Unsupervised dimensionality reduction of hyperspectral images using representations of reflectance spectra

S. Enayat Hosseini Aria[1,2,*], Massimo Menenti[2,3], Ben G. H. Gorte[4], and Saeid Homayouni[5]

[1]Engineering Faculty, Islamic Azad University, Mashhad Branch, Iran;
[2]Department of Geoscience and Remote Sensing, Delft University of Technology, The Netherlands;
M.Menenti@tudelft.nl
[3]State Key Laboratory of Remote Sensing Science, Jointly Sponsored by Institute of Remote Sensing and Digital Earth of Chinese Academy of Sciences and Beijing Normal University, Beijing, P.R. China.
[4]Faculty of Built Environment, University of New South Wales, Australia; b.gorte@unsw.edu.au
[5]Centre Eau Terre Environnement, Institut National de la Recherche Scientifique, Québec Canada;
saeid.homayouni@ete.inrs.ca
* Correspondence: enayat.h.aria@gmail.com; Tel.: +31-633008155

Abstract: Unsupervised feature selection (UFS) is a standard approach to reduce the dimensionality of hyperspectral images (HSIs). The main idea in UFS is to define a similarity metric, and select the features minimising the metric to reduce the data redundancy. In this paper, we proposed a novel criterion for unsupervised dimensionality reduction based on the representation of spectral reflectance to capture dominant reflectance variations. Since capturing all the spectral information from an entire hyperspectral dataset is a time-consuming process, we proposed a greedy search algorithm for spectral representation (GSSR). It divides the spectrum into spectral regions with less spectral variations and merges them. This algorithm, similar to FS techniques, preserves the original data from being distorted or compromised by a transformation. We compared the GSSR algorithm with well-known existing algorithms in different experiments using various datasets. Comparison with the best approximation to represent single spectra as well as entire hyperspectral scene revealed that spectral representation is almost the same. The difference between the best spectral representation and the ones provided by GSSR is less than 0.01%; while on average, GSSR is about 660 times faster to represent single spectra and 37 times faster for a complete hyperspectral scene. Five well-known unsupervised dimensionality reduction methods were also implemented and used for comparison analysis. Based on the image classification accuracy over two hyperspectral datasets, the spectral features identified by the proposed criterion improved the classification accuracy as well.

Keywords: Hyperspectral, spectral representation, unsupervised feature selection, dimensionality reduction, greedy search, image classification.

1. Introduction

Hyperspectral imagers, also termed imaging spectrometers, capture reflected radiance in an image form, where every pixel in the image contains detailed spectral information in hundreds of adjacent narrow spectral channels. Unlike multispectral sensors, with three to ten spectral bands, hyperspectral sensors offer better potential for recognizing particular spectral properties (Manolakis, Marden, and Shaw 2003; Shaw and Burke 2003), such as absorption bands in minerals (Ben-Dor et al. 2008) or the leaf pigment content in vegetation types (Sims and Gamon 2002). On the other hand, the analysis of hyperspectral data may be very challenging because well-known image analysis algorithms are not easily extendable from the multispectral into the hyperspectral data. Classification of hyperspectral images, for example, is often based on notions of distance in the feature space, as in "minimum distance," "minimum Mahalanobis distance," and "k-Nearest Neighbor" classifiers or in variants of k-means clustering algorithms (Gorte 1998). Part of the challenge is that large sets of parameters usually are needed to describe the high-dimensional statistical distributions of attributes. To have a reliable estimation of these parameters, a large number of training samples is indispensable (Hughes 1968). Furthermore, the interpretation of distance metrics in high dimensional spaces is not straightforward, but instead highly unintuitive (Jain and Waller 1978; Jimenez and Landgrebe 1998; Durrant and Kaban 2009; Jia, Kuo, and Crawford 2013).

A solution to the problems raised by the hyper-dimensionality is to reduce the dimensionality while retaining the information required for various applications. In general, dimensionality reduction (DR) is the process of reducing the number of random variables under considerations. DR is categorized into two groups of feature extraction (FE) and feature selection (FS) methods. FE transforms the data into a new data space based on particular criteria (Kumar,

Ghosh, and Crawford 2001; Jimenez-Rodriguez, Arzuaga-Cruz, and Velez-Reyes 2007). Principal Component Analysis (PCA), as a classical and well-known method, eliminates the linear dependency or correlation between the components (new features) in the new feature space (Kaewpijit, Le Moigne, and El-Ghazawi 2002). Zhang et.al (2018) adopted manifold learning and structure sparse learning algorithms to project the spectral and spatial feature into a lower-dimensional subspace (Zhang et al. 2018). Recently, the low-rank matrix factorization techniques showed good potential for FE as well (Zhang et al. 2019).

The FE techniques might have better discriminating potential between the classes in a scene than the FS methods (Zaatour, Bouzidi, and Zagrouba 2017; Hira and Gillies 2015), but the main issue with FE is the loss of some critical and crucial information. Since the original data are no longer represented in the new data space, the information might have been compromised or distorted by the transformation. FS approaches, on the other hand, have the advantage of preserving the original information which is essential to analyse the spectral properties of observed materials (Chein and Su 2006; Martinez-Uso et al. 2007; Carmona et al. 2011; Jia et al. 2014). These techniques, also called band/channel selection, select subsets from original channels and are usually preferable for analysing hyperspectral data.

DR can be applied using both supervised and unsupervised strategies. Having labelled information, i.e., a priori knowledge about land covers in a scene paves the way for supervised DR. In other words, by selecting image samples for each class, a supervised DR algorithm provides a class-specific feature set. An example is the selection of channels maximizing the discrimination between given classes in the feature space (Huang and He 2005; Yang et al. 2011; Hosseini Aria, Menenti, and Gorte 2017). Contrary to supervised algorithms, unsupervised DR or unsupervised FS (UFS) techniques do not require any a priori information. Consequently, all the pixels in an image are considered for analysis. These methods are usually preferable for hyperspectral images

lacking the availability of labelled information (Du and Yang 2008; Cariou, Chehdi, and Le Moan 2011; Jia et al. 2012), which is the main objective of this study.

Different criteria can be applied to obtain features from a given dataset. A frequently used criterion in UFS is to define a similarity metric between the spectral channels and then select those channels with minimum similarity (Martinez-Uso et al. 2007; Du and Yang 2008; Cariou, Chehdi, and Le Moan 2011; Jia et al. 2012). The similarity is considered as the amount of the dependent information between features as well (Mitra, Murthy, and Pal 2002). The more similar the features, the more the dependent. Usually, in hyperspectral data, narrow adjacent spectral channels are highly correlated. As a result, the data suffer from redundancies. Therefore, by selecting less dependent spectral channels, the redundant information will be minimized, and consequently, the dimensionality reduces.

In this paper, unlike classical approaches using similarity metrics, the identification of the spectral features, which accurately represent the spectral reflectance, was applied as an unsupervised dimensionality reduction criterion. It means we developed an algorithm to obtain the most dominant variations of spectral signals of a hyperspectral scene, which can be indicators to distinguish different land covers and targets in a scene. For this purpose, instead of selecting individual channels, the adjacent spectral channels were categorized based on their spectral variations and then averaged; since having wider spectral bands provide more accurate image classifications (Hosseini Aria, Menenti, and Gorte 2017). Hereafter, the spectral features from the original hyperspectral data are called “channels,” and the ones made by averaging the neighbouring channels are named “bands.”

