Phi}, \hat{\boldsymbol{\Phi}}$ and, $\mathbf{J} \in \mathbb{R}^{S\Delta \times L}$ for $S = 3, L = 10$ and $\Delta = 7$. Additionally, note that $\mathbf{J}$ can be viewed as a band-selection matrix which selects neighboring bands on the spectral pixel $\mathbf{f}_j$ and groups them in blocks, as $\mathbf{J}\mathbf{f}_j = \bar{\mathbf{f}}_j = [(\bar{\mathbf{f}}_j^0)^T, (\bar{\mathbf{f}}_j^1)^T, \ldots, (\bar{\mathbf{f}}_j^S)^T]$, with $\bar{\mathbf{f}}_j \in \mathbb{R}^{S\Delta}$ and $\bar{\mathbf{f}}_j^s \in \mathbb{R}^\Delta$, hence if $S\Delta \geq L$, then $\text{Rank}(\mathbf{F}) = \text{Rank}(\bar{\mathbf{F}})$, where $\bar{\mathbf{F}} = \mathbf{J}\mathbf{F}$.

In [5], the authors of SSC provide recovery conditions under which, for data points $\mathbf{f}_j$ that lie in a union of linear subspaces, the optimization program in (15) recovers subspace-sparse representation of the data, where nonzero elements correspond to points belonging to the same subspace. Particularly, denote $\bar{\mathbf{F}}_d$ as the matrix containing all data points $\bar{\mathbf{f}}_j$ from the subspace $\mathcal{S}_d$ with dimension Q_d. Similarly, denote $\bar{\mathbf{F}}_{-d}$ as the matrix containing data points in all subspaces except $\mathcal{S}_d$. Further, let $\mathbb{W}_d$ be the set of all full-rank submatrices $\tilde{\mathbf{F}}_d \in \mathbb{R}^{S\Delta \times Q_d}$ of $\bar{\mathbf{F}}_d$. From

[5, Theorem 3], if the condition

$$\max_{\tilde{\mathbf{F}}_d \in \mathbb{W}_d} \tilde{\sigma}_{Q_d}(\tilde{\mathbf{F}}_d) > \sqrt{Q_d} \max_{d \neq d'} \cos(\theta_{d,d'}), \tag{14}$$

for a column-normalized data holds, then for every $\bar{\mathbf{f}}_j$ in the subspace $\mathcal{S}_d$, the ℓ_1-minimization in (15) recovers a subspace-sparse solution. In (14), $\theta_{j,j'}$ is the first principal angle between $\mathcal{S}_d$ and $\mathcal{S}_{d'}$ and $\tilde{\sigma}_{Q_d}(\tilde{\mathbf{F}}_d) = 1/\|(\tilde{\mathbf{F}}_d^T \tilde{\mathbf{F}}_d)^{-1} \tilde{\mathbf{F}}_d^T\|_{2,2}$ denotes the Q_d-th largest singular value of $\tilde{\mathbf{F}}_d$. Now, the following theorem proves that the condition in (14) still holds for $\tilde{\mathbf{Y}}_d = \hat{\mathbf{\Phi}} \tilde{\mathbf{F}}_d$.

Theorem III.2: Consider a collection of measurements given by (2) drawn from n subspace $\{\mathcal{S}\}_{d=1}^n$ of dimensions $\{Q_d\}_{d=1}^n$. Let $\mathbb{W}_d$ be the set of all full-rank submatrices of the column-normalized data $\tilde{\mathbf{F}}_d \in \mathbb{R}^{S\Delta \times Q_d}$, and $\hat{\mathbf{\Phi}}$ the decomposition given by (12) of the matrix obtained from the Algorithm 1. If the number of shots $S \geq Q_d$ then, the following condition holds

$$\tilde{\sigma}_{Q_d}(\tilde{\mathbf{Y}}_d) \geq \frac{(1-\rho)}{k} \sqrt{Q_d} \max_{d \neq d'} \cos(\theta_{d,d'}),$$

for some $\rho \in (0,1)$ with probability at least $1 - 2e^{-\rho^2/2}$, where $\tilde{\mathbf{Y}}_d = \hat{\mathbf{\Phi}} \tilde{\mathbf{F}}_d$ is a full column rank matrix and $k \geq 1$ is its condition number.

Proof: The proof can be found in Appendix B (see Supplementary Material).

Notice that Theorem III.2 essentially establishes that the sparse solution can be obtained from the compressed measurements with high probability. In addition, note that because one of the $\hat{\mathbf{\Phi}}$ design criteria is to preserve the similarity (a.k.a, cosine of the angle between two vectors), it is expected that $\theta_{j,j'}$ is also preserved. Therefore, it is possible to infer that if the condition in (14) holds for the spectral pixels $\mathbf{F}$, it will also hold for the compressed pixels $\mathbf{Y}$ with high probability.

