UNSUPERVISED HYPERSPECTRAL BAND SELECTION USING CLUSTERING AND SINGLE-LAYER NEURAL NETWORK

Mateus Habermann[1,3], Vincent Frémont[2], Elcio Hideiti Shiguemori[3]

1: Sorbonne Universités, Université de Technologie de Compiègne,
CNRS, Heudiasyc UMR 7253, CS 60319, 60203 Compiègne cedex, France.

2: Ecole Centrale de Nantes, LS2N, UMR CNRS 6004, Nantes, France.
vincent.fremont@ec-nantes.fr

3: Institute for Advanced Studies (IEAv), Brazilian Air Force, Brazil.
hab3rmann@gmail.com, elcio@ieav.cta.br

Résumé

Les images hyperspectrales fournissent des détails de la scène observée en exploitant les bandes spectrales continues. Cependant, le traitement de ces images est long à cause de leurs grandes dimensions. Donc, la sélection des bandes est une pratique commune qui est adoptée avant qu'un traitement soit fait. Ainsi, dans ce travail une nouvelle approche non-supervisée pour la sélection des bandes, basée sur le clustering et les réseaux de neurones, est proposée. Une comparaison avec six autres aproches de sélection de bandes montre la validité de l'algorithme proposé.

Mots clés : Sélection de bandes, Non-supervisé, Clustering.

Abstract

Hyperspectral images provide rich spectral details of the observed scene by exploiting contiguous bands. But, the processing of such images becomes heavy, due to the high dimensionality. Thus, band selection is a practice that has been adopted before any further processing takes place. Therefore, in this paper, a new unsupervised method for band selection based on clustering and neural network is proposed. A comparison with six other band selection frameworks shows the strength of the proposed method.

Keywords : Band selection, Unsupervised, Clustering

1. Introduction

Hyperspectral images (HSI) are composed of many continuous bands of the electromagnetic spectrum (Chang, 2003). Thus, a broad range of this spectrum can be covered, which provides lots of information about the scene under analysis. Consequently, those images are of great interest to Remote Sensing and Pattern Recognition, because they make easier some tasks such as image classification and target detection.

In a Pattern Recognition System (Duda et al., 2001), however, a large amount of bands can bring problems in terms of data storage and transmission. Moreover, the high dimensionality of HSI can cause the curse of dimensionality for the classifiers (Hughes, 1968). Thus, in order to solve those problems, dimensionality reduction is normally adopted.

To reduce the dataset dimensionality, *feature extraction* (FE) is a commonly used method (Silva et al., 2013). In sum, FE makes new features by combining the original ones. The resulting features belong to a feature space with lower dimension, keeping much of the original data variance. One largely used FE technique is Principal Components Analysis (PCA) (Sharma, 1996). PCA, as well as other FE techniques, changes the original dataset information, therefore it should not be used when the physical meaning of individual bands is to be kept (Khalid et al., 2014).

Like FE methods, *band selection* (BS) algorithms also decrease the dimension of the feature space, by selecting bands considered to be relevant, under a certain criterion. The positive aspect, in this case, lies in the fact that the original information is kept (Li and Liu, 2017).

BS algorithms are classified in two categories : *supervised* and *unsupervised*. Supervised band selection approaches normally outperform unsupervised ones (Yang et al., 2011). However, the former needs labeled training samples, which are expensive to be collected. Therefore, unsupervised methods are a wise alternative for band selection.

Unsupervised BS methods perform their job without knowing the class information of the data samples, which means that neither class separability measures nor classification accuracies can be used to assess the quality of the selected bands. Instead, the band selection is made by taking into account the data structure (Chandra and Sharma, 2015).

As k-Means algorithm is an unsupervised technique that clusters data samples according to the dataset structure in the feature space (Theodoridis and Koutroumbas, 2008), it will be used in this paper in the proposed band selection framework.

Thus, we propose an easily implementable unsupervised band selection framework. It uses two tools : *1)* k-Means clustering —following a bisecting k-Means fashion (Banerjee et al., 2015) ; and *2)* single-layer neural networks (SLNN) (Haykin, 2009). Initially, the whole data set is clustered by k-Means into two groups. After that, a SLNN is used to find a hyperplane that separates those two clusters, and the bands linked to the biggest and smallest coefficients of the hyperplane equation are selected. This procedure just described represents one iteration of the method, and two bands are selected at each iteration. If more bands are needed, another iteration is done taking as input data either of the clusters just generated, instead of the whole data set.

The contributions of this paper are the following :

— It is a novel band selection framework based on bisecting k-Means clustering and single-layer neural network ; and
— We compare the proposed method with other six unsupervised band selection algorithms.

This paper is organized as follows : In Section 2, a literature review is presented. Section 3 describes the proposed method. In Section 4, one can find the datasets, classifiers, the competitors and the results. Finally, Section 5 provides a conclusion to this paper.

2. Literature Review

In the literature, we find lots of hyperspectral band selection papers following many different perspectives and methodologies, such as *data manifold*, *data information analysis*, *graph theory*, *evolutionary computation* and *clustering*.

2.1. Data manifold

Due to the high HSI dimensionality, the different classes present in the image may lie in manifolds embedded in subspaces of the original feature space. Furthermore, it is also possible to explore the sparsity of the data set in order to find a more meaningful data representation. For example, in (Wang et al., 2016), the authors propose a new method in which they look for salient bands. The number σ of selected bands is user-defined. Then, the band selection algorithm has two steps. Firstly, β bands are selected by means clone selection algorithm, which seeks to minimize the Euclidean distance amongst elements of the same class, whereas maximizing the distance of elements from different classes. After that, if $\beta < \sigma$, those β bands already chosen will serve as seeds to a Manifold Ranking (MR) algorithm. MR sorts the remaining bands, and the most dissimilar band is added to the β group. This step is repeated until $\beta = \sigma$. In (Wang et al., 2017), the authors propose a BS framework based on sparsity. Initially, the most representative bands are obtained according to the correlation matrix, whereas the block-diagonal structure is measured to segment bands into subspaces. Then, a method for

