CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGING SUBSPACE CLUSTERING

CARLOS ALBERTO HINOJOSA MONTERO

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FISICO-MECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
MAESTRÍA EN INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2018

CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGING SUBSPACE CLUSTERING

CARLOS ALBERTO HINOJOSA MONTERO

Submitted in partial fulfillment of the requirements for the Master degree in System Engineering and Informatics.

Director:
HENRY ARGUELLO FUENTES
Ph.D. in Electrical and Computer Engineering

UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2018

AGRADECIMIENTOS

El autor expresa su agradecimiento:

Al profesor Henry Arguello Fuentes, director del proyecto, por ser mi mentor, brindarme su apoyo y acompañamiento a lo largo de mi formación profesional.

A Ema Soledad Hinojosa, mi madre, por todo su esfuerzo y apoyo incondicional.

A Dios,
A mi madre,
A mis abuelos

Contents

List of Figures

Resumen

TITULO: DISEÑO DE APERTURAS CODIFICADAS PARA REALIZAR CLASIFICACIÓN NO SUPERVISADA EN IMÁGENES ESPECTRALES ADQUIRIDAS MEDIANTE LA TÉCNICA DE MUESTREO COMPRESIVO.(*)

AUTOR: CARLOS ALBERTO HINOJOSA MONTERO (**)

PALABRAS CLAVE: Imágenes espectrales, muestreo compresivo, diseño de patrones de codificación, adquisición compresiva de imágenes espectrales, clasificación no supervisada, clustering, sparse subspace clustering.

Los sistemas de adquisición de imágenes espectrales basados en la técnica de muestreo compresivo (CSI por su sigla en inglés), obtienen proyecciones codificadas de las firmas espectrales aplicando diferentes patrones de codificación. Una vez acquiridas las medidas comprimidas, el paso a seguir conmunmente consiste reconstruir la imagen espectral original. En la literatura de CSI, distintos trabajos se han centrado en mejorar la calidad de la reconstrucción mediante el diseño adecuado de los patrones de codificación. Sin embargo, la reconstrucción de la escena subyacente no es estrictamente necesaria para realizar distintas tareas de procesamiento. Por ejemplo, suponiendo que los píxeles espectrales asociados a clases diferentes conservan su desemejanza despues de ser comprimidos, los métodos de clasificación no supervisada (*clustering*) pueden aplicarse directamente con el objetivo de separar dichos píxeles en grupos diferentes o *clusters*, sin la necesidad de reconstuir la imagen espectral. En este trabajo, se propone un método para realizar *clustering* con medidas comprimidas obtenidas mediante CSI. En particular, se propone el diseño de un conjunto óptimo de patrones de codificación de manera que la desemejanza entre píxeles de diferentes clases se preserve después de la proyección de la escena. Luego, para realizar la clasificación de los datos comprimidos se propone un algoritmo de *clustering* basado en el modelo de *Sparse Subspace Clustering (SSC)*, el cual tiene en cuenta la correlación espacial existente entre firmas espectrales. Se realizaron diferentes simulaciones para validar el método de clasificación propuesto. En general, se obtuvo una precisión global del 73.07%, 80.12% y del 83.81% utilizando las imágenes espectrales de “Indian Pines”, “Salinas” y “Pavia University”, respectivamente, añadiendo 25 dB de relación señal / ruido a las mediciones comprimidas.

(*) Trabajo de Investigación.

(**) Facultad de Ingenierías Fisicomecánicas. Escuela de Ingeniería de Sistemas e Informática. Director, Henry Arguello Fuentes.

Abstract

TITLE: CODED APERTURE DESIGN FOR COMPRESSIVE SPECTRAL IMAGING SUBSPACE CLUSTERING. (*)

AUTHOR:CARLOS ALBERTO HINOJOSA MONTERO (**)

KEYWORDS: Spectral imaging, compressive spectral imaging, coding pattern design, clustering, sparse subspace clustering.

Compressive spectral imaging (CSI) is a spectral imaging approach which acquires compressed observations of the spectral signatures by applying different coding patterns on each spectral signature and then performing an spectral-wise integration. Once acquired the compressed measurements, the common subsequent procedure is the spectral image recovery. In CSI literature, several works has focused on improving the quality of reconstruction by properly designing a set of coding pattern. However, the recovery step is not actually necessary in many signal processing applications. For instance, assuming that spectral pixels from different class material preserve their dissimilarity after being compressed, the clustering methods can be straight applied to separate them into a different group or cluster, without the need of spectral image reconstruction. In this work, a subspace clustering approach for CSI measurements is proposed. In particular, an optimal set of coding patterns is proposed such that the dissimilarity between pixels from different classes is best preserved after the scene projection. Then, the CSI measurements classification is performed using a proposed clustering algorithm based on the sparse subspace clustering (SSC) model, which takes into account the spatial property of spectral images. Different simulations were made in order to validate the proposed CSI subspace clustering approach. In general, an overall accuracy of 73.07%, 80.12% and 83.81% were obtained using the Indian Pines, Salinas and Pavia University hyperspectral images respectively, when 25 dB of signal-to-noise ratio is added to the compressed measurements.

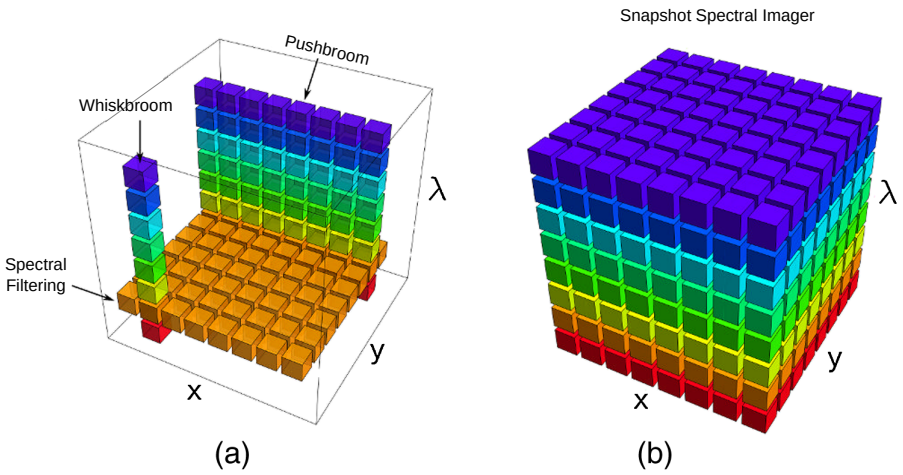
(*) Research Work.

(**) School of Physical-Mechanical Engineering. Department of Systems Engineering and Informatics. Advisor, Henry Arguello Fuentes.

INTRODUCTION

Spectral imaging captures the spectral information of a scene by sensing a large amount of spatial information at different electromagnetic radiation frequencies. Spectral images are regarded as three dimensional datasets or data cubes with two dimensions in the spatial domain (x, y) and one in the wavelength domain (λ). In general, traditional sensing techniques, as illustrated in Fig. 1 (a), 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. On the other hand, snapshot spectral imaging, depicted in Fig. A (b), captures the spatial and spectral information of a scene by mapping all the voxels(*) of the spectral data cube into different regions of the focal plane array (FPA(**)) [1, 2, 3]. In general, traditional techniques require to sense every single voxel of the 3D scene, hence huge storage capacities and computational resources are necessary in order to store and process such high dimensional images.

Figure A: Spectral imaging techniques. (a) Traditional scanning-based techniques using dispersion and spectral filtering. (b) Snapshot spectral imager.



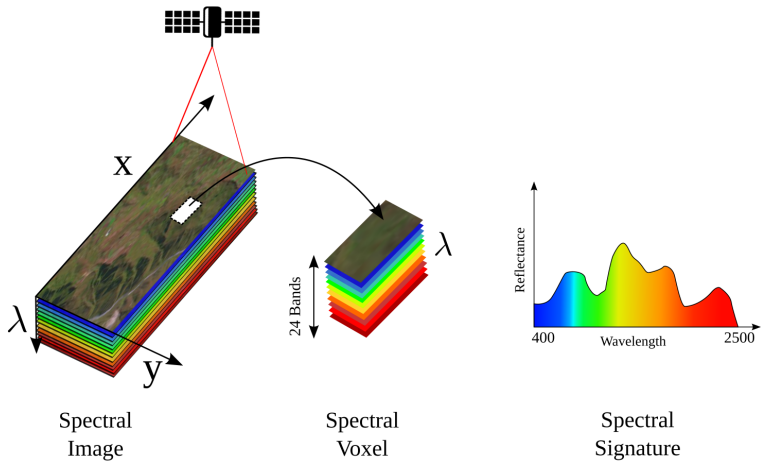
Knowledge of the spectral content at various spatial locations from a scene can be a valuable tool for the detection, identification, and classification of materials and objects with complex compositions [4]. In particular, spectral image classi-

(*) The voxel (volumetric pixel) is the cubic unit composing a three dimensional (3D) object. It constitutes the minimum processable unit of a 3D array, being the equivalent to a pixel in a 2D object.

(**) A Focal Plane Array (FPA) is an imaging device consisting of an array (typically rectangular) of light-sensing pixels at the focal plane of a lens.

fication is an important task for many practical applications, such as precision agriculture [5], monitoring and management of the environment [6], and security and defense issues [7, 8]. As shown in Fig. B, every spatial location in a spectral image is represented by a vector whose values correspond to the intensity at different spectral bands. These vectors are also known as the spectral signature of the pixels. Different materials usually reflect electromagnetic energy differently at specific wavelengths [9], hence the information provided by the spectral signatures enables to distinguish different physical materials and objects, leading to the potential of a more accurate image classification.

Figure B: Spectral voxel and signature of a spectral image [1].



Clustering analysis is a common classification technique which group data points with similar patterns into the same group or cluster, such that the inter-cluster dissimilarity and the intra-cluster similarity are maximized. A recent clustering approach is called subspace clustering which assumes that a given set of data points is drawn from a union of unknown subspaces, with unknown and possibly different dimensions, and aims at finding a low-dimensional subspace to fit each group of data points [10]. Subspace clustering algorithms can be divided into four main categories: iterative [11], algebraic [12, 13], statistical [14, 15] and spectral clustering-based methods [16]. Particularly, spectral clustering finds the cluster membership of the data points by using the spectrum of a symmetric nonnegative affinity matrix whose entries measure the similarities between connected points. Therefore, the most important step in spectral clustering-based methods is the similarity graph construction problem [16]. A widely used approach to build the similarity graph consists on measuring the pairwise distance among the data points. On the other hand, in recent years, a new algorithm named sparse

subspace clustering (SSC) proposes to capture the global geometric relationship among all data points by expressing each data point as a linear combination of all other points and then, the set of solutions is restricted to be sparse by minimizing the ℓ_1 norm of the representation coefficient matrix. Using the sparse representation matrix, a similarity graph is then built, from which the segmentation of the data is obtained [17, 18].

Assuming that spectral signatures with similar spectral characteristics will lie in the same low-dimensional subspace, the subspace clustering theory can be used for modeling the spectral image classification problem [19]. In general, spectral image clustering is a very challenging task due to the inherent data complexity and computational cost, which grows in proportion to the dimensions of the spectral data sets. When the ratio between spectral bands and the number of data samples is greatly different, spectral images suffer from the well known curse of dimensionality [20]. 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 [21].

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. This enables to sense and simultaneously reduce the data dimensionality without any further processing step. Additionally the cost of sensing, storing, transmitting and processing a spectral image acquired using this approach is reduced. In order to acquire the compressed measurements, CSI devices use an optical coding element such as a coded aperture which modulates the scene, and a dispersive element to obtain the spectral component of the spectral image. According to their optical configuration, CSI devices employ different sampling strategies which allow to exploit statistical properties of spectral data, leading to different sensing performance in terms of spectral reconstruction quality [22]. The spatial-spectral coded compressive spectral imager (3D-CASSI) is a CSI sensing scheme which modulates the spectral data cube in both 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 [22]. Although the 3D-CASSI allows to modulate the spectral scene with any 3D coded aperture which entails a higher performance, its physical implementation is not trivial. However, the color-

coded aperture spectral camera imager (CCASSI) [23, 24, 25, 26] and the dual-coded hyper-spectral imager (DCSI) [27] are two approximate implementations of the ideal 3D-CASSI.

In recent work, the CSI theory has been used in conjunction with spectral image classification. Specifically a supervised spectral image classifier that labels each spectral pixel in one of the known classes using a set of compressed CASSI measurements was proposed in [28]. However, this classifier uses the compressed measurements to recover a sparse representation for each spectral signature by solving an optimization problem which incurs in a high computational cost.

Since the structure of the acquired compressed projections directly depends on the applied coding pattern, a set of coded apertures can be designed such that the information, and hence the similarity between the spectral signatures, is approximately preserved after the sensing process. This research work, focuses on the problem of unsupervised spectral image classification directly on the compressed measurements without recovering the original spectral scene. Particularly, 2D projections of the spectral image data are first acquired using the 3D-CASSI sensing approach, which reduces the data dimensionality and hence the storage cost. Then, a subspace clustering algorithm is proposed in order to perform the spectral image clustering. Moreover, since we are only interested in the classification results, the spectral image reconstruction phase is not performed, thus the cost of recovering all the data is avoided. In the following chapters, the spectral image, subspace clustering and CSI theory are described. Then, a coding pattern design and a clustering algorithm based on SSC model are proposed. Finally, simulations and results are included to analyze the performance of the developed spectral image clustering approach.

1. SPECTRAL IMAGING SUBSPACE CLUSTERING

In this chapter, the theoretical background related to the problem addressed in this research is introduced. First, the spectral imaging concept, characteristics and some traditional sensing approaches are presented. In the second part, the subspace clustering theory and some algorithms are briefly introduced.

1.1. SPECTRAL IMAGING

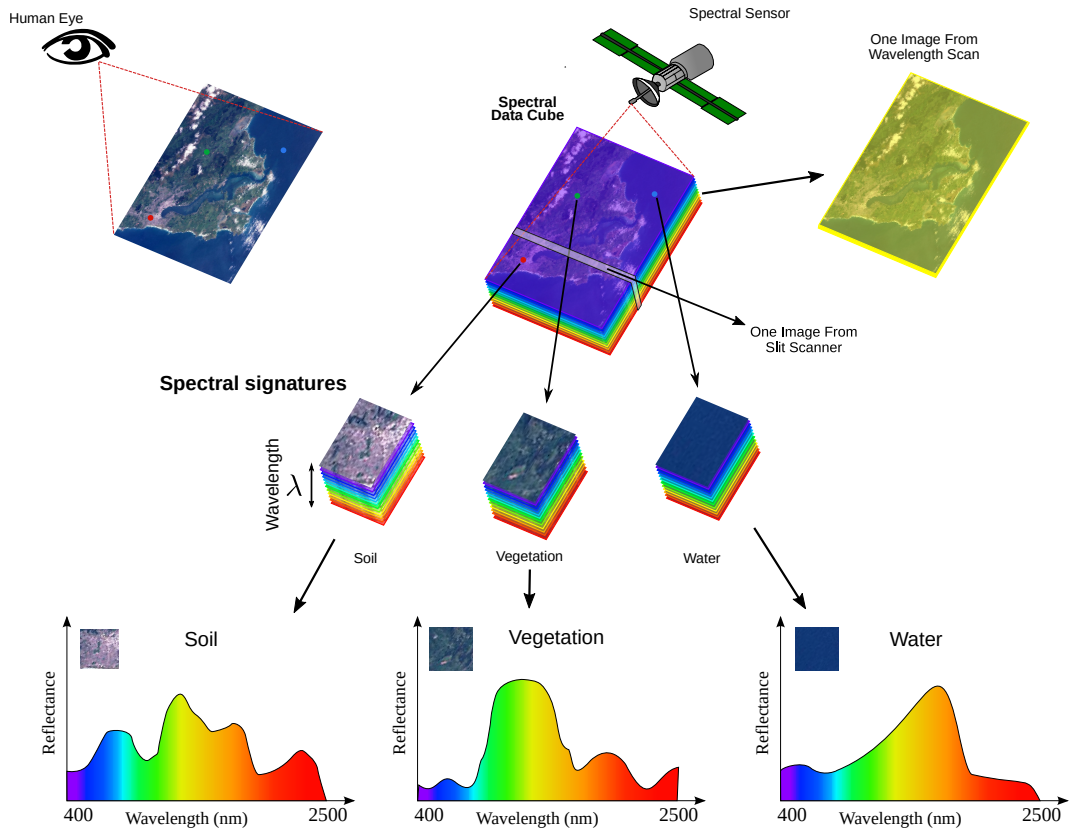
The human visual system builds a representation of the surrounding environment by detecting and interpreting the information from the visible range, roughly from 380 to 750 nm, of the electromagnetic spectrum. This limitation extends to traditional photo cameras which group all the spectral information within the visible spectral range into three broad spectral ranges roughly corresponding to the three primary colors, red, green and blue. Although such information allows to perceive the shape, surface texture and mutual spatial relation in the depth of 3D space, it is insufficient in applications where the spectral information of interest extends to other ranges of the electromagnetic spectrum.