To achieve the objective, the spectral signal sampled by an imaging spectrometer is represented by a few spectral bands approximating the spectra with a required representation accuracy; i.e., the difference between an original spectrum and the approximated one is low enough

to recognize a specific target using its representation spectrum (Price 1994; Jensen and Solberg 2007). By doing this, the most relevant spectral properties of all pixels in an image, e.g., absorption features would be preserved for further analyses. When the spectral properties of pixels in an image are accurately identified, they could be classified correctly with a low number of features. Therefore, in this approach, a spectral band configuration is identified while minimizing the loss in accuracy of representation. One of the challenges here is to find a spectral configuration, i.e. the spectral locations of boundaries between spectral bands which can accurately represent all the pixels in a scene. This process is mainly a very time-consuming process in unsupervised scenarios since all the image pixels have to be considered for analysis. Employing a greedy search algorithm (Bendall and Margot 2006; Cormen 2009) makes a locally optimal choice at each iteration and provides spectral representations in a faster and more efficient fashion. Therefore, we propose an algorithm called GSSR (Greedy Search for Spectral Representation) to represent the spectra and evaluated it by comparisons with well-known existing algorithms for the same proposes in different experiments.

The paper is organized as follows. Section 2 reviews the criteria frequently utilized in UFS, the methods applying them to HSIs, and the algorithms applied for an accurate representation of spectral reflectance. Section 3 articulates the details of the proposed criterion and how it can be applied to a hyperspectral scene. The characteristics of the hyperspectral datasets used to assess the proposed method are given in Section 4. Section 5 describes the evaluation procedures followed by the results of different experiments, including the accuracy of spectral representation and image classification. Section 6 is the conclusion.

2. Related Works

In this section, we first present the criteria and the methods frequently used in UFS of HSIs.

Secondly, we review the algorithms for accurate spectral representation, since as mentioned, our proposed criterion for UFS is to identify the most dominant spectral features from the reflectance spectra of a hyperspectral scene by accurately representing the spectra.

2.1. UFS Criteria and methods

This section reviews the criteria mostly used in UFS and the algorithms applying them to HSIs. These criteria are usually based on similarity (or dependency) between hyperspectral features. Accordingly, the least similar spectral features have to be selected as the ones carrying less redundant information. A group of UFS criteria are obtained by calculating the similarity between just two spectral features and creating a matrix for all the features in a dataset. This matrix is symmetric. So, if $\mathbf{R}$ is a hyperspectral image with n spectral channels, $\mathbf{R} = \{\boldsymbol{R}_1, \boldsymbol{R}_2, \dots, \boldsymbol{R}_n\}$, and every channel ($\boldsymbol{R}_i$) is a vector with m pixels; $\boldsymbol{R}_i = \{\boldsymbol{r}_{i1}, \boldsymbol{r}_{i2}, \dots, \boldsymbol{r}_{im}\}$, where r values are defined in space Ω, i.e. $p \in \Omega$; the similarity matrix is presented as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} f_1(\boldsymbol{R}_1, \boldsymbol{R}_1) & f_1(\boldsymbol{R}_1, \boldsymbol{R}_2) & \dots & f_1(\boldsymbol{R}_1, \boldsymbol{R}_n) \\ f_1(\boldsymbol{R}_2, \boldsymbol{R}_1) & f_1(\boldsymbol{R}_2, \boldsymbol{R}_2) & \dots & f_1(\boldsymbol{R}_2, \boldsymbol{R}_n) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(\boldsymbol{R}_n, \boldsymbol{R}_1) & f_1(\boldsymbol{R}_n, \boldsymbol{R}_2) & \dots & f_1(\boldsymbol{R}_n, \boldsymbol{R}_n) \end{bmatrix}, \quad (1)$$

where $\boldsymbol{\Sigma}$ is the similarity or dependence matrix of $\mathbf{R}$, and $f_1(\boldsymbol{R}_i, \boldsymbol{R}_j)$ is the value of the dependence of the named variables. In this group, a specified search strategy is applied to the matrix and selects the channels with minimum similarity to the other channels in the dataset (Gu and Zhang 2003; Martinez-Uso et al. 2007; Qian, Yao, and Jia 2009; Jihao, Yisong, and Zhanjie 2010; Cariou, Chehdi, and Le Moan 2011; Jia et al. 2012).

The other group of UFS criteria can be calculated as a unique score without making a similarity matrix; i.e. the similarity or dependence score is not obtained just based on two features,

147 but more than two features. So, there is no need to make a similarity matrix. Given a set of features,
148 the score can indicate the amount of dependent or independent information of a feature in a set. In
149 both cases, a model taking into account multiple variables applies to a hyperspectral dataset to
150 calculate the score. So, in a general way, the score of a channel in a hyperspectral dataset can be
151 obtained as follows:

$$D_{R_i} = f_2(\boldsymbol{R}_1, \boldsymbol{R}_2, \dots, \boldsymbol{R}_n), \quad (2)$$

152 where D_{R_i} is the score of the channel $\boldsymbol{R}_i$ which is calculated by the model f_2 taking into account
153 multiple channels. A list of UFS criteria mostly used and the methods of applying them are
154 presented in the next subsections.

155 *2.1.1. Correlation Coefficient*

156 The Pearson coefficient of correlation was utilized in (Gu and Zhang 2003; Jihao, Yisong, and
157 Zhanjie 2010) to automatically subspace hyperspectral data in an unsupervised manner. After
158 making the correlation coefficient matrix, the authors used the local minimum of the correlation
159 coefficient between adjacent channels of the datasets to partition the spectral channels.

160 *2.1.2. Mutual Information*

161 Mutual information (I) is a quantitative measurement of the amount of shared information between
162 two random variables. Despite the correlation coefficient, it takes into account both linear and non-
163 linear dependencies (Dionisio, Menezes, and Mendes 2004). Less mutual information between two
164 random variables indicates more of uncertainty. As a result, zero, as the minimum value of the
165 metric, means the variables are not dependent at all. It is a dimensionless quantity, generally, with
166 units of bits (logarithms of base 2) (Cover and Thomas 2006):

$$I(\boldsymbol{R}_i, \boldsymbol{R}_j) = \sum_{\boldsymbol{r}_i \in \Omega} \sum_{\boldsymbol{r}_j \in \Omega} p(\boldsymbol{r}_i, \boldsymbol{r}_j) log \frac{p(\boldsymbol{r}_i, \boldsymbol{r}_j)}{p(\boldsymbol{r}_i) p(\boldsymbol{r}_j)}, \quad (3)$$

where $p(\boldsymbol{r}_i, \boldsymbol{r}_j)$ is the joint probability distribution function of $\boldsymbol{R}_i$ and $\boldsymbol{R}_j$, and $p(\boldsymbol{r}_i)$ and $p(\boldsymbol{r}_j)$ are the marginal probability distribution function of them.

The mutual information measure was utilized for UFS to cluster spectral channels with minimum-shared information by a recursive binary search algorithm (Cariou, Chehdi, and Le Moan 2011). Martínez-Usó et al. (2007) normalized the mutual information metric and converted it into a dissimilarity metric between two channels (Martinez-Uso et al. 2007). By building a symmetric dissimilarity matrix for the entire hyperspectral dataset, a hierarchical clustering process (Jain and Dubes 1988) was applied, to form clusters of channels as similar as possible within each cluster. After obtaining *k*-desired clusters, a channel was selected by using a weighting method to provide the best representative channel predicting the information content of the other channels in each cluster. Eventually, it selects channels with minimum shared information. This algorithm used Ward's linkage method (Ward 1963) in hierarchical clustering, so it is named WaLuMI (Ward's Linkage strategy using Mutual Information). The advantage of this method is that it is not a ranking or incremental method that selects channels taking into account the previously selected channels, i.e., *k* selected channels in the clustering-based strategy are not equal to the *k-1* selected channels plus another relevant channel.

2.1.3. Kullback-Liebler Divergence

This metric is also based on the information theory, which was applied for UFS of hyperspectral images (Martinez-Uso et al. 2007; Qian, Yao, and Jia 2009). The metric was considered as a dissimilarity distance between two probability distributions and interpreted as the cost of using one of the distributions instead of the other one (Martinez-Uso et al. 2007). Martínez-Usó et al. (2007)

applied an algorithm similar to WaLuMI by replacing the normalized mutual information metric with the Kullback-Liebler criterion, and they named it WaLuDi (Ward's Linkage strategy using Divergence).

