Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis
Qian Du, Senior Member, IEEE, and He Yang, Student Member, IEEE

Abstract—Band selection is a common approach to reduce the data dimensionality of hyperspectral imagery. It extracts several bands of importance in some sense by taking advantage of high spectral correlation. Driven by detection or classification accuracy, one would expect that, using a subset of original bands, the accuracy is unchanged or tolerably degraded, whereas computational burden is significantly relaxed. When the desired object information is known, this task can be achieved by finding the bands that contain the most information about these objects. When the desired object information is unknown, i.e., unsupervised band selection, the objective is to select the most distinctive and informative bands. It is expected that these bands can provide an overall satisfactory detection and classification performance.

In this letter, we propose unsupervised band selection algorithms based on band similarity measurement. The experimental result shows that our approach can yield a better result in terms of information conservation and class separability than other widely used techniques.

Index Terms—Band selection, classification, detection, hyperspectral imagery, similarity measurement.

I. INTRODUCTION

A hyperspectral image cube contains hundreds of spectral bands with very fine spectral resolution. Its abundant spectral information provides the potential of accurate object identification. However, its vast data volume brings about problems in data transmission and storage. In particular, the very high data dimensionality presents a challenge to many traditional image analysis algorithms. One approach of reducing the data dimensionality is to transform the data onto a low-dimensional space by using certain criteria. For instance, the objective of principal component analysis (PCA) is to maximize the variance of the transformed data (or minimize the reconstruction error), whereas that of Fisher’s linear discriminant analysis is to maximize the class separability. However, these methods usually change the physical meaning of the original data because the channels in the low-dimensional space do not correspond to individual original bands but their linear combinations.

Another dimensionality reduction approach is band selection. It is to select a subset of the original bands without losing their physical meaning. In most applications, the goal of hyperspectral image analysis is to detect or classify objects. Driven by detection or classification accuracy, one would expect that, using a subset of original bands, the accuracy is unchanged or tolerably degraded, whereas computational complexity is significantly reduced.

Many band selection methods have been proposed. In terms of object information availability, band selection techniques can be divided into two categories: supervised and unsupervised. Supervised methods are to preserve the desired object information, which is known a priori, whereas unsupervised methods do not assume any object information. For example, canonical analysis was employed for band selection in [1]; the Jeffries–Matusita distance, divergence, and Bhattacharyya distance between classes were used as selection criteria in [2]–[4], respectively.

Although these supervised techniques clearly aim at selecting bands that include important object information and the selected bands can provide better detection or classification than those from unsupervised techniques, the required prior knowledge may be unavailable in practice. Therefore, it is a need to develop reliable unsupervised band selection methods that can generally offer good performance regardless of the types of objects to be detected or classified in the following data analysis.

Unsupervised band selection has also been studied. A series of approaches were compared in [5]. For instance, first spectral derivative (FSD) and uniform spectral spacing (USS) can be easily implemented with superior performance in general. PCA and noise-adjusted PCA were proposed for unsupervised band selection in [1], distance-based measurement was investigated in [6], and information-theory-based band selection can be found in [7] and [8].

Because the basic idea of unsupervised band selection methods is to find the most distinctive and informative bands, the approaches that are proposed to search for distinctive spectral signatures as endmembers can be applied. The major difference is that the algorithms are applied in the spatial domain for band selection, instead of being applied in the spectral domain for endmember extraction. There are quite a few endmember extraction algorithms existing, and a review and comparative study can be found in [9]. In general, endmember extraction algorithms can be divided into the following two categories: one extracting distinctive pixels based on similarity measurement and the other using the geometry concept, such as simplex. The endmember extraction algorithms using unsupervised fully constrained least squares linear unmixing (UFCLSLU) in [10] and orthogonal subspace projection (OSP) in [11] belong to the first category, whereas the well-known pixel purity index [12] and NFINDR algorithms [13] belong to the second category.