Spectral imaging combines spectroscopy and two-dimensional imaging methodologies. Whereas imaging provides the intensity at every pixel of a 2D image, and a typical spectrometer measures a single spectrum, spectral imaging collects 2D images at specific wavebands across the electromagnetic spectrum. This is a three-dimensional (3D) data set and can be viewed as a cube of information, as observed in Fig. 1.1.

Many different techniques for spectral imaging have been developed over the years. For instance traditional spectral imaging methods, such as Whiskbroom[29], Pushbroom[30] and tunable filter imagers[31], scan adjacent zones of the underlying spectral scene and merge the results to construct a spectral 3D data cube. On the other hand, snapshot spectral imaging captures the spatial and spectral information of a scene by mapping all the voxels of the spectral data cube into different regions of a large focal plane array (FPA) [2, 3]. Furthermore, spectral imaging using Fabry-Perot filters or colored mosaic FPA detectors captures small subsets of spectral bands by assigning a particular spectral response to each FPA pixel such that a specific range of wavelengths is captured [32, 33]. In general, these traditional techniques require all voxels of the 3D scene to be sensed. Then, as the spatial or spectral resolution increases, the number of voxels to be sensed increases proportionally, leading to an increment in the cost of sensing, storing and transmitting an spectral image acquired through these methods.

Every spatial location in a spectral image is represented by a vector whose values correspond to the intensity at different spectral bands. These vectors are also known as the spectral signature of the pixels, see Fig. 1.1. All materials have unique spectral characteristics because they absorb, reflect, and emit radiation in a unique way. For instance, in the visible portion of the spectrum, a leaf

Figure 1.1: Spectral Imaging concept.

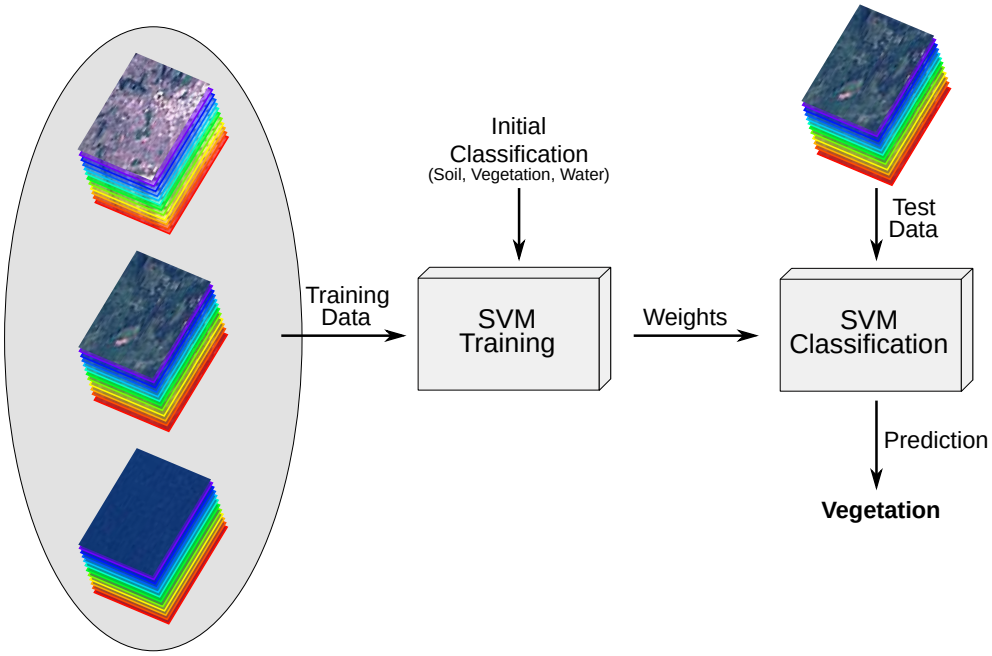


appears green because it absorbs in the blue and red regions of the spectrum and reflects in the green region. These variations in absorption, reflection, and emission are due to the material composition. Differences in spectral responses due to absorption, transmission, and reflection cause materials to have a unique spectral signature. Therefore, the information provided by the spectral signatures can be a valuable tool for the detection, identification, and classification of materials and objects with complex compositions [4]. Particularly, the classification of spectral images consists on labeling individual spectral signatures to one of the classes based on its spectral characteristics. Spectral image classification has found many applications in various fields such as military [34, 35], precision agriculture [36], and mineralogy [37].

In general, there are two main approaches to the classification problem: supervised and unsupervised. Supervised techniques require the availability of a training set for learning the classifier. Among various supervised techniques, support vector machines (SVMs) [38, 39] have shown a good performance for spectral image classification [40, 41]. In particular, a few spectral signatures are used as training samples to train a SVM classifier and then, the remaining spectral signatures are classified. This process is depicted in Fig. 1.2. Variations of the SVM-based algorithms have also been proposed to improve the classification accuracy. These variations include semisupervised learning which exploits both labeled and unlabeled samples [42], postprocessing of the individually labeled

samples based on certain decision rules [43], and incorporating spatial information directly in the SVM kernels [44]. More recent spectral imaging supervised classification techniques can be found in [45, 46].

Figure 1.2: Supervised spectral image classification using SVM.



Unsupervised methods, known also as clustering methods, perform classification just by exploiting information conveyed by the data, without requiring any training sample set. Supervised methods offer a higher classification accuracy compared to the unsupervised ones, but in some applications, it is necessary to resort to unsupervised techniques because training information is not available. When trying to cluster high dimensional data such as spectral images, the given set of data points, i.e., the spectral signatures, could be drawn from an arrangement of an unknown number of subspaces that have unknown and possibly different dimensions. Then, the goal is to simultaneously estimate these subspaces and cluster the points into their corresponding subspaces [47].

1.2. SUBSPACE CLUSTERING THEORY

In recent years, the unprecedented technological advances have lead to an increment in the availability and dimensionality of the data in all areas of science and engineering. These include machine learning, signal and image processing, computer vision, pattern recognition, bioinformatics, etc. For instance, a conventional gray scale image consists of billions of pixels whereas a spectral image consists of hundreds of grayscale images which provide information from different wavelengths. This high dimensionality of the data leads to an increment in memory resources, in order to store such information, and computational cost for processing and data analysis. However, high dimensional data often lies in low

dimensional structures instead of being uniformly distributed across the ambient space. Therefore, different techniques for finding a low-dimensional representation of a high dimensional data set have been developed [10].

A traditional dimensionality reduction technique is the principal component analysis (PCA). PCA assumes that a set of points $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N_e}\} = \{\mathbf{x}_j\}_{j=1}^{N_e}$ in a high dimensional space $\mathbb{R}^D$ is drawn from a single low dimensional affine subspace A of dimension $d \ll D$. This technique is well established in the literature, and has become one of the most useful tools for data modeling, compression, and visualization [10].

In practice, however, the data points could be drawn from the union of $n_e \geq 1$ linear or affine subspaces $\{A_i\}_{i=1}^{n_e}$ of unknown dimensions $d_i = \dim(A_i)$, $0 < d_i < D$. The subspaces can be described as

$$A_i = \{\mathbf{x} \in \mathbb{R}^D : \mathbf{x} = \boldsymbol{v}_i + \mathbf{U}_i \mathbf{y}\}, \quad i = 1, \cdots n, \tag{1.1}$$

where $\boldsymbol{v}_i \in \mathbb{R}^D$ is an arbitrary point in subspace A_i, that can be chosen as $\boldsymbol{v}_i = \mathbf{0}$ for linear subspaces; $\mathbf{U}_i \in R^{D \times d_i}$ is a basis for subspace A_i; and $\mathbf{y} \in \mathbb{R}_i^d$ is a low dimensional representation for point $\mathbf{x}$. Then, the goal of subspace clustering is to find the number of subspaces n_e, their dimensions $\{d_i\}_{i=1}^{n_e}$, the subspace bases $\{\mathbf{U}_i\}_{i=1}^{n_e}$, the points $\{\boldsymbol{v}_i\}_{i=1}^{n_e}$, and cluster the data points into their corresponding subspaces. Since data in a subspace is often distributed arbitrarily and not around a centroid, standard centroid-based clustering methods [48] that take advantage of the spatial proximity of the data in each cluster are not in general applicable to subspace clustering.

Different algorithms for subspace clustering which take into account the multi-subspace structure of the data have been proposed in the past two decades. These algorithms can be divided into four main categories: iterative, algebraic, statistical and spectral clustering-based methods [18]. In particular, spectral clustering-based methods construct a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V} = \{1, \cdots, N\}$ is the set of nodes, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges, and $\mathbf{W} \in \mathbb{R}^{N_e \times N_e}$ is a symmetric nonnegative affinity matrix whose (j,k)-th entry, W_{jk}, measures the affinity between points $\mathbf{x}_j$ and $\mathbf{x}_k$. Ideally, $W_{jk} = 1$ if points j and k are in the same group and $W_{jk} = 0$ if points j and k are in different groups. In practice, a typical affinity is given by

$$W_{jk} = \exp\left(-\frac{1}{2\sigma}\text{dist}(\mathbf{x}_j, \mathbf{x}_k)^2\right), \tag{1.2}$$

where $\text{dist}(\mathbf{x}_j, \mathbf{x}_k)$ is some measure of the distance between points j and k and $\sigma > 0$ is a parameter. Let $\mathcal{D} = \text{diag}(\mathbf{W}\mathbf{1})$, where $\mathbf{1} \in \mathbb{R}^{N_e}$ is an all-one vector, be a diagonal matrix whose j-th diagonal entry gives the degree $d_{jj} = \sum_k W_{jk}$ of node j, and let $\mathcal{L} = \mathcal{D} - \mathbf{W} \in \mathbb{R}^{N_e \times N_e}$ be the graph's Laplacian matrix. Spectral clustering obtains a clustering of the data by applying the K-means algorithm to the columns of the matrix $\mathbf{Y} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_{n_e}]^T \in \mathbb{R}^{n_e \times N_e}$, where $\{\mathbf{u}_i\}_{i=1}^{n_e}$ are the eigenvectors of $\mathcal{L}$ associated with its n_e smallest eigenvalues. However, the distance-based affinity described in Eq. (1.2) is not appropriate for subspace clustering since two points could be very close to each other but lie in different

subspaces (e.g., near the intersection of two subspaces). Conversely, two points could be far from each other but lie in the same subspace. In general, the geometric relationships among multiple points must be considered in order to construct an effective affinity measure for subspace clustering [10, 18].

In fact, the construction of a good affinity matrix is one of the main challenges in applying spectral clustering to subspace clustering problem. Based on the representative method for designing an affinity matrix, the existing spectral clustering-based algorithms can be categorized in local and global methods. Local methods such as Local Subspace Affinity (LSA) [49], Locally Linear Manifold clustering (LLMC)[50] and Spectral Local Best-fit Flats (SLBF)[51, 52] compute an affinity between two points that depends only on the data points in a local neighborhood of each of the two points. These methods have difficulties in dealing with points near the intersection of two subspaces, because the neighborhood of a point can contain points from different subspaces. In addition, they are sensitive to the right choice of the neighborhood size to compute the local information at each point. On the other hand, global methods such as Spectral Curvature Clustering (SCC) [53] and algebraic subspace affinity[16] compute an affinity between two points that depends on all the data points.

Recently, a new spectral clustering-based algorithm named Sparse Subspace Clustering has been proposed [17, 18]. SSC is also based on the idea of writing a data point as a linear or affine combination of neighboring data points. However, while LSA, SLBF, and LLMC use the angular or Euclidean distance between two points to choose the K-NNs, SSC uses the principle of sparsity to choose any of the remaining data points $(N_e - 1 \gg K)$ as a possible neighbor. Specifically, denote the matrix containing all the noise-free data points as

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N_e}] = [\mathbf{X}_1, \cdots, \mathbf{X}_{n_e}]\, \mathbf{\Gamma}, \tag{1.3}$$

where $\mathbf{X}_i \in \mathbb{R}^{D \times N_{e_i}}$ is a rank-d_i matrix of the $N_{e_i} > d_i$ points that lie in A_i, and $\mathbf{\Gamma} \in \mathbb{R}^{N_e \times N_e}$ is an unknown permutation matrix.

The SSC algorithm takes advantage of the self-expressiveness property of the data, i.e., each data point in a union of subspaces can be efficiently reconstructed by a combination of other points in the dataset. Then with the $\mathbf{X}$ matrix itself being used as the dictionary, the SSC algorithm constructs the sparse representation model as follows

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_1 \quad \text{s.t.} \quad \mathbf{X} = \mathbf{X}\mathbf{Z},\ \text{diag}(\mathbf{Z}) = 0, \tag{1.4}$$

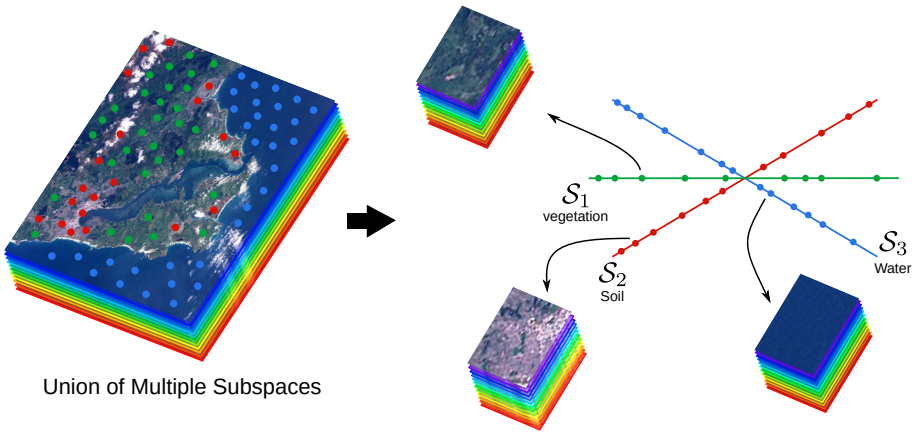
where the ℓ_1-norm regularization in this formulation suggests that a sparse representation of a data point finds points from the same subspace. Using the obtained sparse coefficient matrix $\mathbf{Z}$, the affinity matrix $\mathbf{W}$ is constructed, which defines the weight on the edge between the data nodes as follows

$$\mathbf{W} = |\mathbf{Z}| + |\mathbf{Z}|^T. \tag{1.5}$$

Finally, the clustering result is obtained by applying spectral clustering to the Laplacian $\mathcal{L}$ matrix induced by the affinity matrix $\mathbf{W}$.

Assuming that spectral signatures with similar spectral characteristics lie in the same low-dimensional subspace, the subspace clustering theory can be used for modeling the spectral image classification problem. Figure 1.3 depicts a spectral image, whose spectral signatures belong to different subspaces, i.e it can be viewed as a union of multiple subspaces.

Figure 1.3: Spectral image classification using the subspace clustering approach.



In general, the exponential increment of data features (number of sensed spectral bands) tends to deteriorate the accuracy of classification algorithms. This phenomenon is called the "curse of dimensionality" and was introduced by Richard E. Bellman to describe the problem caused by the exponential increase in volume associated with adding extra dimensions to the Euclidean space [54]. Therefore, the reduction of data dimensionality with no appreciable loss of information is a crucial preprocessing step for learning and recognition tasks such as classification. Furthermore, the dimensionality reduction alleviates the need of high computational and storage resources. In the next section, a spectral imaging approach which senses and simultaneously reduces the data dimensionality is presented. As the data is compressed (dimensionality-reduced), a reconstruction process is typically used in order to recover a 3D approximation of the underlying scene and to further perform different image processing techniques, such as classification. The aim of this work is to avoid the reconstruction step and to perform subspace clustering directly on the dimensionality-reduced observations. In the following sections, the spectral imaging approach and the proposed subspace clustering method are described.

2. COMPRESSIVE SPECTRAL IMAGING

Most of the information acquired by traditional sampling methods of spectral images is discarded during the compression process before being stored. Due to the high dimensionality of the spectral images, recent works have focused on obtaining a compressed signal or spectral image directly avoiding the compression step.

Compressive spectral imaging (CSI) is a new approach which senses and reduces the data dimension in a single step. Specifically, CSI captures 2D coded and dispersed projections of the 3D scene rather than direct measurements of the voxels. Therefore, a far less number of samples are acquired compared to traditional techniques, leading to an improved sensing speed [55].

CSI theory establishes that an estimation of the 3D data cube can be successfully recovered from the compressed measurements. To make this possible, CSI relies on two principles: sparsity, which characterizes the spectral scene of interest, and incoherence, which shapes the sensing structure[56, 57].

