Correntropy-Based Sparse Spectral Clustering for Hyperspectral Band Selection

Weiwei Sun, Member, IEEE, Jiangtao Peng, Member, IEEE, Gang Yang, and Qian Du, Fellow, IEEE

Abstract—This letter presents a correntropy-based sparse spectral clustering (CSSC) method to select proper bands of a hyperspectral image. The CSSC first constructs an affinity matrix with the correntropy measure which considers the nonlinear characteristics of hyperspectral bands and can suppress effects from noise or outliers in measuring band similarity. The CSSC imposes the sparsity and block diagonal constraint on spectral clustering, which can further improve band clustering performance. Bands are finally selected from each cluster on the connected graph. Experimental results on two widely used hyperspectral images show that the CSSC behaves better than spectral clustering and other several state-of-the-art methods in band selection.

Index Terms—Band selection, correntropy, hyperspectral imagery (HSI), remote sensing, sparse spectral clustering (SSC).

I. INTRODUCTION

Hyperspectral imaging techniques simultaneously collect spectral information, spatial information, and radiation information of ground objects with hundreds of bands [1], [2]. The classification of collected hyperspectral imagery (HSI) can be utilized to identify different materials with subtle divergences [3]. However, too many spectral bands bring problems including information redundancy, heavy computation burden, and “Hughes” phenomenon, which seriously limits realistic applications [4], [5]. Dimensionality reduction using band selection is a feasible choice to handle these problems.

Band selection is to select a proper band subset which has maximal information and minimal interband correlation. The selected bands could well inherit initial spectral meanings of ground objects. Until now, many band selection methods have been proposed, such as ranking-based, searching-based, sparsity-based, and clustering-based methods. The ranking-based methods quantify the importance of each band and select the top-ranked bands, e.g., maximum-variance principal component analysis [6] and the manifold ranking-based algorithm [7]. They highly depend on the metric about the distinctiveness of spectral bands and could not alleviate the information redundancy of spectral bands. The searching-based methods choose proper bands by optimizing a given measure, and typical examples are the particle swarm optimization algorithm [8], fire fly algorithm [9], and multigraph determinantal point process model [10]. The nonlinear searching procedure always brings about high computational complexity. The sparsity-based methods originate from the sparsity theory of compressive sensing, which makes full use of sparse coefficients of all bands to select bands. Typical examples are improved sparse subspace clustering [11], low-rank representation [12], and fast and robust self-representation [13]. However, the sparse coefficients are sensitive to the convergence of defined optimization program. The clustering-based methods group all bands into different clusters and select the ones that are closest to their cluster centroids. Typical approaches are the optimal clustering framework [14], dual-cluster band selection [15], fuzzy clustering [16], and spectral clustering (SC) [17]. We focus on SC in this letter, because it is more suitable for high-dimensional data.

For hyperspectral band selection, SC first constructs a full connected graph through an affinity matrix and then partitions all the bands into different clusters to select bands [18]. The selected bands are determined by two key issues: the construction of the affinity matrix and the band clustering result on the connected graph. The affinity matrix can be constructed by many measures, e.g., spectral angle measure (SAM), Euclidean distance, and correlation coefficient. Unfortunately, the Gaussian or linear assumptions of above measures contradict with the nonlinear properties of HSI data [19]. They are accordingly sensitive to noise or outliers and easily bring about large error or bias in measuring band similarity. On the other hand, the affinity matrix is assumed to be sparse and have a block diagonal structure in order to guarantee good clustering result on the connected graph [11].

To handle the aforementioned problems, we present a correntropy-based sparse spectral clustering (CSSC). The CSSC utilizes the correntropy measure [20] to construct the affinity matrix and then imposes the sparse and block diagonal structure into the regular SC to formulate the sparse spectral clustering (SSC) model [21]. Finally, proper bands are selected from each cluster on the connected graph. Compared with other methods, our CSSC method has two main contributions. To the best of author’s knowledge, this is the first time to adopt the correntropy measure for hyperspectral band selection. The correntropy measures nonlinear characteristics of hyperspectral bands and could better suppress negative effects from noise or outliers in quantifying the similarity between two bands. In addition, the CSSC considers the sparse and block diagonal

structure of the affinity matrix and promotes the clustering result of SC.

## II. SPECTRAL CLUSTERING

Let all the HSI bands constitute a matrix \( Y = [y_1, \ldots, y_N] \in \mathbb{R}^{M \times N} \), where \( M \) and \( N \) are the number of pixels and bands, respectively, and \( y_i \) is the \( i \)th band vector. The steps of SC for selecting a band subset \( M = Y(:,k) \in \mathbb{R}^{M \times k} \) can be summarized as follows.

