



Classification of Remote Sensing Images Through Reweighted Sparse Subspace Representation Using Compressed Data

Jianchen Zhu^{1,3}, Shengjie Zhao^{1,2*}, Di Wu³

¹ The Key Laboratory of Embedded System and Service Computing, Ministry of Education (Tongji University), Shanghai 200082, China

² School of Software Engineering, Tongji University, Shanghai 200082, China

³ College of Electronics and Communication Engineering, Tongji University, Shanghai 200082, China

Corresponding Author Email: shengjiezhao@tongji.edu.cn

<https://doi.org/10.18280/ts.380103>

ABSTRACT

Received: 2 November 2020

Accepted: 15 January 2021

Keywords:

coherence-based coded aperture, reweighted sparse subspace clustering (RSSC), spectral image clustering

In many real-world scenarios, subspace clustering essentially aims to cluster unlabeled high-dimensional data into a union of finite-dimensional linear subspaces. The problem is that the data are always high-dimensional, with the increase of the computation, storage, and communication of various intelligent data-driven systems. This paper attempts to develop a method to cluster spectral images directly using the measurements of compressive coded aperture snapshot spectral imager (CASSI), eliminating the need to reconstruct the entire data cube. Assuming that compressed measurements are drawn from multiple subspaces, a novel algorithm was developed by solving a 1-norm minimization problem, which is known as reweighted sparse subspace clustering (RSSC). The proposed algorithm clusters the compressed measurements into different subspaces, which greatly improves the clustering accuracy over the SSC algorithm by adding a reweighted step. The compressed CASSI measurements obtained using the coherence-based coded aperture can improve the performance of the proposed spectral image clustering method. The accuracy of our spectral image clustering approach was verified through simulations on two real datasets.

1. INTRODUCTION

Spectral imaging (SI) combines two-dimensional (2D) imaging with spectroscopy to capture spatial information of a scene across multiple wavelengths. During the SI, the desired three-dimensional (3D) data cube is regarded as spectral images with two coordinates representing the spatial domain, and the third coordinate corresponding to the spectral wavelengths.

Spectral images are widely used in various fields, such as remote sensing, computer vision, and medical diagnosis. In traditional SI technologies, namely, push-broom spectral imaging [1] and spectrometers based on optical bandpass filters [2], the scene must be scanned along spatial lines, or a group of bandpass filters be adjusted to acquire different spectral bands. The disadvantage of these technologies is that the volume of the data cube increases proportionally to the required spatial or spectral resolution, resulting in an exponential increase in cost and time of data acquisition.

In fact, it is wavelength that determines the amount of radiation reflected, scattered, absorbed, or emitted by each material. As a result, spectral signature is highly valued in many real-world applications [3, 4], including but not limited to classification [5, 6], target detection [7, 8], and spectral unmixing [9, 10].

Classification can be performed in a supervised or semi-supervised manner, depending on the availability of labeled information. Unfortunately, unsupervised techniques like subspace clustering only serves as an alternative to groups a set of similar pixels, because many practical applications do

not have labeled samples in advance.

To date, many clustering methods have been developed with different working mechanisms. The existing clustering algorithms roughly fall into four categories [11-14]: (1) centroid-based methods, (2) density-based methods, (3) biological methods, and (4) spectral-based methods [15, 16]. The spectral-based methods stand out for their excellent and robust performance on spectral images. Spectral-based clustering usually covers two stages: (1) establishing an adjacency matrix about the relationship between all points; 2) applying centroid clustering to the Laplacian matrix to complete final segmentation.

Assuming that the spectral signatures with the land cover class belong to the same low-dimensional subspace, Elhamifar et al. and Li et al. [17, 18] proposed spectral-based approaches like sparse subspace clustering (SSC) for unsupervised classification of spectral images by representing every spectral pixel as a linear combination of other spectral signatures in the scene. The SSC is achieved by solving the sparse convex optimization problem, which ensures that the spectral signatures corresponding to these coefficients belong to the same subspace. Nevertheless, spectral image clustering is often a complex and compute-intensive task, owing to the high dimensionality of spectral datasets. To solve the problem, the key is to reduce the dimensionality of the spectral images.

The SSC algorithm has been successfully applied to cluster spectral images obtained by a compressed spectral imaging (CSI) system, which needs far fewer sampling resources compared to traditional spectral imaging sensors. Several approaches can be utilized to complete spectral image

clustering based on compressed data [19-22]. The relevant authors designed a down-sampling strategy to extract samples, whose probability is inversely proportional to the number of samples in their own subspaces, and combined the strengths of different methods.

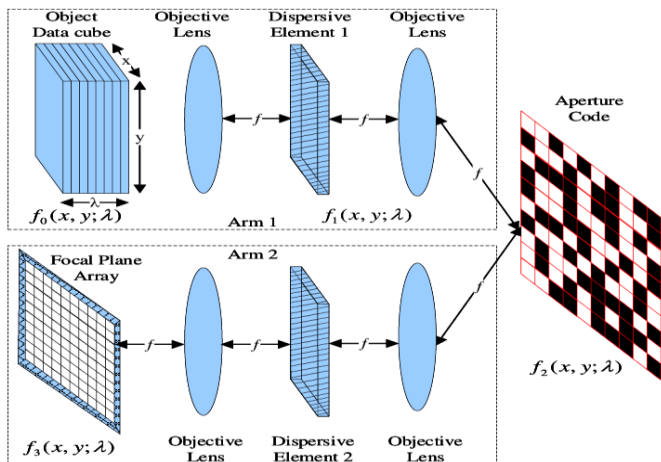


Figure 1. Symmetrical structure of CASSI system

Inspired by the symmetrical structure of compressive coded aperture snapshot spectral imager (CASSI) system (Figure 1), our work is grounded on the data collected by a novel compressive imaging system called 3D-CASSI. The novel system first encodes all the information of the scene induced by the 3D coded aperture, and then integrates the coded information along the spectral dimension. Unlike the conventional systems [23-27], 3D-CASSI ensures that each spatial position of the collected measurements contains compressed information with a uniquely coded spectral signature [28-31].

The focus of this research lies in the unsupervised classification of each pixel in the scene, as one of the known classes from the compressed measurements. The premise is that the measurements with the same feature are drawn from the same low-dimensional subspace, corresponding to a specific land cover class. The proposed approach illustrates each spectral signature, derived from its respective subspace, as a linear/affine combination of other spectral pixels, such that nonzero entries are distributed in the same class. Moreover, the representation of similar materials as adjacent pixels in the spectral image helps to extract more information from the data, and reduce the representation error using a smoothing filter on the sparse matrix [32].

Previous works [33-35] have shown that the inclusion of spatial information in spectral clustering, that is, preserving the rich spatial information in spectral images, can promote clustering accuracy associated with total change (TV) regularization. This means the use of TV regularization eliminates every two consecutive pixels. The modified SSC-based methods, as a strategy of local averaging, is applicable to spatial information in spectral pixel clustering on compressed domain problem. Nonetheless, these methods are extremely heuristic and unrepresentative in more complex areas of land cover distribution. There is still ample room to improve their clustering performance. To improve the clustering accuracy at a low computing cost, this paper replaces the 1-norm minimization approach (SSC and its variants) with l-norm minimization approach associated with an iterative weighting (RSCC).

In general, the performance of compressive spectral image clustering depends on the structure of coded apertures and the targeted spectral-based method. Particularly, the design of the coherence-based coded aperture is more accurate in spectral clustering than the traditional random coded aperture. Hence, the coded aperture is often designed using compressed subspace clustering (SSC) [36] or SC conditions with reduced dimensions [37]. Recent years saw the emergence of a coded aperture design for compressive subspace clustering. However, the designer did not theoretically analyze how to preserve the projected subspaces. Here, the authors propose to find the connection between affinity and distance obtained by introducing a projection F-norm distance after the sensing process, and construct a metric space for a low-dimensional subspace set, which depends on the coded aperture.

The contributions of this paper are summarized follows: First, our general framework can analyze the performance of RSSC algorithm, when it is applied to the data compressed by coherence-based coding pattern, rather than the random matrix adopted by Lin et al. and Arguello et al. [38, 39]; that is, the designed matrix realized by the 3D-CASSI system can compress the data into a low-dimensional space; Second, our reweighted sparse subspace clustering algorithm can perform spectral pixel clustering directly from the CASSI measurements, thereby elevating the clustering accuracy without greatly complicating the computation.