band selection based on trace LASSO and spectral clustering is used. In (Sun et al., 2015), the authors propose a method that initially represents data instances as sparse coefficient vectors by solving a L2-norm optimization using the least squares regression (LSR) algorithm. Then, a correct segmentation of band vectors is made using the resulting LSR matrix with sparse and block-diagonal structure. After that, a similarity matrix is constructed by angular similarity measurement, and then the size of the band subset is calculated by the distribution compactness plot algorithm. In (Gan et al., 2017), the authors state that all HSI bands can be represented by a band subset. Thus, they propose a sparse representation of bands with row-sparsity constraint. Besides, a dissimilarity-weighted regularization term is integrated with the self-representation model, to avoid contiguous bands. The problem is solved by the alternating direction method of multipliers, and the representative bands can be chosen. In (Sun et al., 2017), a fast and robust self-representation framework to select a band subset is proposed. It is assumed the separability structure of the spectral bands, thus the problem may be seen as non-negative matrix factorization. After that, an optimizing convex problem is addressed and augmented Lagrangian multipliers are used to select the band subset. In (Zhu et al., 2017), the authors propose a BS framework that can capture the inter-band redundancy through low-rank modeling. Then, by using an affinity matrix and concepts of data quality, the most representative bands are selected. In (Wang et al., 2015), a BS method based on column subset selection is proposed. By means of column subset selection problem, it is possible to select some bands maximizing the volume of the selected subset of columns. The high dimensionality decreases the contrast amongst bands, thus Manhattan distance is used to get a higher quality in the BS process. In (Cao et al., 2016), the authors propose a framework that removes low-discriminating bands that normally need to be discarded manually. Based on the spatial structure of the data set, it is possible to determine which bands have low-discriminating power. Then, a new clustering algorithm is proposed in order to define the optimum number of bands to be selected.

2.2. Data information analysis

Another criterion that can be used in BS strategies is the HSI data information analysis. For example, in (Sui et al., 2015), the authors propose a framework that integrates both the overall accuracy and redundancy. Thus, an optimization problem using adaptive balance parameter is devised to handle the trade-off between the overall accuracy and redundancy. Furthermore, an unsupervised overall accuracy prediction method was adopted. In (Sun et al., 2014), the authors propose a framework that merges the concept of noiseadjusted principal components with maximum determinant of covariance matrix. A new index to measure the HSI quality is also proposed, taking into account signal-to-noise ratios (SNR) and correlation of bands. Based on the new index, the authors

devise an unsupervised band selection method, which considers the quality of the data set as selection criterion. It selects bands with both high SNR and low correlation. In (dos Santos et al., 2015), the authors propose a BS method based on the dissimilarity amongst neighboring bands. They use an intermediary representation named spectral rhythm, which can take advantage of a pixel sampling strategy, what ends up improving its efficiency without reducing the selected bands quality. In (Luo et al., 2017), the authors propose a method based on information-assisted density peak index. It takes into account the intraband information entropy into the local density and intercluster distance to ensure cluster centers with a high quality. Besides, the channel proximity and band distance are integrated to control the local density compactness. The bands with top-ranked scores may get clear global distinction, good local density and also high informative quality. In (Chang et al., 2017), the authors formulate the BS as a channel capacity problem. After constructing a band channel with the original bands. Then, some bands are selected by Blahut's algorithm, which iteratively finds a feature space that provides the best channel capacity. Thus, neither band prioritization nor interband decorrelation are required. Two iterative methods are devised to find the best band subset, which avoid an exhaustive search.

2.3. Graph theory

Using graph theory, in (Yuan et al., 2017) the authors propose a multigraph determinantal point process (MDPP). The aim is to capture the structure amongst bands and find the optimal band subset. For this, multiple graphs are designed to capture the intrinsic relationship amongst bands. Besides, the proposed MDPP is used to model the multiple dependencies in graphs, providing an efficient search strategy for the BS process.

2.4. Evolutionary computation

Evolutionary computation with optimization have been largely used by BS methods. For example, in (Xu et al., 2017), the authors propose an incorporated rank-based multiobjective band selection framework, to avoid conflicting objective functions, such as Jeffreys-Matusita (JF) and Bhattacharyya distances. During the processing, the spectral bands are transformed into binary vectors, whose elements are subjected to flipping with a certain probability. In (Gong et al., 2016), the authors propose a framework that handles two conflicting objective functions. One function is designed to represent the information contained in the selected bands, by means of entropy. The other function is set as the number of selected bands. Both objective functions are optimized simultaneously by a multiobjective evolutionary algorithm. In (Su et al., 2014), the authors propose a framework for band selection which employs two objective functions using JF. During the search process, the spectral bands are treated as firefly variables. In (Zhang et al., 2017a), a framework for band selection based on fuzzy clustering and swarm optimization in proposed. The authors devise a modified fuzzy clustering

method for band selection, whose drawbacks are alleviated by swarm optimization. In (Zhang et al., 2017b), the authors proposed a BS framework based on memetic algorithms (MA). Firstly, MA is used to select a subset of spectral bands. Also, a objective function is designed to select bands considering both bands information and redundancy deduction. The authors claim that this method is not only computationally faster than exhaustive search approaches, but also has comparable performances.

2.5. Clustering

Finally, clustering techniques can also be used in band selection methods. For instance, in (Datta et al., 2012) the authors propose a framework that removes redundancy amongst bands by means of clustering. Then, from each cluster one representative band is selected. After that, the bands are ranked according to their classification capabilities. In (Yuan et al., 2016), the authors propose a framework based on dual clustering that takes into account the contextual information. For this, a novel descriptor that reveals the image context is devised, in order to select the representatives of each cluster, taking into consideration the mutual effects of each cluster.

3. Proposed Framework

3.1. Definitions

Let $\mathbf{C}^{(0)}$ be a hyperspectral dataset, whose elements are vectors $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$ that contain spectral signatures, where d is the number of bands.