1. An affinity matrix \( W \in \mathbb{R}^{N \times N} \) is constructed to quantify the similarity between two bands \( y_i \) and \( y_j \). The matrix can be regarded as a full connected undirected weighted graph \( G = (V, W) \), where \( w_{ij} \in V \) is the edge between pairwise bands, and the weight \( w_{ij} \in W \) measures their similarity.

2. The normalized Laplacian matrix is computed with \( L = I - D^{-(1/2)}W D^{-(1/2)} \), where \( D \) is a diagonal matrix with each diagonal element as \( d_{ii} = \sum_{j=1}^{N} w_{ij} \).

3. A matrix \( U \) is computed by solving

\[
\min_{U \in \mathbb{R}^{N \times k}} \langle L, UU^T \rangle, \quad s.t., \quad U^T U = I. \tag{1}
\]

The columns of \( U \) are the first \( k \) eigenvectors of the Laplacian matrix \( L \) corresponding to the \( k \) smallest nonzero eigenvalues.

1. The rows of \( U \) are normalized to have unit Euclidean length and then are clustered into \( k \) groups by \( k \)-means.
2. The bands closest to their cluster centroids are selected.

## III. CORRENTROPY-BASED SPARSE SPECTRAL CLUSTERING

### A. Affinity Matrix Construction With Correntropy Measure

The current measures for calculating the affinity matrix are based on Gaussian or linear assumptions and are sensitive to noise or outliers within the HSI bands. Take the Euclidean distance as an example, the errors from noise or outliers quadratically increase in the loss function of least squares as an example, the errors from noise or outliers of hyperspectral bands, both the formed affinity matrix \( W \) directly correlate with the indicator matrix of all the bands, whose row entries indicate to which cluster the bands belong. Accordingly, the matrix \( UU^T \) has the similarly sparse and block diagonal structure with the affinity matrix \( W \), which also indicates the true membership of the HSI clusters [21]. In real-world applications, due to the collecting noise or outliers of hyperspectral bands, the former affinity matrix \( W \) and the matrix \( UU^T \) in SC are not necessarily to be sparse and block diagonal. That would negatively affect band segmentation and degrade the accuracy of band selection. In this letter, we impose the sparse constraint on the matrix \( UU^T \) to improve the discriminability of all the bands. The SSC model is formulated as

\[
\min_{U \in \mathbb{R}^{N \times k}} \langle L, UU^T \rangle + \lambda \|UU^T\|_1, \quad s.t., \quad U^T U = I \tag{3}
\]

where \( \| \cdot \|_1 \) is the \( l_1 \)-norm operation and \( \lambda > 0 \) is the regularization parameter that trades off the objective of SC and the sparsity of \( UU^T \). Generally, the intracluster connections in the affinity matrix \( W \) are relatively strong while the intercluster connections are weak. Accordingly, the entries of \( UU^T \) corresponding to weak intercluster connections tend to be zeros, whereas the ones corresponding to the strong intracluster connections should be maintained. Therefore, the sparsity regularization term improves the sparse and block-diagonal structure of the matrix \( UU^T \) and benefits the following clustering for band selection.

### B. CSSC Model

With the above affinity matrix \( W \), the \( k \) eigenvectors corresponding to the \( k \) smallest nonzero eigenvalues of the normalized Laplacian matrix \( L \) can be computed to find clusters of all the bands. Ideally, the affinity matrix \( W \) is assumed to be sparse and block diagonal to guarantee the true clusters of all the HSI bands. Furthermore, the nonzero entries of \( W \) directly correlate with the indicator matrix of all the bands, whose row entries indicate to which cluster the bands belong. Accordingly, the matrix \( UU^T \) has the similarly sparse and block diagonal structure with the affinity matrix \( W \), which also indicates the true membership of the HSI clusters [21]. In real-world applications, due to the collecting noise or outliers of hyperspectral bands, both the formed affinity matrix \( W \) and the matrix \( UU^T \) in SC are not necessarily to be sparse and block diagonal. That would negatively affect band segmentation and degrade the accuracy of band selection. In this letter, we impose the sparse constraint on the matrix \( UU^T \) to improve the discriminability of all the bands. The SSC model is formulated as