2. PROBLEM DESCRIPTION

This section firstly introduces the 3D-CASSI system model for the acquisition of CASSI measurements, and then formulates the subspaces clustering problems and corresponding rules to handle affine subspaces with noise perturbation.

2.1 General matrix-based 3D CASSI model

As shown in Figure 2, the 3D CASSI system firstly uses a 3D coded aperture (a 2D-coded aperture ensemble or an array of encoded patterns) to modulate the voxels of the spectral scene. Then, the coded spectral pixels are integrated along the spectral axis into the focal plane array (FPA) detector. Let $T_{m,n,k}^s$ be the discrete form of the time-varying 3D coded aperture, and $F_{m,n,k}$ be the discrete form of source, where both m and n are spatial coordinates, k is spectral component, and s is snapshot number. Then, the s -th output on the FPA in discrete form can be expressed as:

$$Y_{m,n}^s = \sum_{k=0}^{L-1} T_{m,n,k}^s F_{m,n,k} + w_{m,n}, \quad (1)$$

where, $Y_{m,n}^s$ is the attained measurement of the $(m, n)^{th}$ position on the detector with $M \times N$ dimensions; $w_{m,n}$ is the additive noise of the system.

Formula (1) can be vectorized as:

$$y^s = \hat{H}^s f + e. \quad (2)$$

where, $y^s \in R^{MN}$ and $f \in R^{MNL}$ are the vectorized representations of $Y_{m,n}^s$ and $F_{m,n,k}$, respectively; $\hat{H}^s \in R^{MN \times MNL}$ is the coded apertures $T_{m,n,k}^s$. Nevertheless, a single

shot CASSI measurement cannot provide enough compressed measurements, if the scenes are very diverse or the spatial scenes contain too many details. To solve the problem, the CASSI imaging system treats multiple measurement shots, instead of a single measurement shot, as separate FPA measurements; each FPA measurement has a distinct code aperture pattern during the fixed integration time of the detector. The output of the multi-shot system can be written as:

$$y^S = \hat{H}^S f + e, \quad (3)$$

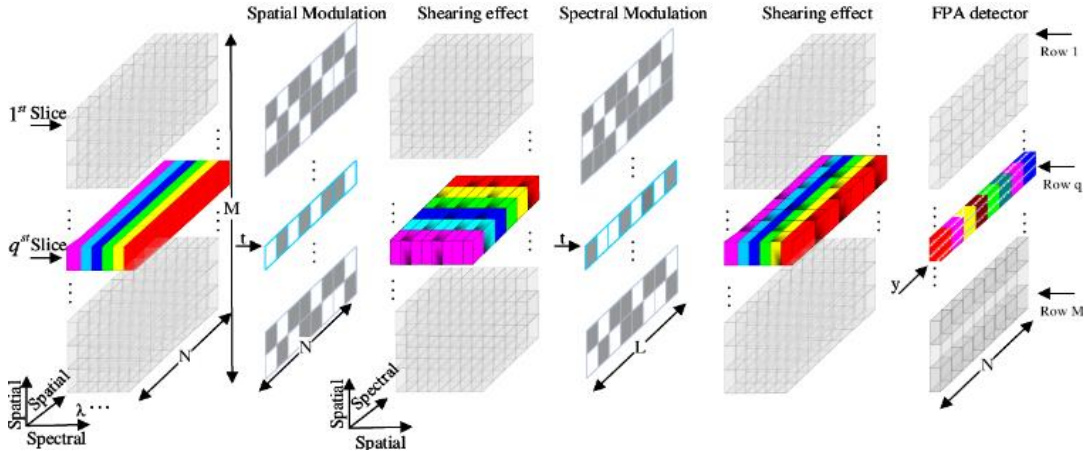
where, $y^S = [(y^0)^T, \dots, (y^{S-1})^T] \in R^{SMN}$; $\hat{H}^S = [(\hat{H}^0)^T, \dots, (\hat{H}^{S-1})^T] \in R^{SMN \times MNL}$ is the set of matrices $\hat{H}^S \in R^{MN \times MNL}$, which represents the effect of each coded aperture. \hat{H}^S can be obtained by $diag[(T_{0,0,k}^S), \dots, (T_{m-1,n-1,k}^S)] = \sum_{k=0}^{L-1} diag[(H^S)_k]$, where $diag(H_k^S)$ is an $MN \times MN$ diagonal matrix with vectorized binary entries of the coded apertures. Specifically, the matrix of the coded apertures (coding patterns) for all shots S can be expressed as $H^S = [(H^0)^T, \dots, (H^{S-1})^T]$.

Figure 3 describes the H structure that contains coded

apertures in CASSI. The vector representation of 2D arrays (dashed circles) is represented by columns and the vectorization (solid line circle) is performed vertically to concatenate the vectorizations of each $F(:, :, k)$ with $k=0, \dots, L-1$ like 3D arrays $f \in R^{M \times N \times L}$.

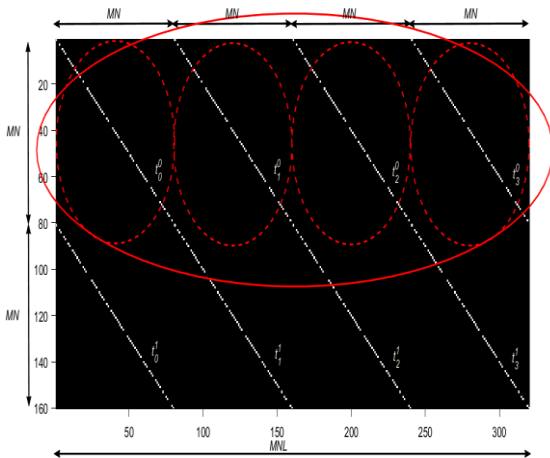
Note that each spatial location of T^S is assigned to one coding pattern $H^S \in R^L$ to modulate a pixel in a specific location. Hence, the rich structure of T^S depends on the coding pattern H^S . As a result, the coded apertures design can be regarded as the coding patterns design. The matrix of S coding patterns can be alternatively given by $H = [H^0, H^1, \dots, H^{S-1}]^T \in R^{S \times L}$.

Let f be an $L \times MN$ matrix whose columns correspond to the spectral signatures f_j of the data cube, with $j=0, \dots, MN-1$. Thus, the ensemble of S measurements can be written as $y^S = [[y_{0,0}^0, \dots, y_{M-1,N-1}^0]^T, \dots, [y_{0,0}^{S-1}, \dots, y_{M-1,N-1}^{S-1}]^T]^T$, where y^S is an $S \times MN$ matrix. Note that each column value and each row value in matrix y^S correspond to a compressed spectral signature, and the compressed information (spectral response) of each pixel obtained at the s^{th} snapshot, respectively. Because of this intrinsic structure, matrix y is applicable to SSC, making it easier to distinguish among all measurements and to improve the clustering results.



Note: The q -th slice of the data cube F with $L=6$ spectral components is coded by a row of the coded aperture t and sheared by the dispersive element; the detector gets the intensity y by integrating the coded light.

Figure 2. The CASSI spatial-spectral optical flow



Note: On the indicated diagonal, the spectral response of the pixels with the coded aperture H for each band is shown separately.

Figure 3. The CASSI matrix H with $S = 2$, $M = 8$, $N = 10$, and $L = 4$

2.2 Clustering of compressive spectral imaging (CSI) measurements

Let $y \in R^{S \times MN}$ be a 2D matrix reorganized as $y = [y_1, \dots, y_{MN}]$, where y_i is the spectral signature of the i^{th} spectral pixels on the compressed domain, i.e., compressed pixels. Assuming that the CSI measurements are in the union of l low-dimensional subspace $\cup_{l=1}^l C$ of SSC, each subspace must be corresponding to a specific class of land cover [40]. It is also assumed that each spectral signature y_i corresponds to a specific land cover class, which belongs to the same subspace with features of independence. Treating the compressed measurements as a dictionary, the SSC optimization problem can be expressed as:

$$\begin{aligned} \min_{c,g} & \|c\|_0 + \frac{\lambda}{2} \|g\|_F^2 \\ \text{s. t. } & y = yc + g, \text{diag}(c) = 0, c^T \mathbf{1} = 1, \end{aligned} \quad (4)$$

where, λ is a parameter for the tradeoff between sparsity and

noise level; $c \in R^{MN \times MN}$ is the coefficient matrix; $g \in R^{S \times MN}$ is the error matrix to model noise; $\text{diag}(c)=0$ is a constraint to eliminate the trivial solution of representing a point as a linear combination of itself; $c^T \mathbf{1}=1$ is a constraint to guarantee the operability in affine subspace.