A signal is K-sparse if at most K of its components are nonzero. However, many natural and man-made signals are not sparse but compressible in the sense that they can be well-approximated as a linear combination of just a few elements from a known basis or dictionary $\mathbf{\Psi}$. Formally, denote $\mathcal{F}$ as the spatio-spectral input data cube, with $M \times N$ spatial dimensions, and L spectral bands. Then, a spectral image can be expanded in an orthonormal basis $\mathbf{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_{MNL}]$ as

$$\bar{\mathbf{f}} = \sum_{i=1}^{MNL} \theta_i \boldsymbol{\psi}_i, \tag{2.1}$$

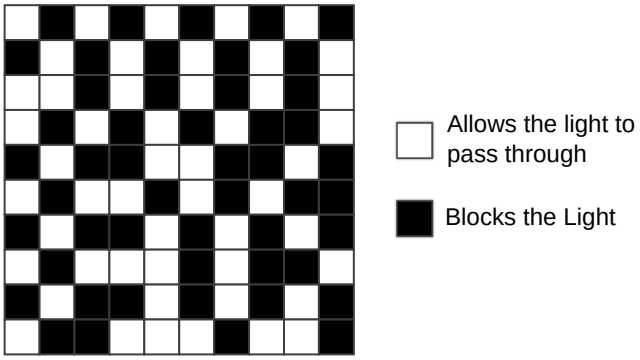
where $\bar{\mathbf{f}} \in \mathbb{R}^{MNL}$ is a vector representation of the source $\mathcal{F}$, $\boldsymbol{\theta}$ is the coefficient sequence of $\mathbf{f}$, such that $\theta_i = \langle \bar{\mathbf{f}}, \boldsymbol{\psi}_i \rangle$. Then, the spectral image $\bar{\mathbf{f}}$ can be expressed as $\bar{\mathbf{f}} = \mathbf{\Psi}\boldsymbol{\theta}$, where $\mathbf{\Psi}$ is an $MNL \times MNL$ matrix with $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_{MNL}$ as columns.

On the other hand, the CSI incoherence property measures how correlated are the elements of $\mathbf{H}$ and $\mathbf{\Psi}$ where a low correlation is desired [58]. Specifically, let $\mathbf{H} \in \mathbb{R}^{MV \times MNL}$, where MV is the number of compressed measurements to acquire, and $\mathbf{\Psi} \in \mathbb{R}^{MNL \times MNL}$ be two orthonormal bases. Following the compressive sensing theory, the first basis is used to measure/sense the signal $\bar{\mathbf{f}}$, and the second basis is used to represent $\bar{\mathbf{f}}$ in a sparse domain. The mutual coherence of the orthonormal basis $\mathbf{H}$ and $\mathbf{\Psi}$ is defined as the maximum absolute value of the inner product between any two columns of the basis, given by

$$\mu(\mathbf{H}, \mathbf{\Psi}) = \sqrt{MN} \max_{1 \leq i,j \leq MN} |\langle \boldsymbol{H}_i, \boldsymbol{\Psi}_i \rangle|, \tag{2.2}$$

where $\boldsymbol{H}_i$ represents the i-th column of the sensing matrix $\mathbf{H}$, whose structure is determined by the sensing architecture. The coherence measures the maximum

Figure 2.1: Example of a coded aperture. The black elements represent the zero values which block the light whereas the white elements represent the one values which allow the light to pass through.



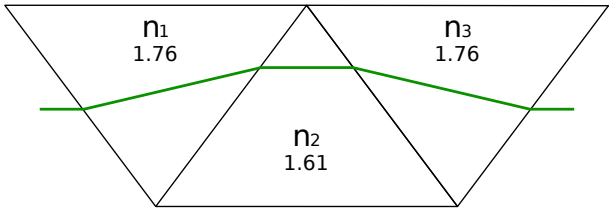
correlation between any two elements of $\mathbf{H}$ and $\mathbf{\Psi}$. In CSI a minimum coherence among the orthonormal base is desired.

CSI-based systems can capture a spectral image with a single snapshot, without the need to linearly scan all the adjacent zones, as observed in Fig. A b). In order to acquire the 2D compressed projections, CSI devices usually employ optical elements like focal plane arrays, optical coding elements, e.g coded apertures, and dispersive elements, e.g prisms.

Coded Apertures or Coded-Aperture Masks are grids, gratings, or other patterns of materials opaque to various wavelengths of light. A coded aperture can be represented as a binary matrix, where the one-valued elements allow the electromagnetic radiation to pass through and the zero-valued elements block the light, generating a coding pattern in the object of interest. Figure 2.1 shows an example of a coded aperture whose elements were generated at random.

Another fundamental optical component of the CSI devices is the dispersive element, whose function is to decompose the light in its spectral components. A commonly used optical dispersive prism in CSI systems is the double Amici prism shown in Fig. 2.2. Such prism is composed by the union of three prisms with different refraction indices. The prisms at the ends have the same refractive index while the prism at the center has a higher refractive index and, therefore, a greater capacity for dispersion. This prism is specially designed to increase the angular dispersion of the set of waves, and also has the property that, when refracted again, the spectral component with central wavelength λ_c forms a straight line with the trajectory of the incident ray.

Figure 2.2: Double Amici prism.



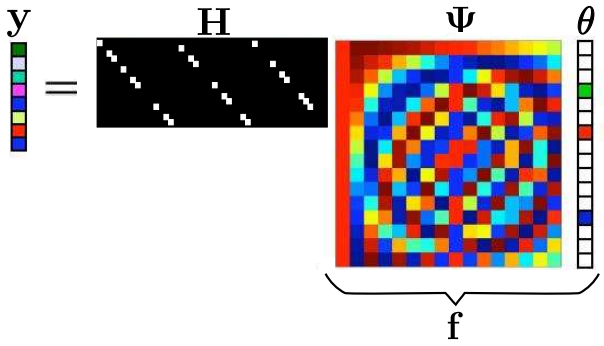
In the CSI mathematical model, the spectral image $\mathcal{F}$ with $M \times N$ spatial dimensions and L spectral bands is discretized as $\mathcal{F}_{m,n,k}$, where m and n index the spatial coordinates, and k determines the k-th spectral band. $\mathcal{F}$ can also be represented in vector form as $\mathbf{f} = \left[\mathbf{f}_0^T, \cdots, \mathbf{f}_{L-1}^T\right]^T$, where each spectral band $\mathbf{f}_k$ can be expressed as $\mathbf{f}_k = \left[\mathcal{F}_{0,0,k}, \mathcal{F}_{1,0,k}, \cdots, \mathcal{F}_{(M-1),0,k}, \cdots, \mathcal{F}_{0,1,k}, \mathcal{F}_{1,1,k}, \cdots, \mathcal{F}_{(M-1),1,k}, \cdots, \mathcal{F}_{(M-1),(N-1),k}\right]^T$.

In general, CSI projections can be written in matrix notation as

$$\mathbf{y} = \mathbf{H}\mathbf{f} = \mathbf{H}\mathbf{\Psi}\boldsymbol{\theta}, \tag{2.3}$$

where $\mathbf{y}$ is the measurement set in vector form and, $\mathbf{H}$ is known as the system sensing matrix whose entries are determined by the CSI optical setup [22, 55]. The spectral image $\mathbf{f}$ is said to have a sparse representation $\boldsymbol{\theta}$ in a basis $\mathbf{\Psi}$, as depicted in Fig. 2.3.

Figure 2.3: CSI linear system representation. A measurement vector $\mathbf{y}$ is obtained by sensing a spectral image $\mathbf{f}$ using the sensing matrix $\mathbf{H}$. The spectral image $\mathbf{f}$ has a sparse representation $\boldsymbol{\theta}$ in a representation basis $\mathbf{\Psi}$.



Source: Adapted image from the paper *Compressive Sensing* [59].

Although CSI allows to reconstruct a spectral image from a considerable smaller number of samples than those required by the Shannon-Nyquist theorem, there is a minimum number of necessary measurements that must be acquired to successfully reconstruct a scene. Considering a spectral image of size MNL which is K-sparse in a basis $\mathbf{\Psi}$, the CSI theory states that $u \geq K\log(MNL)$ measurements are necessary in order to achieve a correct reconstruction. However, for

spectrally rich scenes or very detailed spatial scenes, a single shot CSI measurement may not provide a sufficient number of compressed measurements to satisfy the above constraint. This leads to an excessively undetermined inverse problem hence failing to obtain an adequate reconstruction. Increasing the number of measurement shots will multiply the number of measurements, thus rapidly overcoming such limitations [60, 61, 62].

In CSI it is possible to acquire $S \ll L$ measurement shots employing a different coded aperture at each snapshot, such that different measurements of the spectral data cube are acquired each time. Mathematically, the compressed measurements acquired with the s-th snapshot can be expressed in vector notation as

$$\mathbf{y}^s = \mathbf{H}^s \mathbf{f}, \tag{2.4}$$

where $\mathbf{y}^s \in \mathbb{R}^{MV}$ is the vector form representation of the acquired measurements, and $\mathbf{H}^s$ is the sensing matrix of the system for each snapshot s. Specifically, $\mathbf{H}^s$ is a $MV \times MNL$ sparse matrix whose nonzero entries are determined by the coded aperture. The measurement vectors $\mathbf{y}^s$ acquired at each snapshot can be succinctly expressed in vector form as $\mathbf{y} = \left[\left(\mathbf{y}^0\right)^T, \cdots, \left(\mathbf{y}^{S-1}\right)^T \right]^T$. Therefore, Eq. 2.4 can be rewritten in the standard form of an underdetermined system, described in Eq. 2.3, where $\mathbf{H} = \left[\left(\mathbf{H}^0\right)^T, \cdots, \left(\mathbf{H}^{S-1}\right)^T \right]^T$ is the concatenation of all sensing matrices $\mathbf{H}^s$.

Once acquired the compressed measurements $\mathbf{y}$, the subsequent procedure is spectral image recovery. Given that the amount of the acquired compressed measurements SMV (generally $V \leq N + L - 1$), is far less than the number of 3D data cube entries to be estimated MNL, the reconstruction problem to be solved becomes ill posed. Therefore, it cannot be solved by directly inverting the system in Eq. 2.3. In particular, CSI spectral image recovery consists on finding a K-sparse representation of $\mathbf{f}, \boldsymbol{\theta}$ in a given basis $\boldsymbol{\Psi}$. The sparse representation can be recovered by minimizing the $\ell_2 - \ell_1$ cost function given by $\|\bar{\mathbf{y}} - \mathbf{A}\boldsymbol{\theta}\|_2 + \tau\|\boldsymbol{\theta}\|_1$, where τ is a regularization constant [55, 56]. In other words, it looks for a sparse approximation of the spectral data cube. Formally, the reconstruction optimization problem can be written as

$$\tilde{\mathbf{f}} = \boldsymbol{\Psi} \left\{ \underset{\boldsymbol{\theta} \in \mathbb{R}^{MNL}}{\arg\min} \|\bar{\mathbf{y}} - \mathbf{A}\boldsymbol{\theta}\|_2 + \tau\|\boldsymbol{\theta}\|_1 \right\}, \tag{2.5}$$

where $\mathbf{A} = \mathbf{H}\boldsymbol{\Psi}$.

CSI reconstruction algorithms can be classified in five computational approaches [63]. The *greedy* algorithms obtain an sparse estimation of $\boldsymbol{\theta}$ in an iterative manner by identifying the components which provide the best reconstruction at each iteration. Algorithms such as OMP (orthogonal matching pursuit) [64], StOMP [65] and CoSaMP(*Compressive Sampling Matching Pursuit*) [66] implement this approach. The second approach obtains a sparse representation of the spectral image by solving the optimization problem described in Eq. 2.5. Algorithms within

this approach are SpaRSA [67], TwIST [68] and GPSR [69]. The third category of algorithms employ a Bayesian framework and assume a prior distribution for the unknown coefficients that favors sparsity [70]. The fourth approach consists on relaxing the ℓ_0 problem to a related nonconvex problem and attempts to identify a stationary point [71]. Finally, in the last category are the brute force algorithms which search through all possible support sets, possibly using cutting-plane methods to reduce the number of possibilities [72]. The complete analysis and review of these algorithms can be found in [63].

The matrix $\mathbf{A}$ plays a crucial role in the mathematics of the inverse CSI problem. Indeed, the Restricted Isometry Property (RIP) of $\mathbf{A}$ must be satisfied in order to achieve a reliable estimation of the original spectral image. The RIP establishes the conditions necessary for $\mathbf{A}$ such that the ℓ_2 norm of the underlying 3D spectral image is approximately preserved under the transformation $\mathbf{A}\boldsymbol{\theta}$. Furthermore the RIP determines the minimum number of compressed projections needed for a correct reconstruction [55]. Formally, assuming that $|\boldsymbol{\theta}| = K$, the restricted isometry property of the CSI matrix $\mathbf{A}$ of order K is defined as the smallest δ_s such that

$$(1-\delta_s)\|\boldsymbol{\theta}\|_2^2 \leq \|\mathbf{A}\boldsymbol{\theta}\|_2^2 \leq (1+\delta_s)\|\boldsymbol{\theta}\|_2^2, \tag{2.6}$$

where the constant δ_s is given by

$$\delta_s = \max_{\mathcal{T}\subset[MNL],|\mathcal{T}|\leq K} \|\mathbf{A}_{|\mathcal{T}|}^T\mathbf{A}_{|\mathcal{T}|} - \mathbf{I}\|_2^2, \tag{2.7}$$

the operator $\|\cdot\|_2^2$ is the squared norm from ℓ_2 into ℓ_2, $\mathbf{A}_{|\mathcal{T}|}$ is a $KV \times |\mathcal{T}|$ matrix whose columns are equal to $|\mathcal{T}|$ columns of $\mathbf{A}$ indexed by the set $\mathcal{T}$, and $\mathbf{I}$ is an identity matrix [73].

Since the structure of the matrix $\mathbf{A}$ is directly determined by the coded aperture, different works have proposed to improve the spectral image reconstruction quality by designing a set of coded apertures such that the RIP is better satisfied [26, 74, 75]. For instance, in [26] the traditional block-unblock coded apertures are replaced with a set of colored coded apertures. Then, an optimal design of such coded apertures is developed based on the RIP.

2.1. SPATIAL-SPECTRAL CODED COMPRESSIVE SPECTRAL IMAGER (3D-CASSI)

The spatial-spectral coded compressive spectral imager (3D-CASSI) is a CSI sensing scheme which 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, see Fig. 2.5. 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 [22]. As shown in Fig. 2.4, the 3D-CASSI first modulates the scene using a 3D coded aperture $\mathbf{C}$, whose entries are indexed as $C_{m,n,k}$ and then, the coded spectral scene is integrated along the spectral axis. The output

Figure 2.4: CSI sensing approach used for compressed measurements acquisition at snapshot s.

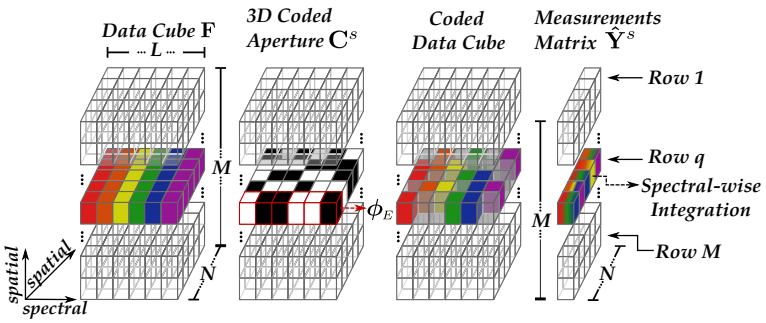
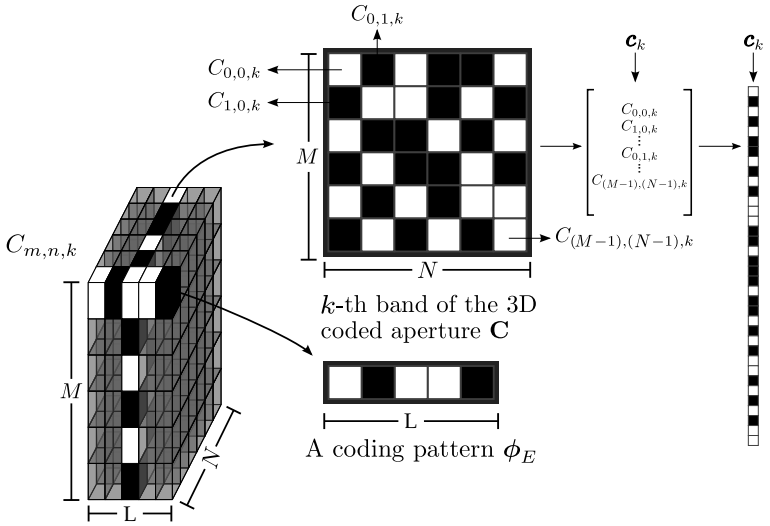


Figure 2.5: Representation of the 3D coded aperture. This ensemble of block-unblock 2D coded apertures can be seen as a set of different coding patterns ϕ_E.



of the sensing process, at a specific snapshot s, can be expressed as

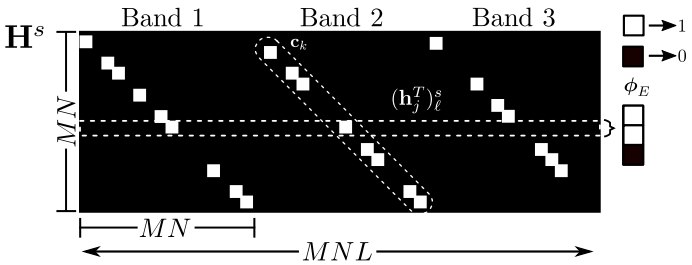
$$\hat{Y}^s_{m,n} = \sum_{k=0}^{L-1} \mathcal{F}_{m,n,k} C^s_{m,n,k}, \tag{2.8}$$

where $\hat{\mathbf{Y}}^s$ is the matrix containing the compressed information of all spectral signatures at a specific spatial position (m, n) and snapshot s. In addition, note that in the 3D-CASSI sensing approach $V = N$, i.e., the number of acquired compressed measurements is $MV = MN$.