Let $\mathbf{S}$ be the set that contains the selected bands, and let $\mathbf{G}$ be set with bands highly correlated to the bands in $\mathbf{S}$. Let $\mathbf{A}$ be the set containing the original spectral bands a_k, with $k = 1, 2, ..., d$. Let γ be the number of bands to be selected.

Let $f : \mathbf{F} \longrightarrow t$ be a single-layer neural network, where $t = \{0, 1\}$, and $\mathbf{F}$ is the feature space given by $\mathbf{A} \setminus (\mathbf{S} \cup \mathbf{G})$. The input to f is $\mathbf{x}$ and its output is a scalar given by

$$\hat{t} = f(z) = \frac{1}{1 + e^{-z}}, \tag{1}$$

where $z = \mathbf{w}^T \mathbf{x} + b$, where $\mathbf{w}$ and b are, respectively, the weights and bias of the neural net.

According to Formula 1, $\hat{t} \in [0, 1]$, and we adopt the following criteria to give it a binary value :
— if $z < 0 \implies f < 0.5 \implies \hat{t} \leftarrow 0$,
— if $z \geq 0 \implies f \geq 0.5 \implies \hat{t} \leftarrow 1$.

Thus, the signal of z determines whether an input vector is to be assigned to class 0 or 1.

The input data is normalized into $[0, 1]$, consequently the biggest weights $w_l \in \mathbf{w}$, with $l = 1, ..., p$, in the hyperplane equation

$$z = x_i^1 w_1 + x_i^2 w_2 + ... + x_i^p w_p + b \tag{2}$$

cause a major impact on the signal of z, and, consequently, the estimate $\hat{t}_i$ for x_i. In Formula 2, p represents the cardinality of the feature space $\mathbf{F} = \mathbf{A} \setminus (\mathbf{S} \cup \mathbf{G})$.

As for the cost-function of f, we adopt the cross-entropy. The parameters learning is done by means of stochastic gradient descent and back-propagation algorithm (Theodoridis and Koutroumbas, 2008).

Finally, let $\mathbf{C}_g^{(l)}$ be a partition of $\mathbf{C}^{(0)}$, where l is the partition level, and $\mathbf{C}_1^{(l)} \cup \mathbf{C}_2^{(l)} \cup \cdots \cup \mathbf{C}_g^{(l)} = \mathbf{C}^{(0)}$, and $\mathbf{C}_p^{(l)} \cap \mathbf{C}_q^{(l)} = \emptyset, \forall p \neq q$. Depending on the problem, there may be several levels, *i.e.*, $l = 1, 2, 3, \ldots$, and for each level l the number of partitions g is calculated by

$$g = 2^l.$$

3.2. Description

3.2.1. General view

Our method starts with an empty set, that is, $\mathbf{S} = \emptyset$, to which the bands selected from $\mathbf{A}$ will be added. Initially, at the first iteration, $\mathbf{C}^{(0)}$ is clustered by k-Means into two groups, $\mathbf{C}_1^{(1)}$ and $\mathbf{C}_2^{(1)}$, using Euclidean distance. After that, a single-layer neural network is used to find a separating hyperplane between those two partitions. Then, two bands are selected and removed from the feature space $\mathbf{F}$. If more bands are needed, this procedure must be repeated using the just generated clusters.

Since the proposed framework is based on clustering and single-layer neural network, it will be called CSLN. In the sequel, deeper details will be given.

3.2.2. Iterations

CSLN is an iterative BS framework. At each iteration, a two-class classification problem between the clusters generated by k-Means $\mathbf{C}_p^{(l)}$ and $\mathbf{C}_q^{(l)}$ is to be solved by means of the function f. Two bands are selected at each iteration, thus it is necessary to repeat this process until the number of bands γ is reached. These criteria are adopted :

— If γ is even, $\gamma/2$ iterations are necessary ; or
— If γ is an odd number, we need $(\gamma + 1)/2$ iterations. Then, only the first γ selected bands are selected.

3.2.3. Selection of bands

When the training of the neural network ends, it is possible to assess the importance of all $a \in \mathbf{F}$ by analyzing $\mathbf{w}$. Every element $x^l \in \mathbf{x}$ is associated to w_l—see Figure 1—, for $l = 1, ..., p$, thus the magnitude of w_l is a indicator for the band a_l. In Formula 2, the largest and the smallest weights are the most important contributors to the signal of z. Consequently, the bands attached to those weights are also considered the most important, thus they are added to the set $\mathbf{S}$. Then, the feature space $\mathbf{F}$ is updated by $\mathbf{A} \setminus (\mathbf{S} \cup \mathbf{G})$.

Since the dimensionality of the feature space $\mathbf{F}$ changes after each iteration, the neural network must be trained from scratch at each iteration, because the sizes of $\mathbf{w}$ and b also change.
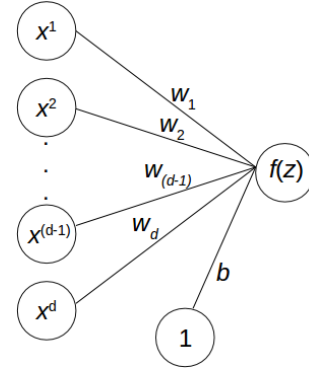

FIGURE 1: An example of the single-layer neural network used in this thesis. This architecture permits that each band $\mathbf{x}^i$ be linked to only one weight w_i.

3.2.4. Avoiding highly correlated bands

Since the neighboring HSI bands are highly correlated, we devise a method to avoid the selection of highly correlated bands. For each band $a_k \in \mathbf{F}$, we make a vector $\mathbf{v}_k$, whose elements are the bands indices in a descending order in relation to the correlation to the band a_k. Thus, $\mathbf{v}_k(1)$ is the index of the band $a_{\mathbf{v}_k(1)}$, which is the band with the highest correlation with a_k. The correlation ρ between two spectral bands a_α and a_β is calculated by

$$\rho_{(a_\alpha, a_\beta)} = \frac{cov(a_\alpha, a_\beta)}{\sigma_{a_\alpha} \sigma_{a_\beta}},$$

where $cov()$ is the covariance and σ stands for the standard-deviation.