2.1.4. Euclidean Distance

The negative Euclidean distance was used as a similarity measure in an affinity propagation-based channel selection algorithm. It is used to indicate how well a spectral channel represents other channels by making a similarity matrix. The channels are, then, clustered based on their similarities (Jia et al. 2012).

2.1.5. *Dependent Information Metric*

Sotoca. et al. (2007) defined a metric to obtain the dependent information of a set of random spectral channels. The set can have more than two variables, and therefore, there is no need to make a dependence matrix of pairs of variables (Sotoca, Pla, and Sanchez 2007). The metric measures the dependent information of a set of channels by employing the joint entropy and the conditional entropy. Applying the metric in a greedy search algorithm, the authors selected channels having the minimum-shared information (Sotoca, Pla, and Sanchez 2007).

2.1.6. *Linear Prediction*

Du et al. (2007) applied two similarity-based endmember extraction algorithms to select spectral channels in an unsupervised manner (Plaza et al. 2004; Du and Yang 2007, 2008). They searched the most distinctive channels based on linear unmixing methods. These methods model an HSI pixel as a linear mixture of a set of finite image endmembers and select the most distinctive channels in a sequential forward selection searching strategy (Du, Ren, and Chang 2003). They can

jointly evaluate the similarity between a given channel and multiple channels.

The first one, linear prediction (LP) criterion, makes a linear estimation of the selected channels and searches for the most dissimilar one. To find a channel that is the most dissimilar to channel $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$, one can first estimate $\boldsymbol{R}'$ as the linear prediction of the two channels by solving the following linear model using the least-squares solution:

$$\alpha_0 + \alpha_1 \boldsymbol{R}_1 + \alpha_2 \boldsymbol{R}_2 = \boldsymbol{R}', \quad (4)$$

where α_0, α_1, and α_2 are the parameters that can minimize the LP error. Then they searched for a channel having the maximum dissimilarity with the estimated channel, i.e., having the maximum $e = \|\boldsymbol{R}' - \boldsymbol{R}_i\|$, where $\|.\|$ can be the Euclidean norm of the vector.

2.1.7. Orthogonal Subspace Projection

Du et al. (2007) used orthogonal subspace projection (OSP) as the second criterion to map a channel onto an orthogonal subspace (Du and Yang 2007). With this criterion, first, an orthogonal subspace based on the pre-selected channels is constructed. Then, each channel is projected into the orthogonal space. The channel yielding the maximum orthogonal component is considered as the most dissimilar one to the pre-selected channels.

2.1.8. Independent Component Analysis

Independent component analysis (ICA) is a method that extracts independent source signals by searching for a linear transformation that minimizes the statistical dependence between the components (Comon 1994). Unlike PCA, ICA imposes higher-order dependence than the second-order one, so the components are not always orthogonal (Comon 1994; Hyvärinen and Oja 2000). ICA is used as a feature extraction technique in several approaches (Chiang, Chang, and Ginsberg

2000; Lennon et al. 2001; Robila and Varshney 2004; Liu et al. 2017), while Du et. al. (2003) presented an ICA-based method for feature selection as well to reduce the dimensionality of HSIs (Du et al. 2003). The authors weighed the spectral channels using the independent components and selected those having maximum information.

The ICA-based model represents a hyperspectral image, **R**, as the multiplication of an unmixing matrix, **A**, and a number of independent components called "sources", **S** :

$$\mathbf{R}_{n\times m} = \mathbf{A}_{n\times c} \times \mathbf{S}_{c\times m}, \quad (5)$$

where c is the number of source signals. Following this model, the ICA aims at deriving the best possible estimation of **S** by approximating the unmixing matrix **A** under some constraints. Based on the approximation of the unmixing matrix, Du et. al. (2003) calculated a mean absolute weight per spectral channel as the indicator of the information content of each channel. These weights were then sorted, and the channels with the highest weights were selected as the most informative ones.

2.2. Spectral representation

There are several methods for spectral approximation and representation (Price 1975; Price 1990; Li et al. 1999; Wang et al. 2007; Huynh and Robles-Kelly 2008; Angelopoulou 2000; Angelopoulou, Molana, and Daniilidis 2001), mostly used in colourimetric sciences. Their scope is to represent spectra accurately with a limited number of samples e.g. Discrete Fourier Transform (DFT) (Agrawal, Faloutsos, and Swami 1993), Singular Value Decomposition (SVD) (Keogh et al. 2001) or Discrete Wavelet Transform (DWT) (Kahveci and Singh 2001). A common way for the approximation is the replacement of local variations in a spectrum with a constant value over a small range in wavelength. Chakrabarti et. al. (2002) presented such a technique, named Adaptive

Piecewise Constant Approximation (APCA), and proved that this technique yields a better representation than other existing methods such as DFT and DWT for approximating signals in time series analyses (Chakrabarti et al. 2002). The APCA algorithm degrades a curve into a constant segment-based approximation, where the user specifies the number of segments. It includes two main steps. At first, it converts the signal approximation issue into a wavelet compression problem, for which there are well-known optimal solutions; and next, it converts the solution back to the APCA representation and makes minor modifications. The term 'segment' is equivalent to 'band' in our approach. More details on APCA can be found in (Chakrabarti et al. 2002). Approximating spectra by piecewise constant functions has also been used in other fields, taking into account the physical characteristics of the spectra to determine the location of the spectral segments (Thomson, Lue, and Bannerman 2014; Zehentbauer and Kiefer 2012).

Konno and Kuno (1998) proposed a method that provides the best piecewise approximation (Konno and Kuno 1988). They used the maximum norm and the Euclidean norm to find the approximation of a function of a single variable with less than a predefined number of constant-value segments. Another study applied the Bayesian approach for piecewise smoothing of one-dimensional signals (Winkler and Liebscher 2002). Later, an extension of this method was used for multiple spectral curves to reduce the dimensionality of hyperspectral scenes (Jensen and Solberg 2007). The goal of the last approach was also to partition the spectra of a hyperspectral scene into a fixed number of contiguous intervals with fixed intensities using the piecewise constant function approximations (PCFA) algorithm. The intensity in a spectral band is the mean value of the signal in its constitutive channels per pixel. Considering the number of bands, the algorithm examines all the possible spectral locations for the breakpoints and finds the best approximation having the lowest error of representation.

3. Spectral Representation Criterion

As reviewed, the criteria used in UFS algorithms applied to HSIs are based on defining a similarity metric between spectral features and selecting the ones having minimum similarity with other features. Following this procedure, the redundant information is decreasing while the dimensionality of the original dataset is reducing. In this approach, we suggest a criterion for UFS with regards to capturing the most relevant spectral information concerning all the pixels in a hyperspectral image. It means all the spectral reflectance of an entire scene are approximated in a way that the approximated spectra represent the original ones accurately.

For this purpose, we partition the reflectance signals of a hyperspectral scene into a predefined number of adjacent bands with fixed intensities. Figure 1 schematically illustrates the idea for a spectral signature with 195 channels, which is divided into six bands. In this figure, the blue line is the representation (approximation) of the original spectral signal (the red line). In fact, using this method, the channels are transformed into a new feature space but the transformation is in a way that the relations between the reflectance spectra and their wavelength are retained, similar to the FS techniques.