The l_0 -norm aims to find the total number of nonzero entries of c . However, formula (4) generally leads to a nondeterministic polynomial-time (NP) hard problem, due to the nature of combinatorial optimization. In fact, the l_0 -norm is usually replaced by its tightest convex relaxation, namely, l_1 -norm constraint, which transforms formula (4) into:

$$\begin{aligned} \min_{c,g} & \|c\|_1 + \frac{\lambda}{2} \|g\|_F^2 \\ \text{s.t. } & y = yc + g, \text{diag}(c) = 0, c^T \mathbf{1} = 1, \end{aligned} \quad (5)$$

where, the l_0 -norm regularization promotes the sparsity of c resulting in subspace preservation. The optimization problem (5) can be effectively solved by using the alternating direction method of multiplier (ADMM). Besides, the solution of formula (5) is such that $z_{ij}=0$ when points i and j exist in distinct subspaces. Therefore, it is possible to adopt z to define a data adjacency matrix w as $|z| + |z|^T$. The final segmentation of the data can be achieved by applying spectral clustering-based methods to Laplacian matrix caused by w .

3. MULTIFRAME CASSI CODING PATTERN OPTIMIZATION FOR CSI SUBSPACE CLUSTERING

Considering its dimensionality reduction effect on ambient signal space, the compression property can effectively reduce the computational cost of using the self-expression features of the data in the SC. According to the affinity concept of similarity between the two subspaces [41], many scholars have theorized and verified the conditions for several subspace clustering-based methods on compressed data [42-44]. Due to the high-dimensionality of data, all of them randomly projected the samples onto a low-dimensional subspace using an off-the-shelf dimensionality reduction (DR) method, and then called a clustering algorithm in this subspace. Recent works [44, 45] have studied the distance preserving properties of compressed data points, using random projections. More recently, Jiao et al. [46] theoretically concluded that the distance between the two subspaces is almost constant after random projection. However, the traditional coded apertures cannot preserve the spectral signature similarities between any pair of vectors on the compressed domain, for the random projection cannot retain orthogonality, and the vectors that define the orthogonal basis of the subspace cannot be normalized.

This section mainly aims to solve these challenges with the projection determined by the coded aperture design, thereby improving the clustering results.

3.1 Coding pattern formulation and analysis of the principal angles, affinity, and distance

Let $f = [f_1^T, f_2^T, \dots, f_L^T]^T \in R^{L \times MN}$ be a matrix representing; $\{G_u\}_{u=1}^u$ be a set of u subspaces, where $\text{dim}[G_u] = d'_u < d'$, $\forall u = 1, \dots, u$. Whereas each column $f_u \in R^L$ is of dimension $1 \times MN$ corresponding to a given spectral band in a spectral image, linear combinations of samples can be taken from separate bands applying the

sampling matrix $H = [H^1, H^2, \dots, H^S]^T \in R^{S \times L}$, where $S < L$. This yields $y = [(y^1)^T, (y^2)^T, \dots, (y^S)^T]^T \in R^{S \times MN}$, where each $y_u \in R^S$ is of dimension $1 \times MN$.

With the growing dimension u of the ambient space, it is necessary to reduce the time and memory costs of SC, that is, to reduce the data dimensionality. Moreover, the available data have been compressed or incomplete constraint to data acquisition methods, which leads to the use of SC method for data compression. Heckel et al. [47] theoretically analyzed the performance of different types of compression matrices H to validate the performance of SC algorithms. Unfortunately, their work does not exploit the data structure. Therefore, this paper chooses the encoding mode matrix $H \in R^{S \times L}$ for compression and dimensionality reduction. The full data cube f , which lies in the G -dimensional space, was compressed by H to yield $y=Hf$.

Because the principal angle between two different subspaces provides a fundamental way to measure their similarity in term of cosine definition, the similarity between two columns of matrix representation of the spectral cube can be recursively defined as:

$$\begin{aligned} \cos(\theta_{uu'}) &= \max_{f_u \in S_u} \max_{f_{u'} \in S_{u'}} \frac{f_u^T f_{u'}}{\|f_u\|_2 \|f_{u'}\|_2} := f_u^T f_{u'} \\ \text{s.t. } & \|f_u\|_2 = 1 \text{ and } \|f_{u'}\|_2 = 1 \end{aligned} \quad (6)$$

where, $\theta_{uu'}$ is the principal angle between two subspaces G_u and $G_{u'}$ of dimensions $d'_u \leq d'_{u'}$; $f_u \in R^L$ is the submatrix of f indexed by u .

Recently, the most efficient low-rank matrix estimation technique has emerged as a powerful estimation tool in principal component decomposition (PCA), such as singular value decomposition (SVD). The estimator is mainly decided by dimension reduction of the matrix with respect to the orthogonal basis. This paper adopts a similar dimension reduction to calculate the principal angles. Specifically, given the columns of orthogonal bases U_u for subspace G_u of dimension d'_u , $\lambda_1 \geq \lambda_2 \dots \geq \lambda_{d'_u} \geq 0$ are the singular values of $U_u^T U_{u'}$, $u \neq u'$. This motivates the use of the singular values of $U_u^T U_{u'}$ for the cosine of the principal angles, such that $\cos \theta_{\{1, \dots, d'_u\}} = \lambda_{\{1, \dots, d'_u\}}$. The clustering result, which relies on the affinity between pairs of subspaces, can be obtained by imposing the principal angles for measuring the similarity between two subspaces. The affinity between two subspaces G_u and $G_{u'}$ of dimension $d'_u < d'_{u'}$ can be defined as

$$\text{aff}(G_u, G_{u'}) := \frac{1}{\sqrt{d'_u}} \|U_u^T U_{u'}\|_F \quad (7)$$

It can also be represented by the principal angles $\theta_1 \leq \theta_2 \dots \leq \theta_{d'_u}$ between G_u and $G_{u'}$:

$$\begin{aligned} \text{aff}(G_u, G_{u'}) &= \sqrt{\frac{\cos^2(\theta_1) + \cos^2(\theta_2) + \dots + \cos^2(\theta_{d'_u})}{d'_u}}, \end{aligned} \quad (8)$$

where, $0 \leq \text{aff}(G_u, G_{u'}) \leq 1$ uses a smaller value to show that one subspace is further apart from the other subspaces, if G_u and $G_{u'}$ for subspaces intersect in t dimensions.

More accurate distance measures were introduced for the subspace's separability, due to their advantages in space

measurement. In addition, the Forbenius norm (F-norm) distance between pair of subspaces in term of principal angle was considered as a special case of formulas (6), (7) and (8) constraint to the same dimension. The projected F-norm distance can be written in the case where the two subspaces have different dimensions:

$$D(G_u, G_{u'}) := \frac{1}{\sqrt{2}} \|P_u - P_{u'}\|_F, \quad (9)$$

where, $D(G_u, G_{u'})$ is the general F-norm distance with two subspaces G_u and $G_{u'}$. Formula (9) is general enough to encompass a wide applications, thanks to its properties like non-negativity, positive-definiteness, symmetry, and triangular inequality. Capable of adapting to different dimensional subspaces in a metric space, formula (9) provides an effective way to measure the similarity between two subspaces of different dimensions.

Combining formulas (7) and (9), the proposed relationship between distance and affinity can be expressed as:

$$D^2(G_u, G_{u'}) = \frac{d'_u + d_{u'}}{2} - \frac{1}{\sqrt{d'_u}} aff^2(G_u, G_{u'}). \quad (10)$$