We follow this procedure :
— At a given iteration r, a band a_k is selected, so $\mathbf{S} \leftarrow a_k$;
— $\mathbf{G} \leftarrow a_{\mathbf{v}_k(1)}$; and
— After iteration r, $\mathbf{F}$ is updated by $\mathbf{A} \setminus (\mathbf{S} \cup \mathbf{G})$.

It is worth-noting that only $a_k \in \mathbf{S}$ are the selected bands. The bands $a_{\mathbf{v}_k(1)} \in \mathbf{G}$ are discarded.

Algorithm 1 lists the steps adopted by the proposed CSLN method.

Algorithm 1 Proposed band selection framework

1: Input : $\mathbf{C}^{(0)}$, $\mathbf{A}$, γ, $\mathbf{S} = \emptyset$ and $\mathbf{G} = \emptyset$
2: **for** $r = 1 :$ maxIterations **do**
3:     Train a single-layer neural network f to find a separating hyperplane between $\mathbf{C}_p^{(l)}$ and $\mathbf{C}_q^{(l)}$, previously clustered by k-Means
4:     Identify the bands $a_k \in \mathbf{F}$ associated to the largest and smallest $w \in \mathbf{w}$
5:     $\mathbf{S} \leftarrow a_k$, and $\mathbf{G} \leftarrow a_{\mathbf{v}_k(1)}$
6:     Update $\mathbf{F}$ by $\mathbf{A} \setminus (\mathbf{S} \cup \mathbf{G})$
7: **end for**
8: Return : $\mathbf{S}$

Figure 2 illustrates the proposed method. Initially, the whole dataset is clustered into two groups by k-Means. Then, a separating hyperplane between them is calculated by means of a single-layer neural network. This pro-
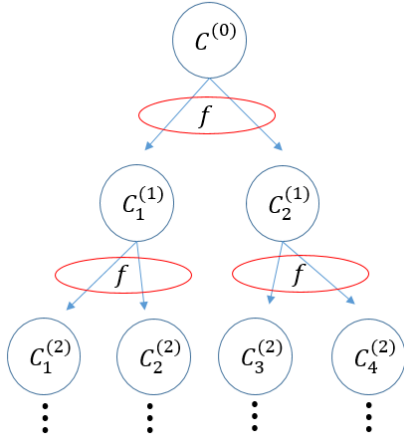
FIGURE 2: A general view of the proposed BS framework. At each binary clustering, a single-layer neural net f is used to select the bands.

cess is repeated until the previously determined quantity of bands is selected.

4. Results

The results of the proposed framework are shown in this Section. We compare our results with other six band selection approaches by taking into account the accuracy of supervised classifiers.

4.1. Datasets and classifiers

In this paper, we use two hyperspectral images [1], which are

— **Indian Pines :** It consists of 224 spectral reflectance bands in the 0.4-2.5 μm wavelength range. There are 16 classes.
— **Pavia University :** The Pavia University image has 103 spectral bands, and 9 classes.

For the the Indian Pines dataset, we use Classification and Regression Trees (CART) and k-Nearest Neighbors.

For the Pavia University image, we use KNN, CART and Support Vector Machine (SVM) (Theodoridis and Koutroumbas, 2008).

All the classifiers are run in MATLAB. For KNN, we use the `fitcknn` command, $k = 3$ in all experiments, for a fair comparison. For CART, the command `fitctree` is used, and to run SVM for multiple classes, we use `fitcecoc`, with polynomial kernel. All the commands above belong to the Statistics and Machine Learning toolbox.

The proposed BS method is also implemented in MATLAB. We use the `trainSoftmaxLayer`, from Neural Network Toolbox, for single-layer neural network, with 2000 training epochs. We emphasize that normally the training

phase stopped before the 2000th training epoch, so bigger training epochs quantities were not tested. As for the k-Means algorithm, we use `kmeans`, from the Statistics and Machine Learning toolbox.

The classifiers input data are the images with the selected bands. Moreover, the number of bands selected by our framework are for comparison purposes. Thus, it does not mean they are the best quantity for any given task.

For the classification step, each dataset is divided in two subsets. The first one is used for the classifier training, using 70% of the data. The second subset has the remaining 30%, and is used during the test phase of the classifier. The results shown in Section 4.4 are calculated based on this second data subset.

4.2. Competitors

The performance of the proposed CSLN framework is compared with six other band selection methods.

Four of them are used with the Indian Pines image :
— This method uses both ranking and clustering for band selection (Datta et al., 2015), and we will call it CR ;
— This competitor relies on information divergence, and it will be called ID (Chang and Wang, 2006) ;
— This framework is also clustering-based (Martinez-Uso et al., 2007), and it will be referred to as WaLuDi ; and
— This method resorts to band elimination with partitioned image correlation (Datta et al., 2014), and it will be referred to as EM.

For the Pavia University image, there are two competitors :
— It is a framework that handles two conflicting objective functions. One function is designed to represent the information contained in the selected bands, by means of entropy (Gong et al., 2016). It will be called MOBS ;
— This framework constructs a band channel with the original bands. Then, some bands are selected by means of Blahut's algorithm, which iteratively finds a feature space that provides the best channel capacity (Chang et al., 2017). This competitor will be referred to as CC.

As already stated in Section 3.2.1, the proposed BS algorithm will be called CSLN.

4.3. Selected bands

The bands selected by the CSLN method are shown in Table 4.3. For the Indian Pines dataset, we have only the first 18 best-ranked bands of the competitors, consequently, the analyses of results are restricted to this quantity of bands. For the Pavia University image, 21 bands are selected for the same reason.