In this letter, we propose the application of those similarity-based endmember extraction algorithms for band selection. To ensure that the selected bands are not only distinctive but also...
informative, data preprocessing, including bad band preremoval and data whitening, is needed. Practical considerations, such as algorithm initialization, the number of pixels to be involved in the band selection process, and the determination of the number of bands to be selected, are investigated. Several widely used band selection algorithms, such as the USS and FSD methods, serve as the primary comparison basis due to their computational efficiency and robustness; experimental results demonstrate that our approach can compete with them in terms of classification accuracy. The NFINDR-based band selection method is also compared, and our data preprocessing step can improve its performance.

II. SIMILARITY-BASED BAND SELECTION

A. Data Preprocessing

To select the most distinctive but informative bands, water absorption and low SNR bands need to be preremoved. This is because they can be very distinctive but not informative. Instead of manual selection, we compute the spectral correlation coefficients between original bands; those bands that have very low correlation coefficients with adjacent bands are considered as bad bands and will be preremoved [15].

The noise component in different bands is varied. If the noise component is larger, a band may look more different from others, although it may not be informatively distinct. Thus, noise whitening is needed, which requires noise estimation. It is known that noise estimation is a difficult task. In [16], it was demonstrated that the net effect of noise whitening and data whitening is similar. Therefore, we apply data whitening to the original bands (after bad band removal), which can be easily achieved by the eigendecomposition of the data covariance matrix. Then, the whitened bands actually participate in the following band selection process. Note that the selected bands are the original ones, not the whitened ones.

B. Properties of Similarity-Based Band Selection

To select the distinctive bands or the most dissimilar bands, a similarity metric needs to be designated. The widely used metrics include distance, correlation, etc. The measurement is taken on each pair of bands. Here, we prefer to use the approaches where band similarity is evaluated jointly instead of pairwise. The proposed band selection algorithms use the same concept in endmember extraction have this property.

In addition, due to the large number of original bands, the exhaustive search for optimal band combinations is computationally prohibitive. The straightforward criteria that can be employed for similarity comparison include linear prediction (LP) and OSP, which can jointly evaluate the similarity between a single band and multiple bands. The concept in the LP-based band selection was originally used in the UFCLSLU for endmember pixel selection in [10], which means that a pixel with the maximum reconstruction error, using the linear combination of existing endmember pixels, is the most distinctive pixel. The difference here is that, for band selection, there is no constraint imposed on the coefficients of linear combination. The OSP-based band selection is the same as the one for endmember extraction algorithm in [11].

1) LP: Assume that there are two bands $B_1$ and $B_2$ in $\Phi$ with $N$ pixels each. To find a band that is the most dissimilar to $B_1$ and $B_2$, $B_1$ and $B_2$ are used to estimate a third band $B$, i.e.,

$$a_0 + a_1B_1 + a_2B_2 = B'$$

where $B'$ is the estimate or linear prediction of band $B$ using $B_1$ and $B_2$, and $a_0$, $a_1$, and $a_2$ are the parameters that can minimize the linear prediction error: $e = \|B - B'|$. Let the parameter vector be $a = (a_0a_1a_2)^T$. It can be determined using a least squares solution

$$a = (X^TX)^{-1}X^Ty$$

where $X$ is an $N \times 3$ matrix whose first column is one, second column includes all the $N$ pixels in $B_1$, and third column includes all the pixels in $B_2$, and $y$ is an $N \times 1$ vector with all the pixels in $B$. The band that yields the maximum error $e_{\text{min}}$ (using the optimal parameters in $a$) is considered as the most dissimilar band to $B_1$ and $B_2$ and will be selected as $B_3$ for $\Phi$. Obviously, the similar procedure can be easily conducted when the number of bands in $\Phi$ is larger than two.

2) OSP: Assume that there are two bands $B_1$ and $B_2$ in $\Phi$. To find a band that is the most dissimilar to $B_1$ and $B_2$, an orthogonal subspace of $B_1$ and $B_2$ is constructed as

$$P = I - Z(Z^TZ)^{-1}Z^T$$

where $I$ is an $N \times N$ identity matrix, and $Z$ is an $N \times 2$ matrix whose first column includes all the pixels in $B_1$ and second column includes all the pixels in $B_2$. Then, the projection $y_o = P^Ty$ is computed, where $y$ includes all the pixels in $B$
and \( y_o \) is the component of \( B \) in the orthogonal subspace of \( B_1 \) and \( B_2 \). The band that yields the maximum orthogonal component \( \| y_o \| \) is considered as the most dissimilar band to \( B_1 \) and \( B_2 \) and will be selected as \( B_3 \) for \( \Phi \). The similar procedure can be easily conducted when the number of bands in \( \Phi \) is larger than two.