Figure 1 ---- >>>

We use the square error between the original reflectance spectrum (r) and the approximated (a) one to identify the error of representation. Since the intensity is a constant value in every spectral band of the representation; to minimize the error, the intensity of a spectral band should be the mean value of its constitutive channels. So, the root means square error (RMSE) is used to calculate the error of spectral representation. For a complete hyperspectral scene, the following procedure is performed.

let the hyperspectral dataset defined in Section 2 be divided into *k+1* bands where $k \leq n$ and k is the number of breakpoints. The set of the spectral locations of breakpoint is $\boldsymbol{S} = \{\boldsymbol{s}_0, \boldsymbol{s}_1, \dots, \boldsymbol{s}_k, \boldsymbol{s}_{k+1}\}$; where $\boldsymbol{s}_0 = 0$, $\boldsymbol{s}_{k+1} = n$, and $\boldsymbol{s}_1, \dots, \boldsymbol{s}_k$ indicate the channel numbers in an ascending order where the breakpoints are placed after them. Therefore, a new band set with m pixels in each band is $\boldsymbol{A} = \{\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_k\}$ where

$$\boldsymbol{A}_i = \frac{\sum_{s_{i-1} < t \leq s_i} \boldsymbol{R}_t}{\boldsymbol{s}_i - \boldsymbol{s}_{i-1}}. \tag{6}$$

To compute the error of representations for the entire scene, first, the reduced spectral configuration is expanded at each pixel back into the original channel configuration. Then, the values of each band are duplicated in the adjacent channels covered by the band. Next, the reconstructed and full spectra per pixel are compared by computing the RMSE between the two spectra. The difference between the expanded and the original spectra, then, can be calculated and averaged over all pixels to obtain the error (E_{rep}) of the representation for the entire scene:

$$E_{\text{rep}} = \frac{\sum_{j=1}^{m} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{r}_{ij} - \boldsymbol{a}_{ij}\right)^2}}{m}, \tag{7}$$

where $\boldsymbol{r}_{ij}$ and $\boldsymbol{a}_{ij}$ are the i^{th} signal value in the j^{th} pixel of the original and the approximated spectra respectively.

Having the representation error of different band configurations, we can select the optimal band set representing the signals with adequate accuracy. There is a huge number of combinations to select the location of the breakpoints and every band configuration gives different representation errors. Ideally, the best locations are the places where the total error (E_{rep}) is minimum. It can be achieved by an exhaustive search (Nievergelt 2000), i.e. all band configurations with the given number of bands are considered and evaluated. In a practical situation, however, the computational

cost for large datasets is prohibitive. This method can be used for a limited number of spectral signatures (Jensen and Solberg 2007).

In our approach, we apply a greedy search strategy to determine the spectral location in a sequential manner. This strategy was used in the FS algorithms to apply different criteria for DR of hyperspectral images as well (Pudil, Novovicova, and Kittler 1994; Sotoca, Pla, and Sanchez 2007; Le Moan et al. 2011; Yang et al. 2011; Han, Lee, and Bien 2013; Hosseini Aria, Menenti, and Gorte 2017). Using the spectral representation as a criterion, we named the algorithm as Greedy Search for Spectral Representation (GSSR). In this procedure, the algorithm iteratively selects a spectral location of a breakpoint that appears to be the best with regards to the representation error and the previously selected subset of breakpoints. The method significantly reduces the complexity by progressively ranking the evaluated subset.

3.1. Greedy search for spectral representation (GSSR).

Given a hyperspectral dataset, **R**, and the number of bands, *k+1*, the algorithm scans all possible spectral locations by taking into account the previously selected breakpoints to determine a new breakpoint in each iteration. For every tentative breakpoint, it creates the band set *(A)* (Eq. (6)) based on the preselected breakpoints and the new one, and then calculates the representation error (E_{rep}). Therefore, all possible locations for a new breakpoint are examined, and the best approximation having the lowest error of representation is identified. Then, the determined breakpoint is added to set *S*. This procedure terminates when the number of bands reaches the predefined value (*k+1*). Figure 2 illustrates the flowchart of the procedure.

Figure 2 ---- >>>

The GSSR algorithm complexity for a single spectrum is of the order $O(kn)$, where k is the number of breakpoints, and n is the number of channels in the original hyperspectral dataset. The complexity of the metric for spectra representation is of the order $O(mn)$, where m is the number of pixels. Hence, the overall computation time of the GSSR algorithm is $O(kmn^2)$.

In the next section, the data sets used for the assessment of the proposed algorithm are presented.

4. Hyperspectral Datasets

The GSSR algorithm was evaluated by applying it to different hyperspectral datasets, including a spectral library. We have done the necessary pre-processing steps before using the datasets, including atmospheric correction and removal of the noisy channels for the scenes. The noisy channels are those that do not have any signal, located at water absorption spectral regions, and the ones having a low signal to noise ratio (SNR). The channels with low SNR were identified by estimating the SNR using the geostatistical method described in (Curran and Dungan 1989), and visual inspection. We used the following datasets for the experiments:

- A spectral library: it comprised 1365 spectra from different materials and was developed by researchers at the Spectroscopy Lab, USGS, in 2007. The library is divided into six chapters: 1. Minerals, 2. Mixtures, 3. Coatings, 4. Volatiles, 5. Man-Made; and 6. Plants, Vegetation Communities, Mixtures with Vegetation, and Microorganisms. There is more than one spectrum for many of materials since different factors have been considered for the collection of the spectra such as the type of the spectrometer, the spectral resolution, the purity of the materials, the grain size, the presence of other elements in the sample, etc. The chapters contain 881, 138, 12, 24, 110, 200 spectra respectively. The library is used as a

reference for material identification in remote sensing images. The database is over 6000 webpages. More details of the spectral library can be found at https://speclab.cr.usgs.gov/spectral-lib.html. We used the convolved version of the library corresponding to the AVIRIS channels. After analyzing the spectral library, we found out that three pairs of the spectra are the same, and they cannot be distinguished from each other. These spectral pairs are from Mixtures and Plants chapters. The duplicated spectra were removed. Consequently, the final number of spectra in Chapter 2 and 6 became 136 and 199 respectively.

- Moffett Field: AVIRIS has acquired this dataset in California with 224 bands. The band set covers the spectrum from 365nm to 2497nm continuously with approximately 10nm-wide channels. The channels located at 366-385, 1353-1433, 1811-1948, 2337-2497 nm wavelength were removed due to noise and water absorption. As a result, the final dataset has 177 channels (Figure 3).

Figure 3 ---- >>>

- Indian Pines: the scene consists of 145*145 pixels with a spatial resolution of about 20m. Two-thirds of the Indian Pines scene is covered by agriculture, and one-third by forest and other natural perennial vegetation (Figure 4). The ground truth available documents sixteen classes, not mutually exclusive. Since three classes in the scene contain less than 50 samples, we do not use them for the experiments. After the atmospheric correction and the removal of noisy channels, the number of channels was reduced to 178. We removed water absorption channels (104-108, 150-163, and 220), noisy bands (1-4, 103, 109-111, 148-149, 164-166, and 217-219), and seven channels that are spectrally overlapping channels (32, 33, 95, 96, 158, 191, and 192). The Indian Pines dataset is available free of charge via

Purdue University website: https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html.

Figure 4 ---- >>>

- Salinas: This scene (Figure 5) is characterized by high spatial resolution (3.7m). The area covered comprises 512 lines by 217 pixels. The dataset is available at http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes only as at-sensor radiance. So, it has been atmospherically corrected, and the noisy and duplicated channels have been removed. The final dataset has 190 channels. The ground-truth is also available and documents 16 classes, including vegetables, bare soils, and vineyard fields, which we used in the experiments.

Figure 5 ---- >>>

5. Evaluation of the Proposed Method

Two types of experiments were performed to evaluate the GSSR algorithm: a) single signal representation and b) unsupervised dimensionality reduction of hyperspectral scenes. In both types of experiments, the accuracy of representation and the running time were evaluated, while for the second type of experiment, the image classification accuracy was also considered. We performed the assessments by comparing the algorithm with well-known existing algorithms for the same purposes.

5.1. Evaluation of the algorithm using single spectra

Two experiments were performed in order to assess the GSSR algorithm for representing a single spectrum. In the first experiment, the GSSR algorithm is compared with APCA and PCFA (Section

2.2) using various spectra. Both algorithms represent spectra with a set of constant signal value segments as GSSR (Figure 1). In this experiment, the three algorithms were compared in terms of the spectral representation and the running time.