The bands in Table 4.3 are displayed according to the order they were selected. For instance, at the first iteration, the bands 2 and 42 were selected for the Indian Pines image.
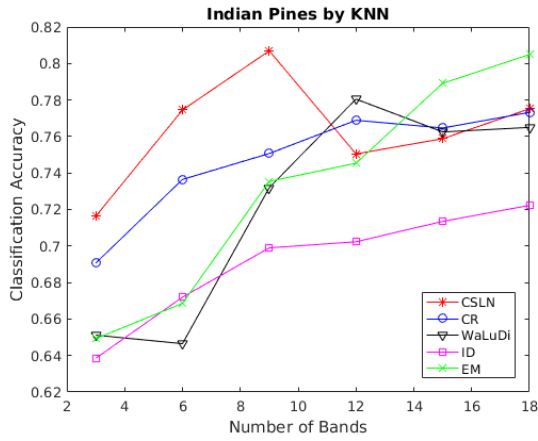
FIGURE 3: The Indian Pines dataset under the KNN classification.

FIGURE 4: The Indian Pines image classified by CART.

The comparison of results is made taking into account different quantities γ of selected bands, that is, $\gamma_s = s \times 3$, with $s = 1, 2, 3, 4, 5, 6$, for Indian Pines. For example, for $\gamma_2 = 6$, the first six bands of Table 4.3 are used. For the Pavia University dataset, $s = 1, 2, 3, 4, 5, 6, 7$.

Selected bands for Indian Pines	2, 42, 6, 39, 22, 58, 25, 62, 71, 101, 94, 151, 111, 203, 156, 183, 171, 215.
Selected bands for Pavia University	69, 1, 68, 3, 92, 6, 77,18, 101, 24, 99, 51, 14, 74, 9, 75, 29, 97, 95, 8, 103.

TABLE 1: The selected bands according to the order of selection by the proposed method.

4.4. Results comparison

All the classification results shown in this paper are the mean values over ten runs.

The results for the Indian Pines image are shown in Table 4.4. Using KNN, the proposed method `CSLN` has the best results using 3, 6 and 9 bands. This fact is illustrated in Figure 3.

The results achieved by the CART classifier using Indian Pines are also shown in Table 4.4. The proposed method has the best results with 3 and 9 bands. Figure 4 gives a visual idea of the results.

For the Pavia University dataset, the results are given in Table 4.4. Using the KNN classifier, the `CSLN` method achieves the best results using 3, 6, 9, 12 and 15 bands. This is shown in Figure 5.

For the CART classifier, the proposed method has the best results with 3, 6, 9 and 12 bands. In Figure 6 it is possible to have a visual perspective.

With the SVM classifier, our framework has the best results with 3, 6, 9, 12 and 15 bands, which is shown in Figure 7.

4.4.1. Visual inspection of the selected bands

The spectral signatures of the different classes give us an idea of the features—or bands—that provide a good
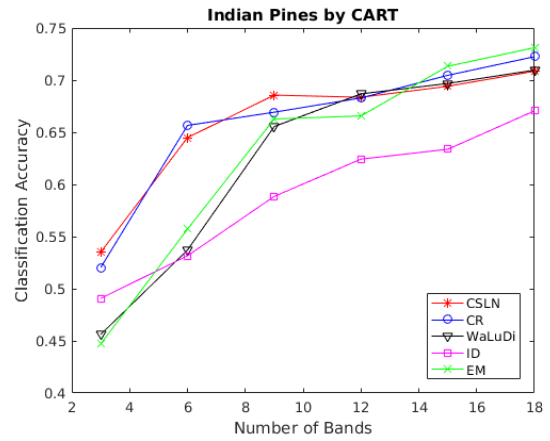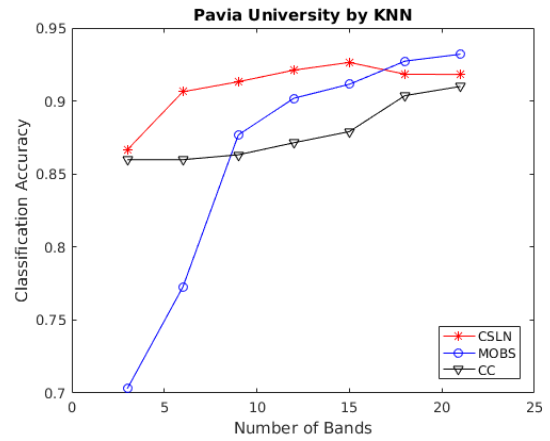

FIGURE 5: The Pavia University dataset classified by KNN.
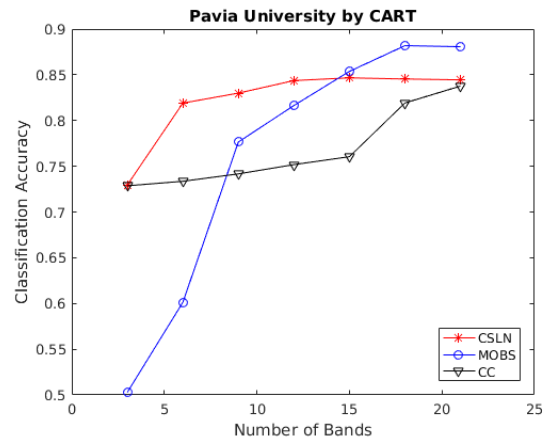
FIGURE 6: The Pavia University image classified by CART.

KNN results

	3 bands	6 bands	9 bands	12 bands	15 bands	18 bands
Method	acc.	acc.	acc.	acc.	acc.	acc.
CSLN	**71.63%**	**77.45%**	**80.69%**	75.05%	75.88%	77.52%
WaLuDi	65.12%	64.65%	73.19%	**78.05%**	76.25%	76.50%
CR	69.06%	73.65%	75.07%	76.89%	76.47%	77.32%
EM	64.92%	66.86%	73.54%	74.54%	**78.92%**	**80.50%**
ID	63.85%	67.20%	69.90%	70.23%	71.35%	72.23%

CART results

	3 bands	6 bands	9 bands	12 bands	15 bands	18 bands
Method	acc.	acc.	acc.	acc.	acc.	acc.
CSLN	**53.51%**	64.49%	**68.56%**	68.36%	69.41%	70.85%
WaLuDi	45.62%	53.71%	65.55%	**68.68%**	69.68%	70.96%
CR	52.03%	**65.66%**	66.93%	68.29%	70.46%	72.25%
EM	44.72%	55.72%	66.28%	66.57%	**71.33%**	**73.12%**
ID	49.07%	53.16%	58.85%	62.43%	63.37%	67.06%

TABLE 2: Classification results for Indian Pines.