It can be proved that the least squares solution to the unconstrained linear unmixing problem is mathematically equivalent to the OSP solution. Therefore, the LP-based approach (which can be considered as the unconstrained linear unmixing problem) actually yields the identical selected bands from the OSP-based approach. Compared to the OSP solution, the LP-based approach is computationally more efficient because it involves matrices with relatively smaller sizes. Thus, in this letter, we will only present the results from the LP-based approach due to its computational efficiency.

### D. Band Used as the Initial for Band Selection

The initial band pair is critical to the performance of the proposed algorithms. Intuitively, we should use the two bands whose dissimilarity is the largest. Instead of the exhaustive search, the following algorithm can be applied to an original \( M \)-band data set.

1) Randomly select a band \( A_1 \), and project all the other \( M - 1 \) bands to its orthogonal subspace \( \langle A_1 \rangle^\perp \).
2) Find the band \( A_2 \) with the maximum projection in \( \langle A_1 \rangle^\perp \), which is considered as the most dissimilar to \( A_1 \).
3) Project all the other \( M - 1 \) bands to the orthogonal subspace \( \langle A_2 \rangle^\perp \), and find the band \( A_3 \) with the maximum projection.
4) If \( A_3 = A_1, A_1 \) and \( A_2 \) are confirmed to be the pair with the most significant dissimilarity, and the algorithm is terminated; if \( A_3 \neq A_1 \), go to the next step.
5) Continue the algorithm until \( A_{i+1} = A_{i-1} \), then either \( A_{i-1} \) or \( A_i \) can be used as the band selection initial \( B_1 \) (or \( A_{i-1} \) and \( A_i \) are used as the initial band pair).

We find out that this algorithm can always extract the two most distinctive bands, regardless of its initial \( A_1 \), although it will result in a suboptimal set in the following band selection.

### E. Practical Considerations

1) Number of Pixels Involved in the Band Selection Process:
The size of the \( X \) and \( Z \) matrices is very large due to the large number of pixels in an image. Fortunately, we find out that using a small subset of pixels in the band selection process will not change the results in most cases. This is because each band image is spatially highly correlated. Therefore, in the experiments, we randomly choose 10% or 1% of \( N \) pixels for band similarity assessment.

2) Number of Bands to be Selected: In practice, it is difficult to know how many bands should be selected. The rule of thumb is that more bands need to be selected if an image scene is complicated and contains many classes. This is because the data dimensionality should be high enough to accommodate these classes for detection or classification. The number of classes present in an image scene can be estimated by using a virtual dimensionality (VD) estimation approach proposed in [17], which can be considered as a reference value for the number of bands to be selected.

### F. Performance Evaluation

In order to evaluate the amount of information and class separability in the selected bands, a supervised classification algorithm, such as constrained linear discriminant analysis [18], can be applied. When the class information is unknown, an unsupervised algorithm, such as independent component analysis (ICA), can be applied.

When the pixel-level ground truth is unavailable, the classification maps from all the original bands can be considered as ground truth, and those from the selected bands are compared with them using spatial correlation coefficient \( \rho \). An average \( \rho \) that is closer to one generally means better performance. This is under the assumption that, using all the original spectral bands (after band selection), the best or at least satisfying classification performance can be provided. For classes with similar but separable spectra, this is a reasonable assumption [19]. Such a method based on image similarity provides quantitative evaluation even in an unsupervised situation or in the lack of pixel-level ground truth.