In the second experiment, the reduced spectral configurations obtained by the algorithms were evaluated for material detection, i.e. different materials were detected by comparing the approximated spectra with the full spectra available in the spectral library.

5.1.1. Experiment 1: single spectra representation

This evaluation was performed as a benchmark to identify the error of representation of single spectra using the three methods: GSSR, PCFA, APCA. At first, three dominant reflectance spectra; soil, water, and vegetation, were compared, and the three algorithms were applied to represent the spectra with 5, 10, and 15 bands. These reflectance spectra were obtained from the pixels with the same land cover in the Moffett Field scene. Figure 6 illustrates the results, and Table 1 gives the error of the estimate by different algorithms for all the spectral configurations.

Figure 6 ---- >>>

Table 1 ---- >>>

As expected, the PCFA algorithm gave the smallest error of estimate in all cases, while APCA gave the largest error (Table 1). Interestingly, GSSR represented the spectra almost twice as accurately as APCA and with an accuracy comparable with the PCFA algorithm. For example, using ten bands to approximate the soil spectrum, the error of the estimate was 0.0230 with APCA, 0.0115 with GSSR, and 0.0096 with the PCFA algorithm. The difference in the error between GSSR and PCFA is less than 0.002, and it became lower when 15 bands were used, with the

difference in RMSE being 0.0005 only. It is also observed (Figure 6) that the locations of the breakpoints determined by the PCFA and GSSR algorithms are almost identical.

The GSSR and PCFA algorithm always divide the spectrum into the exact predefined number of bands, while APCA does not, as shown in this experiment. This situation occurred, for example, when seeking to approximate the water spectrum, with ten bands by the APCA algorithm. In this case, the reduced spectral configuration had one spectral band less than the prescribed number of bands, while the error of estimate would have been lower with one additional band. APCA is based on the Haar wavelet transform, so the number of samples in the original signal fed into the algorithm has to be a power of two. In the case that the signal does not have enough samples, it is padded with zeros, and later truncated. This process sometimes may yield fewer bands than expected.

We repeated the same experiment using more than 1000 pixels with different reflectance spectra derived from the Moffett Field AVIRIS image. The pixels were chosen to sample various land cover types, including different types of water, soil, vegetation, man-made features such as buildings, roads, etc. The reflectance spectra were represented separately for each pixel with a different number of bands starting from 5 to 30, in steps of 5. Figure.7 shows the results.

Figure 7 ---- >>>

The mean RMSEs decrease with an increasing number of bands. Similar to the previous results, the APCA error was the largest one, while the error for GSSR and PCFA algorithms were very similar and lower than when using APCA. The mean RMSE difference between PCFA and GSSR over all the spectra samples was about 0.0003 in the unit of the spectral reflectance, i.e. in [0,1], with the 5-band representation to 0.0001 with the 30-band one. The difference between GSSR and PCFA representation errors decreased with an increasing number of bands.

We also considered the run time required to carry out the numerical experiment on the dataset with more than 1000 spectra (Table.2). The time estimate is based on the implementation of the algorithms on a desktop computer that has the following characteristics: Operating system: Windows 7, Processor: Intel Core 2 and 16 GB RAM. The algorithms were written in IDL programing language, version 8.2.

Table 2 ---- >>>

The APCA algorithm is fast. The running time was less than two seconds for all the spectral configurations, while it was increasing with the number of bands for GSSR: started at less than 14 seconds for the 5-band representations and reached more than two and half minutes for the 30-band ones. PCFA consumed much more time than the two other algorithms. In the worst case, i.e. the 30-band representation, the run time was more than two days to find the representations for the 1089 spectra, which was about 1200 times slower than GSSR. On average, GSSR ran 660 times faster than PCFA. The main issue affecting the running time of the PCFA algorithm is that the algorithm recursively calls itself with respect to the number of bands, and checks all the possible situations. When the number of bands increases, the run time increases dramatically.

5.1.2. Experiment 2: material detection using approximated spectral signatures

The previous experiment showed that the GSSR algorithm yields comparable spectral configurations to the best representation provided by PCFA with much shorter run time. In the second experiment, we evaluated the spectral configurations obtained by GSSR by applying them to material detection.

This experiment reveals the number of bands needed to correctly identify a target spectral signature using the reduced spectral configurations derived by the algorithms GSSR and PCFA. The APCA algorithm was omitted since the spectral representations provided by this algorithm are

not as accurate as of the representations obtained by the other algorithms. For this experiment, we used the spectral library that contains different and well–defined spectra.

In this experiment, a spectrum from the library is selected as a "target spectrum." Then the GSSR and PCFA algorithms were applied to identify the breakpoints in such a way that is the reduced spectral configuration represents the full spectrum with increasing accuracy. At each iteration, the reduced spectral signature of the target spectrum was compared with all the spectra in the spectral library to check whether the approximated spectrum could correctly be identified, i.e. the approximated target spectrum and the full detailed one has the least difference. The iterative procedure was ended when the reduced spectral configuration of the target signature had been correctly identified, or the number of bands was more than 30. We used a distance-based identifier and a spectral angle based identifier to measure the difference between the known and unknown spectra (Kruse et al. 1993; Price 1994; Cochrane 2000). Finally, we calculated the percentage of spectra correctly identified vs. the number of bands (Figure 8).

Figure 8 ---- >>>

In general, the accuracy of the distance-based identifier is higher than the angle-based one. For instance, the 10-, 15-, and 20-band spectral configurations achieved correct identification of materials in 97%, 99%, and 100% of cases with the distance-based identifier, while with the angle-based identifier, the correct identification reached to 61%,79%, 86% respectively. The latter normalises the spectra and removes the signal intensity dependence, i.e., reflectance in this experiment.

Using either identifier, the reduced spectral configurations obtained with PCFA and GSSR gave a comparable accuracy in material detection. The spectral configurations obtained with PCFA gave slightly more identifications than the ones obtained with GSSR. The difference in

performance between PCFA and GSSR was higher when the number of bands is small, and it decreased with an increasing number of bands. However, if the spectral configuration obtained with GSSR has just one band more than the PCFA configuration, the detection accuracy for GSSR is higher. For example, using the angle–based identifier, the 16-band configurations obtained by GSSR were correctly identified in 80.7% of the cases, while the 15-band configuration obtained with PCFA was accurate in 80.6 % of cases. It should be noted that the computational cost of the 15-band PCFA configuration is much higher than the one of the 16-band configuration obtained by GSSR. In the example mentioned, GSSR was more than 2000 times faster than PCFA.

Both experiments (Section 5.1.1 and 5.1.2) revealed that the spectral representation of single spectra using GSSR is almost identical with the best spectral representation and has comparable accuracy with it in the representation and in detecting materials using the reduced band configurations. Meanwhile, GRRS provides the representations in a much faster way than obtaining the best representation. In the second type of experiment, we evaluated the algorithm applying to an entire hyperspectral scene.

5.2. Evaluation of the algorithms using the entire scene

In the second type of experiment, the GSSR algorithm was evaluated by two experiments using the entire hyperspectral scenes. In this case, the spectral locations of the breakpoints must be the same for all pixels to reduce the dimensionality of the image. At first, it was again compared with PCFA to assess the error of representation and the running time, since PCFA provides the best spectral approximations for all the spectral reflectance in a hyperspectral scene with the same situation as GSSR. It means that both algorithms take the average of adjacent channels to form wider spectral bands. Therefore, the band configurations, i.e., the spectral locations of breakpoints over the spectrum identified by the algorithms, exert the principal influence on the spectral

representations; having the same spectral configurations would provide similar spectral representations.