KNN results

	3 bands	6 bands	9 bands	12 bands	15 bands	18 bands	21 bands
Method	acc.	acc.	acc.	acc.	acc.	acc.	acc.
CSLN	**86.66%**	**90.63%**	**91.32%**	**92.11%**	**92.64%**	91.83%	91.82%
MOBS	70.30%	77.26%	87.68%	90.19%	91.15%	**92.73%**	**93.21%**
CC	85.98%	85.98%	86.32%	87.14%	87.91%	90.37%	91.00%

CART results

	3 bands	6 bands	9 bands	12 bands	15 bands	18 bands	21 bands
Method	acc.	acc.	acc.	acc.	acc.	acc.	acc.
CSLN	**72.96%**	**81.91%**	**83.00%**	**84.37%**	84.66%	84.54%	84.45%
MOBS	50.28%	60.13%	77.65%	81.64%	**85.37%**	**88.19%**	**88.07%**
CC	72.87%	73.36%	74.17%	75.18%	76.05%	81.92%	83.77%

SVM results

	3 bands	6 bands	9 bands	12 bands	15 bands	18 bands	21 bands
Method	acc.	acc.	acc.	acc.	acc.	acc.	acc.
CSLN	**78.81%**	**87.17%**	**90.33%**	**93.03%**	**94.36%**	83.76%	92.58%
MOBS	61.30%	71.50%	85.96%	91.65%	91.50%	**98.23%**	**98.88%**
CC	77.99%	79.47%	80.71%	82.33%	83.75%	84.63%	91.49%

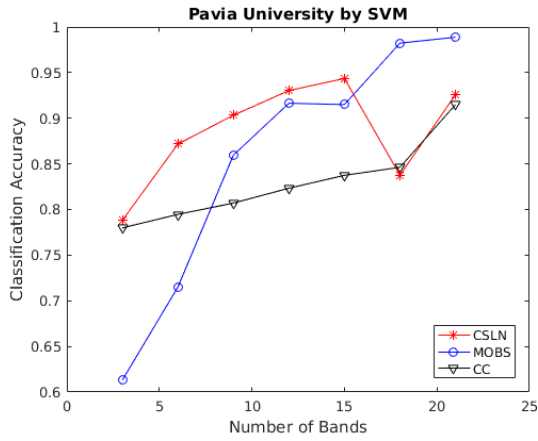**TABLE 3:** Classification results for Pavia University.

FIGURE 7: The Pavia University dataset under SVM classification.

FIGURE 8: Mean spectral signature values of the Indian Pines image classes. The vertical lines indicate the location of the first 6 bands selected by the proposed `CSLN` method.
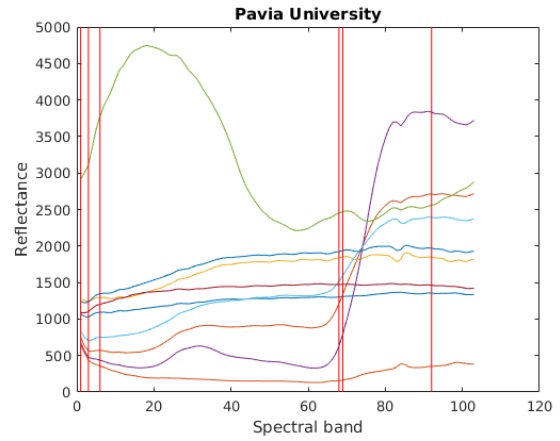

FIGURE 9: Mean spectral signature values of the Pavia University image classes. The vertical lines indicate the location of the first 6 bands selected by the proposed `CSLN` method.
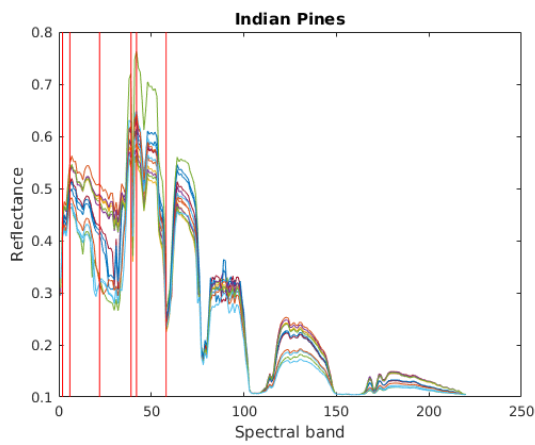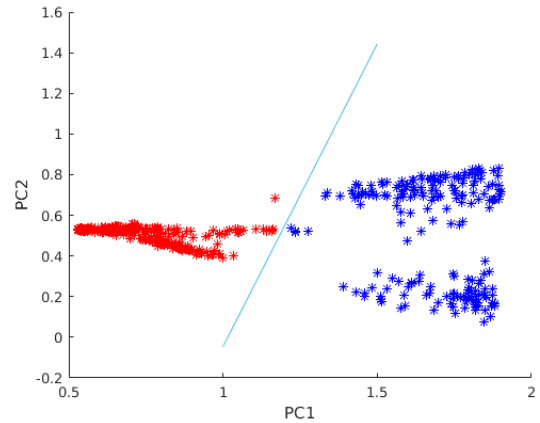

FIGURE 10: Two Indian Pines dataset clusters. The straight line is calculated by a single-layer neural network.

separation amongst classes. The more the signatures are far from one another, the better it is for the classifier.

Figures 8 and 9 show the mean spectral signatures of the classes present in Indian Pines and Pavia University datasets, respectively. In order to avoid excessive visual information, the location of only the first 6 selected bands is displayed, in vertical lines.