IV. Compressed Sparse Subspace Clustering With Spatial Regularizer

Assuming that compressed pixels of the same land-cover class lie in one independent subspace, subspace clustering methods can be used in order to separate them into the same group or cluster. In particular, SSC builds the similarity matrix, which describes the data points membership, by finding a sparse representation for each compressed pixel whose nonzero elements ideally correspond to points from the same subspace. Given the designed matrix $\mathbf{\Phi}$ and the compressive measurements $\mathbf{Y} = \mathbf{\Phi F}$, the SSC sparse representation model is formulated as the following optimization problem:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{R}} \quad & \|\mathbf{Z}\|_1 + \frac{\lambda}{2}\|\mathbf{R}\|_F^2 \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{YZ} + \mathbf{R},\ \text{diag}(\mathbf{Z}) = 0,\ \mathbf{Z}^T\mathbf{1} = \mathbf{1}, \end{aligned} \tag{15}$$

where $\mathbf{1}$ is a one-valued vector, $\mathbf{Z} \in \mathbb{R}^{MN \times MN}$ refers to the representation coefficient matrix and the ℓ_1-norm regularization in this formulation suggests that a sparse representation of a data point finds points from the same subspace. The matrix $\mathbf{R}$ stands for the representation error, and the regularization parameter λ for the sparsity trade-off. The constraint $\text{diag}(\mathbf{Z}) = 0$ is used to eliminate the trivial solution of writing a point as an affine combination of itself and the constraint $\mathbf{Z}^T\mathbf{1} = \mathbf{1}$ ensures that it is a case of affine subspaces [4], [5].

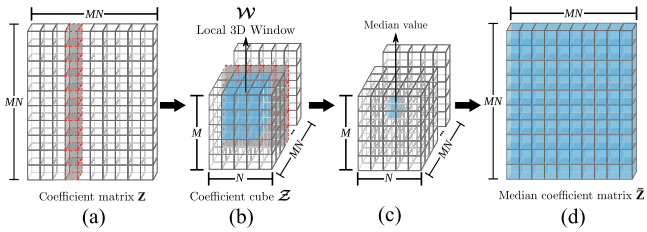

Fig. 6. Visual representation of the median filter step, (a) Sparse Coefficient matrix $\mathbf{Z}$, then it is reshaped as in (b) and a median filter is applied to obtain the new values (c) and finally it is reshaped to its initial size (d).

Taking into account that neighboring pixels in a spectral image usually consist of similar materials, a smoothing filter can be applied to the sparse coefficient matrix, in order to reduce the representation error, being able to extract more information from the data [3]. Specifically, the smoothing filter will reduce the noise trying to assign the same representation value to neighboring pixels. In this work, such spatial information is effectively incorporated into the similarity matrix by first rearranging the 2D sparse coefficient matrix $\mathbf{Z} \in \mathbb{R}^{MN \times MN}$ into a 3D cube $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{M \times N \times MN}$, treating each coefficient vector as a "pixel" in the cube. Then, unlike previous work in [3] which performs a 2D average filtering in each slice of the cube $\boldsymbol{\mathcal{Z}}$, we propose to perform the smooth filtering using a 3D median filter with a 3D moving window $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{3\times3\times3}$. Specifically, $\boldsymbol{\mathcal{W}}$ is moved through $\boldsymbol{\mathcal{Z}}$, on each band, pixel by pixel and replacing each value with the median value of neighboring pixels. Finally, the filtered cube $\boldsymbol{\mathcal{Z}}$ is rearranged back to form the matrix $\bar{\mathbf{Z}} \in \mathbb{R}^{MN \times MN}$. Using $\bar{\mathbf{Z}}$, a regularization term $\|\mathbf{Z} - \bar{\mathbf{Z}}\|_F^2$ is incorporated in the original SSC optimization program shown in (15). Then, the problem of finding a sparse representation coefficient matrix exploiting the spatial information of the scene is formulated as the following optimization program

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{R},\bar{\mathbf{Z}}} \quad & \|\mathbf{Z}\|_1 + \frac{\lambda}{2}\|\mathbf{R}\|_F^2 + \frac{\alpha}{2}\|\mathbf{Z} - \bar{\mathbf{Z}}\|_F^2 \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{YZ} + \mathbf{R},\ \text{diag}(\mathbf{Z}) = 0,\ \mathbf{Z}^T\mathbf{1} = \mathbf{1}, \end{aligned} \tag{16}$$

where α is a regularization parameter denoting the weight of the spatial information in the subspace clustering algorithm. In the subsequent sections, we also refer to the optimization problem in (16) as S-SSC. The minimization in (16) can be efficiently solved by the alternating direction method of multipliers (ADMM), which is described in detail in the Appendix C (see Supplementary Material). The solution of (16) corresponds to subspace-sparse representation of the data points, which is used by spectral clustering (SC) to infer the clustering of the data. Specifically, the clustering result is obtained by applying SC to the Laplacian matrix induced by the similarity matrix $\mathbf{W} \in \mathbb{R}^{MN \times MN}$ which is defined as $\mathbf{W} = |\mathbf{Z}| + |\mathbf{Z}|^T$ [4], [5]. The complete CSI subspace clustering algorithm (CSI-SSC) is summarized in Algorithm 2.

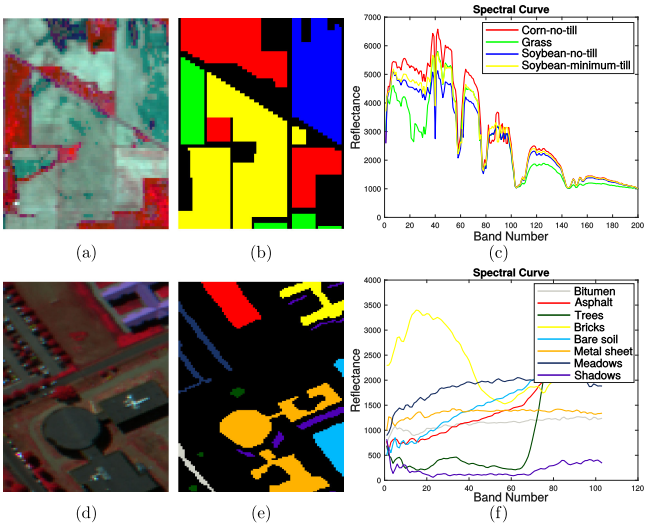

Fig. 7. False-color images, ground truth images and spectral curves of each land-cover classes for AVIRIS Indian Pines (a)–(c) and ROSIS Pavia University (d)–(f), respectively.