\[
\min_{U \in \mathbb{R}^{N \times k}} \langle L, UU^T \rangle + \lambda \|UU^T\|_1, \quad s.t., \quad U^T U = I \tag{3}
\]

where \( \| \cdot \|_1 \) is the \( l_1 \)-norm operation and \( \lambda > 0 \) is the regularization parameter that trades off the objective of SC and the sparsity of \( UU^T \). Generally, the intracluster connections in the affinity matrix \( W \) are relatively strong while the intercluster connections are weak. Accordingly, the entries of \( UU^T \) corresponding to weak intercluster connections tend to be zeros, whereas the ones corresponding to the strong intracluster connections should be maintained. Therefore, the sparsity regularization term improves the sparse and block-diagonal structure of the matrix \( UU^T \) and benefits the following clustering for band selection.

### C. Inexact Augmented Lagrange Multiplier Optimization

We employ the inexact augmented Lagrange multiplier (IALM) method [25] to optimize model (3). The IALM introduces proper auxiliary variables to augment the constraints and iteratively minimizes the Lagrangian with respect to primal variables. First, an auxiliary variable \( P = UU^T \) is introduced, and the model (3) is equally transformed as

\[
\min_{U \in \mathbb{R}^{N \times k}} \langle L, UU^T \rangle + \lambda \|UU^T\|_1, \quad s.t., \quad P = UU^T, U^T U = I \tag{4}
\]

The augmented Lagrange function of (4) can be written as

\[
\mathcal{L}(P, U, A, \mu) = \langle L, UU^T \rangle + \lambda \|P\|_1 + \langle A, P - UU^T \rangle + \frac{\mu}{2} \|P - UU^T\|_2^2, \quad s.t., \quad U^T U = I \tag{5}
\]

where \( A \in \mathbb{R}^{N \times N} \) is the Lagrange multiplier and \( \mu > 0 \) is the penalty parameter.

The IALM optimizes the four variables with iterative procedures and updates each variable at iteration \( t + 1 \) using the following schemes. When fixing other variables, the primal variable \( U^{(t+1)} \) can be updated as

\[
U^{(t+1)} = \text{argmin}_{U \in \mathbb{R}^{N \times k}} \mathcal{L}(P^{(t)}, U, A^{(t)}, \mu^{(t)})
\]

\[
U^{(t+1)} = \text{argmin}_{U \in \mathbb{R}^{N \times k}} \|UU^T - P^{(t)} + (L - A^{(t)})/\mu^{(t)}\|_2^2, \quad s.t., \quad U^T U = I. \tag{6}
\]
Algorithm 1 CSSC for Hyperspectral Band Selection

Input: the HSI band matrix \( Y = \{y_i\}_{i=1}^{M} \in \mathbb{R}^{M \times N} \), the subset size \( k \), the error tolerance \( \tau = 10^{-6} \) and the maximum iteration time \( t_{\text{max}} = 10^6 \).

1. Compute the correntropy similarity matrix \( W \) using (3);  

2. Compute the normalized Laplacian matrix \( L = I - D^{-\frac{1}{2}}WD^{-\frac{1}{2}} \), where \( D \) is a diagonal matrix with \( d_{ii} = \sum_{j=1}^{N} \text{Corren}_{ij} \).

3. Construct the CSSC model in (3) and implement the IALM to optimize the augmented Lagrange program (5) 
\[
\mathcal{L}(P, U, \Lambda, \mu) = (L, UU^T) + \lambda \|P\|_1 + (A, P) - UU^T) + \frac{\mu}{2} \|P - UU^T\|_2^2, s.t., U^TU = I
\]

Initialization: \( A^{(0)} = 0, \tau = 10^{-6}, \mu^{(0)} = 10^{-6} \)

While \( \|U^{(t+1)} - U^{(t+1)}\|_\infty > \tau, \|U^{(t+1)} - U^{(t+1)}\|_\infty > \tau \) do 

a) Update \( U^{(t+1)} \) as 
\[
U^{(t+1)} = \arg\min_{U \in \mathbb{R}^{N \times k}} \|UU^T - P^{(t)}(L - A^{(t)}) / \mu^{(t)}\|^2, \text{s.t.}, U^TU = I
\]

b) Update \( P^{(t+1)} \) as 
\[
P^{(t+1)} = \arg\min_{P \in \mathbb{R}^{N \times N}} \|P - U^{(t+1)}U^{(t+1)^T} + A^{(t)}\|_1 + \|P - U^{(t+1)}U^{(t+1)^T} + A^{(t)}\|^2 + \frac{\mu^{(t)}}{\mu^{(t)}}\|U^{(t+1)}U^{(t+1)^T}\|^2.
\]

c) Update \( A^{(t+1)} \) and \( \mu^{(t+1)} \) using equations (8)–(9) 
\[
A^{(t+1)} = A^{(t)} + \mu^{(t)}(P - U^{(t+1)}U^{(t+1)^T})
\]
\[
\mu^{(t+1)} = \min(1.1 \mu^{(t)}, \mu_{\text{max}})
\]

End While

Output: the matrix \( \hat{U} = U^{(t+1)} \).