We notice that in both cases the selected bands fall in regions where the spectral signatures are far from one another. It denotes that our BS framework proposed in this chapter is capable of selecting appropriate spectral bands.

4.4.2. Considerations about the single-layer neural net choice

As already stated in Section 3.2.3, our rationale for the band selection is based on Formula 2, which, in turn, is the separating hyperplane calculated by the single-layer neural network used in this chapter. If the neural net does not converge to a good local minimum, or if its architecture is not appropriate for the problem at hand, no conclusion based on Formula 2 would be reliable.

According to the framework proposed in this chapter, the binary classification problem addressed by the single-layer neural network comes from the clustering performed by the k-Means algorithm. Thus, the two groups are linearly separable, as shown in Figures 10 and 11, where the straight lines that separate the two clusters are calculated using a single-layer neural network. The dimension of the datasets was reduced by the Principal Components Analysis, whose the first two principal components—PC1 and PC2, respectively—are kept, for a 2D illustration.

It is worth-noting that the two groups are linearly separable not only in a sparse feature space—Figure 10—, but also in a more dense situation, such as in Figure 11. Consequently, it is reasonable to use a single-layer neural network in such situations.

4.5. Remarks about the results

4.5.1. KNN versus CART

For the Indian Pines image, KNN results are, in general, superior than that of CART : 73.26% and 63.13%, respectively. This may be attributed to the fact that CART
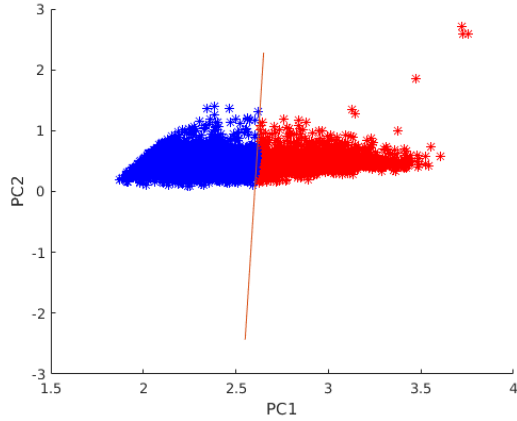
FIGURE 11: Two clusters from the Pavia University image. The straight line that separates the groups is calculated by a single-layer neural network.
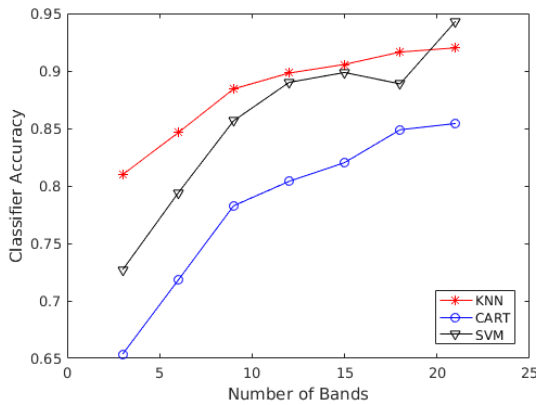

FIGURE 12: Mean results of the three classifiers used for the Pavia University image.

splits the feature space into regions that correspond to the classes. Therefore, if $\mathbf{x}_i$ is found in a region corresponding to a class α, for example, it will be classified as α, even if it belongs to class β. Whereas, in this same situation, KNN would analyze the k nearest neighbors of $\mathbf{x}_i$ before assigning it a label. Consequently, KNN outperforms CART when the class boundaries are highly non-linear.

4.5.2. KNN, CART and SVM

Concerning the Pavia University image, the mean results for the KNN, CART and SVM are, respectively, 88.30%, 78.31% and 85.69%. This shows a slight superiority of KNN in relation to SVM, what is somehow unexpected in high dimensional feature spaces, as shown in Figure 12. It is worth mentioning that our objective in this thesis is the classification comparison amongst different BS methods, and not the best attainable classification result. For this, further studies on the classifiers hyperparameters would be necessary.

4.5.3. Band selection methods

As for the band selection methods, using the Indian Pines image, the proposed BS framework achieves the best results in 5 out of 12 experiments, whereas the competitor have $4/12$, $2/12$, $1/12$ and $0/12$. For the Pavia University image, our method gets $14/21$, and the competitors $7/21$ and $0/21$. The CSLN framework not only gets superior results than its competitors, but it is also easily implementable. Thus, one can conclude that it is a good method for band selection.

5. Conclusion

The spectral information provided by HSIs allow for a rich characterization of objects in the scene. However, the large amount of data can also bring problems in terms of processing and storage. Furthermore, features spaces with high dimensions can cause the curse of dimensionality. Thus, we proposed a band selection method to decrease the dataset dimensionality.

The proposed unsupervised BS framework is based on k-Means clustering and single-layer neural network. It begins by splitting the whole dataset into two clusters. After that, a single-layer neural network is used to calculate a separating hyperplane between the two groups. The bands associated to the biggest and smallest hyperplane coefficients are selected. After that, one needs to repeat this procedure using the generated clusters until the desired number of bands is achieved.

According to the the results, we saw that the proposed CSLN method outperformed its competitors in both images analyzed.

As for the future works, we will investigate other clustering algorithms and binary classifiers to check their performance in the proposed framework.

Acknowledgements

Références

Banerjee, S., Choudhary, A., Pal, S., 2015. Empirical evaluation of k-means, bisecting k-means, fuzzy c-means and genetic k-means clustering algorithms. In : 2015 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE). pp. 168–172.

Cao, X., Wu, B., Tao, D., Jiao, L., 2016. Automatic band selection using spatial-structure information and classifier-based clustering. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 9 (9), 4352–4360.

Chandra, B., Sharma, R. K., 2015. Exploring autoencoders for unsupervised feature selection. In : 2015 International Joint Conference on Neural Networks (IJCNN). pp. 1–6.

Chang, C. I., 2003. Hyperspectral Imaging : Techniques for Spectral Detection and Classification. Vol. 1. Springer.