Algorithm 2: Compressive Spectral Subspace Clustering.

Input: A set of CSI measurements acquired as $\mathbf{Y} = \mathbf{\Phi F}$, where the coding pattern matrix $\mathbf{\Phi}$ is obtained with Algorithm 1.

1: Solve the sparse optimization problem in (16).
2: Normalize the columns of $\mathbf{Z}$ as $\mathbf{z}_j \leftarrow \frac{\mathbf{z}_j}{\|\mathbf{z}_j\|_\infty}$
3: Form a similarity graph representing the data points. Set the weights on the edges between the nodes as $\mathbf{W} = |\mathbf{Z}| + |\mathbf{Z}|^T$.
4: Apply SC [31] to the similarity graph.

Output: Segmentation of the data: $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\ell$

V. Experimental Results

The proposed compressed spectral image clustering approach was tested on two real hyperspectral data sets, with different imaging environments. The **Indian Pines** hyperspectral data set was acquired by the AVIRIS sensor from the Northwestern Indian Pines test site in June 1992. The spatial dimensions of this image are 145×145 pixels. A total of 20 water absorption and noisy bands were removed from the original 220 bands, leaving 200 spectral features for the experiment [3]. Considering the computational efficiency, a subimage with size 70×70 pixels, which includes four main land-cover classes: corn-no-till, grass, soybeans-no-till, and soybeans-minimum-till, was used in the experiments.

The second scene, **University of Pavia**, was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy. The spatial dimensions of the image are 610×340 pixels, with 103 bands used in the experiments. A typical area for the test data with a size of 140×80 pixels, containing eight main land-cover classes: Bitumen, asphalt, trees, bricks, bare soil, metal sheet, meadows and shadows, was used.

Clustering of these selected images is a challenging task because the spectral signatures of the land-cover classes are very similar and some of the spectral curves are mixed, as observed in Fig 7. (c) and (f). A false-color image and the ground truth for the Indian Pines and Pavia University are also provided in Fig. 7(a)–(b) and (d)–(e), respectively. In the experiments, the number of clusters was set as a manual input for the subspace clustering algorithm. Furthermore, the parameters of the algorithm were manually adjusted. Specifically, the regularization parameter λ, which acts as the trade-off between the sparsity of the coefficient matrix and the magnitude of the noise, was set using the following formulation [5]:

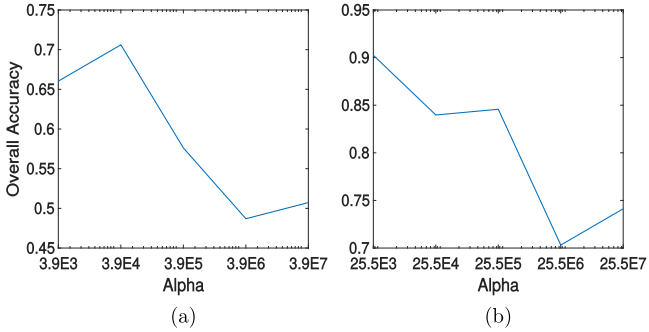

Fig. 8. Analysis of parameter α: Change in the overall accuracy with various values of α. (a) Indian Pines image, (b) University of Pavia image.

$$\lambda = \frac{\beta}{\gamma}, \quad \gamma = \min_j \max_{j \neq j'} |\mathbf{y}_{j'}^T \mathbf{y}_j|, \tag{17}$$

where β is the adjustment coefficient, γ is a parameter related to the data set, which can be explicitly determined, and $\mathbf{y}_j, \mathbf{y}_{j'}$ are columns of $\mathbf{Y}$. The regularization parameter α in (16) denotes the weight of the spatial information in S-SSC. In order to analyze the sensitivity of α, experiments for each data set were conducted. In these experiments, the coding patterns $\mathbf{\Phi}$ were generated using Algorithm 1 with $\Delta = 20$ and $S = 25$. Further, white Gaussian noise with a signal-to-noise ratio (SNR) of 25 dB was added to the acquired compressed measurements, simulating the CSI acquisition system noise.

The change in the overall accuracy of the proposed S-SSC algorithm corresponding to different α values, with other parameters fixed, is shown in Fig. 8. It can be seen that the precision changes significantly with different values of α, which suggests that the spatial information plays a very important role in the clustering process. For simplicity this parameter was fixed for all experiments: $\alpha = 3.9 \times 10^4$ for the Indian Pines image and $\alpha = 25.5 \times 10^5$ for the University of Pavia image. Similarly, the parameter λ for all the experiments is calculated using (17) with $\beta = 1000$. Since the structure of the acquired compressed measurements is determined by the generated coding pattern, the γ parameter is determined at the beginning of each experiment using (17).