Note that each voxel of the spectral data cube is coded by one voxel of the 3D coded aperture at the same position (m, n, k). More specifically, each spatial location $C_{m,n}$ contains a coding pattern $\boldsymbol{\phi}_s \in \mathbb{R}^L$, with $\phi_k \in \{0, 1\}$, that modulates a spectral pixel in that particular position (m, n), see Fig. 2.5.

A snapshot of the 3D-CASSI can be described in vector form using Eq. 2.4, where $\mathbf{y}^s$ is the vectorization of the matrix $\hat{\mathbf{Y}}^s$ and $\mathbf{H}$ is a $MN \times MNL$ matrix whose structure is determined by the 3D coded aperture. Formally, the j-th row

Figure 2.6: Sensing matrix $\mathbf{H}$ for $N = 4, M = 4, L = 3$ and $S = 2$.



entries of the sensing matrix $\mathbf{H}^s$ can be written as

$$(\mathbf{h}_j)_\ell^s = \begin{cases} (\boldsymbol{\phi}_E)_{\lfloor \ell/MN \rfloor} & \text{if} \quad j = \ell - \lfloor \ell/MN \rfloor MN \\ 0 & \text{otherwise,} \end{cases} \tag{2.9}$$

for $j = 0, \cdots, MN-1$, $\ell = 0, \cdots, MNL-1$ and $s = 0, \cdots, S$, where $\boldsymbol{\phi}_E \in \{\boldsymbol{\phi}_0, \cdots, \boldsymbol{\phi}_{P-1}\}$ is selected at random among P coding patterns, such that a different one is assigned for each snapshot s at a specific row j. An example of the $\mathbf{H}^s$ matrix is shown on Fig. 2.6 for $N = 4, M = 4$ and $L = 3$.

Considering that one spectral signature is coded by a different coding pattern at each measurement shot, there are essentially three cases to analyze. The first case is when the number of measurement shots is greater than the number of coding patterns i.e., $S > P$. In this case, some pixels are oversampled thus redundant information is acquired. The second case is when $S < P$. In such case, the pixels are coded by a subset of coding patterns. Since a specific pattern $\boldsymbol{\phi}_s$ encodes a determined group of spectral bands, every subset must be designed such that all the spectral bands are sensed in order to avoid losing or discarding important information. Furthermore, clustering two spectral signatures becomes harder if they were coded by different subsets of coding patterns with no elements in common. Finally, in the case $S = P$, all spectral signatures are coded by the same set of coding patterns and each $\boldsymbol{\phi}_s$ can be designed such that no redundant information is acquired. Through this research work, the case $S = P$ is assumed for simplicity.

2.2. CSI AND VERY SPARSE RANDOM PROJECTION

A dimensionality reduction technique that is capable to reduce the data into a lower-dimensional model, while preserving the reconstructive or discriminative properties of the original data can be marked as ideal. However, in practice information is lost as the dimensionality is reduced. Therefore, a method which efficiently reduces dimensionality, while preserving as much as possible information from the original data is needed. One solution is to reduce the dimensionality of data by projecting it onto a lower-dimensional subspace [76].

Principal component analysis (PCA) is a classic dimensionality reduction (DR) method that finds the low-dimensional linear subspace that minimizes the mean-

squared error between the original data and the data projected onto the subspace. A low-dimensional representation of the data is constructed in such a way that it describes as much of the variance in the data as possible. This is achieved by finding a linear basis of reduced dimensionality for the data (a set of eigenvectors) in which the variance of the data is maximal [77].

PCA, however, is based on a linear subspace model that is generally not capable of capturing the geometric structure of real-world datasets [78]. In addition, most of the computationally efficient approaches to compute the principal components assume ready access to the stored full data samples. However, this full data access is not always possible in modern data settings. For instance, in the traditional spectral imaging approaches, the full data is obtained by gradually scanning adjacent zones of the spectral scene hence the PCA analysis must be strictly performed after the sensing process. Furthermore, considering a spectral image with $M \times N$ spatial dimension and L spectral bands, computing the Q principal components takes $O(MNL^2 + L^3 + MNQL) = O(MNL^2)$ which is computational expensive when the spatial resolution or the number of spectral bands grow exponentially [79].

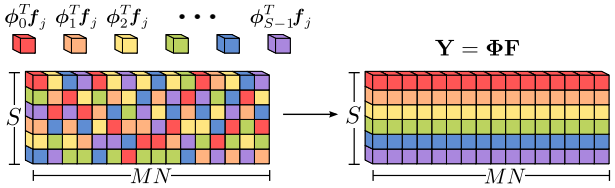
One promising strategy to reduce the data dimensionality in an efficient way, which also allows for rigorous theoretical analysis, is to use Random Projections (RP). In RP the data dimensionality reduction is computationally simple: the original high-dimensional data is projected onto a low-dimensional subspace using a random matrix, thereby saving memory and computation. For example, reducing the dimension of a $M \times N \times L$ spectral image using RP will take only $O(MNLS)$, which is a significant computation time reduction in comparison with PCA.

There exist theoretical results supporting that RP preserves volumes and affine distances [80] or the data structure, hence it is possible to directly learn from the low dimensional data [81]. In fact, based on the concept of affinity [82], which characterizes the similarity between two subspaces, it has been theoretically proved and numerically verified that several dominant subspace clustering algorithms could successfully perform clustering on the compressed data [83, 84, 85, 86]. In addition, recent works have shown that RP improves the algorithm performance when dealing with high dimensional data. Specifically, compression reduces the dimension of ambient signal space, hence the computational cost of finding the similarity representation in subspace clustering can be efficiently reduced [83, 87].

Generally, RP uses a random matrix with elements generated by a normal distribution $N(0, 1)$ due to its simplicity in terms of analysis. The problem of this type of RP matrix is the computational complexity due to the dense nature of the projection matrix and hardware implementation difficulties [88]. However, different authors have suggested the use of simpler distributions that generate sparse projections matrices [89, 90]. The introduction of sparsity has shown to significantly reduce computational requirements and to allow an efficient hardware implementation [91].

Given a high dimensional data set, the random projection is essentially performed

Figure 2.7: Rearrangement of the matrix $\mathbf{Y}$ such that the s-th row contains the compressed measurements acquired with the s-th coding pattern ϕ_s. In this figure, colors represent a specific codification, e.g, red pixels denote the compressed measurements acquired with the ϕ_0 coding pattern.



by applying the same realization of a random projection matrix to each data point in order to obtain the set of dimensionality reduced or compressed measurements. In this sense, the CSI sensing mechanism can be rewritten in a RP scheme. Particularly, the measurement vectors $\mathbf{y}^s$ acquired with each snapshot, can be arranged in a matrix $\mathbf{Y} = \left[\mathbf{y}^0, \mathbf{y}^1, \cdots, \mathbf{y}^{S-1}\right]^T$ where each column contains different spectral signature codifications, acquired with distinct coding patterns. Since the case $S = P$ is assumed, it is possible to rearrange the entries of $\mathbf{Y}$ such that each row contains a compressed spectral signature acquired with a specific coding pattern ϕ_s. Formally, this rearrangement can be expressed as

$$Y_{sj} \leftrightarrow Y_{s'j} \quad \text{if } Y_{s'j} = \boldsymbol{\phi}_s^T \boldsymbol{f}_j,$$

for $s, s' = 0, \cdots, S-1$ ($s' \neq s$) and $j = 0, \cdots, MN-1$, where $\leftrightarrow$ stands for the swap operation and $\boldsymbol{f}_j$ is the j-th spectral signature. The described acquisition model and the rearrangement is depicted in Fig. 2.7. Alternatively, define the matrix of S coding patterns as $\mathbf{\Phi} = \left[\boldsymbol{\phi}_0^T, \boldsymbol{\phi}_1^T, \cdots \boldsymbol{\phi}_{S-1}^T\right]^T$, the problem of acquiring and rearranging the measurements $\mathbf{y}$ can be succinctly expressed as

$$\mathbf{Y} = \mathbf{\Phi F}, \tag{2.10}$$

where $\mathbf{F} = [\mathbf{f}_0, \mathbf{f}_1, \cdots, \mathbf{f}_{L-1}]^T$ is a $L \times MN$ matrix whose columns, denoted as $\boldsymbol{f}_j$, are the spectral signatures of the data cube.

As described in the first part of this chapter, spectral image recovery in Eq. 2.5 is achieved using nonlinear and relatively expensive optimization-based or iterative algorithms. For instance, trying to recover a sparse representation of the underlying scene, the GPSR algorithm takes $O(SM^2N^2L)$ operations at each iteration [69].

Then, most of the CS literature has focused on improving the speed and accuracy of such algorithms. However, signal recovery is not actually necessary in many signal processing applications [92]. In particular, the aim of this work is to perform all the spectral image clustering directly on the compressed measurements by designing a set of coding patterns such that the similarity among the spectral signatures is approximately preserved. In this work, all the data processing is performed with the compressed measurements, thus the cost of recovering all the data is avoided.

3. COMPRESSIVE SPECTRAL IMAGING SUBSPACE CLUSTERING

In CSI, the structure of the acquired compressed projections directly depends on the coding pattern design. Indeed, due to the rapt attention of the community on signal recovery, different recent works in CSI have focused on properly designing the coding patterns in order to better reconstruct the underlying spectral scene [26, 93]. These coding pattern designs use the restricted isometry property (RIP) as the main optimization criteria. On the other hand, in this research work the aim is to perform classification on the compressed measurements. Therefore the design of the coding patterns must preserve the information and hence the similarity among the spectral signatures. In this chapter, some coding pattern design considerations are first presented. Then, the optimization problem, whose solution provides an optimal set of coding patterns for CSI measurement clustering, and a solution algorithm are proposed. Finally, a subspace clustering algorithm based on the SSC model is proposed and explained.

3.1. CODING PATTERN DESIGN CONSIDERATIONS

In order to design the coding pattern matrix $\mathbf{\Phi}$, the following three design criteria are considered.

Preserving Similarities

The success of subspace clustering on the compressed measurements depends fundamentally on how the coding matrix $\mathbf{\Phi}$ 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 has unit length, the similarity between two compressed measurements $\boldsymbol{y}_j = \mathbf{\Phi}\boldsymbol{f}_j,\ \boldsymbol{y}_{j'} = \mathbf{\Phi}\boldsymbol{f}_{j'}$ is defined as

$$\text{similarity}(\boldsymbol{y}_j, \boldsymbol{y}_{j'}) = \boldsymbol{y}_j^T \boldsymbol{y}_{j'} = \boldsymbol{f}_j^T \mathbf{\Phi}^T \mathbf{\Phi} \boldsymbol{f}_{j'} \quad j \neq j', \tag{3.1}$$

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

$$\mathbf{\Phi}^T\mathbf{\Phi} = \mathbf{I} + \boldsymbol{\epsilon}, \tag{3.2}$$

where

$$\epsilon_{j,j'} = \boldsymbol{\phi}_j^T \boldsymbol{\phi}'_j \quad j \neq j', \tag{3.3}$$

and $\epsilon_{j,j} = 0$. Observe that the matrix $\boldsymbol{\epsilon}$ collects all the entries outside the diagonal of $\mathbf{\Phi}^T\mathbf{\Phi}$. Therefore, if $\epsilon_{j,j'} = 0\ \forall j, j'$, the matrix $\mathbf{\Phi}^T\mathbf{\Phi}$ would be equal to $\mathbf{I}$ and the similarities of the spectral signatures would be preserved exactly in the compressed measurements. However, given that the matrix $\mathbf{\Phi}$ has more columns than rows, all the entries of $\boldsymbol{\epsilon}$ could be mostly small but not zero [94]. Considering that a linear mapping such as that in Eq.2.10 can cause significant distortions

in the compressed measurements if $\mathbf{\Phi}^T\mathbf{\Phi}$ is not close to $\mathbf{I}$, the proposed coded aperture design should minimize the entries of $\boldsymbol{\epsilon}$.

Sensing Neighboring Spectral Bands

The entries of the $\mathbf{\Phi}$ matrix are usually chosen from a Bernoulli distribution $\mathbf{\Phi} \sim Be(p)$. Formally, following such distribution, the entries of the s-th coding pattern can be expressed as

$$(\boldsymbol{\phi}_s)_k = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } q, \end{cases} \tag{3.4}$$

for $k = 0, 1, \cdots, L-1$, where $q = 1-p$. A projection matrix with this structure simply carries out a random sampling on the data vectors before performing element-wise addition. Considering that surface-emitted spectral signatures are, in general, relatively smooth functions of wavelengths [95], the intuition is to perform the random sampling of neighboring spectral bands instead of randomly sampling all the spectral data, which could add outliers to the measurements.

Let $\{\lambda_0, \lambda_1, \cdots, \lambda_{L-1}\}$ be a set of cutoff wavelengths, where $\lambda_k \in \{0, 1, \cdots, L-1\}$. For each coding pattern $\boldsymbol{\phi}_s$, select two cutoff wavelengths $(\lambda_{k_1}, \lambda_{k_2})$ at random such that $\lambda_{k_1} < \lambda_{k_2}$ and $\lambda_{k_2} - \lambda_{k_1} + 1 = \Delta$, where Δ is defined as the coding pattern bandwidth. Then, the banded-structure random matrix can be expressed as

$$(\boldsymbol{\phi}_s)_k = \begin{cases} 1, & \text{with prob. } \frac{1}{2} \iff \lambda_{k_1} \leq k \leq \lambda_{k_2} \\ 0, & \text{otherwise.} \end{cases} \tag{3.5}$$

Equation 3.5 can be alternatively written as

$$(\boldsymbol{\phi}_s)_k = \delta\left(\lfloor \lambda_{k_1}/k \rfloor\right) \delta\left(\lfloor k/\lambda_{k_2} \rfloor\right) \nu_k, \tag{3.6}$$

where $\delta(\cdot)$ is the Kronecker delta function and $\boldsymbol{\nu} \in \mathbb{R}^L$ is a random vector whose entries follow a Bernoulli distribution $Be(p)$ with $p = \frac{1}{2}$.

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 [90], 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 could be expected. On the other hand, decreasing κ (increasing Δ), the entries of $\boldsymbol{\epsilon}$ will increase leading to significant distortions in the acquired compressed measurements. Then, the selection of Δ is a fundamental step. In the simulations section, an analysis in this regard is presented. Specifically, the effect of variations of Δ and the number of measurement shots S.

Information Acquisition

In order to better discriminate among the classes, new information from the underlying spectral scene should be acquired in each measurement shot. Therefore, the coding patterns should be linearly independent, i.e. the matrix $\mathbf{\Phi}$ should be full rank. Additionally, the number of measurements acquired from each spectral band should be approximately the same, i.e. the matrix $\mathbf{\Phi}\mathbf{\Phi}^T$ should approximate the identity matrix $\mathbf{I}$. Specifically, decomposing the matrix $\mathbf{\Phi}\mathbf{\Phi}^T$ as

$$\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{I} + \boldsymbol{\mu}, \tag{3.7}$$

where

$$\mu_{ij} = \boldsymbol{\phi}\boldsymbol{\phi}^T \quad i \neq j, \tag{3.8}$$

and $\mu_{ii} = 0$. Therefore, the minimization of the $\boldsymbol{\mu}$ entries is considered in the coded aperture design described next.

3.2. CODING PATTERN OPTIMIZATION ALGORITHM

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

$$\begin{aligned} \underset{\{\boldsymbol{\phi}_0,\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_{S-1}\}}{\arg\min} \quad & \|\boldsymbol{\epsilon}\|_F^2 + \|\boldsymbol{\mu}\|_F^2 \\ \text{subject to} \quad & \boldsymbol{\epsilon} = \mathbf{\Phi}^{\mathbf{T}}\mathbf{\Phi} - \mathbf{I}, \\ & \boldsymbol{\mu} = \mathbf{\Phi}\mathbf{\Phi}^{\mathbf{T}} - \mathbf{I}, \\ & \text{Rank}(\mathbf{\Phi}) = S, \\ & (\boldsymbol{\phi}_s)_k = \delta\left(\lfloor \lambda_{k_1}/k \rfloor\right) \delta\left(\lfloor k/\lambda_{k_2} \rfloor\right) \nu_k, \end{aligned} \tag{3.9}$$

for $s = 0, \cdots, S-1$ and $k = 0, \cdots, L$.

This optimization problem can be efficiently solved using the proposed Algorithm 1.