The second experiment was a standard methodology to compare different feature sets obtained by various algorithms in an image classification procedure (Shaw and Burke 2003; Martinez-Uso et al. 2007; Sotoca, Pla, and Sanchez 2007; Cariou, Chehdi, and Le Moan 2011; Jia et al. 2012). A better image classification generally means that the process of assigning a label to a pixel using its spectral information is more accurate, which leads to better recognition of objects and land covers in the image.

Therefore, to validate the performance of the proposed method, we presented a comparison with five other unsupervised DR algorithms by evaluating the image classification accuracy. Four of them are in the FS category, and the last one is in the FE category. The FS algorithms are WaLuMI, LP, OSP, and ICA-based model. Similar to GRRS, they preserve the physical relationship between the selected features and their wavelength. The FE algorithm is the Principal Component Analysis (PCA). We compared the GSSR algorithm with PCA since GSSR, similar to PCA, transforms the data into a new feature space but without using a rotation.

5.2.1. Experiment 1: HSI Spectra Representation:

In this experiment, we applied the PCFA and GSSR algorithm to an entire Moffett Field dataset. The dataset contains various spectral reflectance with different variations in the spectrum. We computed the mean RMSE of the spectral configurations provided by the PCFA and GSSR algorithms with respect to the number of bands in the reduced spectral configuration. The error is the average of the spectral approximation error of all pixels in the scene (Eq. (7)).

It should be noted that PCFA was developed to minimize the sum of squared error (SSE) as a performance metric. However, minimizing SSE concerning the approximated spectrum with k predefined number of breakpoints is equivalent to minimizing RMSE with the same conditions.

The running time of the algorithms is also computed. Figure 9 shows the results and Table 3 presents more details about six band sets. The error of spectral representation using the same configuration for an entire scene reveals an almost complete overlap between the two graphs showing the mean representation error obtained by the PCFA and GSSR algorithm. Table 3 indicates that the difference between the mean errors of the two methods is about 0.0001. On the other hand, the PCFA was about 37 times slower than GSSR, on average. The higher the number of bands, the slower the PCFA than GSSR, as clearly illustrated in Figure 9 (left).

Figure 9 ---- >>>

Table 3 ---- >>>

Jensen et. al. (2007) applied PCFA to a set of sampled spectra of a hyperspectral scene derived from the classes in a scene and identified a single spectral configuration for the spectra to reduce the dimensionality of the data (Jensen and Solberg 2007). The PCFA algorithm is applicable when the number of spectra is low; however, when the number increases, PCFA is slow. The complexity of the algorithm is $O(kmn^3)$ (Jensen and Solberg 2007), i.e. the order of the algorithm has a direct relation with the cube of the number of spectral samples multiplied by the number of pixels. Therefore, if the number of pixels increases, the time consumption of the algorithm drastically goes up. As a consequence, the PCFA algorithm becomes a prohibitive method in an unsupervised DR situation, where applied to all pixels in a hyperspectral scene.

On the other hand, GSSR provides a spectral configuration for the entire scene as accurate as of the best spectral configuration supplied by the PCFA algorithm but in a much faster way.

570 *5.2.2. Experiment 2: Image Classification*

571 The evaluation has been done by examining the number of features selected by the proposed and
572 reference methods vs. the classification overall accuracy using different classifiers to check the
573 relevance of the features selected. We applied the five mentioned methods to compare the results
574 with the band sets obtained by the GSSR method. These comparisons were performed using two
575 datasets: the Indian Pines and Salinas scene. We used two types of classifiers: maximum likelihood
576 classifier (MLC) and support vector machine (SVM), a parametric classifier, and a non-parametric
577 classifier, respectively. Figure 10 shows the results.

578 *Figure 10 ---- >>>*
579

580 As observed, the proposed method gives better overall accuracy than its competitors from
581 the FS category. It means that the criterion used in our approach in an unsupervised manner, i.e.,
582 extraction the most relevance spectral features by spectral reflectance representation provides
583 higher accuracy of classification than the frequently used criteria based on the similarity between
584 the spectral features in UFS techniques.

585 Comparison with PCA, an FE algorithm, shows that GSSR has better accuracy in
586 classification when using MLC while using the SVM classifier, PCA provided higher accuracy.
587 The reason is that the feature extraction techniques often have a higher potential in distinguishing
588 between different classes in a scene, leading to better accuracy in image classification. However,
589 the problem of the FE algorithms is that the critical information of the reflectance spectra can be
590 distorted. One of the main objectives of this study is to keep the critical information of the
591 reflectance spectra like the FS methods. This information, e.g. the absorption spectral features of a
592 specific target, is of interest to a wide range of HSI users. The GSSR algorithm, while retaining the

key spectral information, classified more accurately than PCA when using MLC. In addition, it obtained a better result than its competitors in the UFS category as well.

Using SVM, WaLuMI sometimes provided channel sets with comparable classification accuracy to GSSR. The channel selection algorithms based on ICA and the linear unmixing methods (LP and OSP) mostly gave less accurate results than GSSR.

6. Conclusions

This study showed the advantage of applying the representation of reflectance spectra of HSIs as the criterion to the unsupervised dimensionality reduction purpose. The typical rules applied in unsupervised feature selection techniques are based on finding the most dissimilar spectral channels, while the proposed criterion focuses on the extraction of the most spectral variations from the spectral reflectance. Since obtaining accurate spectral representations for all the pixels in a scene; i.e., in an unsupervised manner, is a time-consuming process, we applied the proposed criterion to a greedy algorithm, GSSR; to create spectral bands approximating the original reflectance. The final band configurations obtained by GSSR are sets of continuous spectral bands covering the whole spectrum, which preserves the physical meaning of the features like the FS techniques. Compared with the PCFA method providing the best spectral configuration minimizing the representation error, GSSR yields band configurations almost as accurate as PCFA, but in a much faster way. Applying the algorithms to more than 1000 diverse spectra to provide spectral configurations with 5 to 30 bands, GSSR was 50 to 1200 times faster than PCFA, while the mean difference in RMSE was 0.0002 on reflectance scale, i.e. in [0,1]. The difference was even less when both algorithms provided the same spectral configuration for an entire hyperspectral scene, i.e. 0.0001. Using the reduced band configurations in a target detection experiment showed that the bands provided by GSSR are more correctly identified than by PCFA if the spectral configuration

has one band more, while the former can be obtained 2000 times faster. The classification overall accuracy over two hyperspectral datasets using two classifiers revealed that the proposed criterion provides a band configuration giving a higher classification accuracy than its FS competitors.

Acknowledgement: The authors would like to appreciate NASA, Jet Propulsion Laboratory (JPL) for providing free AVIRIS hyperspectral datasets used in this study, and the anonymous reviewers for their constructive comments and corrections. Massimo Menenti also acknowledges the support of the MOST High Level Foreign Expert program (Grant No. G20190161018).

References

Agrawal, Rakesh, Christos Faloutsos, and Arun N. Swami. 1993. "Efficient Similarity Search In Sequence Databases." In *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, 69-84. Springer-Verlag.

Angelopoulou, E. 2000. "Objective colour from multispectral imaging." *Computer Vision - Eccv 2000, Pt I, Proceedings* 1842:359-74.

Angelopoulou, E., R. Molana, and K. Daniilidis. 2001. "Multispectral skin color modeling." *2001 Ieee Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, Proceedings*:635-42.

Ben-Dor, E., R. G. Taylor, J. Hill, J. A. M. Dematte, M. L. Whiting, S. Chabrillat, and S. Sommer. 2008. "Imaging spectrometry for soil applications." *Advances in Agronomy, Vol 97* 97:321-+. doi: Doi 10.1016/S0065-2113(07)00008-9.

Bendall, Gareth, and François Margot. 2006. "Greedy-type resistance of combinatorial problems." *Discrete Optimization* 3 (4):288-98.

Cariou, C., K. Chehdi, and S. Le Moan. 2011. "BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing." *Ieee Geoscience and Remote Sensing Letters* 8 (3):565-9. doi: 10.1109/Lgrs.2010.2091673.