In the next experiments, the random coding patterns are generated from a Bernoulli distribution $\mathbf{\Phi} \sim B_e(p)$, using (3) with $p = \Delta/L$ in order to use a similar transmittance (the number of nonzero elements) to the designed coding patterns. The results presented throughout the document are the average

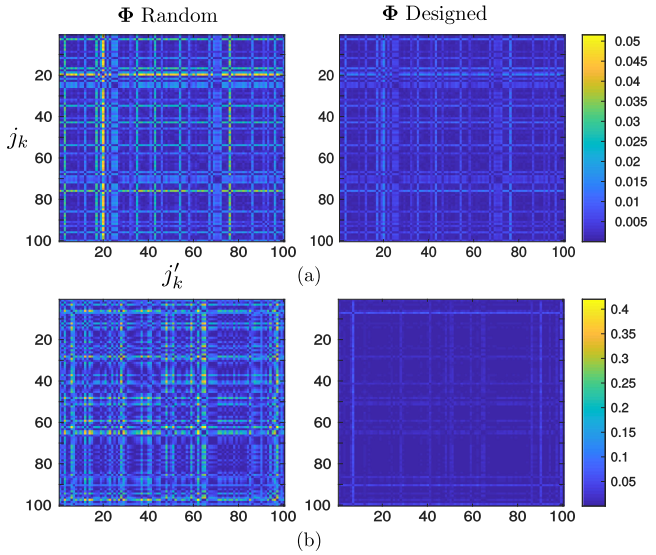

Fig. 9. Absolute error between the spectral signatures similarities and the compressed measurements similarities acquired with the random and designed coding patterns for (a) Indian Pines and (b) Pavia University datasets.

of five experiments, each with a different set of coding patterns.[1]

A. Similarity Preservation

In this experiment, the performance of the proposed coding pattern design is tested. Specifically, this experiment is intended to show how well the similarity between two spectral signatures is approximately preserved when the designed coding patterns are used. For this experiment, 100 spectral pixels from each image are randomly chosen and then are compressed using both random and the designed coding pattern matrix $\mathbf{\Phi}$, generated with parameters $\Delta = 20$ and $S = 25$. Then, all the spectral signatures and their compressed versions are normalized to have unit ℓ_2 norm, i.e., $\|\mathbf{f}\|_2 = 1$ and $\|\mathbf{y}\|_2 = 1$. Using the definition of similarity, presented in (6), the absolute error is calculated as

$$\left|\text{sim}(\mathbf{f}_{j_k}, \mathbf{f}_{j'_k}) - \text{sim}(\mathbf{y}_{j_k}, \mathbf{y}_{j'_k})\right| = \left|\mathbf{f}_{j_k}^T \mathbf{f}_{j'_k} - \mathbf{y}_{j_k}^T \mathbf{y}_{j'_k}\right|, \quad (18)$$

where $j, j' \in \{0, 1, \ldots, MN\}$, index a spectral signature chosen at random among MN possibilities, and $k = 1, \ldots, 100$. Fig. 9 shows the obtained results for the two spectral images. As observed, the absolute errors obtained with the designed coding patterns are significantly smaller than those obtained with a random-designed matrix $\mathbf{\Phi}$. Therefore, the proposed coding pattern design approximately preserves the similarities among the spectral signatures after the scene projection.

B. Noise Analysis

It is important to note that the acquired measurements described by (2) are noise free. However, in real CSI architectures, the acquired compressed measurements are contaminated with noise due to the physical limitations of the sensor. Therefore,

[1]A MatLab implementation of Algorithm 1 and some experiments presented in the next sections can be found at https://git.io/vpvZS

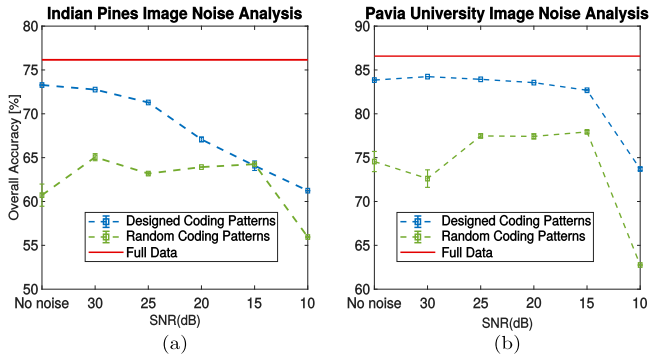

Fig. 10. Overall clustering accuracy as a function of the aggregated noise using the two types of coding patterns.

(2) should be rewritten as $\mathbf{Y} = \mathbf{\Phi F} + \mathbf{\Omega}$, where $\mathbf{\Omega} \sim N(0, \sigma^2)$ represents the noise of the system. In order to analyze the impact of noise, different experiments varying SNR are performed. Fig. 10 presents the classification accuracy results obtained with the designed and the random coding patterns for the two hyperspectral data sets. Additionally, the overall clustering accuracy achieved when using the spectral image data cube (Full data) as input for the optimization problem in (16) is shown as reference. The designed patterns are generated with the fixed parameters $\Delta = 20$ and $S = 25$.

The results show that the proposed coding pattern design outperforms the random-generated patterns even when white Gaussian noise is added to the CSI measurements. As expected, when SNR decreases the clustering overall accuracy is affected. However, the accuracy curve, obtained with the proposed coding patterns, achieves a slow decrease rate in comparison with the random coding patterns.

C. Analysis of the Coding Pattern Design Parameters

The parameters S and Δ determine the structure of the proposed coding pattern design. In order to analyze how the accuracy is affected by those parameters, experiments for each pair (Δ, S) were performed varying each parameter. In this experiment, the noise added to the system was fixed to 25 dB of SNR. Fig. 11(a) and (c) present the obtained overall accuracy and Fig. 11(b) and (d) show the variance of the obtained accuracy for each spectral image, respectively.