In Algorithm 1, steps 2 to 4 generate the first filter, which has a banded structure with a predefined bandwidth Δ. Then, steps 6-9 are intended to minimize the number of times in which a spectral band is sensed. Specifically, the algorithm counts how many spectral bands have been sensed in a certain bandwidth and then the banded section with less information is chosen (expressed in step 9) complying with the criteria of subsection 3.1. Finally, the algorithm chooses the position in which the inner products are approximately minimized. This is attained by minimizing the elements outside the diagonal of $\mathbf{\Phi}^T\mathbf{\Phi}$, i.e, by minimizing the sum of the values in the neighborhood (step 12) expressed in steps from 15 to 18, see Fig. 3.1 a). As observed in Fig. 3.1 b), a random design of the $\mathbf{\Phi}$ entries

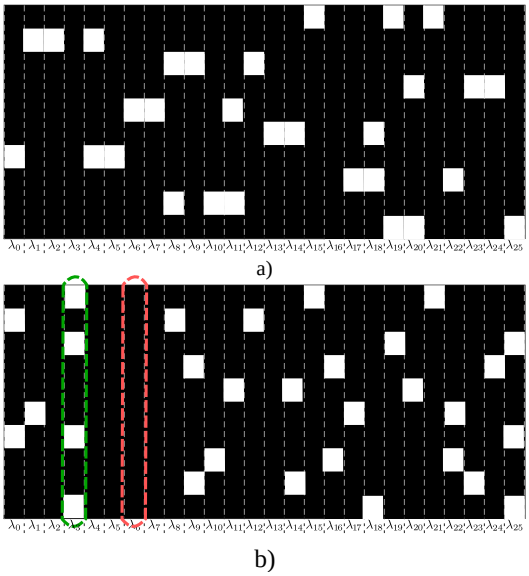
Algorithm 1 Coding Pattern Design

Input: number of bands L, number of shots S, bandwidth $\Delta > 0$, probability $0 < p < 1$.

$\mathbf{\Phi} \leftarrow \mathbf{0}_{S,L}$
Select $(\lambda_{k_1}, \lambda_{k_2})$ randomly, such that $\lambda_{k_2} > \lambda_{k_1}$ with $\lambda_{k_2} - \lambda_{k_1} + 1 = \Delta$
Select $\boldsymbol{\nu} \in \mathbb{R}^L$ such that $\nu_k \sim \mathit{Be}(p)$
$(\mathbf{\Phi}_1)_k \leftarrow \delta(\lfloor \lambda_{k1}/k \rfloor)\delta(\lfloor k/\lambda_{k2} \rfloor)\nu_k$
for $s \leftarrow 2$ to S **do**
  for $i \leftarrow 1$ to $(L - \Delta + 1)$ **do**
    $\mathbf{u}_i \leftarrow \sum_{j=1}^{s} \sum_{k=i}^{i+\Delta-1} (\mathbf{\Phi}_j)_k$
  end for
  $\hat{i} \leftarrow \arg\min_i \mathbf{u}_i$
  $\ell = 1$
  for $i \leftarrow \hat{i}$ to $(\hat{i} + \Delta - 1)$ **do**
    $\mathbf{b}_\ell \leftarrow \sum_{j=1}^{s} \sum_{k=(i-1)}^{i} (\mathbf{\Phi}_j)_k$
    $\ell = \ell + 1$
  end for
  for $j \leftarrow 1$ to $\lfloor p\Delta \rfloor$ **do**
    $\hat{\ell} \leftarrow \arg\min_\ell \mathbf{b}_\ell$
    $(\mathbf{\Phi}_s)_{\hat{\ell}+\hat{i}-1} \leftarrow 1$
    $\mathbf{b}_{\hat{\ell}} \leftarrow \infty$
  end for
end for

Output: $\mathbf{\Phi}$

Figure 3.1: Examples of coding patterns. a) obtained by the proposed design b) random coding pattern.



may lead to sense more information from a specific spectral band (green dashed region) while leaving some spectral bands unsampled (red dashed region).

3.3. SPARSE SUBSPACE CLUSTERING WITH SPATIAL REGULARIZER (S-SSC)

Assuming that compressed pixels of the same land-cover class lie in one independent subspace, the spectral clustering methods can be used in order to separate them into the same group or cluster. In particular, SSC algorithm builds the similarity matrix, which describes the membership of the data, finding a sparse representation for each compressed pixel that ideally corresponds to selecting a few points from the same subspace. The sparse representation model can be described as follows

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{R}} \quad & \|\mathbf{Z}\|_1 + \frac{\lambda}{2}\|\mathbf{R}\|_F^2 \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{Y}\mathbf{Z} + \mathbf{R},\ \text{diag}(\mathbf{Z}) = 0,\ \mathbf{Z}^T\mathbf{1} = \mathbf{1}, \end{aligned} \tag{3.10}$$

where $\mathbf{Y} = \mathbf{\Phi}\mathbf{F}$, $\mathbf{1}$ 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 indicates 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 an affine subspaces [17, 18].

Taking into account that neighboring pixels in a spectral image usually consist of

similar material that has a very high probability of belonging to the same class, a smoothing filter can be applied to the representation coefficient matrix, in oder to reduce the representation error, being able to extract more information from the data [19]. Specifically, the smoothing filters will reduce the noise trying to assign the same representation value to neighboring pixels. This information can be effectively incorporated into the similarity matrix by rearranging the 2-D sparse coefficient matrix $\mathbf{Z} \in \mathbb{R}^{MN \times MN}$ into a 3-D cube $\tilde{\mathbf{Z}} \in \mathbb{R}^{M \times N \times MN}$, treating each coefficient vector as a pixel in the 3-D cube. In this work, the median filter is expressed as

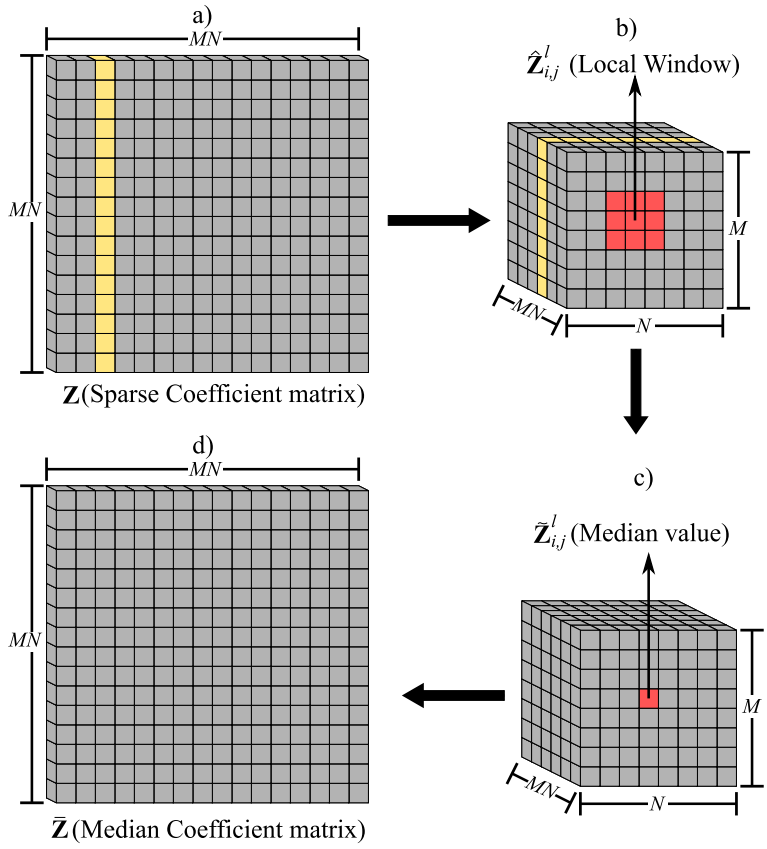
$$\tilde{\mathbf{Z}}^l_{i,j} = \text{Median}(\hat{\mathbf{Z}}^l_{i,j}), \tag{3.11}$$

where $\hat{\mathbf{Z}}^l_{i,j}$ is a 3D window of size $3 \times 3 \times 3$ (see Fig.3.2 b), and Median$(\cdot)$ is the Median operator. The optimization problem of rearranging $\tilde{\mathbf{Z}} \in \mathbb{R}^{M \times N \times MN}$ to the new $\bar{\mathbf{Z}} \in \mathbb{R}^{MN \times MN}$ is expressed as

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{R},\bar{\mathbf{Z}}} \quad & \|\mathbf{Z}\|_1 + \frac{\lambda}{2}\|\mathbf{R}\|_F^2 + \frac{\alpha}{2}\|\mathbf{Z} - \bar{\mathbf{Z}}\|_F^2 \\ \text{s.t.} \quad & \mathbf{Y} = \mathbf{Y}\mathbf{Z} + \mathbf{R},\ \text{diag}(\mathbf{Z}) = 0,\ \mathbf{Z}^T\mathbf{1} = \mathbf{1}, \end{aligned} \tag{3.12}$$

where $\bar{\mathbf{Z}}$ is the rearranged matrix after applying the median filter and α is a regularization parameter denoting the weight of the spatial information in the subspace clustering algorithm. After solving the optimization problem in Eq. 3.12, the segmentation of the data points into different subspaces is inferred using the sparse coefficients $\mathbf{Z}$. Specifically, the clustering result is obtained by applying spectral clustering to the Laplacian matrix induced by the similarity matrix $\mathbf{W} \in \mathbb{R}^{MN \times MN}$ which is defined as $\mathbf{W} = |\mathbf{Z}| + |\mathbf{Z}|^T$ [17, 18].

Figure 3.2: Visual representation of the median filter step, a) Sparse Coefficient matrix $\mathbf{Z}$, then it is reshaping in b) and a medium filter is applied to obtain the new values c) and finally it reshaped to its initial size.



The complete CSI subspace clustering algorithm is summarized in Algorithm 2.

Algorithm 2 Compressive Spectral Imaging Subspace Clustering

Input: A set of CSI $\{\mathbf{y}\}_{j=0}^{MN-1}$ measurements acquired with $\mathbf{Y} = \mathbf{\Phi F}$. The coding patterns $\mathbf{\Phi}$ is acquired with the Algorithm. 1.

1: Solve the sparse optimization problem in Eq. 3.12 using the ADMM algorithm described in the Appendix section.
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 by $\mathbf{W} = |\mathbf{Z}| + |\mathbf{Z}|^T$.
4: Apply spectral clustering [16] to the similarity graph.

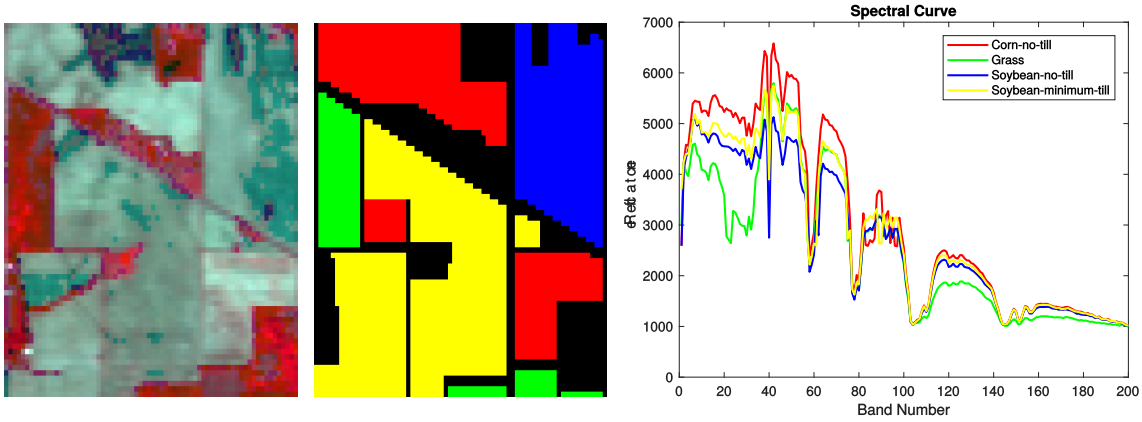
Output: Segmentation of the data: $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_\ell$

4. SIMULATIONS AND RESULTS

The proposed compressed spectral image clustering approach was tested on three real hyperspectral data sets, with different imaging environments, acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) and the Reflective Optics System Imaging Spectrometer (ROSIS). Specifically, the Indian Pines data set, the Salinas data set and the Pavia University data set, were used through the experiments.

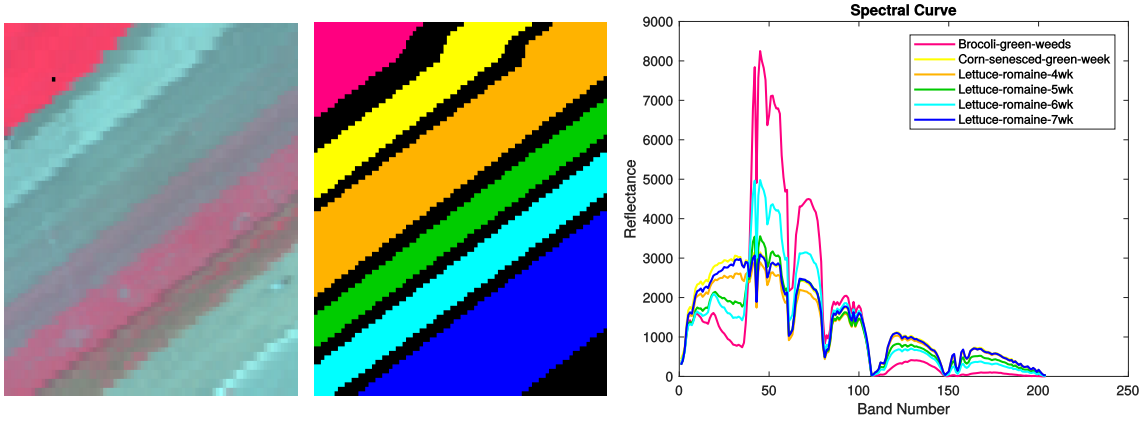
- **AVIRIS Data set: Indian Pines Image**: This hyperspectral data set was acquired by the AVIRIS sensor from the Northwestern Indian Pines test site in June 1992 [19]. The size of this image is 145×145. A total of 20 water absorption and noisy bands ($104 - 108, 150 - 163$, and 220) were removed from the original 220 bands, leaving 200 spectral features for the experiment [19]. Considering the computational efficiency, a subimage with the size of 70×70, which included four main land-cover classes: corn-no-till, grass, soybeans-no-till, and soybeans-minimum-till, was used in the experiments. The clustering was a challenging task due to the spectral signatures of the land-cover classes in this area are very similar and some of the spectral curves are mixed, as shown in Fig. 4.1(c). The false-color image and the ground truth are provided in Fig. 4.1(a) and (b).
- **AVIRIS Data set: Salinas Image**: The second hyperspectral data set was acquired by the 224-band AVIRIS sensor over the Salinas Valley, CA, USA. The size of the image is 512×217. As with the first data set, a total of 20 water absorption bands ($108 - 112, 154 - 167$, and 224) were removed. A subimage with the size of 83×86, containing six land-cover classes: Brocoli-green-weeds, corn-senesced-green-week, lettuce-romaine-4wk, lettuce-romaine-5wk, lettuce-romaine-6wk and lettuce-romaine-7wk. This image is also difficult for clustering because of the high similarity between the spectral signatures of some land-cover classes, which can be seen in Fig. 4.2(c). The false-color image and the ground truth are provided in Fig. 4.2(a) and (b), respectively.
- **ROSIS Urban Data: University of Pavia, Italy**: This scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy. The size of the image is 610×340, with 103 bands used in the experiments. A typical area for the test data with a size of 140×80, containing eight main land-cover classes: Bitumen, asphalt, trees, bricks, bare soil, metal sheet, meadows and shadows, was used. This data set contains more complex land-cover classes, and the spectral signatures of some of the classes are very similar, which results in the clustering being a more challenging task. The spectral curves of the eight land-cover classes are shown in Fig. 4.3. The false-color image and the ground truth are also provided.

Figure 4.1: AVIRIS Indian Pines test image. (a) False-color image (RGB 40,30,20). (b) Ground truth. (c) Spectral curves of the four land-cover classes.



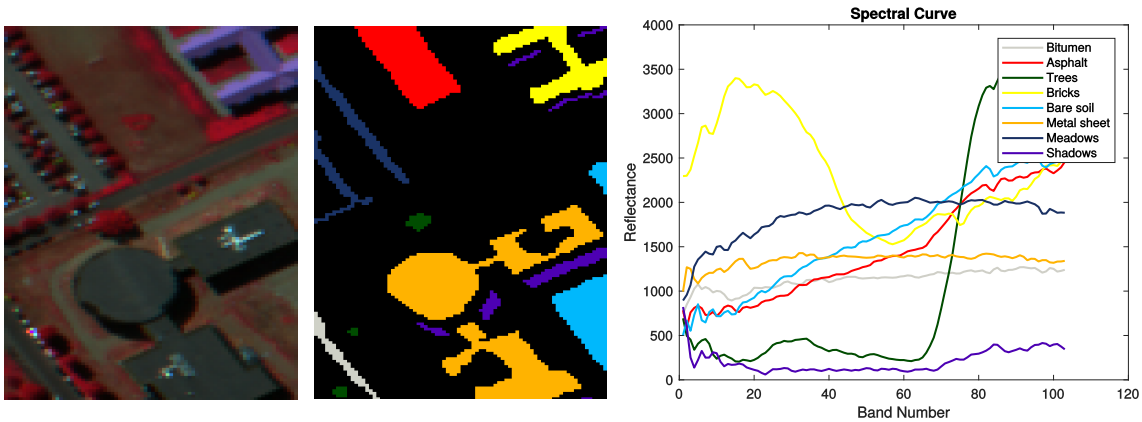
(a) (b) (c)

Figure 4.2: AVIRIS Salinas test image. (a) Salinas false-color image (RGB 70,27,17). (b) Ground truth. (c) Spectral curves of the six land-cover classes.