4. Compute the normalized row of \( U^{(t+1)} \) and cluster all rows into \( k \) groups using k-means.

5. Select bands with normalized row vectors in \( U^{(t+1)} \) that are closest to their cluster centroids

Output: the band subset \( M = Y(:,k) \).

When fixing other variables, the primal variable \( P^{(t+1)} \) can be updated in the following equation:
\[
P^{(t+1)} = \arg\min_{P \in \mathbb{R}^{N \times N}} \mathcal{L}(P, U^{(t+1)}, A^{(t)}, \mu^{(t)})
\]
\[
= \arg\min_{P \in \mathbb{R}^{N \times N}} \|P\|_1 + \|P - U^{(t+1)}U^{(t+1)^T} + A^{(t)}\|^2
\]
\[
= \arg\min_{P \in \mathbb{R}^{N \times N}} \|P\|_1 + \|P - U^{(t+1)}U^{(t+1)^T} + A^{(t)}\|^2
\]
\[
= \arg\min_{P \in \mathbb{R}^{N \times N}} \|P\|_1 + \|P - U^{(t+1)}U^{(t+1)^T} + A^{(t)}\|^2
\]
\[
\mu^{(t+1)} = \min(1.1 \mu^{(t)}, \mu_{\text{max}}).
\]

When fixing other variables, the Lagrange multiplier \( A^{(t)} \) and the penalty parameter \( \mu^{(t)} \) can be updated as 
\[
A^{(t+1)} = A^{(t)} + \mu^{(t)}(P - U^{(t+1)}U^{(t+1)^T})
\]
\[
\mu^{(t+1)} = \min(1.1 \mu^{(t)}, \mu_{\text{max}}).
\]

The above iterations from (6) to (9) are repeated until satisfying the convergence conditions \( \|P^{(t+1)} - P^{(t+1)}\|_\infty \leq \tau \) and \( \|U^{(t+1)}U^{(t+1)^T} - I\|_\infty \leq \tau \) or the maximum number of iterations \( t_{\text{max}} \geq 10^6 \), where \( \tau \) is a predefined residual error. With the optimal matrix \( \hat{U} = U^{(t+1)} \), the rows of the matrix are normalized to have the unit Euclidean length. After that, the \( k \)-means algorithm is implemented on the normalized rows to cluster them into \( k \) groups. Finally, the bands closest to all \( k \) cluster centroids are chosen to constitute the final band subset. Algorithm I lists the procedure of CSSC for hyperspectral band selection.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experimental Data Sets

The first data set is Indian Pines from the Multispectral Image Data Analysis System group at Purdue University. The data set was acquired from the Airborne Visible / Infrared Imaging Spectrometer sensor from JPL. It has 20-m spatial resolutions and 10-nm spectral resolutions covering a spectrum range of 400–2500 nm. The image scene of size 145 \( \times \) 145 pixels was implemented. The data set in Fig. 1 has 200 bands after preprocessing, and 16 classes of ground objects exist in the image scene.

The second data set is PaviaU taken from the Computational Intelligence Group in the Basque University. It was obtained from the Reflective Optics System Imaging Spectrometer sensor with 1.3-m spatial resolutions and 115 bands. After removing low SNR bands, 103 bands were implemented. The small subset shown in Fig. 2 contains 610 \( \times \) 340 pixels and covers the area of Pavia University. The image scene in Fig. 2 has nine classes of ground objects.

B. Parameter Setting

For the IALM in CSSC, the matrix \( U^{(0)} \) is initialized as the top \( k \) smallest nonzero eigenvectors of normalized graph Laplacian matrix \( L \), \( A^{(0)} = 0, \tau = 10^{-6}, \mu^{(0)} = 10^{-6} \), and \( \mu_{\text{max}} = 10^6 \). Considering its robust performance in handng high-dimensional HSI data with a limited number of training samples, the support vector machine (SVM) is implemented as the classifier, which adopts the radial basis function as the kernel function. Other popular classifiers [3], e.g., joint collaborative representation, random forest and deep learning-based classifiers, can also be considered to testify the classification performance of CSSC. The overall classification

Fig. 1. Image of Indian pines.