Note that the number of shots S determines the ambient space of the projected vectors. It can be shown that, when the ambient dimension after random projection is sufficiently large, the distance between two subspaces almost remains unchanged after random projection [26]. This behavior is observed in the presented results where increasing the measurement shots leads to better classification accuracies. In addition, notice that the sparsity of the projection matrix $\mathbf{\Phi}$ and the selected bandwidth Δ are directly related. Specifically, denote the sparsity of the matrix $\mathbf{\Phi}$ as $K = 1/\kappa$, where $\kappa \approx \frac{2L}{\Delta}$. As described in [32], sparse random projections are robust for $\kappa \approx \sqrt{L}$. However, as κ increases (decreasing Δ), variances for sparse random projections will also increase and large errors are expected. On the other hand, by decreasing κ (increasing Δ), the entries outside

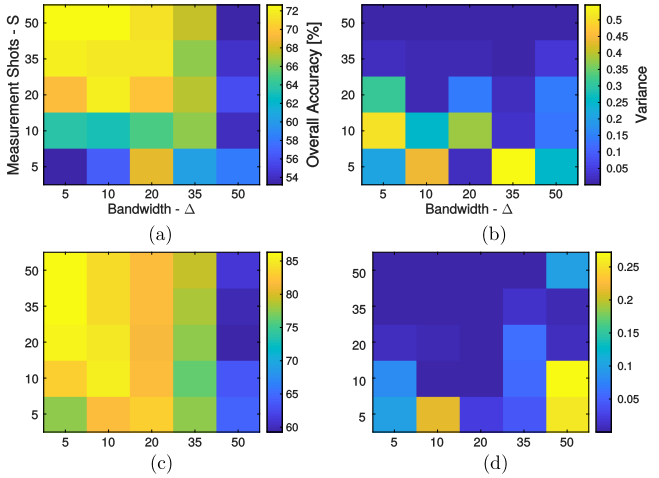

Fig. 11. Analysis of the coded aperture design parameters. (a) and (c) show the overall accuracy varying the bandwidth Δ and the measurement shots S for the hyperspectral datasets Indian Pines and Pavia University, respectively. (b) and (d) present the variance of the obtained accuracy.

diagonal of $\boldsymbol{\Theta}$ will increase leading to significant distortions in the acquired compressed measurements. In all the other experiments we fix $S = 25$ and $\Delta = 10$ since, as observed in Fig. 11, the OA and variance stabilize when $S > 20$ in both hyperspectral datasets. Note that $S = 25$ yields a compression of 87.5 % and 75.7% for Indian Pines and University of Pavia data set, respectively.

D. Visual Maps and Quantitative Results

In order to validate the clustering performance of the proposed coding pattern design, cluster maps and quantitative results are presented for the two hyperspectral scenes. In all the experiments, the coding patterns were generated for parameters $\Delta = 20$ and $S = 25$. Further, white Gaussian noise with 25 dB of SNR was added to the acquired compressed measurements. The S-SSC algorithm described in section IV was used to perform the clustering on both compressed measurements and the complete spectral data cube (Full-data). Additionally, the results obtained with the original sparse subspace clustering algorithm (SSC), when the complete spectral data cube is used as input (Full-data-SSC), are also shown for comparison purposes. Note that, in the experiments, we refer to "Proposed-design" as the clustering results obtained when the S-SSC algorithm is directly applied on the compressed measurements acquired with the proposed pattern designs. Similarly, "Random-design" corresponds to the results obtained when the S-SSC is directly applied on the compressed measurements acquired with the random patterns. Finally, we refer to "Full-data" when using the proposed S-SSC method with the complete spectral image, i.e., no compression was performed.

Fig. 12 presents the obtained visual clustering results on Indian Pines. The quantitative evaluations corresponding to the accuracy for each class, overall accuracy (OA), average accuracy(AA) and Kappa coefficients are shown in Table I, where all values are given in percentage. Similarly, Fig. 13 and Table II present the visual clustering results and quantitative evaluation

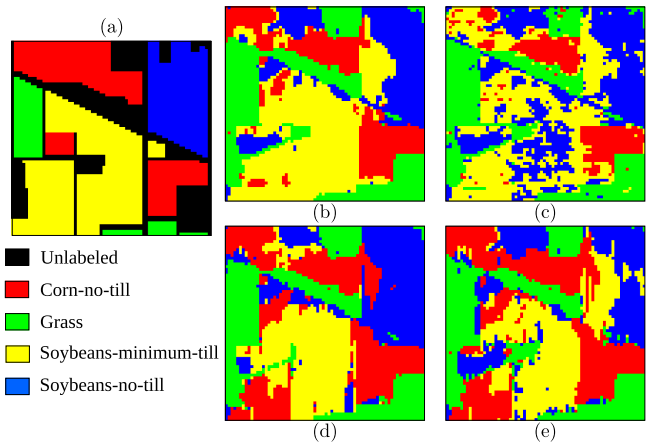

Fig. 12. Visual clustering results on AVIRIS Indian Pines image: (a) Ground truth. (b) Full-data, (c) Full-data-SSC, (d) Proposed-design and (e) Random-design.

TABLE I
QUANTITATIVE EVALUATION OF THE DIFFERENT CLUSTERING RESULTS FOR THE AVIRIS INDIAN PINES IMAGE

Class	Random-design	Proposed-design	Full-data-SSC	Full-data
Corn-no-till	**73.13**	70.45	48.96	66.77
Grass	95.25	**100**	98.60	**100**
Soybeans-no-till	52.87	**88.80**	70.63	69.54
Soybeans-minimun-till	55.29	60.52	59.23	**80.05**
OA	63.83	73.07	62.62	**76.16**
AA	69.14	**79.94**	69.35	79.09
Kappa	49.26	62.65	47.58	**65.89**