(a) (b) (c)

Figure 4.3: ROSIS University of Pavia test image. (a) False-color image (RGB 102,56,31). (b) Ground truth. (c) Spectral curves of the eight land-cover classes.



(a) (b) (c)

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 to a local optimum. Specifically, the regularization parameter λ in Eq. 3.12, which acts as the tradeoff between the sparsity of the coefficient and the magnitude of the noise, was set using the following formulation [18]

$$\lambda = \frac{\beta}{\gamma} \tag{4.1}$$

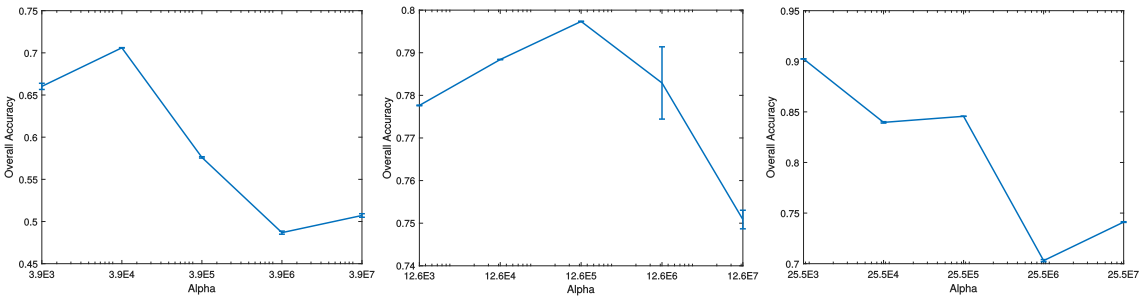
$$\gamma = \min_{j} \max_{j \neq j'} |\mathbf{y}_{j'}^{T} \mathbf{y}_{j}|, \tag{4.2}$$

where β is the adjustment coefficient and γ is a parameter related to the data set, which can be explicitly determined.

The regularization parameter α in Eq. 3.12 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 the Algorithm. 1 with $\Delta = 20$ and $S = 25$ measurement shots. Further, a 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 the other parameters fixed, is shown in Fig. 4.4. As can be seen from Fig. 4.4, the precision changes significantly with different values of α, which suggest 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, $\alpha = 12.6 \times 10^5$ for the Salinas 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 the Eq. 4.1 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 begining of each experiment using the Eq. 4.2

Figure 4.4: Analysis of parameter α: Change in the overall accuracy with various values of α. (a) Indian Pines image. (b) Salinas image. (c) University of Pavia image.



(a) (b) (c)

In the next experiments, the random-designed coding patterns are generated

from a Bernoulli distribution $\mathbf{\Phi} \sim B_e(p)$ with $p = \Delta/L$ in order to use a similar transmittance (the number of nonzero elements) to the designed coding patterns. Since both random and designed coding patterns are randomly generated, the presented results are the average of five experiments, each with a different coding pattern. All the simulations were performed using two computers: the first with an Intel(R) Xeon(R) E5-2697 v3 of 2.60 GHz, 56 cores and 196 GB of RAM. The second, with an Intel(R) Xeon(R) E5-1603 v3 of 2.80 GHz, 4 cores and 128 GB of RAM

4.1. SIMILARITY PRESERVATION

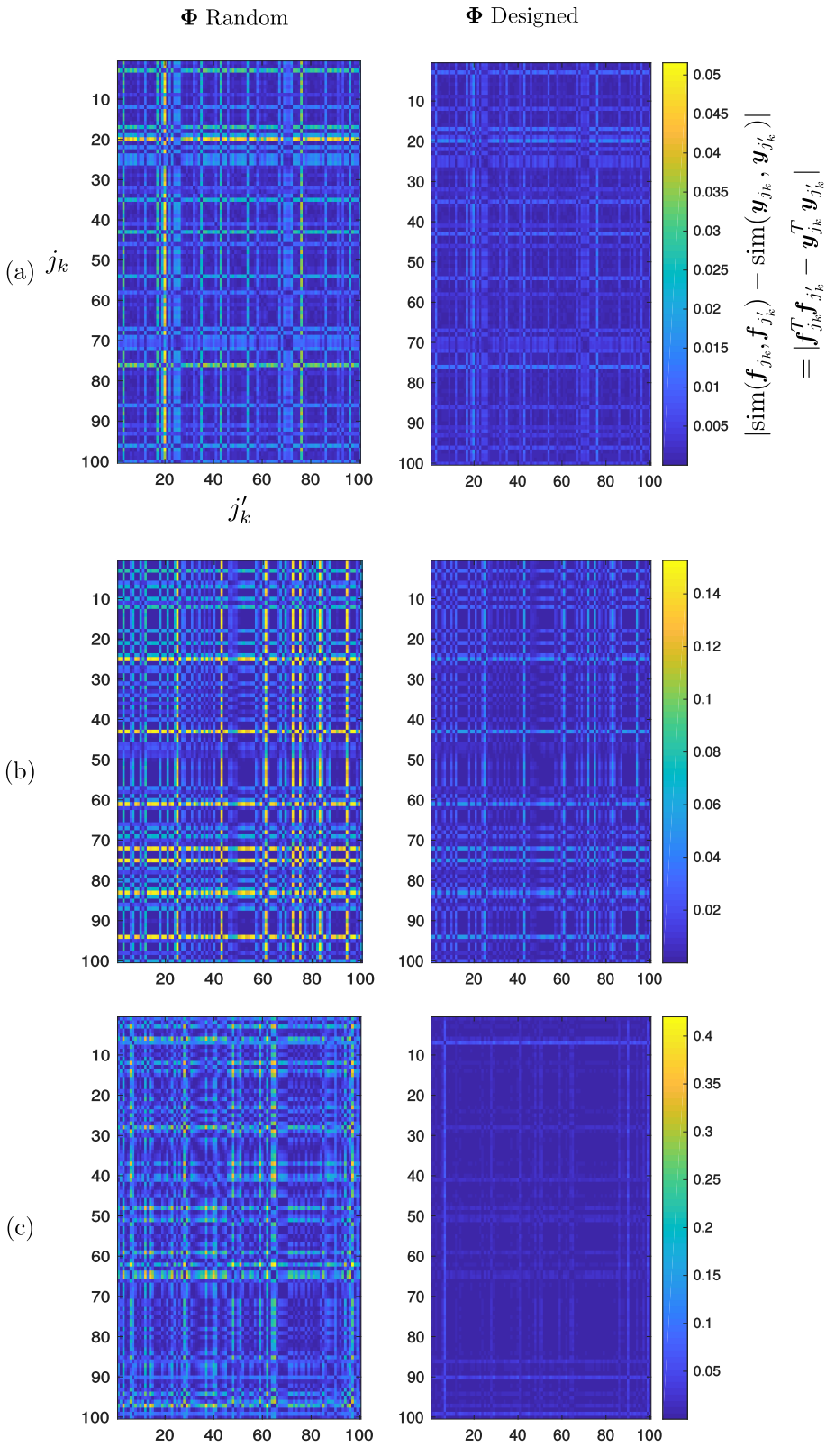
In this experiment, the performance of the proposed coding patterns design is tested. Specifically, this experiment is intended to show how well the similarity (cosine of the angle) between two spectral signatures is approximately preserved when the designed coding patterns are used. For this experiment, 100 spectral signatures are chosen at random and then are compressed using a random and designed coding pattern matrix $\mathbf{\Phi}$, generated with parameters $\Delta = 20$ and $S = 25$. Then all the spectral signatures and its compressed versions are normalized to have unit length, i.e., $\|\boldsymbol{f}\|_2 = 1$ and $\|\boldsymbol{y}\|_2 = 1$. Using the definition of similarity, presented in Eq. 3.1, the absolute error is calculated as

$$|\text{sim}(\boldsymbol{f}_{j_k}, \boldsymbol{f}_{j'_k}) - \text{sim}(\boldsymbol{y}_{j_k}, \boldsymbol{y}_{j'_k}) = |\boldsymbol{f}_{j_k}^T \boldsymbol{f}_{j'_k} - \boldsymbol{y}_{j_k}^T \boldsymbol{y}_{j'_k}|, \tag{4.3}$$

where $j, j' \in \{0, 1, \cdots, MN\}$, index a spectral signature chosen at random among MN possibilities, and $k = 1, \cdots, 100$. Figure 4.5 shows the obtained results for the three spectral images, (a)Indian Pines, (b) Salinas and (c) Pavia University. The provided results are the average of ten experiments, each with a different coding pattern.

As observed, the absolute error obtained with the designed coding patterns are significantly smaller than the obtained with a random-designed matrix $\mathbf{\Phi}$. Therefore, the proposed coding pattern design approximately preserve the similarities among the spectral signatures after the scene projection.

Figure 4.5: Absolute error between the spectral signatures similarities and the compressed measurements similarities acquired with the random and designed coding pattern design.



4.2. NOISE ANALYSIS

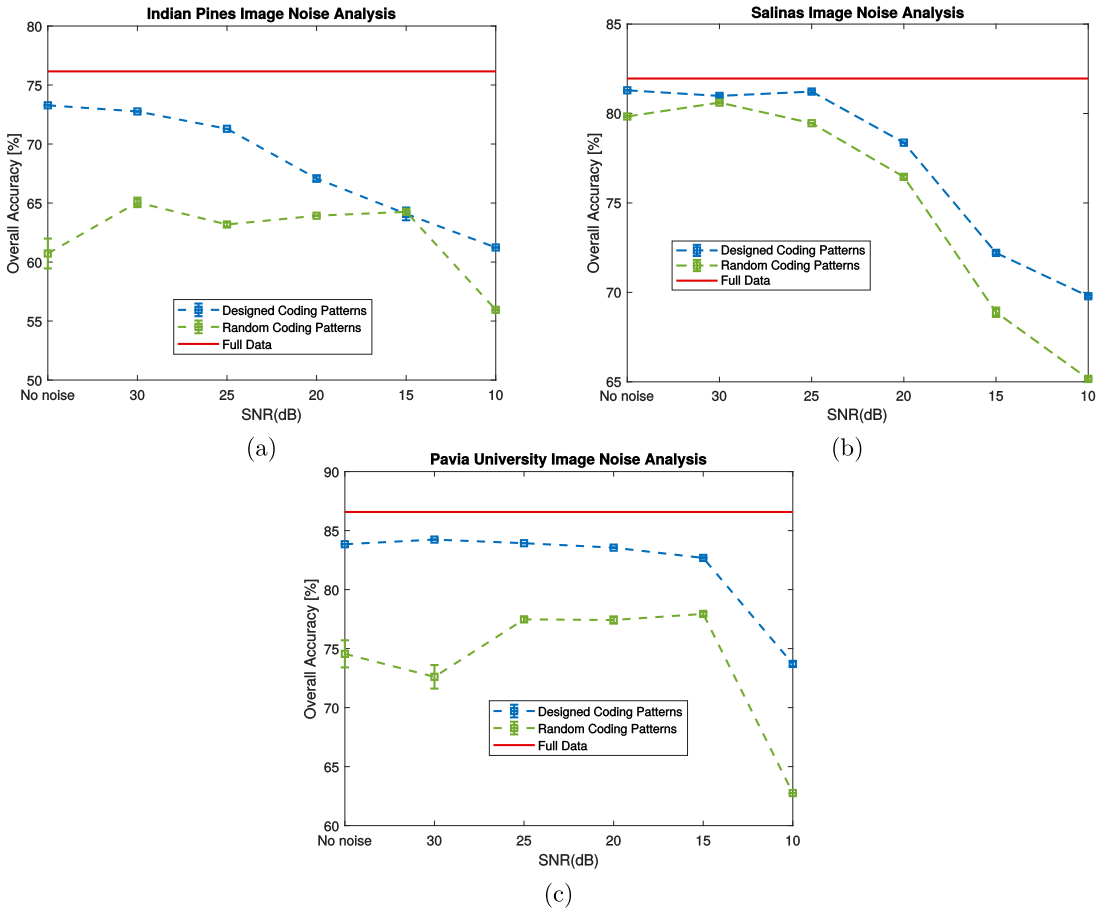
It is important to note that the acquired measurements described by the Eq. 2.10 are noise free. However in real CSI architectures, the acquired compressed measurements are contaminated with noise due to the physical limitation of the sensor. Therefore, the Eq. 2.10 should be rewritten as

$$\mathbf{Y} = \mathbf{\Phi F} + \mathbf{\Omega}, \tag{4.4}$$

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 the signal-to-noise ratio (SNR) are performed. Figure 4.6 presents the classification accuracy results obtained with the designed and the random coding patterns for the three hyperspectral data sets. Additionally, the overall clustering accuracy achieved when used the spectral image data cube (Full data) as input for the optimization problem in Eq. 3.12 is showed 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 obvious, when SNR decreases the clustering overall accuracy is affected. However, the accuracy curve, obtained with the proposed coding patterns, achieve a slow decrease rate in comparison with the random coding patterns.

Figure 4.6: Overall clustering accuracy as a function of the aggregated noise using the two coding patterns design

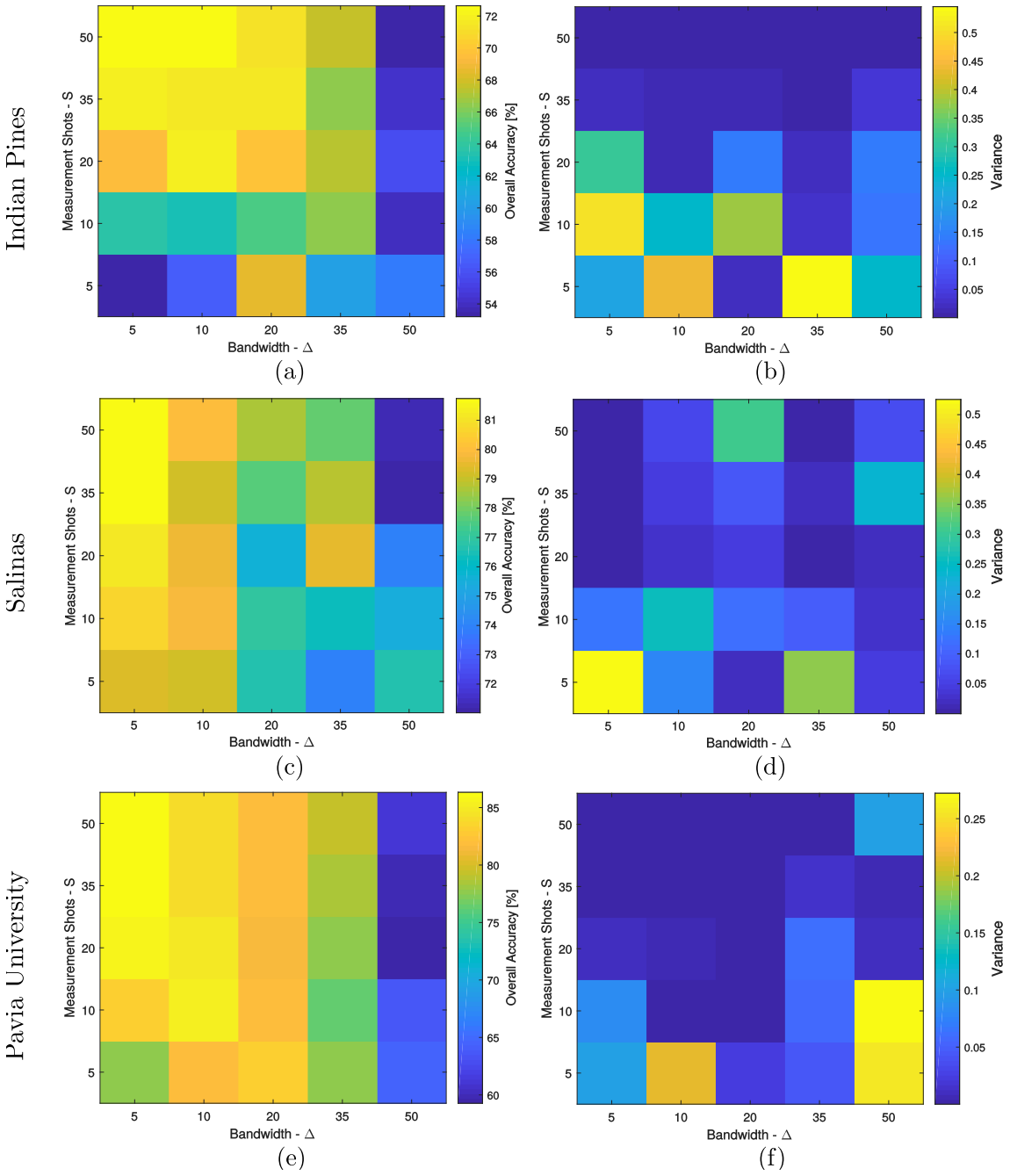


4.3. ANALYSIS OF THE CODING PATTERN DESIGN PARAMETERS

The parameters S and Δ determines 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. Figure 4.7 a) presents the obtained overall accuracy and Fig. 4.7 b) shows the variance of the obtained accuracy.