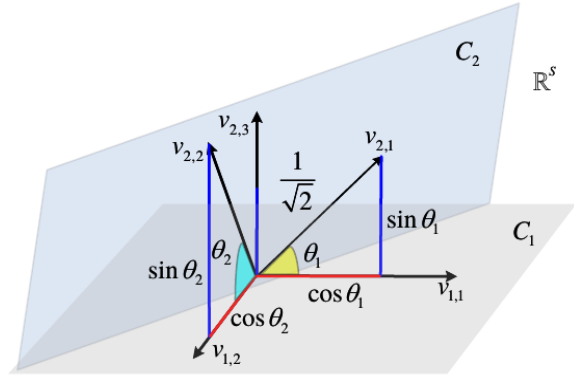
Next, it is possible to access dimensionality reduction set G only, which is resulted from down-sampling. The set increases proportion to G , in terms of computational and complexity cost. More importantly, it is usually desirable to work on a dimensionality-reduced version of G , because the original version weakens storage capacities and computational resources, even when the points in G are simply accessible. The key idea of reducing computational complexity is to apply the realization of $H \in R^{S \times L}$ with $S \geq \max_u d_u$ to each pixel, which guarantees that such projections can preserve orthogonality and normalization of the vectors in formula (5). All the entries inside the sets C_l in $C = C_1 U \dots U C_l$ can be detailed by:

$$G_u \xrightarrow{H} C_l = \{y | y = Hf\}, l \neq l'. \quad (11)$$

Assuming that each class (projected subspaces) exactly preserves the similarities of the spectral signatures from other classes, the projected distance $D(C_l, C_{l'})$ between classes l and l' of dimension $d_l > d_{l'}$ obtains the new data, which come from the underlying scene within the difference of snapshots, so that the entanglement of the class is best described. In this case, the projected distance can be described by:

$$\begin{aligned} D^2(C_l, C_{l'}) &= \frac{d_l + d_{l'}}{2} - \frac{1}{\sqrt{d_l}} aff^2(C_l, C_{l'}), \\ \text{s. t. } aff^2(C_l, C_{l'}) &= \| (H_k^T) H_{k'} + I \|_F^2 + \| \\ H^s (H^{s'})^T + I \|_F^2, \\ k \neq k', s \neq s', \end{aligned}$$

where, $(H_k^T) H_{k'}$ and $H^s (H^{s'})^T$ are all the entries outside the diagonal of $H^T H$ and $H H^T$, respectively; I is the identification matrix. Later, the proposed model (11) will be proved to reveals the link between affinity and distance initialized by the definition of concise, which allows to deal with the constrained optimization problem. Figure 4 displays an example of the relations among principal angles, affinity, and distance after projection.



Note: Red and blue bars stand for the pixels $y_1 \in C_1$ and $y_2 \in C_2$, respectively; all pixels are normalized to unit l_2 -norm prior to clustering.

Figure 4. Geometric interpretation of the main angles, affinity and F-norm distance after projection in the case of two subspaces C_1 and C_2 of dimensions 2 and 3, respectively

3.2 Our algorithm for coding pattern optimization

To improve the clustering effect, the desired values of $D^2(C_l, C_{l'})$ should be as larger as possible. The problem of designing H to maximize $D^2(C_l, C_{l'})$ can be formulated as

$$\max_H D^2(C_l, C_{l'}) \text{ s. t. } H \in \mathcal{C}_{L,S}, \quad (12)$$

where, $H \in \mathcal{C}_{L,S}$ is the set of matrices whose entries are drawn from a Bernoulli distribution $(H)S, L \sim Be(p)$, under the constraint of binary nonnegativity, that strengthens the spectral pixel clustering from the measurements. Although the expression of $D^2(C_l, C_{l'})$ is convex, the direct solution to the non-trivial problem (12) is not workable.

Coherence is a direct measure for the quality of these projections associated with compressive sensing. The coherence of any pair of the measurement matrix H depicts the maximum absolute value of the inner product in term of the l_2 -norm. Thus, the coherence should be minimized obtain the unique solution. More specifically, the matrix H can be expressed as $H = \sum_{s=0}^{S-1} \sum_{k=0}^{L-1} h_k^s$. Formally, the inner product between the columns of A at a different shot can be firstly defined as:

$$\langle H(s^1, \cdot), H(s^2, \cdot) \rangle = \sum_0^{S-1} \langle h^{s^1}, h^{s^2} \rangle. \quad (13)$$

The coherence of H can be expressed as

$$\mu_1(H) = \max_{s^1 \neq s^2} \frac{|\langle h^{s^1}, h^{s^2} \rangle|}{\|h^{s^1}\|_2 \|h^{s^2}\|_2}, \quad (14)$$

where,

$$\mu_1(H) = \frac{|\sum_0^{S-1} \varphi_{s^1, s^2}|}{(\sum_0^{S-1} \varphi_{s^1, s^2})^{\frac{1}{2}} (\sum_0^{S-1} \varphi_{s^1, s^1})^{\frac{1}{2}}} \quad (15)$$

and $\varphi_{s^1, s^2} = \langle h^{s^1}, h^{s^2} \rangle$. It is possible to eliminate the value outside its diagonal inner product, that is, the desired value

should be zero to distinguish all classes. Thus, formula (15) can be rewritten as

$$\mu_1(H) = \frac{|q_{s^1, s^2}|}{(q_{s^1, s^1})^{\frac{1}{2}}(q_{s^1, s^2})^{\frac{1}{2}}}, \quad (16)$$

where,

$$q_{s^1, s^2} = \sum_0^{s-1} \varphi_{s^1, s^2} + \sum_{s^1 \neq s^2} \varphi_{s^1, s^2}, \quad (17)$$

where, the coherence $\mu_1(H)$ depends on the variables φ_{s^1, s^2} and φ_{s^1, s^1} . Similarly, the inner product between any pair of columns of A can be expressed as

$$\mu_2(H) = \frac{|q_{k_1, k_2}|}{(q_{k_1, k_1})^{\frac{1}{2}}(q_{k_1, k_2})^{\frac{1}{2}}}, \quad (18)$$

where,

$$q_{k_1, k_2} = \sum_0^{s-1} \varphi_{k_1, k_1} + \sum_{k_1 \neq k_2} \varphi_{k_1, k_2}. \quad (19)$$

According to the quantities from formula (19), $\mu_1(H)$ and $\mu_2(H)$ require that the value of $D^2(C_l, C_{l'})$ must be nonnegative. To be more specific, given that d_l and $d_{l'}$ are assumed to be fixed, the optimization can be limited to minimize $\mu(H) = \mu_1(H) + \mu_2(H)$, implying the expanding range of values of $D^2(C_l, C_{l'})$, that is, an indirect maximization of $D^2(C_l, C_{l'})$ to identify the local maximum.

Under these considerations, the problem (12) can be equivalently transformed into:

$$\min \sum_{s^1, s^2} \varphi_{s^1, s^2} + \sum_{k_1, k_2} \varphi_{k_1, k_2}, \quad s. t. H \in \mathcal{C}_{L, S}. \quad (20)$$

where, $\mathcal{C}_{L, S}$ is the set of matrices directly clustering multispectral images with L bands using S shots. Due to the NP-hardness of a direct solution of problem (20), the following formulation was further introduced to illustrate the relation between μ and φ :

$$\begin{aligned} \mu(H) \leq & \varepsilon_1 \left[\left(\sum_0^{s-1} \varphi_{s^1, s^2} \right) + \sum_{s^1 \neq s^2} \varphi_{s^1, s^2} \right] \\ & + \varepsilon_2 \left[\left(\sum_0^{s-1} \varphi_{k_1, k_2} \right) + \sum_{k_1 \neq k_2} \varphi_{k_1, k_2} \right]. \end{aligned} \quad (21)$$

The quantities from formula (21) bounds the value of $\mu(H)$. This still likens the minimization of φ to the maximization of $D^2(C_l, C_{l'})$. The $\mu(H)$ is minimized indirectly by reducing the value range of $\mu(H)$ to obtain the local minimums.

Considering this, problem (20) can be rewritten as:

$$\min \sum_{s^1, s^2} \varphi_{s^1, s^2} + \sum_{k_1, k_2} \varphi_{k_1, k_2}, \quad s. t. H \in \mathcal{C}_{L, S}. \quad (22)$$

Figure 5 compares a coherence-based coding pattern (a)

compared with a random coding pattern (b). The coding pattern (a) was realized by implementing Algorithm 1 with different sets of pairs of cut-off wavelengths. Both coding patterns were generated with $S=50$, and $L=220$. For the value of function $\mu(H)$, the smaller value belongs the desired coding pattern, which finds optimal projections to guarantee the spectral pixel clustering directly from the CASSI measurements. Further, the block-unblock entries in Figure 5(a) for the coherence-based coding pattern show a uniform spectral distribution, which provides a more ideal sampling environment and a better condition of H to satisfy the SSC or dimensionality-reduced SC. By contrast, the random coding pattern in Figure 5(b) did not satisfy the previously described conditions, which results in oversampling or subsampling of part of all spectral bands.