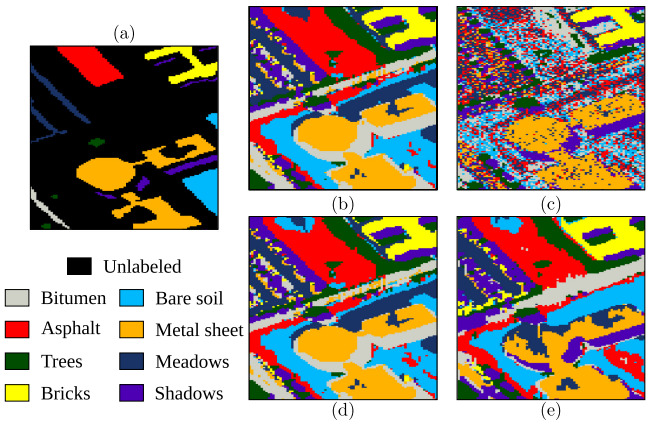

Fig. 13. Visual clustering results on ROSIS Pavia University image. (a) Ground truth. (b) Full-data, (c) Full-data-SSC, (d) Proposed-design and (e) Random-design.

on the Pavia University, respectively. In the tables, the optimal value of each row is shown in bold and the second-best result is underlined. From Tables I and II, it can be clearly observed that the proposed clustering approach (S-SSC), using the proposed coding patterns, provides comparable results to applying clustering directly on the full spectral data cube. Furthermore, it is observed from the visual clustering maps that, although the reconstruction is avoided, the results obtained with the proposed coding patterns are very similar to the results obtained with the Full-data-S-SSC. This behavior was expected

TABLE II
QUANTITATIVE EVALUATION OF THE DIFFERENT CLUSTERING RESULTS WITH THE AVIRIS PAVIA UNIVERSITY IMAGE

Class	Random-design	Proposed-design	Full-data-SSC	Full-data
Bitumen	18.60	88.37	0	**90.70**
Asphalt	71.37	67.25	33.84	**80.26**
Trees	90.38	88.46	**100**	90.38
Bricks	**100**	99.68	99.68	99.68
Bare Soil	46.78	61.40	36.26	**66.67**
Metal sheet	82.90	**97.73**	91.00	**97.73**
Meadows	91.16	**100**	55.02	**100**
Shadows	**99.48**	24.35	98.45	24.35
OA	78.72	83.81	71.45	**86.58**
AA	75.09	78.41	64.28	**81.22**
Kappa	72.63	78.89	62.95	**82.50**

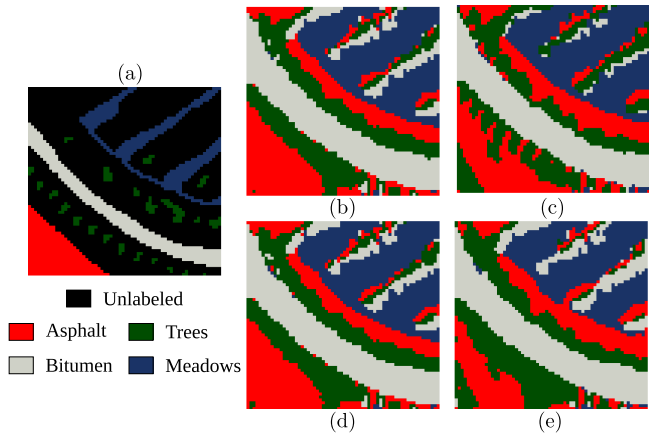

Fig. 14. Visual clustering results on a 64 × 64 region of Pavia University. (a) Ground truth. (b), (c) clustering results on reconstructed images using the random and proposed coding patterns, respectively. (d), (e) clustering results by directly using the compressed measurements acquired with the random and proposed coding patterns respectively.

since the proposed coding patterns approximately preserve the similarities among spectral pixels, as it was theoretically shown in Section III-C and experimentally verified in Section V-A.

E. Clustering Time and Spectral Image Reconstruction

In this section, the effectiveness of applying clustering directly on the compressed domain is evaluated. For this purpose, a subimage of ROSIS Pavia university dataset with the size of 64 × 64 pixels, which contains four land-cover classes: asphalt, meadows, trees, and bricks was used, see Fig. 14(a). CSI measurements were acquired using the random and proposed coding patterns. Then, 300 iterations of the gradient projection for sparse reconstruction algorithm (GPSR) [18] were used to reconstruct the underlying spectral scene. Fig. 14 presents the obtained visual clustering results on the selected subimage. In Table III, the time, quality of the reconstruction and the result of clustering for the types of coding patterns, are shown. From this table it is possible to observe that the proposed coding pattern design shows a gain of up to 6 dB in terms of peak signal-to-noise ratio (PSNR) in comparison with the random patterns. Further, it can be seen that the designed coding pattern not only improves the reconstruction quality but also the clustering result for the two scenarios, i.e., when the subspace clustering is applied after reconstruction and when it is applied directly on the compressed data. Note that, the total clustering time of the reconstructions is greater than the time of directly applying clustering on the compressed measurements because it takes into account the reconstruction time. In the simulations, when using the proposed coding patterns, the total clustering time of the reconstructed spectral image was 143.66 [s], while applying clustering directly on the compressed measurements takes only 103.70 [s], obtaining very similar classification results.