Carmona, P. L., A. Mart, x00Ed, Us nez, x00F, J. M. Sotoca, F. Pla, P. Garc, and Sevilla a. 2011. Band selection in spectral imaging for classification and regression tasks using information theoretic measures. Paper presented at the Information Optics (WIO), 2011 10th Euro-American Workshop on, 19-24 June 2011.

Chakrabarti, Kaushik, Eamonn Keogh, Sharad Mehrotra, and Michael Pazzani. 2002. "Locally adaptive dimensionality reduction for indexing large time series databases." *ACM Trans. Database Syst.* 27 (2):188-228. doi: 10.1145/568518.568520.

Chein, I. Chang, and Wang Su. 2006. "Constrained band selection for hyperspectral imagery." *Ieee Transactions on Geoscience and Remote Sensing* 44 (6):1575-85. doi: 10.1109/tgrs.2006.864389.

Chiang, Shao-Shan, Chein-I Chang, and Irvin W Ginsberg. 2000. Unsupervised hyperspectral image analysis using independent component analysis. Paper presented at the IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120).

Cochrane, MA. 2000. "Using vegetation reflectance variability for species level classification of hyperspectral data." *International Journal of Remote Sensing* 21 (10):2075-87.

Comon, Pierre. 1994. "Independent component analysis, A new concept?" *Signal Processing* 36 (3):287-314. doi: https://doi.org/10.1016/0165-1684(94)90029-9.

Cormen, Thomas H. 2009. *Introduction to algorithms*. 3rd ed. Cambridge, Mass.: MIT Press.

Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*: Wiley-Interscience.

Curran, P. J., and J. L. Dungan. 1989. "Estimation of Signal-to-Noise - a New Procedure Applied to Aviris Data." *Ieee Transactions on Geoscience and Remote Sensing* 27 (5):620-8.

Dionisio, Andreia, Rui Menezes, and Diana A. Mendes. 2004. "Mutual information: a measure of dependency for nonlinear time series." *Physica A: Statistical Mechanics and its Applications* 344 (1):326-9. doi: https://doi.org/10.1016/j.physa.2004.06.144.

Du, Hongtao, Hairong Qi, Xiaoling Wang, Rajeev Ramanath, and Wesley E Snyder. 2003. Band selection using independent component analysis for hyperspectral image processing. Paper presented at the 32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings.

Du, Q., H. Ren, and C. I. Chang. 2003. "A comparative study for orthogonal subspace projection and constrained energy minimization." *Ieee Transactions on Geoscience and Remote Sensing* 41 (6):1525-9. doi: 10.1109/Tgrs.2003.813704.

Du, Q., and H. Yang. 2007. "Unsupervised band selection for hyperspectral image analysis." *Igarss: 2007 Ieee International Geoscience and Remote Sensing Symposium, Vols 1-12*:282-5. doi: 10.1109/Igarss.2007.4422785.

———. 2008. "Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis." *Ieee Geoscience and Remote Sensing Letters* 5 (4):564-8. doi: 10.1109/Lgrs.2008.2000619.

Durrant, R. J., and A. Kaban. 2009. "When is 'nearest neighbour' meaningful: A converse theorem and implications." *Journal of Complexity* 25 (4):385-97. doi: 10.1016/j.jco.2009.02.011.

Gorte, Ben. 1998. Probabilistic segmentation of remotely sensed images.

Gu, Y. F., and Y. Zhang. 2003. "Unsupervised subspace linear spectral mixture analysis for hyperspectral images." *2003 International Conference on Image Processing, Vol 1, Proceedings*:801-4.

Han, J. S., S. W. Lee, and Z. Bien. 2013. "Feature subset selection using separability index matrix." *Information Sciences* 223:102-18. doi: DOI 10.1016/j.ins.2012.09.042.

Hira, Zena M, and Duncan F Gillies. 2015. "A review of feature selection and feature extraction methods applied on microarray data." *Advances in bioinformatics* 2015.

Hosseini Aria, S Enayat , Massimo Menenti, and Ben GH Gorte. 2017. "Spectral region identification versus individual channel selection in supervised dimensionality reduction of hyperspectral image data." *Journal of Applied Remote Sensing* 11 (4):046010.

Huang, R., and M. Y. He. 2005. "Band selection based on feature weighting for classification of hyperspectral data." *Ieee Geoscience and Remote Sensing Letters* 2 (2):156-9. doi: 10.1109/Lgrs.2005.844658.

Hughes, G. 1968. "On the mean accuracy of statistical pattern recognizers." *Ieee Transactions on Information Theory* 14 (1):55-63. doi: 10.1109/tit.1968.1054102.

Huynh, C. P., and A. Robles-Kelly. 2008. "A NURBS-based spectral reflectance descriptor with applications in computer vision and pattern recognition." *2008 Ieee Conference on Computer Vision and Pattern Recognition, Vols 1-12*:3345-52.

Hyvärinen, Aapo, and Erkki Oja. 2000. "Independent component analysis: algorithms and applications." *Neural networks* 13 (4-5):411-30.

Jain, A. K., and W. G. Waller. 1978. "On the optimal number of features in the classification of multivariate Gaussian data." *Pattern recognition* 10 (5–6):365-74. doi: http://dx.doi.org/10.1016/0031-3203(78)90008-0.

Jain, Anil K., and Richard C. Dubes. 1988. *Algorithms for clustering data*: Prentice-Hall, Inc.

Jensen, A. C., and A. S. Solberg. 2007. "Fast hyperspectral feature reduction using piecewise constant function approximations." *Ieee Geoscience and Remote Sensing Letters* 4 (4):547-51. doi: 10.1109/Lgrs.2007.896331.

Jia, S., Z. Ji, Y. T. Qian, and L. L. Shen. 2012. "Unsupervised Band Selection for Hyperspectral Imagery Classification Without Manual Band Removal." *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 5 (2):531-43. doi: Doi 10.1109/Jstars.2012.2187434.

Jia, S., Z. Zhu, L. Shen, and Q. Li. 2014. "A Two-Stage Feature Selection Framework for Hyperspectral Image Classification Using Few Labeled Samples." *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (4):1023-35. doi: 10.1109/jstars.2013.2282161.

Jia, X. P., B. C. Kuo, and M. M. Crawford. 2013. "Feature Mining for Hyperspectral Image Classification." *Proceedings of the Ieee* 101 (3):676-97. doi: Doi 10.1109/Jproc.2012.2229082.

Jihao, Yin, Wang Yisong, and Zhao Zhanjie. 2010. Optimal Band Selection for Hyperspectral Image Classification Based on Inter-Class Separability. Paper presented at the Photonics and Optoelectronic (SOPO), 2010 Symposium on, 19-21 June 2010.

Jimenez-Rodriguez, L. O., E. Arzuaga-Cruz, and M. Velez-Reyes. 2007. "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data." *Ieee Transactions on Geoscience and Remote Sensing* 45 (2):469-83. doi: Doi 10.1109/Tgrs.2006.885412.

Jimenez, L. O., and D. A. Landgrebe. 1998. "Supervised classification in high-dimensional space: Geometrical, statistical, and asymptotical properties of multivariate data." *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews* 28 (1):39-54. doi: Doi 10.1109/5326.661089.

Kaewpijit, S., J. Le Moigne, and T. El-Ghazawi. 2002. "A wavelet-based PCA reduction for hyperspectral imagery." *Igarss 2002: Ieee International Geoscience and Remote Sensing Symposium and 24th Canadian Symposium on Remote Sensing, Vols I-Vi, Proceedings*:2581-3.

Kahveci, T., and A. Singh. 2001. Variable length queries for time series data. Paper presented at the Data Engineering, 2001. Proceedings. 17th International Conference on, 2001.