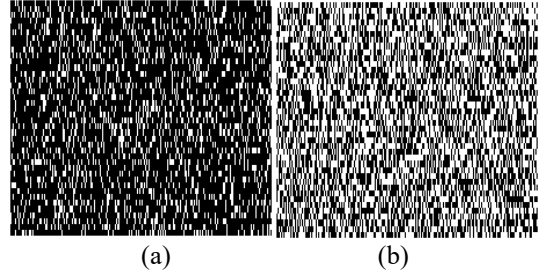


Figure 5. Coherence-based coding pattern (a) compared with a random coding pattern (b)

4. REWEIGHTED SPARSE SUBSPACE CLUSTERING ALGORITHM FOR CSI

Under a high-dimensional ambient space, of the whole subspace clustering processing can be forbidding it terms of time and memory costs. One efficient way to reduce the data dimensionality is to apply the compressed or incomplete data via some practical methods. More specifically, induced by a sampling matrix, the original data were compressed by H , which leads to $y=Hx$ distributed in a low dimensional subspace. The final clustering result was realized by the compressed data using the RSSC algorithm.

Compressed measurements with similar features can be allocated into the same cluster (group) via the spectral clustering-based methods such as SSC algorithm, provided that they contain a small part of the spectral pixels with the same land-cover class. However, applying the classical SSC algorithm adapted to CSI cannot fully exploit the spatial correlation of the spectral image, which leads to a low representation accuracy of the learned sparse coefficient matrix. What is worse, their presentation coefficients are very close, if specific land-cover materials are distributed locally, i.e., two spatially adjacent pixels are very likely to be allocated to the same subspace.

Suppose each band of \tilde{c} represents the distribution of the representation coefficients of each image realized by rearranging the 2D sparse matrix $c \in R^{MN \times MN}$ into the 3D cube $\tilde{c} \in R^{M \times N \times MN}$ along the rows. In other words, the adjacent elements of each row of \tilde{c} are generally smaller and each row of \tilde{c} is piecewise smooth. As a result, it is natural to constrain a coefficient vector with adjacent coefficient vectors, in order to reduce the representation bias.

The mean of a spatial window can be used to regularize the coefficient vector in center representation, under the

assumption that pixels can share the equal dominant subspace in the window. More specifically, the authors in [48] opened a moving window with 3×3 at each coefficient vector and limit the difference between them. The mean of the neighboring pixels was defined as $\|c - \bar{c}\|_F^2 < \varepsilon$, where ε is the restriction and $\bar{c} \in R^{MN \times MN}$ is the mean coefficient matrix obtained by rearranging the mean of the cube $\tilde{c} \in R^{M \times N \times MN}$ to a 2D matrix, with the aim to greatly improve the SSC performance. Therefore, a spatial regularizer [49] capable of promoting piecewise smoothness and preserving edges was introduced into the original SSC model, and used to formulate the l_1 -norm minimization problem to better utilize spatial neighborhood information and boost piecewise smoothness of the coefficient matrix c

$$\begin{aligned} \min_{c, g, \bar{c}} & \|c\|_1 + \frac{\lambda}{2} \|g\|_F^2 + \frac{\alpha}{2} \|c - \bar{c}\|_F^2 \\ \text{s. t. } & y = yc + g, \text{diag}(c) = 0, c^T \mathbf{1} = 1, \end{aligned} \quad (23)$$

where, α is the regularization coefficient representing the contribution of each related spatial constraint. However, the SSC algorithm coupled with such a regularizer might not obtain the optimal solution of formula (23). Inspired by the iterative update of weighting matrix, the reweighted l_1 -norm minimization approach was applied to enhance the compressive clustering performance. The, the problem can be modeled as:

$$\begin{aligned} \min_{c, g, \bar{c}} & \|w \odot c\|_1 + \frac{\lambda}{2} \|g\|_F^2 + \frac{\alpha}{2} \|c - \bar{c}\|_F^2 \\ \text{s. t. } & y = yc + g, \text{diag}(c) = 0, c^T \mathbf{1} = 1, \end{aligned} \quad (24)$$

where, \odot is the element-wise operator. The problem (24) can be solved by ADMM. The numerous variables in the objection function will hinder the convergence of the reweighted ADMM algorithm, which is still a ubiquitous problem. To overcome this difficulty, the maximum number of iterations and necessary convergence conditions should be configured to ensure the convergence, whiling guaranteeing the algorithm performance.

Assuming that the compressed data are corrupted by the system, the following convex program can be considered alternatively to solve problem (24):

$$\begin{aligned} \min_{c, a, \bar{c}} & \|w \odot c\|_1 + \frac{\lambda}{2} \|y - ya\|_F^2 + \frac{\alpha}{2} \|\bar{c} - a\|_F^2 \\ \text{s. t. } & a^T \mathbf{1} = 1, a = c - \text{diag}(c). \end{aligned} \quad (25)$$

By introducing two penalty terms $a^T \mathbf{1} = 1$ and $a = c - \text{diag}(c)$ into the constraint function of (25), the objective can be reformulated as:

$$\begin{aligned} \min_{c, a, \bar{c}} & \|w \odot c\|_1 + \frac{\lambda}{2} \|y - ya\|_F^2 + \frac{\alpha}{2} \|\bar{c} - a\|_F^2 \\ & + \frac{\rho}{2} \|a^T \mathbf{1} - 1\|_2^2 + \frac{\rho}{2} \|a - (c - \text{diag}(c))\|_F^2 \\ \text{s. t. } & a^T \mathbf{1} = 1, a = c - \text{diag}(c). \end{aligned} \quad (26)$$

By introducing a vector $\bar{\delta} \in R^{MN}$ and a matrix $\bar{\Delta} \in R^{MN \times MN}$ for the two equality constraints in (26), the augmented Lagrangian can be expressed as:

$$\begin{aligned} L(c, a, \bar{c}, w, \bar{\delta}, \bar{\Delta}) = & \|w \odot c\|_1 + \frac{\lambda}{2} \|y - ya\|_F^2 + \frac{\alpha}{2} \|\bar{c} - a\|_F^2 \\ & + \frac{\rho}{2} \|a^T \mathbf{1} - 1\|_2^2 + \frac{\rho}{2} \|a - (c - \text{diag}(c))\|_F^2 + \bar{\delta}^T (a^T \mathbf{1} - 1) \\ & + \text{tr}(\bar{\Delta}(a - c + \text{diag}(c))) \text{ s. t. } a^T \mathbf{1} = 1, a = c - \text{diag}(c). \end{aligned} \quad (27)$$

where, $\text{tr}(\bullet)$ is the trace operator of a given matrix. The variables a and c were updated by setting their gradient, namely, a and c to 0 per iteration. Furthermore, at each iteration, the dual variables $\bar{\delta}$ and $\bar{\Delta}$ were updated via gradient ascent. The update of a at iteration $(k+1)$ can be expressed as:

$$\begin{aligned} (\lambda y^T y + aI + \rho \mathbf{1} \mathbf{1}^T + \rho I) a^{(k+1)} = & \lambda y^T y + a \bar{c}^{(k)} \\ & + \rho (\mathbf{1} \mathbf{1}^T + c^{(k)}) - \mathbf{1} \bar{\delta}^{(k)T} - \bar{\Delta}^{(k)}. \end{aligned}$$

Accordingly, the update for c at iteration $(k+1)$ can be expressed as:

$$c^{(k+1)} = J - \text{diag}(J), J \stackrel{\Delta}{=} S_{1/\rho}^w(a^{(k+1)} + \frac{\bar{\Delta}^{(k)}}{\rho}), \quad (28)$$

where, $S_{1/\rho}^w(\bullet)$ is a shrinkage-thresholding operator; $S_{\eta}^w(v) = \max(|v| - \eta, 0) \odot \text{sgn}(v)$. The operator $(\bullet)_+$ returns its arguments, if it is nonnegative, and returns zero, if otherwise. Then, w in iteration $(k+1)$ can be updated by:

$$w_{ij}^{(k+1)} = \frac{\varepsilon_2}{|c_{ij}^{(k+1)}| + \varepsilon_1}, \quad (29)$$

where, ε_1 and ε_2 are the absolute parameters; w_i is the sum of the weighted matrix w at the i -th row.

Finally, $\bar{\delta}^{(k+1)}$ and $\bar{\Delta}^{(k+1)}$ can be respectively updated with the step size $\rho=500$ as

$$\begin{aligned} \bar{\delta}^{(k+1)} &= \bar{\delta}^{(k)} + \rho(c^{(k+1)} \mathbf{1} - 1) \\ \bar{\Delta}^{(k+1)} &= \bar{\Delta}^{(k)} + \rho(a^{(k+1)} - c^{(k+1)}). \end{aligned} \quad (30)$$

These three steps were repeated until the convergence or the maximum number of iterations is reached. The iteration was terminated by verifying whether the following constraints holds for the primal and dual residuals at each iteration k :

$$\begin{aligned} \|a^{(k)T} \mathbf{1} - 1\|_{\infty} &\leq \bar{\varepsilon}, \|a^{(k)T} - c^{(k)}\|_{\infty} \leq \bar{\varepsilon} \\ \|a^{(k)T} - a^{(k-1)T}\|_{\infty} &\leq \bar{\varepsilon} \end{aligned} \quad (31)$$

where, $\bar{\varepsilon}$ is the error tolerance.