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 projection is sufficiently large, the distance between two subspaces almost remains unchanged after random projection [86]. This behavior is observed in the presented results where increasing the measurement shots leads to better classification accuracies. As described in Section 3.1, when Δ decreases, the variances for sparse random projections will also increase and large errors could be expected. On the other hand, increasing Δ, the entries of ϵ will increase leading to significant distortions in the acquired compressed measurements.

Figure 4.7: Analysis of the coded aperture design parameters. (a),(c) and (e) show the overall accuracy varying the bandwidth Δ and the measurement shots S for the hyperspectral datasets Indian Pines, Salinas and Pavia University, respectively. (b), (d) and (f) present the variance of the obtained accuracy.



4.4. 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 three hyperspectral scenes. In all the experiments, the coding patterns were generated with the parameters $\Delta = 20$ and $S = 25$. Further, white Gaussian noise with 25 dB of SNR was added to the acquired compressed measurements. The regularization parameter α for subspace clustering optimization problem in Eq. 3.12 was set for each datasets as described in the first part of this chapter. In addition, the results obtained with the sparse subspace clustering algorithm (SSC), explained in section 1.2 and with the optimization problem described in Eq. 3.10 , when the complete spectral data cube is used as input (Full-data-SSC). The λ parameter for the SSC and SSSC algorithms was set as described in the first part of this chapter.

Figure 4.8 presents the obtained visual clustering results on Indian Pines. The corresponding quantitative evaluations are shown in Table 4.1. Similarly, Fig. 4.9 and Table 4.2 present the visual clustering results and quantitative evaluation on the Salinas Valley, respectively. Finally, Fig. 4.10 and Table 4.3 presents the visual clustering results and quantitative evaluation on the Pavia University respectively. In the tables, the optimal value of each row is shown in bold and the second-best results are underlined. From Tables 4.1, 4.2 and 4.3, it can be clearly observed that the proposed clustering approach using the proposed coding patterns provides comparable results outperforming the clustering applied directly on the full spectral data cube. Furthermore, it is observed, from the visual clustering maps, that the results obtained with the proposed coding patterns are very similar to the results obtained with the Full-data. This behavior was expected since the proposed coding patterns approximately preserves the similarities among spectral pixels, as shown in section 4.1.

Figure 4.8: 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.

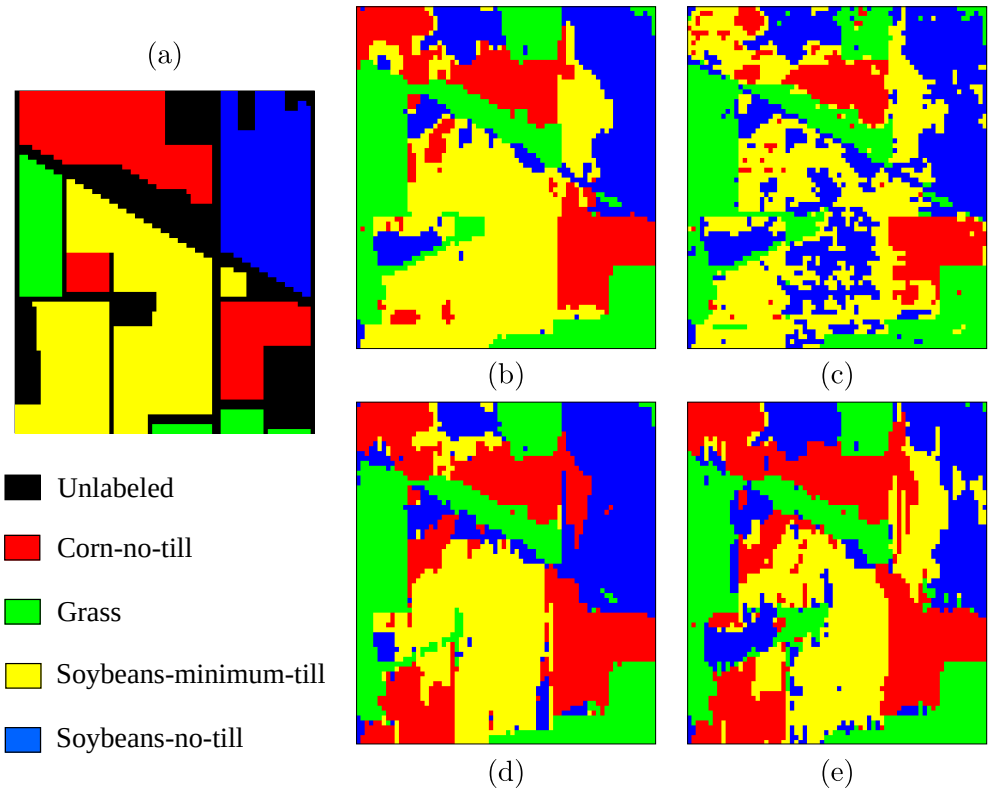


Table 4.1: Quantitative evaluation of the different clustering results for the AVIRIS Indian Pines Image.

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**
Overall Accuracy	63.83	73.07	62.62	**76.16**
Average Accuracy	69.14	**79.94**	69.35	79.09
Kappa	49.26	62.65	47.58	**65.89**

Figure 4.9: Visual clustering results on AVIRIS Salinas Valley image: (a) Ground truth. (b) Full-data, (c) Full-data-SSC , (d) Proposed-design and (e) Random-design.

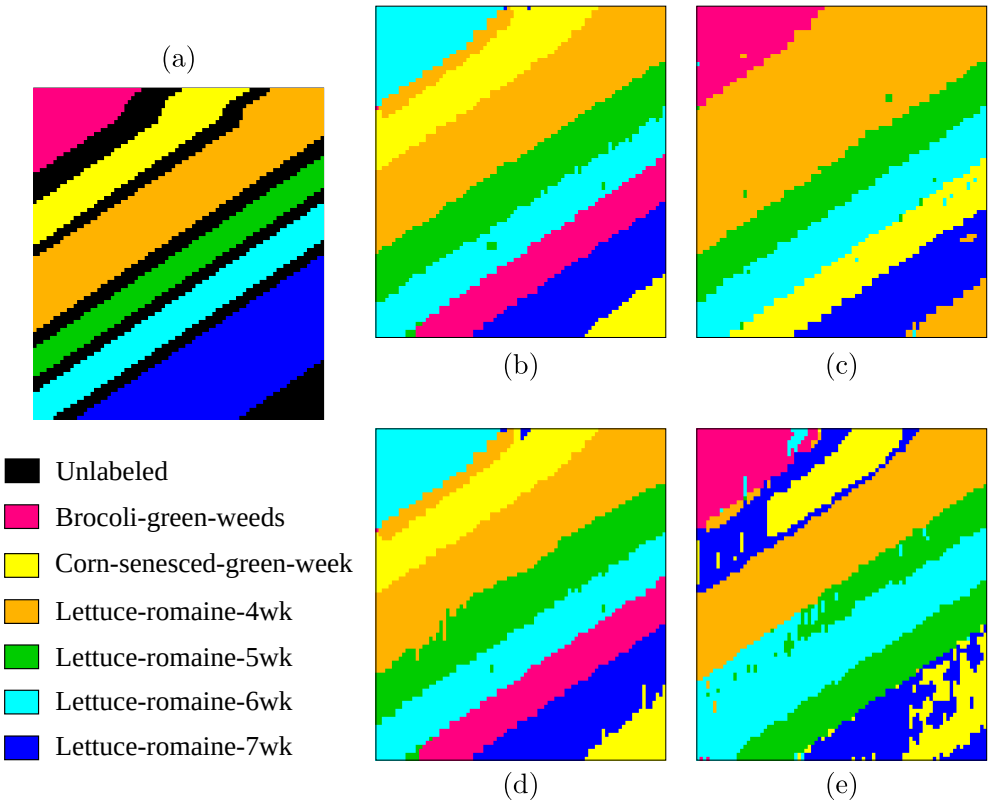


Table 4.2: Quantitative evaluation of the different clustering results with the AVIRIS Salinas Valley Image.

Class	Random-design	Proposed-design	Full-data-SSC	Full-data
Brocoli-green-weeds	97.70	0	**99.23**	0
Corn-senesced-green-week	40.51	58.30	56.66	**59.94**
lettuce-romaine-4wk	61.85	97.08	0	**98.21**
lettuce-romaine-5wk	**100**	95.41	99.48	**100**
lettuce-romaine-6wk	55.19	99.85	99.55	**100**
lettuce-romaine-7wk	98.62	97.12	**99.50**	96.87
Overall Accuracy	74.64	80.12	77.26	**81.96**
Average Accuracy	75.65	74.63	75.74	**75.84**
Kappa	68.72	75.56	71.35	**77.75**

Figure 4.10: 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.

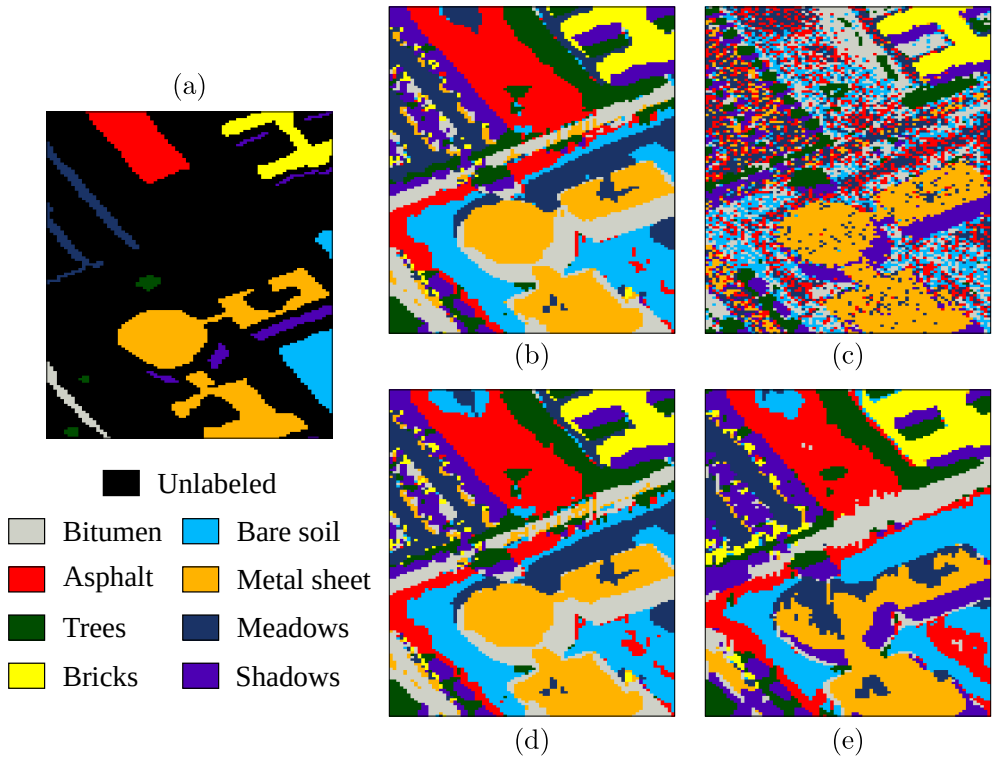


Table 4.3: Quantitative evaluation of the different clustering results with the AVIRIS Pavia University Image.

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
Overall Accuracy	78.72	83.81	71.45	**86.58**
Average Accuracy	75.09	78.41	64.28	**81.22**
Kappa	72.63	78.89	62.95	**82.50**

5. CONCLUSIONS

A new spectral image subspace clustering method, which performs all the clustering task directly on the compressed measurements, was developed in this research work. The 3D-CASSI CSI acquisition model was used since it allows to encode the spectral pixels individually before performing an spectral-wise integration. In the proposed method, a coding pattern set was first designed in order to acquire the compressed measurements. After rearranging the acquired measurements, a proposed subspace clustering algorithm based on the SSC model was used to obtain the image segmentation.

The main contribution of this work is the coding pattern design. The proposed design aims at preserving the spectral signatures separability as much as possible after the scene projection. On the other hand, the proposed subspace clustering algorithm 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.

The coding pattern design and the spectral image subspace clustering approach was validated trough several experiments. Two hyperspectral remote sensing scenes from the AVIRIS sensor and one 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 comparison with those provided by performing the clustering on the full 3D spectral image, when a properly designed coding pattern was used. Particularly, a difference of at most 4% in terms of overall accuracy was observed when comparing the clustering results obtained by SSSC with the full 3D data and SSSC with CSI measurements acquired with the proposed coding pattern design.

ADMM ALGORITHM TO SOLVE THE PROPOSED SUBSPACE CLUSTERING OPTIMIZATION PROBLEM

In this section, the solving process of the sparse optimization problem in Eq. 3.12 with the well-known ADMM algorithm is introduced. First, an auxiliary matrix $\mathbf{U} \in \mathbb{R}^{MN \times MN}$ with the same size as the sparse coefficient matrix $\mathbf{Z}$ to separate the variables. In this way, we only need to solve the following optimization problem

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

Two penalty terms corresponding to $\mathbf{U^T 1} = \mathbf{1}$ and $\mathbf{U} = \mathbf{Z} - \text{diag}(\mathbf{Z})$ are then added to the penalty function of Eq. 1 to obtain the following new optimization problem:

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{U},\bar{\mathbf{Z}}} \quad & \|\mathbf{Z}\|_1 + \frac{\lambda}{2}\|\mathbf{Y} - \mathbf{Y}\mathbf{U}\|_F^2 + \frac{\alpha}{2}\|\bar{\mathbf{Z}} - \mathbf{U}\|_F^2 \\ & + \frac{\rho}{2}\|\mathbf{U}^T\mathbf{1} - \mathbf{1}\|_2^2 + \frac{\rho}{2}\|\mathbf{U} - (\mathbf{Z} - \text{diag}(\mathbf{Z}))\|_F^2 \\ \text{s.t.} \quad & \mathbf{U}^T\mathbf{1}, \mathbf{U} = \mathbf{Z} - \text{diag}(\mathbf{Z}) \end{aligned} \tag{2}$$

It can be easily proved that the solutions to Eq. 3.12 and Eq. 1 coincide with that of Eq. 2. Next, we introduce a vector $\delta \in \mathbb{R}^{MN}$ and a matrix $\mathbf{\Delta} \in \mathbb{MN} \times \mathbb{MN}$ as Lagrange multipliers for the two equality constraints in Eq. 2 to obtain the Lagrange function as

$$\begin{aligned} \min_{\mathbf{Z},\mathbf{U},\bar{\mathbf{Z}}} \quad & \|\mathbf{Z}\|_1 + \frac{\lambda}{2}\|\mathbf{Y} - \mathbf{Y}\mathbf{U}\|_F^2 + \frac{\alpha}{2}\|\bar{\mathbf{Z}} - \mathbf{U}\|_F^2 \\ & + \frac{\rho}{2}\|\mathbf{U}^T\mathbf{1} - \mathbf{1}\|_2^2 + \frac{\rho}{2}\|\mathbf{U} - (\mathbf{Z} - \text{diag}(\mathbf{Z}))\|_F^2 \\ & + \delta^T(\mathbf{U}^T\mathbf{1} - \mathbf{1}) + \text{tr}(\mathbf{\Delta}^T(\mathbf{U} - \mathbf{Z} + \text{diag}(\mathbf{Z}))) \end{aligned} \tag{3}$$

where tr(·) denotes the trace operator of a given matrix.

The aforementioned optimization problem can then be divided intro three subproblems:

1. Updating $\mathbf{U}$ with the other four variables fixed;
2. Updating $\mathbf{Z}$ by fixing the other variables and then updating $\bar{\mathbf{Z}}$ with $\mathbf{Z}$
3. Updating δ and $\mathbf{\Delta}$ using $\mathbf{U}$ and $\mathbf{Z}$.

Specifically, the ADMM utilizes an iterative procedure as follows

1. For subproblem 1., $\mathbf{U}^{(k+1)}$ is obtained by minimizing the L with respect to $\mathbf{U}$, while $(\mathbf{Z}^{(k)}, \bar{\mathbf{Z}}^{(k)}, \delta^{(k)}, \mathbf{\Delta}^{(k)})$ are fixed. We calculate the derivative of L with respect to $\mathbf{U}$ and set it to zero to obtain the calculation formula of $\mathbf{U}$ as follows:

$$(\lambda \mathbf{Y}^T\mathbf{Y} + \alpha\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T + \rho\mathbf{I})\mathbf{U}^{(k+1)} = \lambda \mathbf{Y}^T\mathbf{Y} + \alpha\bar{\mathbf{Z}}^{(k)} \tag{4}$$

$$+ \rho(\mathbf{1}\mathbf{1}^T + \mathbf{Z}^{(k)}) - \mathbf{1}\delta^{(k)^T} - \mathbf{\Delta}^{(k)}. \tag{5}$$

2. For subproblem 2., $\mathbf{Z}^{(k+1)}$ can be obtained by minimizing L with respect to $\mathbf{Z}$, while $(\mathbf{U}^{(k+1)}, \bar{\mathbf{Z}}^{(k)}, \delta^{(k)}, \mathbf{\Delta}^{(k)})$ are fixed

$$\mathbf{Z}^{(k+1)} = \mathbf{J} - \text{diag}(\mathbf{J}), \mathbf{J} \triangleq \Gamma_{\frac{1}{\rho}}\left(\mathbf{U}^{(k+1)} + \frac{\mathbf{\Delta}^{(k)}}{\rho}\right), \tag{6}$$

where $\Gamma_{1/\rho}(\cdot)$ is a shrinkage-thresholding operator, $\Gamma_{1/\rho}(v) = (|u| - (1/\rho))_+ \text{sgn}(v)$, and the operator $(\cdot)_+$ returns its arguments if it is nonnegative and returns zero otherwise. We then update $\bar{\mathbf{Z}}^{(k+1)}$ utilizing $\mathbf{Z}^{(k+1)}$ with Eq. 3.2.