Keogh, Eamonn, Kaushik Chakrabarti, Michael Pazzani, and Sharad Mehrotra. 2001. "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases." *Knowledge and Information Systems* 3 (3):263-86. doi: 10.1007/pl00011669.

Konno, H., and T. Kuno. 1988. "Best Piecewise Constant Approximation of a Function of Single Variable." *Operations Research Letters* 7 (4):205-10. doi: Doi 10.1016/0167-6377(88)90030-2.

743 Kruse, F. A., A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and
744 A. F. H. Goetz. 1993. "The Spectral Image-Processing System (Sips) - Interactive
745 Visualization and Analysis of Imaging Spectrometer Data." *Remote Sensing of*
746 *Environment* 44 (2-3):145-63. doi: Doi 10.1016/0034-4257(93)90013-N.
747 Kumar, S., J. Ghosh, and M. M. Crawford. 2001. "Best-bases feature extraction algorithms for
748 classification of hyperspectral data." *Ieee Transactions on Geoscience and Remote Sensing*
749 39 (7):1368-79.
750 Le Moan, S., A. Mansouri, Y. Voisin, and J. Y. Hardeberg. 2011. "A Constrained Band Selection
751 Method Based on Information Measures for Spectral Image Color Visualization." *Ieee*
752 *Transactions on Geoscience and Remote Sensing* 49 (12):5104-15. doi:
753 10.1109/Tgrs.2011.2158319.
754 Lennon, M, G Mercier, MC Mouchot, and L Hubert-Moy. 2001. Independent component analysis
755 as a tool for the dimensionality reduction and the representation of hyperspectral images.
756 Paper presented at the IGARSS 2001. Scanning the Present and Resolving the Future.
757 Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat.
758 No. 01CH37217).
759 Li, Zhao-Liang, F. Becker, M. P. Stoll, Zhengming Wan, and Yulin Zhang. 1999. "Channel
760 selection for soil spectrum reconstruction in 8–13 μm region." *Journal of Geophysical*
761 *Research: Atmospheres* 104 (D18):22271-85. doi: 10.1029/1999jd900479.
762 Liu, Lan, Cheng-fan Li, Yong-mei Lei, Jing-yuan Yin, and Jun-juan Zhao. 2017. "Feature
763 extraction for hyperspectral remote sensing image using weighted PCA-ICA." *Arabian*
764 *Journal of Geosciences* 10 (14):307.
765 Manolakis, Dimitris, David Marden, and Gary A Shaw. 2003. "Hyperspectral image processing for
766 automatic target detection applications." *Lincoln Laboratory Journal* 14 (1):79-116.
767 Martinez-Uso, A., F. Pla, J. M. Sotoca, and P. Garcia-Sevilla. 2007. "Clustering-based
768 hyperspectral band selection using information measures." *Ieee Transactions on*
769 *Geoscience and Remote Sensing* 45 (12):4158-71. doi: Doi 10.1109/Tgrs.2007.904951.
770 Mitra, P., C. A. Murthy, and S. K. Pal. 2002. "Unsupervised feature selection using feature
771 similarity (vol 24, pg 301, 2002)." *IEEE Transactions on pattern analysis and machine*
772 *intelligence* 24 (6):721-. doi: Doi 10.1109/Tpami.2002.1008379.
773 Nievergelt, Jürg. 2000. Exhaustive search, combinatorial optimization and enumeration: Exploring
774 the potential of raw computing power. Paper presented at the International Conference on
775 Current Trends in Theory and Practice of Computer Science.
776 Plaza, A., P. Martinez, R. Perez, and J. Plaza. 2004. "A quantitative and comparative analysis of
777 endmember extraction algorithms from hyperspectral data." *Ieee Transactions on*
778 *Geoscience and Remote Sensing* 42 (3):650-63. doi: 10.1109/Tgrs.2003.820314.
779 Price, J. C. 1990. "On the Information-Content of Soil Reflectance Spectra." *Remote Sensing of*
780 *Environment* 33 (2):113-21. doi: Doi 10.1016/0034-4257(90)90037-M.
781 ———. 1994. "How Unique Are Spectral Signatures." *Remote Sensing of Environment* 49
782 (3):181-6. doi: Doi 10.1016/0034-4257(94)90013-2.
783 Price, John C. 1975. "Information content of Iris spectra." *Journal of Geophysical Research* 80
784 (15):1930-6. doi: 10.1029/JC080i015p01930.
785 Pudil, P., J. Novovicova, and J. Kittler. 1994. "Floating Search Methods in Feature-Selection."
786 *Pattern Recognition Letters* 15 (11):1119-25. doi: Doi 10.1016/0167-8655(94)90127-9.
787 Qian, Y., F. Yao, and S. Jia. 2009. "Band selection for hyperspectral imagery using affinity
788 propagation." *Iet Computer Vision* 3 (4):213-22. doi: 10.1049/iet-cvi.2009.0034.

Robila, Stefan A, and Pramod K Varshney. 2004. "Feature extraction from hyperspectral data using ICA." In *Advanced image processing techniques for remotely sensed hyperspectral data*, 199-216. Springer.

Shaw, Gary A., and Hsiao-hua K. Burke. 2003. "Spectral Imaging for Remote Sensing." *Lincoln Laboratory Journal* 14 (1):3-28.

Sims, D. A., and J. A. Gamon. 2002. "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages." *Remote Sensing of Environment* 81 (2-3):337-54. doi: Pii S0034-4257(02)00010-X Doi 10.1016/S0034-4257(02)00010-X.

Sotoca, J. M., F. Pla, and J. S. Sanchez. 2007. "Band selection in multispectral images by minimization of dependent information." *Ieee Transactions on Systems Man and Cybernetics Part C-Applications and Reviews* 37 (2):258-67. doi: Doi 10.1109/Tsmcc.2006.876055.

Thomson, C., L. Lue, and M. N. Bannerman. 2014. "Mapping continuous potentials to discrete forms." *Journal of Chemical Physics* 140 (3). doi: Artn 034105 10.1063/1.4861669.

Wang, C., M. Menenti, M. P. Stoll, E. Belluco, and M. Marani. 2007. "Mapping mixed vegetation communities in salt marshes using airborne spectral data." *Remote Sensing of Environment* 107 (4):559-70. doi: DOI 10.1016/j.rse.2006.10.007.

Ward, J. H. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 58 (301):236-&. doi: Doi 10.2307/2282967.

Winkler, G., and V. Liebscher. 2002. "Smoothers for discontinuous signals." *Journal of Nonparametric Statistics* 14 (1-2):203-22. doi: 10.1080/10485250290003407.

Yang, H., Q. A. Du, H. J. Su, and Y. H. Sheng. 2011. "An Efficient Method for Supervised Hyperspectral Band Selection." *Ieee Geoscience and Remote Sensing Letters* 8 (1):138-42. doi: Doi 10.1109/Lgrs.2010.2053516.

Zaatour, Rania, Sonia Bouzidi, and Ezzeddine Zagrouba. 2017. Impact of Feature Extraction and Feature Selection Techniques on Extended Attribute Profile-based Hyperspectral Image Classification.

Zehentbauer, F. M., and J. Kiefer. 2012. "Detection of unexpected species in soft modelling of vibrational spectra." *Chimica Oggi-Chemistry Today* 30 (3):54-6.

Zhang, Lefei, Liangpei Zhang, Bo Du, Jane You, and Dacheng Tao. 2019. "Hyperspectral image unsupervised classification by robust manifold matrix factorization." *Information Sciences* 485:154-69. doi: 10.1016/j.ins.2019.02.008.

Zhang, Lefei, Qian Zhang, Bo Du, Xin Huang, Yuan Yan Tang, and Dacheng Tao. 2018. "Simultaneous Spectral-Spatial Feature Selection and Extraction for Hyperspectral Images." *IEEE Transactions on Cybernetics* 48 (1):16-28. doi: 10.1109/tcyb.2016.2605044.