TABLE III
TIME AND CLASSIFICATION ACCURACY WHEN CLUSTERING THE RECONSTRUCTED SPECTRAL IMAGE AND THE CSI MEASUREMENTS

	Reconstruction		**No Reconstruction**	
	Random Patterns	Proposed Patterns	Random Patterns	Proposed Patterns
CSI Recovery				
PSNR [dB]	27.92	**34.38**	-	-
Time [s]	**28.56**	26.23	-	-
Subspace Clustering				
Asphalt	**100**	**100**	**100**	**100**
Meadows	71.65	**99.07**	69.78	97.76
Trees	56.05	63.06	18.85	**65.35**
Bricks	88.04	**99.24**	98.68	98.83
OA	83.43	**94.88**	81.70	94.65
AA	78.94	90.34	71.89	**90.45**
Kappa	77.35	**92.90**	74.74	92.60
Solving (16) [s]	16.62	**15.78**	16.28	10.15
SC Time [s]	118.25	**101.65**	106.11	93.55
Total Time [s]	163.43	143.66	122.39	**103.70**

F. Comparison With Other CSI Sensing Approaches

In this section, the proposed CSI sensing design was compared with two different CSI approaches described in [9] and [10]. In order to perform such comparison, some considerations were taken into account. Specifically, [9] uses the measurement matrices $\mathbf{A} \in \mathbb{R}^{S_1 \times L}$ and $\mathbf{B}_k \in \mathbb{R}^{S_2 \times L}$ whose elements are independently drawn at random from Gaussian, Rademacher, or Bernoulli distribution. The measurements $\mathbf{Y}_a = \mathbf{AF}$ obtain S_1 projections per pixel and $\mathbf{y}_{b,k} = \mathbf{B}_k \mathbf{f}_{i_k}$ for $k = 1, \ldots, n_v$, obtains S_2 measurements per sample pixel n_v. Because the main goal of this paper is to apply clustering directly from compressive measurements and, as explained in section II, the similarity among two spectral pixels decrease when they are encoded (projected) differently, we only considered $\mathbf{Y}_a$ with $S_1 = S$ in order to apply clustering such that fair comparisons were performed. Furthermore, Gaussian distribution was chosen, since $\mathbf{\Phi}$ with 0, 1 entries was already considered in previous sections.

To perform a comparison with the work in [10], which assumes that only a few voxels of the image are known (incomplete hyperspectral imagery), we considered that all spectral signatures had the same number of missing voxels in order to be able to apply subspace clustering directly on the compressed measurements. Then, two different scenarios were evaluated for incomplete images: the first is when some voxels are randomly removed in each spectral signature; the second is when complete spectral bands are removed from the image i.e. $\mathbf{y}_i = \mathbf{A}_i \mathbf{f}_i$,

TABLE IV
COMPARISON OF CLUSTERING RESULTS FOR THE AVIRIS INDIAN PINES IMAGE WHEN USING THE PROPOSED CSI SENSING DESIGN AND OTHER CSI SENSING APPROACHES

Class	Gaussian	RRV	RRB	Proposed-design
Corn-no-till	27.36	45.01	52.76	**70.45**
Grass	57.21	86.70	**100**	**100**
Soybeans-no-till	34.18	42.76	72.65	**88.80**
Soybeans-minimun-till	47.48	36.31	53.66	**60.52**
OA	40.06	45.14	61.95	**73.07**
AA	41.56	52.70	69.77	**79.94**
Kappa	16.34	24.02	47.43	**62.65**

TABLE V
COMPARISON OF CLUSTERING RESULTS FOR THE AVIRIS PAVIA UNIVERSITY IMAGE WHEN USING THE PROPOSED CSI SENSING DESIGN AND OTHER CSI SENSING APPROACHES

Class	Gaussian	RRV	RRB	Proposed-design
Bitumen	57.91	12.10	48.37	**88.37**
Asphalt	43.99	68.63	**72.62**	67.25
Trees	41.54	84.62	87.69	**88.46**
Bricks	96.66	99.29	99.55	**99.68**
Bare Soil	34.15	46.08	34.85	**61.40**
Metal sheet	68.82	97.47	97.67	**97.73**
Meadows	88.92	99.12	99.76	**100**
Shadows	**100**	75.65	44.87	24.35
OA	66.78	82.73	81.66	**83.81**
AA	66.50	71.36	73.17	**78.41**
Kappa	59.22	77.35	76.09	**78.89**

where $\mathbf{A}_i \in \{0,1\}^{S\times L}$ and the rows of $\mathbf{A}_i$ are all zero except one corresponding to the observed voxels. In the results, we refer to these scenarios as *randomly-removed-voxels (RRV)* and *randomly-removed-bands (RRB)*, respectively.

In the experiments we used $S = 25$ and white Gaussian noise with 25 dB of SNR was added to the measurements. The results shown in Table IV and V are the average of five experiments. All the clustering results were obtained using the proposed subspace clustering method (S-SSC). At each experiment, different measurements were obtained. From Tables IV and V it can be clearly observed that the adopted 3D-CASSI sensing scheme, using the designed coding patterns, outperforms the CSI compression schemes proposed in [9], [10].

VI. CONCLUSION

This work presented a coding pattern design, which aims at preserving the spectral signatures separability as much as possible after the scene projection. This design allows to improve clustering results on the compressive domain. Furthermore, a subspace clustering algorithm which takes into account the spatial information of the spectral images in order to correct the representation bias and obtain a more accurate representation coefficient matrix was also developed. The coding pattern design and the spectral image subspace clustering approach were validated trough several experiments. The Indian pines hyperspectral remote sensing scene, from the AVIRIS sensor, and the Pavia University, from the ROSIS sensor, were used in the experiments. In general, the results show that performing the clustering directly with the compressed measurements provides similar accuracy results, in a lesser time, compared with those provided by performing the clustering on the full 3D spectral image, when a properly designed coding pattern was used. Particularly, a maximum difference of just 4% in terms of overall accuracy was observed when comparing the clustering results obtained by the full 3D data with those achieved using CSI measurements acquired with the proposed coding pattern design.