3. For subproblem 3., the Lagrange multipliers $\delta^{(k+1)}$ and $\mathbf{\Delta}^{(k+1)}$ are obtained through a gradient ascent update with the step size $\rho = 300$.

$$\delta^{(k+1)} = \delta^{(k)} + \rho\left(\mathbf{Z}^{(k+1)}\mathbf{1} - \mathbf{1}\right) \tag{7}$$

$$\mathbf{\Delta}^{(k+1)} = \mathbf{\Delta}^{(k)} + \rho\left(\mathbf{U}^{(k+1)} - \mathbf{Z}^{(k+1)}\right). \tag{8}$$

These three steps are repeated until convergence is achieved or the number of iterations exceeds the maximum iteration number. Specifically, the iteration is terminated when we have $\|\mathbf{U}^{(k)^T}\mathbf{1} - \mathbf{1}\|_\infty \leq \epsilon, \|\mathbf{U}^{(k)} - \mathbf{Z}^{(k)}\|_\infty \leq \epsilon, \|\mathbf{U}^{(k)} - \mathbf{U}^{k-1}\|_\infty \leq \epsilon$, where ϵ denotes the error tolerance for the primal and dual residuals.

Referencias

[1] Ariolfo Camacho Velasco, César Augusto Vargas García, Fernando Antonio Rojas Morales, Sergio Fernando Castillo Castelblanco, and Henry Arguello Fuentes. Aplicaciones y retos del sensado remoto hiperespectral en la geología colombiana. *Facultad de Ingeniería*, 24(40):17–29, 2015.

[2] Liang Gao, Robert T Kester, Nathan Hagen, and Tomasz S Tkaczyk. Snapshot image mapping spectrometer (ims) with high sampling density for hyperspectral microscopy. *Optics express*, 18(14):14330–14344, 2010.

[3] John P Kerekes and John R Schott. Hyperspectral imaging systems. *Hyperspectral data exploitation: Theory and applications*, pages 19–45, 2007.

[4] Wei He, Hongyan Zhang, Liangpei Zhang, and Huanfeng Shen. Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):3050–3061, 2015.

[5] Yannick Lanthier, Abdou Bannari, Driss Haboudane, John R Miller, and Nicolas Tremblay. Hyperspectral data segmentation and classification in precision agriculture: A multi-scale analysis. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages II–585. IEEE, 2008.

[6] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013.

[7] Dimitris Manolakis, David Marden, and Gary A Shaw. Hyperspectral image processing for automatic target detection applications. *Lincoln Laboratory Journal*, 14(1):79–116, 2003.

[8] X Briottet, Y Boucher, A Dimmeler, A Malaplate, A Cini, M Diani, HHPT Bekman, P Schwering, T Skauli, I Kasen, et al. Military applications of hyperspectral imagery. In *Defense and Security Symposium*, pages 62390B–62390B. International Society for Optics and Photonics, 2006.

[9] Gary A Shaw and Hsiao-Hua K Burke. Spectral imaging for remote sensing. *Lincoln Laboratory Journal*, 14(1):3–28, 2003.

[10] René Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, pages 52–68, 2011.

[11] Teng Zhang, Arthur Szlam, and Gilad Lerman. Median k-flats for hybrid linear modeling with many outliers. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 234–241. IEEE, 2009.

[12] João Paulo Costeira and Takeo Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision*, 29(3):159–179, 1998.

[13] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1945–1959, 2005.

[14] Yi Ma, Harm Derksen, Wei Hong, and John Wright. Segmentation of multivariate mixed data via lossy data coding and compression. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):1546–1562, 2007.

[15] Michael E Tipping and Christopher M Bishop. Mixtures of probabilistic principal component analyzers. *Neural computation*, 11(2):443–482, 1999.

[16] Andrew Y Ng, Michael I Jordan, Yair Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 2:849–856, 2002.

[17] Ehsan Elhamifar and René Vidal. Sparse subspace clustering. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2790–2797. IEEE, 2009.

[18] Ehsan Elhamifar and Rene Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2765–2781, 2013.

[19] Hongyan Zhang, Han Zhai, Liangpei Zhang, and Pingxiang Li. Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6):3672–3684, 2016.

[20] Michael Steinbach, Levent Ertöz, and Vipin Kumar. The challenges of clustering high dimensional data. In *New directions in statistical physics*, pages 273–309. Springer, 2004.

[21] Wei He, Hongyan Zhang, Liangpei Zhang, Wilfried Philips, and Wenzhi Liao. Weighted sparse graph based dimensionality reduction for hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 13(5):686–690, 2016.

[22] Xun Cao, Tao Yue, Xing Lin, Stephen Lin, Xin Yuan, Qionghai Dai, Lawrence Carin, and David J Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5):95–108, 2016.

[23] Hoover Rueda, Daniel Lau, and Gonzalo R Arce. Multi-spectral compressive snapshot imaging using rgb image sensors. *Optics express*, 23(9):12207–12221, 2015.

[24] Claudia V Correa, Carlos A Hinojosa, Gonzalo R Arce, and Henry Arguello. Multiple snapshot colored compressive spectral imager. *Optical Engineering*, 56(4):041309–041309, 2017.

[25] Claudia V Correa, Henry Arguello, and Gonzalo R Arce. Snapshot colored compressive spectral imager. *JOSA A*, 32(10):1754–1763, 2015.

[26] Henry Arguello and Gonzalo R Arce. Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4):1896–1908, 2014.

[27] Xing Lin, Gordon Wetzstein, Yebin Liu, and Qionghai Dai. Dual-coded compressive hyperspectral imaging. *Optics letters*, 39(7):2044–2047, 2014.

[28] Ana Ramirez, Henry Arguello, Gonzalo R Arce, and Brian M Sadler. Spectral image classification from optimal coded-aperture compressive measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 52(6):3299–3309, 2014.

[29] Robert O Green, Michael L Eastwood, Charles M Sarture, Thomas G Chrien, Mikael Aronsson, Bruce J Chippendale, Jessica A Faust, Betina E Pavri, Christopher J Chovit, Manuel Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote Sensing of Environment*, 65(3):227–248, 1998.

[30] Esko Herrala, Jukka T Okkonen, Timo S Hyvarinen, Mauri Aikio, and Jorma Lammasniemi. Imaging spectrometer for process industry applications. In *Optics for Productivity in Manufacturing*, pages 33–40. International Society for Optics and Photonics, 1994.

[31] Hannah R Morris, Clifford C Hoyt, and Patrick J Treado. Imaging spectrometers for fluorescence and raman microscopy: acousto-optic and liquid crystal tunable filters. *Applied spectroscopy*, 48(7):857–866, 1994.

[32] Bert Geelen, Murali Jayapala, Nicolaas Tack, and Andy Lambrechts. Low-complexity image processing for a high-throughput low-latency snapshot multispectral imager with integrated tiled filters. In *SPIE Defense, Security, and Sensing*, pages 87431E–87431E. International Society for Optics and Photonics, 2013.

[33] Jason M Eichenholz, Nick Barnett, Yishung Juang, Dave Fish, Steve Spano, Erik Lindsley, and Daniel L Farkas. Real-time megapixel multispectral bioimaging. In *BiOS*, pages 75681L–75681L. International Society for Optics and Photonics, 2010.

[34] Dimitris Manolakis and Gary Shaw. Detection algorithms for hyperspectral imaging applications. *IEEE Signal Processing Magazine*, 19(1):29–43, 2002.

[35] Michael T Eismann, Alan D Stocker, and Nasser M Nasrabadi. Automated hyperspectral cueing for civilian search and rescue. *Proceedings of the IEEE*, 97(6):1031–1055, 2009.

[36] NK Patel, C Patnaik, S Dutta, AM Shekh, and AJ Dave. Study of crop growth parameters using airborne imaging spectrometer data. *International Journal of Remote Sensing*, 22(12):2401–2411, 2001.

[37] B Hörig, F Kühn, F Oschütz, and F Lehmann. Hymap hyperspectral remote sensing to detect hydrocarbons. *International Journal of Remote Sensing*, 22(8):1413–1422, 2001.

[38] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152. ACM, 1992.

[39] Vladimir Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[40] J Anthony Gualtieri and Robert F Cromp. Support vector machines for hyperspectral remote sensing classification. In *The 27th AIPR Workshop: Advances in Computer-Assisted Recognition*, pages 221–232. International Society for Optics and Photonics, 1999.

[41] Farid Melgani and Lorenzo Bruzzone. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8):1778–1790, 2004.

[42] Lorenzo Bruzzone, Mingmin Chi, and Mattia Marconcini. A novel transductive svm for semisupervised classification of remote-sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(11):3363–3373, 2006.

[43] Francesca Bovolo, Lorenzo Bruzzone, Mattia Marconcini, and Claudio Persello. A novel context-sensitive svm for classification of remote sensing images. 2006.

[44] Gustavo Camps-Valls, Luis Gomez-Chova, Jordi Muñoz-Marí, Joan Vila-Francés, and Javier Calpe-Maravilla. Composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 3(1):93–97, 2006.

[45] Hsiao-Yun Huang and Bor-Chen Kuo. Double nearest proportion feature extraction for hyperspectral-image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4034–4046, 2010.

[46] Begüm Demir and Sarp Erturk. Empirical mode decomposition of hyperspectral images for support vector machine classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4071–4084, 2010.

[47] Lance Parsons, Ehtesham Haque, and Huan Liu. Subspace clustering for high dimensional data: a review. *ACM SIGKDD Explorations Newsletter*, 6(1):90–105, 2004.

[48] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[49] Jingyu Yan and Marc Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *European conference on computer vision*, pages 94–106. Springer, 2006.

[50] Alvina Goh and René Vidal. Segmenting motions of different types by unsupervised manifold clustering. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007.

[51] Teng Zhang, Arthur Szlam, Yi Wang, and Gilad Lerman. Hybrid linear modeling via local best-fit flats. *International journal of computer vision*, 100(3):217–240, 2012.

[52] Lihi Zelnik-Manor and Michal Irani. Degeneracies, dependencies and their implications in multi-body and multi-sequence factorizations. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–287. IEEE, 2003.

[53] Guangliang Chen and Gilad Lerman. Spectral curvature clustering (scc). *International Journal of Computer Vision*, 81(3):317–330, 2009.

[54] Richard E Bellman. *Adaptive control processes: a guided tour*. Princeton university press, 2015.

[55] Gonzalo R Arce, David J Brady, Lawrence Carin, Henry Arguello, and David S Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Processing Magazine*, 31(1):105–115, 2014.

[56] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[57] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[58] David L Donoho and Xiaoming Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

[59] Richard G Baraniuk. Compressive sensing. *IEEE signal processing magazine*, 24(4), 2007.

[60] Henry Arguello Gonzalo R. Arce Carlos A. Hinojosa, Claudia V. Correa. Compressive spectral imaging using multiple snapshot colored-mosaic detector measurements, 2016.

[61] David Kittle, Kerkil Choi, Ashwin Wagadarikar, and David J Brady. Multiframe image estimation for coded aperture snapshot spectral imagers. *Applied Optics*, 49(36):6824–6833, 2010.

[62] Henry Arguello and Gonzalo R Arce. Code aperture optimization for spectrally agile compressive imaging. *JOSA A*, 28(11):2400–2413, 2011.

[63] Joel Tropp, Stephen J Wright, et al. Computational methods for sparse solution of linear inverse problems. *Proceedings of the IEEE*, 98(6):948–958, 2010.

[64] Joel Tropp, Anna C Gilbert, et al. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007.

[65] DL Donoho, Y Tsaig, I Drori, and JL Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit 2006. *Preprint*, 2006.

[66] Deanna Needell and Joel A Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

[67] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *Signal Processing, IEEE Transactions on*, 57(7):2479–2493, 2009.

[68] José M Bioucas-Dias and Mário AT Figueiredo. A new twist: two-step iterative shrinkage/thresholding algorithms for image restoration. *Image Processing, IEEE Transactions on*, 16(12):2992–3004, 2007.

[69] Mário AT Figueiredo, Robert D Nowak, and Stephen J Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *Selected Topics in Signal Processing, IEEE Journal of*, 1(4):586–597, 2007.

[70] Shihao Ji, Ya Xue, and Lawrence Carin. Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, 56(6):2346–2356, 2008.

[71] Rick Chartrand. Exact reconstruction of sparse signals via nonconvex minimization. *Signal Processing Letters, IEEE*, 14(10):707–710, 2007.

[72] Alan Miller. *Subset selection in regression*. CRC Press, 2002.

[73] Holger Rauhut. Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92, 2010.

[74] Hoover Rueda, Henry Arguello, and Gonzalo R Arce. Dmd-based implementation of patterned optical filter arrays for compressive spectral imaging. *JOSA A*, 32(1):80–89, 2015.

[75] Hoover Rueda, Henry Arguello, and Gonzalo R Arce. Compressive spectral imaging based on colored coded apertures. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 7799–7803. IEEE, 2014.

[76] Jessica Lin and Dimitrios Gunopulos. Dimensionality reduction by random projection and latent semantic indexing. In *proceedings of the Text Mining Workshop, at the 3rd SIAM International Conference on Data Mining*, 2003.

[77] Laurens Van Der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative. *J Mach Learn Res*, 10:66–71, 2009.

[78] Richard G Baraniuk, Volkan Cevher, and Michael B Wakin. Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective. *Proceedings of the IEEE*, 98(6):959–971, 2010.

[79] Qian Du and James E Fowler. Low-complexity principal component analysis for hyperspectral image compression. *The International Journal of High Performance Computing Applications*, 22(4):438–448, 2008.

[80] Avner Magen. Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications. *Randomization and approximation techniques in computer science*, pages 953–953, 2002.

[81] Robert Calderbank, Sina Jafarpour, and Robert Schapire. Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain. *preprint*, 2009.

[82] Mahdi Soltanolkotabi, Emmanuel J Candes, et al. A geometric analysis of subspace clustering with outliers. *The Annals of Statistics*, 40(4):2195–2238, 2012.

[83] Reinhard Heckel, Michael Tschannen, and Helmut Bölcskei. Dimensionality-reduced subspace clustering. *arXiv preprint arXiv:1507.07105*, 2015.

[84] Michael Tschannen and Helmut Bölcskei. Noisy subspace clustering via matching pursuits. *arXiv preprint arXiv:1612.03450*, 2016.

[85] Yining Wang, Yu-Xiang Wang, and Aarti Singh. A theoretical analysis of noisy sparse subspace clustering on dimensionality-reduced data. *arXiv preprint arXiv:1610.07650*, 2016.

[86] Gen Li and Yuantao Gu. Restricted isometry property of gaussian random projection for finite set of subspaces. *arXiv preprint arXiv:1704.02109*, 2017.

[87] Xianghui Mao and Yuantao Gu. Compressed subspace clustering: A case study. In *Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on*, pages 453–457. IEEE, 2014.

[88] Santosh S Vempala. *The random projection method*, volume 65. American Mathematical Soc., 2005.

[89] Dimitris Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of computer and System Sciences*, 66(4):671–687, 2003.

[90] Ping Li, Trevor J Hastie, and Kenneth W Church. Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 287–296. ACM, 2006.

[91] Sean Fox, Stephen Tridgell, Craig Jin, and Philip HW Leong. Random projections for scaling machine learning on fpgas. In *Field-Programmable Technology (FPT), 2016 International Conference on*, pages 85–92. IEEE, 2016.

[92] Mark A Davenport, Petros T Boufounos, Michael B Wakin, and Richard G Baraniuk. Signal processing with compressive measurements. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):445–460, 2010.

[93] Henry Arguello and Gonzalo R Arce. Rank minimization code aperture design for spectrally selective compressive imaging. *IEEE Transactions on Image Processing*, 22(3):941–954, 2013.

[94] Samuel Kaski. Dimensionality reduction by random mapping: Fast similarity computation for clustering. In *Neural Networks Proceedings, 1998. IEEE World Congress on Computational Intelligence. The 1998 IEEE International Joint Conference on*, volume 1, pages 413–418. IEEE, 1998.

[95] Degui Gu, Alan R Gillespie, Anne B Kahle, and Frank D Palluconi. Autonomous atmospheric compensation (aac) of high resolution hyperspectral thermal infrared remote-sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 38(6):2557–2570, 2000.

[96] Ella Bingham and Heikki Mannila. Random projection in dimensionality reduction: applications to image and text data. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 245–250. ACM, 2001.