Chang, C. I., Lee, L. C., Xue, B., Song, M., Chen, J., 2017. Channel capacity approach to hyperspectral band subset selection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (10), 4630–4644.

Chang, C.-I., Wang, S., 2006. Constrained band selection for hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing 44 (6), 1575–1585.

Datta, A., Ghosh, S., Ghosh, A., 2012. Clustering based band selection for hyperspectral images. In : 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS). pp. 101–104.

Datta, A., Ghosh, S., Ghosh, A., 2014. Band elimination of hyperspectral imagery using partitioned band image correlation and capacitory discrimination. International Journal of Remote Sensing 35 (2), 554–577.

Datta, A., Ghosh, S., Ghosh, A., 2015. Combination of clustering and ranking techniques for unsupervised band selection of hyperspectral images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8 (6), 2814–2823.

dos Santos, L. C. B., Guimaraes, S. J. F., dos Santos, J. A., 2015. Efficient unsupervised band selection through spectral rhythms. IEEE Journal of Selected Topics in Signal Processing 9 (6), 1016–1025.

Duda, R. O., Hart, P. E., Stork, D. G., 2001. Pattern Classification. Wiley.

Gan, L., Xia, J., Du, P., Xu, Z., 2017. Dissimilarity-weighted sparse representation for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters 14 (11), 1968–1972.

Gong, M., Zhang, M., Yuan, Y., 2016. Unsupervised band selection based on evolutionary multiobjective optimization for hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing 54 (1), 544–557.

Haykin, S. S., 2009. Neural networks and learning machines, 3rd Edition. Pearson Education, Upper Saddle River, NJ.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Transactions on Information Theory 14 (1), 55–63.

Khalid, S., Khalil, T., Nasreen, S., 2014. A survey of feature selection and feature extraction techniques in machine learning. In : 2014 Science and Information Conference. pp. 372–378.

Li, J., Liu, H., 2017. Challenges of feature selection for big data analytics. IEEE Intelligent Systems 32 (2), 9–15.

Luo, X., Xue, R., Yin, J., 2017. Information-assisted density peak index for hyperspectral band selection. IEEE Geoscience and Remote Sensing Letters 14 (10), 1870–1874.

Martinez-Uso, A., Pla, F., Sotoca, J. M., Garcia-Sevilla, P., 2007. Clustering-based hyperspectral band selection using information measures. IEEE Transactions on Geoscience and Remote Sensing 45 (12), 4158–4171.

Sharma, S., 1996. Applied Multivariate Techniques. John Wiley & Sons, Inc., New York, NY, USA.

Silva, W. D., Habermann, M., Shiguemori, E. H., d. L. Andrade, L., d. Castro, R. M., 2013. Multispectral image classification using multilayer perceptron and principal components analysis. In : 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence. pp. 557–562.

Su, H., Li, Q., Du, P., 2014. Hyperspectral band selection using firefly algorithm. In : 2014 6th Workshop on Hyperspectral Image and Signal Processing : Evolution in Remote Sensing (WHISPERS). pp. 1–4.

Sui, C., Tian, Y., Xu, Y., Xie, Y., 2015. Unsupervised band selection by integrating the overall accuracy and redundancy. IEEE Geoscience and Remote Sensing Letters 12 (1), 185–189.

Sun, K., Geng, X., Ji, L., Lu, Y., 2014. A new band selection method for hyperspectral image based on data quality. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 7 (6), 2697–2703.

Sun, W., Tian, L., Xu, Y., Zhang, D., Du, Q., 2017. Fast and robust self-representation method for hyperspectral band selection. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 (11), 5087–5098.

Sun, W., Zhang, L., Du, B., Li, W., Lai, Y. M., 2015. Band selection using improved sparse subspace clustering for hyperspectral imagery classification. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8 (6), 2784–2797.

Theodoridis, S., Koutroumbas, K., 2008. Pattern Recognition, Fourth Edition, 4th Edition. Academic Press.

Wang, C., Gong, M., Zhang, M., Chan, Y., 2015. Unsupervised hyperspectral image band selection via column subset selection. IEEE Geoscience and Remote Sensing Letters 12 (7), 1411–1415.

Wang, J., Zhang, K., Wang, P., Madani, K., Sabourin, C., 2017. Unsupervised band selection using block-diagonal sparsity for hyperspectral image classification. IEEE Geoscience and Remote Sensing Letters 14 (11), 2062–2066.

Wang, Q., Lin, J., Yuan, Y., 2016. Salient band selection for hyperspectral image classification via manifold ranking. IEEE Transactions on Neural Networks and Learning Systems 27 (6), 1279–1289.

Xu, X., Shi, Z., Pan, B., 2017. A new unsupervised hyperspectral band selection method based on multiobjective optimization. IEEE Geoscience and Remote Sensing Letters 14 (11), 2112–2116.

Yang, H., Du, Q., Su, H., Sheng, Y., 2011. An efficient method for supervised hyperspectral band selection. IEEE Geoscience and Remote Sensing Letters 8 (1), 138–142.

Yuan, Y., Lin, J., Wang, Q., 2016. Dual-clustering-based hyperspectral band selection by contextual analysis. IEEE Transactions on Geoscience and Remote Sensing 54 (3), 1431–1445.

Yuan, Y., Zheng, X., Lu, X., 2017. Discovering diverse subset for unsupervised hyperspectral band selection. IEEE Transactions on Image Processing 26 (1), 51–64.

Zhang, M., Ma, J., Gong, M., 2017a. Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization. IEEE Geoscience and Remote Sensing Letters 14 (5), 773–777.

Zhang, M., Ma, J., Gong, M., Li, H., Liu, J., 2017b. Memetic algorithm based feature selection for hyperspectral images classification. In : 2017 IEEE Congress on Evolutionary Computation (CEC). pp. 495–502.

Zhu, G., Huang, Y., Li, S., Tang, J., Liang, D., 2017. Hyperspectral band selection via rank minimization. IEEE Geoscience and Remote Sensing Letters 14 (12), 2320–2324.