REFERENCES

[1] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.

[2] G. A. Shaw and H.-H. K. Burke, "Spectral imaging for remote sensing," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 3–28, 2003.

[3] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.

[4] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 2790–2797.

[5] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[6] J. Bacca, C. Hinojosa, and H. Arguello, "Kernel sparse subspace clustering with total variation denoising for hyperspectral remote sensing images," in *Proc. Math. Imag.*, 2017, paper MTu4C–5.

[7] W. He, H. Zhang, L. Zhang, W. Philips, and W. Liao, "Weighted sparse graph based dimensionality reduction for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 686–690, May 2016.

[8] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 105–115, Jan. 2014.

[9] G. Martín and J. M. Bioucas-Dias, "Hyperspectral blind reconstruction from random spectral projections," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2390–2399, Jun. 2016.

[10] Z. Xing, M. Zhou, A. Castrodad, G. Sapiro, and L. Carin, "Dictionary learning for noisy and incomplete hyperspectral images," *SIAM J. Imag. Sci.*, vol. 5, no. 1, pp. 33–56, 2012. [Online]. Available: https://doi.org/10.1137/110837486

[11] X. Cao *et al.*, "Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world," *IEEE Signal Process. Mag.*, vol. 33, no. 5, pp. 95–108, Sep. 2016.

[12] H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1896–1908, Apr. 2014.

[13] C. V. Correa, C. Hinojosa, G. R. Arce, and H. Arguello, "Multiple snapshot colored compressive spectral imager," *Opt. Eng.*, vol. 56, no. 4, 2016, Art. no. 041309.

[14] M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[15] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, 2014, Art. no. 233.

[16] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," *Optics Lett.*, vol. 39, no. 7, pp. 2044–2047, 2014.

[17] X. Mao and Y. Gu, "Compressed subspace clustering: A case study," in *Proc. Global Conf. Signal Inf. Process.*, 2014, pp. 453–457.

[18] M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586–597, Dec. 2007.

[19] J. Bacca, H. Vargas, and H. Arguello, "A constrained formulation for compressive spectral image reconstruction using linear mixture models," in *Proc. Conf. IEEE 7th Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, 2017, pp. 1–5.

[20] M. Soltanolkotabi *et al.*, "A geometric analysis of subspace clustering with outliers," *Ann. Statist.*, vol. 40, no. 4, pp. 2195–2238, 2012.

[21] R. Heckel, M. Tschannen, and H. Bölcskei, "Dimensionality-reduced subspace clustering," *Inf. Inf., A J. IMA*, vol. 6, no. 3, pp. 246–283, 2017.
[22] M. Tschannen and H. Bölcskei, "Noisy subspace clustering via matching pursuits," *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4081–4104, Jun. 2018.
[23] Y. Wang, Y.-X. Wang, and A. Singh, "A theoretical analysis of noisy sparse subspace clustering on dimensionality-reduced data," 2016, arXiv:1610.07650.
[24] S. S. Vempala, *The Random Projection Method*, vol. 65. Providence, RI, USA: Amer. Math. Soc., 2005.
[25] R. Calderbank, S. Jafarpour, and R. Schapire, "Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain," 2009. [Online]. Available: http://dsp.rice.edu/files/cs/cl.pdf
[26] G. Li and Y. Gu, "Restricted isometry property of Gaussian random projection for finite set of subspaces," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1705–1720, 2018.
[27] Y. Jiao, G. Li, and Y. Gu, "Principal angles preserving property of Gaussian random projection for subspaces," in *Proc. IEEE Global Conf. Signal Inf. Process.*, 2017, pp. 318–322.
[28] H. Arguello and G. Arce, "Rank minimization code aperture design for spectrally selective compressive imaging," *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 941–954, Mar. 2013.
[29] D. Gu, A. R. Gillespie, A. B. Kahle, and F. D. Palluconi, "Autonomous atmospheric compensation (AAC) of high resolution hyperspectral thermal infrared remote-sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 6, pp. 2557–2570, Nov. 2000.
[30] S. Kaski, "Dimensionality reduction by random mapping: Fast similarity computation for clustering," in *Proc. Conf. IEEE Joint Conf. Neural Netw. Proc.*, 1998, vol. 1, pp. 413–418.
[31] A. Y. Ng *et al.*, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, vol. 2, pp. 849–856.
[32] P. Li, T. J. Hastie, and K. W. Church, "Very sparse random projections," in *Proc. Conf. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 287–296.

Carlos Hinojosa (S'17) received the B.Sc. and M.Sc. degrees in computer science from Universidad Industrial de Santander, Bucaramanga, Colombia, in 2015 and 2018, respectively. He is currently working toward the Ph.D. degree in computer science with Universidad Industrial de Santander. His main research areas include computational imaging, optical code design, compressive imaging, computer vision, and sparse representation.

Jorge Bacca (S'17) received the B.S. degree in computer science from the Universidad Industrial de Santander, Bucaramanga, Colombia, in 2017. He is currently working toward the Ph.D. degree with the Department of Computer Science, Universidad Industrial de Santander. His research interests include phase retrieval, compressive sensing, hyperspectral imaging, and spectral clustering.

Henry Arguello (SM'17) received the M.Sc. degree in electrical engineering from the Universidad Industrial de Santander, Bucaramanga, Colombia, and the Ph.D. degree in electrical and computer engineering from the University of Delaware, Newark, DE, USA. He is currently a Titular Professor with the School of Systems Engineering, Universidad Industrial de Santander. His research interests include statistical signal processing, high-dimensional signal coding, optical imaging, and compressed sensing.