The matrix c can be created with the adjacency matrix (similarity graph) $w \in R^{MN \times MN}$, in which each entry expresses the similarity between two pixels, in the same way as the original SSC

$$\omega = |c| + |c|^T. \quad (32)$$

Table 1. Compressed subspace clustering

Algorithm 1: Subspace clustering directly from the 3D-CASSI measurements.
Input: y , the selected RSSC algorithm with the corresponding parameters
Initialization:
1: Generate a sampling matrix for compression Generate random matrix or another matrix
2: Compress the data Compute the compressed data as $y = Hx$
3: Conduct subspace clustering Cluster y using the RSSC algorithm
Output: $\gamma_1, \dots, \gamma_K$

The formula (32) with symmetrization can be used to enhance the connection of the graph. The final result can then be obtained by applying spectral clustering algorithm such as RSSC to the similarity graph. The entire process is listed in Algorithm 1 (Table 1).

5. SIMULATION AND RESULTS ANALYSIS

This section realizes the proposed compressed spectral image method on two real-world datasets with different imaging environments: Indian Pine dataset [50, 52] and University of Pavia dataset [51, 52]. The multi-shot CASSI system model was used to obtain a set of CASSI measurements with the number of shot and the coding pattern bandwidth: $S = 20$, for Indian Pines dataset and for University of Pavia dataset. The CSI measurements were obtained with two different types of codes apertures, i.e., random coding pattern, and designed coding pattern. The RSSC algorithm was proposed to validate the clustering directly from the CSI measurements (Algorithm 2). All parameters of the RSSC algorithm were set as follows: $\alpha=3.9 \times 10^4$ for the Indian Pines images; $\alpha=25.5 \times 10^5$ for the University of Pavia images; the positive parameter λ that trades off sparsity of coefficient against the magnitude of the noise can be determined by:

$$\lambda = \frac{\beta}{\mu}, \mu \triangleq \min_{j \neq i} \max_j |y_j^T y_i|, \quad (33)$$

where, β is the adjustment coefficient fixed for all the experiments at $\beta=1,000$; μ is a parameter with the dataset that can be explicitly determined; y_j is the column of y . Once μ is fixed for a certain dataset, the value of λ is solely dependent on β .

To fully evaluate the performance of our algorithm, the original SSC was taken as the benchmark. The RSSC using designed coding pattern was compared with that using random coding pattern. Furthermore, the results of our approach for subspace clustering were also compared with the algorithms RSSC and SSC, respectively, using the complete spectral data cube, for both types of data.

For simplicity, the direct clustering by RSSC on the set of CASSI measurements associated with the coherence-based codes is denoted as ‘‘Coherence-codes-RSSC’’; the direct clustering by RSSC on the set of CASSI measurements associated with the random codes is denoted as ‘‘Random-codes-RSSC’’; the clustering by RSSC on the complete spectral data cube (Full-data) is denoted as ‘‘Full-data-RSSC’’; the clustering by the original SSC on the complete spectral data cube (Full-data) is denoted as ‘‘Full-data-SSC’’.

A common clustering strategy is to treat the number of clusters as a prior, and then manually determinate the thematic information for each cluster through cross-referencing between the clustering results and the images. Here, all the parameters of each clustering algorithm are manually adjusted to the optimum. Further, a 30dB Gaussian noise was added to the measurements compressed (incomplete pixels) in the real CSI architecture. The mean of 10 trial runs was taken as the final result. Both visual clustering results and quantitative evaluations were provided, including overall accuracy (OA), mean accuracy (AA), Kappa coefficients, and computational time (Time). Every experiment was carried out in MATLAB 2019b (640bit) on a 3.33GHz desktop computer with 32GB of RAM.

5.1 AVIRIS dataset: Indian Pines images

The first experiment was conducted on the images of Indian Pines collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor during a flight campaign on June 12, 1992, over West Lafayette, Indiana, US. In the dataset, each spatial information involves 145×145 pixels and 224 spectral bands. After discarding the 20 noisy and water absorption bands, 200 bands remained for the spectral analysis. This scene covers an agricultural field with 16 main classes. There are 10, 249 labeled samples for this dataset using the distribution as shown in Table 2. Combined with the computational efficiency, a sub image of 70×70 was cropped, including four main land-cover classes: corn-no-till (2), grass (7), soybeans-no-till (10), and soybeans-minimum-till (11). Clustering is a challenging work for the following factors: the highly similar spectral signatures of the land-cover classes in this area leads to some seriously mixed spectral curves (Figure 6(c)). The false-color image and the growth truth are shown in Figures 6(a) and (b), respectively.

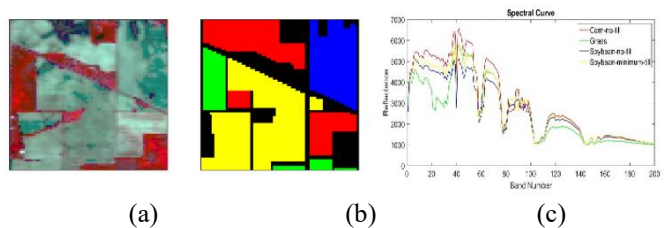


Figure 6. AVIRIS Indian Pines images, (a) False-color image (RGB 40, 30, 20). (b) Ground truth. (c) Spectral curves of the four land-cover classes

Table 2. Class labels with sample number for each class of the set of indian pines data

Label	Class	Samples	Label	Class	Samples
1	Alfalfa	46	9	Oats	20
2	Corn-noill	1428	10	Soy-notill	927
3	Corn-mintill	830	11	Soy-mintill	2455
4	Corn	237	12	Soy-clean	593
5	Pasture	483	13	Wheat	205
6	Tree	830	14	Woods	1265
7	Grass	28	15	Bigd-drives	386
8	Hay-window	478	16	Stone-tower	93
Total		10249			

Figure 7 show the results of different clustering methods on 25% of the samples. Obviously, Coherence-codes-RSSC and Full-data-RSSC obtained better accuracy associated with the infomraiton of spatial neighborhood. Coherence-codes-RSSC performed better than Random-codes-RSSC, thanks to its higher clustering accuracy. Coherence-codes-RSSC effectively separated the Corn-no-till class, Grass class, and Soybeans-minimum-till class, but slightly underperformed on the Soybeans-no-till classes. Meanwhile, due to a reweighted matrix of the pixels in the spectral images, RSSC guaranteed that the correlated signals provide a more accurate coefficient matrix, which brings a high clustering accuracy and a small sparse representation bias. Compared to Random-codes-RSSC, the our method achieved a high precision. Finally, our approach obtained clustering results comparable to those of Full-data-RSSC and Full-data-SSC on full data, yet only consumed 83.65% and 83.23% of the computational time of Full-data-RSSC and Full-data-SSC, respectively. Overall, our method realized relatively good performance on spectral image clustering tasks, both visually and quantitatively.

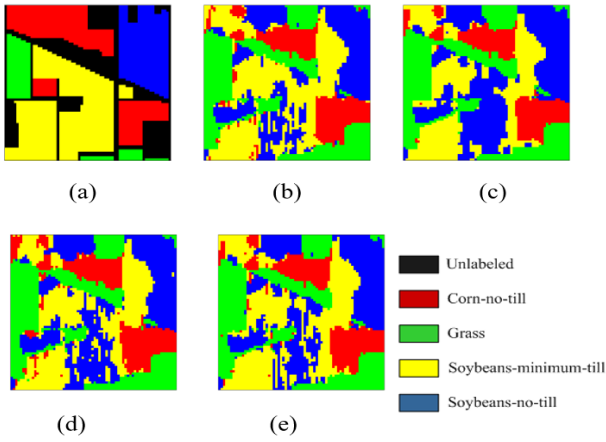


Figure 7. Unsupervised classification maps using different approaches for the Indian Pines images: (a) Ground truth, (b) Full-data-RSSC, (c) Full-data-SSC, (d) Coherence-codes-RSSC, (e) Random-codes-RSSC

5.2 ROSIS Pavia Data: University of Pavia, Italy

The second dataset, the images of the University of Pavia, was captured by the Reflective Optics System Imaging Spectrometer System (ROSIS) sensor over the University of Pavia, Pavia, Italy. The images in the dataset are of the size 610×340, involving 115 spectral bands. A total of 103 band remained for further analysis after the removal of low signal-to-noise ratio (SNR) bands and water absorption bands.