Fig. 2. Image of PaviaU.
accuracy (OCA) is used to quantify the classification accuracy. For the Indian Pines and PaviaU data sets, 10% of labeled samples on each class are randomly selected for training and the rest for predicting.

C. Classification Results

In this experiment, we compare the classification performance of CSSC with those of “SAM + SC,” “SAM + SSC,” “Correntropy + SC,” and two popular methods WaluDI [26] and linear prediction (LP) [27]. The “SAM + SC” utilizes the SAM to construct the affinity matrix and adopts the SC to select proper bands. The “SAM + SSC” implements both the SAM and SSC. The “Correntropy + SC” utilizes the correntropy measure and the regular SC. We manually change the number of selected bands \( k \) from 5 to 80. The parameter of SSC on Indian Pines and PaviaU are set to be 0.1 and 0.005, respectively. The parameters \( \lambda \) of CSSC on Indian Pines and PaviaU are set to be 0.005 and 0.01, respectively.

Fig. 3 shows the OCA curves of six different band selection methods and all the bands. The black dotted lines illustrate the classification accuracy of using all the original bands on both data sets. All the OCA curves from different methods increase with the number of selected bands \( k \) and become relatively stable after a certain value. It can be seen that the LP and WaluDI curves are always inferior to those of “SAM + SSC,” “Correntropy + SC,” and CSSC, and the CSSC performs the best. The “SAM + SC” behaves worse than those of “SAM + SSC,” “Correntropy + SC,” and CSSC. The observations explain that the correntropy measure improves the construction of affinity matrix in “SAM + SC,” and the SSC could promote the segmentation or clustering results of the affinity matrix in “SAM + SC.” The correntropy measure and SSC accordingly have a significantly positive contribution in CSSC for selecting more appropriate bands.

Fig. 4. Classification maps of all six methods on the Indian Pines data set. (a) LP (OCA = 73.50%). (b) WaluDI (OCA = 77.15%). (c) SAM + SC (OCA = 77.45%). (d) SAM + SSC (OCA = 79.71%). (e) Correntropy + SC (OCA = 80.59%). (f) CSSC (OCA = 81.58%).

Fig. 5. Classification maps of all six methods on the PaviaU data set. (a) LP (OCA = 93.90%). (b) WaluDI (OCA = 92.70%). (c) SAM + SC (OCA = 93.34%). (d) SAM + SSC (OCA = 93.97%). (e) Correntropy + SC (OCA = 94.17%). (f) CSSC (OCA = 94.77%).
The slightly better performance of “Correntropy + SC” over “SAM + SSC” indicates that the correntropy measure plays a better or at least similar role in improving the OCAs of the CSSC band subset. On the other hand, Figs. 4 and 5 illustrate the classification maps of all six methods on the Indian Pines and PaviaU data sets, where the number of selected bands \( k \) is manually set to be 30. The observations coincide with those of Fig. 3. We can conclude that using the CSSC bands can yield the best classification performance among all the six methods and the “Correntropy + SC” is the second one.

**D. Impacts of Regularization Parameter**

To investigate the impact of the regularization parameter \( \lambda \) of CSSC on band selection, the ranges of \( \lambda \) on both data sets are manually set to be [0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50]. Fig. 6 illustrates the OCA results of CSSC with different \( \lambda \). For the Indian Pines, the OCA rises with the increasing \( \lambda \) from 0.0001 and achieves the highest at \( \lambda = 0.005 \). After that, the OCA gradually falls down until the end of \( \lambda = 50 \). Similarly, the OCA curve of PaviaU has the highest value when \( \lambda = 0.001 \). Therefore, a small \( \lambda \) is recommended to guarantee good classification performance of CSSC in HSI band selection.

**V. CONCLUSION**

This letter presents a CSSC method to promote the performance of regular SC in selecting proper hyperspectral bands. The CSSC utilizes the correntropy measure and sparse representation to improve the construction of the affinity matrix and produce a more accurate selection of bands on the connected graph. Experimental results on two HSI data sets demonstrate that both the correntropy measure and SSC contribute significantly to band selection of CSSC. Moreover, the CSSC shows its superiority over several state-of-the-art band selection methods.

**REFERENCES**