### III. Experiment

The 224-band AVIRIS Cuprite image was used in the experiment. As shown in Fig. 1, the image scene of size 350 \( \times \) 350 is well understood mineralogically [20]. At least five minerals were present: alunite (A), buddingtonite (B), calcite (C), kaolinite (K), and muscovite (M). Their approximate spatial locations of these minerals are marked in Fig. 1. After water absorption and low SNR bands were removed, 189 bands were left for band selection.

Fig. 2(a) shows the supervised classification result using all the original 189 bands, and Fig. 2(b) is the result using only 15 bands selected with the proposed LP-based band selection algorithm (using the whitened data). We can see that they are comparable, although the classification maps in Fig. 1(a) have clearer background. The average correlation coefficient between the two classification results is about 0.64. However, this reduces the data dimensionality from 189 to 15.

The FSD, USS, NFINDR, and joint entropy in [8] were used for comparison purposes.

### A. Impact of the Data Whitening Process

The results from using the original and whitened bands during band selection were compared. The initial band was exhaustively searched for the LP and USS methods, and the results are marked as “optimal.” As shown in Fig. 3 for supervised
classification, the LP-based approach with the whitened bands yielded a much better result than the one with the original bands. By using the whitened bands, the performance of the FSD method improved in some cases; the USS method was not influenced. Fig. 4 is the case using unsupervised classification for evaluation, where the observation is the same. Figs. 5 and 6 show the comparison among the LP, NFINDR, and entropy methods using supervised and unsupervised classifications, respectively. We can see that the LP method still outperformed the NFINDR and entropy methods. In any case, the performance using the whitened data is better than that using the original data for all the methods (except that the USS method is not affected);

the LP-based band selection using the whitened data provided the best results in terms of classification accuracy in both supervised and unsupervised fashion; the NFINDR provided satisfactory results in both supervised and unsupervised cases; surprisingly, the simplest USS method is very robust in both supervised and unsupervised cases.

B. Impact of Suboptimal Initials

Figs. 7 and 8 show the comparison about the impact of the suboptimal initials in supervised and unsupervised cases, respectively. According to the algorithm in Section II-D, the best initial is band 120 for the LP-based method. Here, the comparison is only between the LP and USS because both of them are sensitive to the selection of initials. We can see that the suboptimal initials did not influence the performance of the LP method very much but significantly degraded the performance of the USS method because its performance is oscillated with the number of bands to be selected (and the best initial band for the LP-based method may not belong to the USS-selected band subset).

C. About the Number of Bands to Be Selected

The three VD methods in [17] provided different estimates. In general, the noise subspace projection (NSP) method yielded the largest estimate. To be conservative, the value of VD from
but also informative, which can also improve the performance of other existing band selection approaches;
3) developing a fast algorithm to search a suboptimal set, which does not significantly degrade the performance.

According to the experimental results, our methods can yield superior performance with less sensitivity to band initial as the USS approach does. In addition, the VD estimate may be used as a reference for the number of bands to be selected.

FIG. 7. Comparison between optimal and suboptimal initial bands on supervised classification accuracy (using the whitened data).

Fig. 8. Comparison between optimal and suboptimal initial bands on unsupervised classification accuracy (using the whitened data).

the NSP method was used as a reference. For this scene, VD = 23 when the probability of false alarm pf = 10^{-3}, and VD = 21 when pf = 10^{-4}. When 25 bands were selected, the correlation coefficients from the LP-based band selection were 0.78 in the optimal case using supervised classification, 0.69 in the optimal case using unsupervised classification, 0.76 in the suboptimal case using supervised classification, and 0.67 in the suboptimal case in unsupervised classification. If 40 bands were selected, these values became 0.85, 0.74, 0.83, and 0.72, respectively. Although all the corresponding values were slightly increased, the number of selected bands was almost doubled. Thus, VD may be a reasonable indicator on the appropriate number of bands to be selected, which can achieve the balance between classification accuracy and data dimensionality reduction.

IV. CONCLUSION

We developed new unsupervised band selection algorithms for hyperspectral imagery. The major contributions are the following:
1) employing the ideas of similarity assessment originally developed for distinctive pixel identification in endmember extraction to unsupervised band selection;
2) applying a data whitening process before band selection to ensure that the selected bands are not only distinctive

REFERENCES