Similar to the former experiment, a representative region with a size of 140×80 pixels was cropped, which contains eight main land cover categories: asphalt (1), meadows (2), trees (3), metal sheet (4), bare soil (5), bitumen (6), bricks (7), and shadows (8). The diversity of land cover classes makes the test a challenging clustering task. The spectral curves of the eight land-cover classes are depicted in Figure 8. The false-color image and the ground truth are also given. From the ground truth, there are approximately 40,677 pixels. The number of samples in each class is listed in Table 3.

Table 3. Class Labels with Sample Number for Each Class of The Set of The University of Pavia Data

Label	Class	Samples	Label	Class	Samples
1	Asphalt	6631	5	Bare soil	5029
2	Meadows	18649	6	Bitumen	1330
3	Tree	3064	7	Bricks	3682
4	Metal sheets	1345	8	Shadows	947

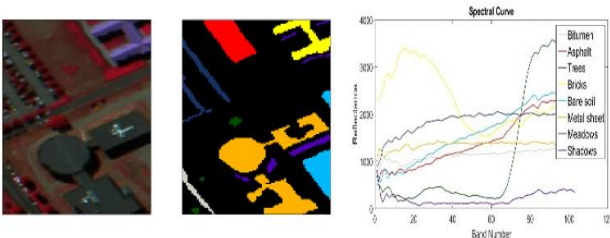


Figure 8. ROSIS University of Pavia images, (a) False-color image (RGB 102, 56, 31). (b) Ground truth. (c) Spectral curves of the eight land-cover classes

Figure 9 presents the clustering results using different methods on 25F% of the samples. It can be seen that Full-data-SSC outputted poor clustering results, which contain lots of

salt-and-pepper noises and many misclassifications, with an OA of 71.45%. Random-codes-RSSC generated smooth clustering results and outshined Full-data-SSC in this scene with an OA of 78.72%. However, there were still some misclassifications in the results of Random-codes-RSSC. For example, the bitumen and bare soil classes were not effectively identified, leading to the misclassification of most bare soil classes as asphalt. Coherence-codes-RSSC achieves an improvement of almost 6% in term of OA by significantly reducing misclassification. In addition, Coherence-codes-RSSC consumed the least classification time while maintaining a good performance, as it solved the clustering problem 7 times faster than other approaches. This further verifies the effectiveness and advantages of this strategy. Finally, the full data-RSSC effectively distinguished most of the land cover classes, achieving the best visual result and highest accuracy with little salt-and-pepper noise in each class; Nevertheless, the full data-RSSC consumed too much time in computation.

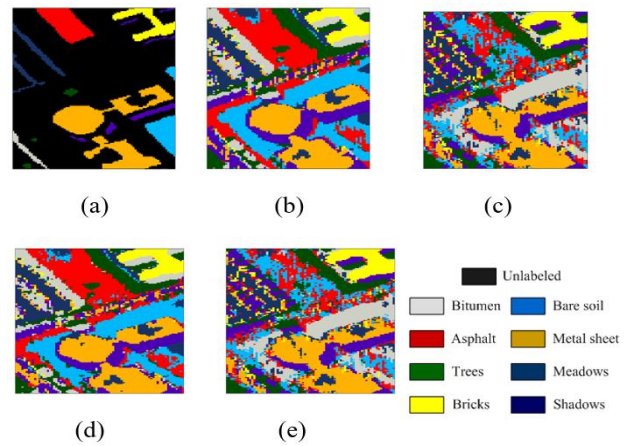


Figure 9. Unsupervised classification maps using different approaches for the University of Pavia image: (a) Ground truth, (b) Full-data-RSSC, (c) Full-data-SSC, (d) Coherence-codes-RSSC, (e) Random-codes-RSSC

6. CONCLUSIONS

This paper derives a compressed spectral image clustering approach, which replaces the compute-intensive task of direct clustering the full spectral data cube with direct clustering a set of CASSI measurements. Our approach fully utilizes the characteristic that each spectral signature is separability preserved after the scene projection using the coherence codes, which allows to improve clustering results on the compressed domain. Simulation results on two spectral image datasets show the proposed approach is up to 7 times faster to achieve similar accuracy than other approaches that cluster the full 3D spectral images.

ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China (Grant No.: 61936014), the National Key Research and Development Project (Grant No.: 2019YFB2102300; 2019YFB2102301), and the Fundamental Research Funds for the Central Universities.

REFERENCES

- [1] Sellar, R.G., Boreman, G.D. (2005). Classification of imaging spectrometers for remote sensing applications. *Optical Engineering*, 44(1): 013602. <https://doi.org/10.1117/1.1813441>
- [2] Gat, N. (2000). Imaging spectroscopy using tunable filters: A review. In *Wavelet Applications VII*. International Society for Optics and Photonics, 4056: 50-64. <https://doi.org/10.1117/12.381686>
- [3] Crocco, M., Martelli, S., Trucco, A., Zunino, A., Murino, V. (2017). Audio tracking in noisy environments by acoustic map and spectral signature. *IEEE Transactions on Cybernetics*, 48(5): 1619-1632. <https://doi.org/10.1109/TCYB.2017.2711497>
- [4] Zhang, C., Cheng, J., Tian, Q. (2019). Unsupervised and semi-supervised image classification with weak semantic consistency. *IEEE Transactions on Multimedia*, 21(10): 2482-2491. <https://doi.org/10.1109/TMM.2019.2903628>
- [5] Zhang, M., Li, W., Du, Q. (2018). Diverse region-based CNN for hyperspectral image classification. *IEEE Transactions on Image Processing*, 27(6): 2623-2634. <https://doi.org/10.1109/TIP.2018.2809606>
- [6] Miao, X., Zhao, W., Li, X., Yang, X. (2019). Structure descriptor based on just noticeable difference for texture image classification. *Applied Optics*, 58(24): 6504-6512. <https://doi.org/10.1364/AO.58.006504>
- [7] Di Simone, A., Park, H., Riccio, D., Camps, A. (2017). Sea target detection using spaceborne GNSS-R delay-Doppler maps: Theory and experimental proof of concept using TDS-1 data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(9): 4237-4255. <https://doi.org/10.1109/JSTARS.2017.2705350>
- [8] Lei, J., Li, Y., Zhao, D., Xie, J., Chang, C.I., Wu, L., Li, W. (2018). A deep pipelined implementation of hyperspectral target detection algorithm on FPGA using HLS. *Remote Sensing*, 10(4), 516. <https://doi.org/10.3390/rs10040516>
- [9] Hong, D., Yokoya, N., Chanussot, J., Zhu, X.X. (2018). An augmented linear mixing model to address spectral variability for hyperspectral unmixing. *IEEE Transactions on Image Processing*, 28(4): 1923-1938. <https://doi.org/10.1109/TIP.2018.2878958>
- [10] Drumetz, L., Chanussot, J., Jutten, C., Ma, W.K., Iwasaki, A. (2020). Spectral variability aware blind hyperspectral image unmixing based on convex geometry. *IEEE Transactions on Image Processing*, 29: 4568-4582. <https://doi.org/10.1109/TIP.2020.2974062>
- [11] Lyu, F., Wu, Q., Hu, F., Wu, Q., Tan, M. (2019). Attend and imagine: Multi-label image classification with visual attention and recurrent neural networks. *IEEE Transactions on Multimedia*, 21(8): 1971-1981. <https://doi.org/10.1109/TMM.2019.2894964>
- [12] Melgani, F., Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on geoscience and remote sensing*, 42(8): 1778-1790. <https://doi.org/10.1109/TGRS.2004.831865>
- [13] Ravi, D., Fabelo, H., Callic, G.M., Yang, G.Z. (2017). Manifold embedding and semantic segmentation for intraoperative guidance with hyperspectral brain imaging. *IEEE Transactions on Medical Imaging*, 36(9): 1845-1857. <https://doi.org/10.1109/TMI.2017.2695523>
- [14] Luo, F., Du, B., Zhang, L., Zhang, L., Tao, D. (2018). Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image. *IEEE Transactions on Cybernetics*, 49(7): 2406-2419. <https://doi.org/10.1109/TCYB.2018.2810806>
- [15] Chen, X., Sun, W., Wang, B., Li, Z., Wang, X., Ye, Y. (2018). Spectral clustering of customer transaction data with a two-level subspace weighting method. *IEEE Transactions on Cybernetics*, 49(9): 3230-3241. <https://doi.org/10.1109/TCYB.2018.2836804>
- [16] Su, L., Wang, W., Zhang, Y. (2019). Strong consistency of spectral clustering for stochastic block models. *IEEE Transactions on Information Theory*, 66(1): 324-338. <https://doi.org/10.1109/TIT.2019.2934157>
- [17] Elhamifar, E., Vidal, R. (2013). Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(11): 2765-2781. <https://doi.org/10.1109/TPAMI.2013.57>
- [18] Li, C.G., You, C., Vidal, R. (2017). Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework. *IEEE Transactions on Image Processing*, 26(6): 2988-3001. <https://doi.org/10.1109/TIP.2017.2691557>
- [19] Liu, G., Zhang, Z., Liu, Q., Xiong, H. (2019). Robust subspace clustering with compressed data. *IEEE Transactions on Image Processing*, 28(10): 5161-5170. <https://doi.org/10.1109/TIP.2019.2917857>
- [20] Wang, Y., Lin, X., Wu, L., Zhang, W., Zhang, Q., Huang, X. (2015). Robust subspace clustering for multi-view data by exploiting correlation consensus. *IEEE Transactions on Image Processing*, 24(11): 3939-3949. <https://doi.org/10.1109/TIP.2015.2457339>
- [21] Chen, J., Mao, H., Sang, Y., Yi, Z. (2017). Subspace clustering using a symmetric low-rank representation. *Knowledge-Based Systems*, 127: 46-57. <https://doi.org/10.1016/j.knosys.2017.02.031>
- [22] Liu, W., Ye, M., Wei, J., Hu, X. (2017). Compressed constrained spectral clustering framework for large-scale data sets. *Knowledge-Based Systems*, 135: 77-88. <https://doi.org/10.1016/j.knosys.2017.08.003>
- [23] Martín, G., Bioucas-Dias, J.M. (2016). Hyperspectral blind reconstruction from random spectral projections. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6): 2390-2399. <https://doi.org/10.1109/JSTARS.2016.2541541>
- [24] Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Brady, D.J. (2016). Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5): 95-108. <https://doi.org/10.1109/MSP.2016.2582378>
- [25] Arguello, H., Arce, G.R. (2014). Colored coded aperture design by concentration of measure in compressive spectral imaging. *IEEE Transactions on Image Processing*, 23(4): 1896-1908. <https://doi.org/10.1109/TIP.2014.2310125>
- [26] Wu, Y., Mirza, I.O., Arce, G.R., Prather, D.W. (2011). Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *Optics Letters*, 36(14): 2692-2694. <https://doi.org/10.1364/OL.36.002692>
- [27] Arguello, H., Rueda, H., Wu, Y., Prather, D.W., Arce, G.R. (2013). Higher-order computational model for coded aperture spectral imaging. *Applied Optics*, 52(10):

- D12-D21. <https://doi.org/10.1364/AO.52.000D12>
- [28] Gehm, M.E., John, R., Brady, D.J., Willett, R.M., Schulz, T.J. (2007). Single-shot compressive spectral imaging with a dual-disperser architecture. *Optics Express*, 15(21): 14013-14027. <https://doi.org/10.1364/OE.15.014013>
- [29] Lin, X., Liu, Y., Wu, J., Dai, Q. (2014). Spatial-spectral encoded compressive hyperspectral imaging. *ACM Transactions on Graphics (TOG)*, 33(6): 1-11. <https://doi.org/10.1145/2661229.2661262>
- [30] Lin, X., Wetzstein, G., Liu, Y., Dai, Q. (2014). Dual-coded compressive hyperspectral imaging. *Optics Letters*, 39(7): 2044-2047. <https://doi.org/10.1364/OL.39.002044>
- [31] Cao, X., Yue, T., Lin, X., Lin, S., Yuan, X., Dai, Q., Brady, D.J. (2016). Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Processing Magazine*, 33(5): 95-108. <https://doi.org/10.1109/MSP.2016.2582378>
- [32] Zhang, H., Zhai, H., Zhang, L., Li, P. (2016). Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 54(6): 3672-3684. <https://doi.org/10.1109/TGRS.2016.2524557>
- [33] Yuan, Q., Zhang, L., Shen, H. (2013). Regional spatially adaptive total variation super-resolution with spatial information filtering and clustering. *IEEE Transactions on Image Processing*, 22(6): 2327-2342. <https://doi.org/10.1109/TIP.2013.2251648>
- [34] Zhang, F., Ye, X., Liu, W. (2008). Image decomposition and texture segmentation via sparse representation. *IEEE Signal Processing Letters*, 15: 641-644. <https://doi.org/10.1109/LSP.2008.2002722>
- [35] Hait, E., Gilboa, G. (2018). Spectral total-variation local scale signatures for image manipulation and fusion. *IEEE Transactions on Image Processing*, 28(2): 880-895. <https://doi.org/10.1109/TIP.2018.2872630>
- [36] Meng, L., Li, G., Yan, J., Gu, Y. (2018). A general framework for understanding compressed subspace clustering algorithms. *IEEE Journal of Selected Topics in Signal Processing*, 12(6): 1504-1519. <https://doi.org/10.1109/JSTSP.2018.2879743>
- [37] Lin, X., Wetzstein, G., Liu, Y., Dai, Q. (2014). Dual-coded compressive hyperspectral imaging. *Optics Letters*, 39(7): 2044-2047. <https://doi.org/10.1364/OL.39.002044>
- [38] Lin, X., Wetzstein, G., Liu, Y., Dai, Q. (2014). Dual-coded compressive hyperspectral imaging. *Optics Letters*, 39(7): 2044-2047. <https://doi.org/10.1364/OL.39.002044>
- [39] Arguello, H., Arce, G.R. (2012). Restricted isometry property in coded aperture compressive spectral imaging. In *2012 IEEE Statistical Signal Processing Workshop (SSP)*, IEEE, 716-719. <https://doi.org/10.1109/SSP.2012.6319803>
- [40] Zhai, H., Zhang, H., Zhang, L., Li, P., Plaza, A. (2016). A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(1): 43-47. <https://doi.org/10.1109/LGRS.2016.2625200>
- [41] Soltanolkotabi, M., Candes, E.J. (2012). A geometric analysis of subspace clustering with outliers. *The Annals of Statistics*, 40(4): 2195-2238. <https://doi.org/10.1214/12-AOS1034>
- [42] Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y. (2012). Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1): 171-184. <https://doi.org/10.1109/TPAMI.2012.88>
- [43] Patel, V.M., Van Nguyen, H., Vidal, R. (2015). Latent space sparse and low-rank subspace clustering. *IEEE Journal of Selected Topics in Signal Processing*, 9(4): 691-701. <https://doi.org/10.1109/JSTSP.2015.2402643>
- [44] Qiu, Q., Sapiro, G. (2015). Learning transformations for clustering and classification. *J. Mach. Learn. Res.*, 16(1): 187-225.
- [45] Blumensath, T., Davies, M.E. (2009). Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Transactions on Information Theory*, 55(4): 1872-1882. <https://doi.org/10.1109/TIT.2009.2013003>
- [46] Jiao, Y., Li, G., Gu, Y. (2017). Principal angles preserving property of Gaussian random projection for subspaces. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, IEEE, 318-322. <https://doi.org/10.1109/GlobalSIP.2017.8308656>
- [47] Heckel, R., Tschannen, M., Bölcskei, H. (2017). Dimensionality-reduced subspace clustering. *Information and Inference: A Journal of the IMA*, 6(3): 246-283. <https://doi.org/10.1093/imaiai/iaw021>
- [48] Zhang, H.Y., Zhai, H., Zhang, L.P., Li, P.X. (2016). SpectralCSpatial sparse subspace clustering for hyperspectral remote sensing images. *IEEE Transactions on Geoscience & Remote Sensing*, 54(6): 3672-3684.
- [49] Engl, H.W., Ramlau, R. (2015). *Regularization of Inverse Problems*. Springer Berlin Heidelberg.
- [50] Ahmed, M.N., Cooper, B.E. (2005). Content-based document enhancement by fuzzy clustering with spatial constraints. In *Applications of Neural Networks and Machine Learning in Image Processing IX*. International Society for Optics and Photonics, 5673: 57-64. <https://doi.org/10.1117/12.585543>
- [51] Luo, F., Du, B., Zhang, L., Zhang, L., Tao, D. (2018). Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image. *IEEE Transactions on Cybernetics*, 49(7): 2406-2419.
- [52] http://www.ehu.es/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes.