Efficient Unsupervised Band Selection
Through Spectral Rhythms

Lilian Chaves B. dos Santos, Silvio Jamil F. Guimarães, and Jefersson A. dos Santos

Abstract

The main goal of remote sensing image classification is to associate land cover classes to each pixel in the monitored area. In this sense, hyperspectral images play a key role by providing detailed spectral information per pixel. On the other hand, although the huge amount of spectral bands enables the creation of more accurate thematic maps, they can compromise the quality of results due to data redundancy, high-dimensionality problems and noisy bands. Many dimensionality reduction techniques have been proposed in order to better use the available information. An effective strategy is to perform a band selection, which aims at selecting the best bands for classification. This process decreases the dimensionality without degrading information, i.e., it keeps the physical properties acquired by the sensors. As a drawback, the dimensionality reduction process can take a lot of time to be performed. In this paper, we propose a new unsupervised band selection method based on the dissimilarity among neighboring bands by exploiting an intermediary representation called spectral rhythm. Our approach can take advantage of a pixel sampling strategy to improve its efficiency without significant reduction on the quality of selected bands. Experimental results reveal that our method can efficiently select suitable bands to represent the whole data by producing accuracy results as good as the baselines in the classification problem.

Index Terms

hyperspectral imagery; band selection; spectral rhythms; visual rhythms; graph matching; remote sensing; random pixel sampling.

This work was supported by CNPq (grant 449638/2014-6), FAPEMIG (grant APQ-00768-14), and CAPES (grant 064965/2014-01).

Lilian Chaves B. dos Santos and Jefersson A. dos Santos are with the Department of Computer Science of Universidade Federal de Minas Gerais (UFMG). Silvio Jamil F. Guimarães is with Instituto de Ciências Exatas e Informática of Pontifícia Universidade Católica de Minas Gerais (PUC Minas).
I. INTRODUCTION

Hyperspectral images encode information from several frequency ranges (bands) along the electromagnetic spectrum [1]. They provide a vast amount of information which is useful for identifying materials in earth surface but requires substantial computational efforts. Moreover, it can affect many supervised classification methods due to the “curse of dimensionality” [2].

Many approaches have been developed and/or applied to dimensionality reduction and make better use of the amount of information provided by hyperspectral images [3]–[20]. These methods can be divided into two main groups. The first one comprises the feature transformation methods, which aim at finding a transformation of the available data that reduces the dimensionality preserving the most salient information. A well studied branch of methods of this group is based on Fisher Discriminant Analysis [3], [21], [22] whose purpose is to made a transformation that optimizes the Fisher score. This kind of method is supervised, which means that they require some prior knowledge about the hyperspectral image content that may not be available. Concerning unsupervised methods on feature transformation group, a widely used approach is the principal component analysis (PCA) [3]–[7]. PCA is a dimension reduction technique based on an orthogonal transformation in which the components must have the highest possible variance. According to [23], hyperspectral data exhibit large within-class variance, which dominates the intraclass variance, thereby PCA may not be an appropriate choice. Nevertheless, by proceeding a transformation, the original bands are converted into a new feature space resulting in a loss of physical meaning or interpretation.

The second group of hyperspectral image reduction methods are the band selection-based ones. They typically use a metric to select the most informative and distinctive bands in order to provide a subset that could be the most appropriate. This approach is based on the fact that some bands may contain noisy or redundant information, being bad candidates to be selected. It is very useful in many applications because it can not only reduce dimensionality but also preserves relevant original information of the spectral bands. Once again, if prior knowledge is available, supervised methods [8]–[11] can be used to achieve good results.

Regarding the unsupervised band selection domain, some approaches were proposed. Cluster-based methods use clustering techniques, in which the first step is to compute a distance measure for each pair of bands. With these metrics, the bands are grouped in disjoint clusters such that bands in a given cluster tend to be similar to each other according to these metrics, and bands in different clusters tend to be dissimilar. After the grouping, a representative band from each cluster is chosen [12]–[15]. Search-
based methods aim at finding a good set of band by evaluating sub-sets features. Using exhaustive search strategies to find the best sub-set is normally unmanageable for this kind of data, however, several sub-optimal search strategies like sequential backward selection [24] or evolutionary techniques are used in this kind of problem [25], [26]. Rank-based methods use metrics, such as entropy, mutual information and correlation criterion, to sort the bands by their importance in the selection. Some examples of such metrics can be found in [16]–[20]. Although there are many dimensionality reduction techniques in the literature, both the effectiveness and the efficiency are still challenging issues for most of them [3], [12], [18].

In this paper, we present a novel approach for unsupervised hyperspectral band selection. It can automatically and efficiently select bands that provide good classification results preserving physical meaning of the data. Our method is based on a simplification of the hyperspectral image content, the spectral rhythms, and a dissimilarity measure among the bands and their neighbours. We have used dissimilarity measures based on a bipartite graph matching introduced in [27], which have been successfully and efficiently used for identifying abrupt and gradual video transition. However, our focus is to use the dissimilarity measure to find the most dissimilar and informative bands of a neighbourhood from the spectral rhythm. The use of spectral rhythm and bipartite graph-based dissimilarity measure for band selection were first introduced in our previous work [28]. The main differences of the proposed band selection approach in comparison with ones found in the literature are highlighted as follows:

- A new content simplification for hyperspectral images, called spectral rhythms. This computationally simple representation opens a new area of investigation related to the use of dissimilarity measures that are successfully applied to video analysis;
- A strategy for selecting bands from a dissimilarity curve. It produces competitive accuracy rates, outperforming the results obtained by the baselines;
- A random pixel sampling procedure. The proposed approach drastically reduces the overall computational cost without reducing the classification accuracy. Therefore, it allows our method to be more efficient, in terms of time, than three well-known methods from the literature.

The paper is organized as follows. A technical background with some important concepts used in our band selection are presented in Section II. Section III introduces the proposed band selection method. Experimental protocol, datasets and baselines used are presented in Section IV. Experiments results and discussion are presented in Section V. Final discussion and our perspective for further work are presented in Section VI.
II. Technical Background

In this section, we describe some concepts necessary for the understanding of this work. We firstly present the notion of visual rhythm, which consists in an abstraction of a video. This concept is adapted for representing spectral data, in the so-called Spectral Rhythm (SR) representation. Furthermore, we present the dissimilarity measure which we have used for analyzing gradual transitions among spectral bands.

A. Visual rhythms

Visual rhythms were first used in [29] for video transition detection. It is widely explored in several works in order to study temporal properties on video data [27], [30]–[32].

A visual rhythm consists of an abstraction of a video that encodes the temporal change of pixel values according to a specific sampling strategy [30]. Formally, a visual rhythm is a simplification of a video \( V = \{ f_t \}, t \in [1, T] \), in domain \( 2D + t \), with \( T \) frames of dimensions \( W_V \times H_V \), in which each frame \( f_t \) is transformed into a vertical line on an image \( R \), in domain \( 1D + t \), such that,

\[
R(t, z) = f_t(r_x \times z + a, r_y \times z + b), t \in [1, W_R], z \in [1, H_R],
\]

where \( W_R \) (\( W_R = T \)) and \( H_R \) are its width and height, respectively; \( r_x \) and \( r_y \) are the sampling rates along the horizontal and vertical directions; \( a \) and \( b \) are the horizontal and vertical offsets on each frame, respectively.

B. Dissimilarity measure based on bipartite graph matching

A dissimilarity measure based on bipartite graph matching for transition detection in videos was proposed in [27], in which abrupt and gradual transitions can easily be identified by a strategy based on this measure. It is important to note that this measure can be computed on either the frame video sequence or the visual rhythm. Moreover, two points \( P_{l_1} \) and \( P_{l_2} \) at locations \( l_1 \) and \( l_2 \), respectively, belong to different columns of the visual rhythm are considered spatially close if \( |l_1 - l_2| \) is smaller than a specified threshold (\( \lambda \)). For simplification, and without loss of generality, we will omit the point distance hereafter, but all pair of points must be within the distance \( \lambda \) which is fixed to 5. Furthermore, a list of frame points \( L_{f_t} \) is used for representing the point values extracted from a visual rhythm at location (or column) \( t \).

To formally define the dissimilarity measure, some definitions are introduced as follows.
Definition 1 (Point similarity – $\text{PS}(P_{l_1}, P_{l_2}, \delta)$): Let $P_{l_1}$ and $P_{l_2}$ be two close points at locations $l_1$ and $l_2$, respectively. Two points are similar if a (dis)similarity measure $\mathcal{D}(P_{l_1}, P_{l_2})$ between them is smaller than a specified threshold ($\delta$). The point similarity is defined as

$$
\text{PS}(P_{l_1}, P_{l_2}, \delta) = \begin{cases} 
  1, & \text{if } \mathcal{D}(P_{l_1}, P_{l_2}) \leq \delta \\
  0, & \text{otherwise}
\end{cases}
$$

There are several choices for $\mathcal{D}(P_{l_1}, P_{l_2})$, i.e., the (dis)similarity measure between two points, such as pair-wise difference of histogram (grayscale) values, difference of color average calculated for point neighborhoods, among others. In this work, the euclidean color distance is used as a dissimilarity measure between two points. After selecting a value for $\delta$, it is possible to construct a point similarity graph based on two images $f_{t_1}$ and $f_{t_2}$. The frames are represented by the columns $t_1$ and $t_2$ of the visual rhythm. The point similarity graph is computed directly from the visual rhythm instead of the frames and it is built upon a list of frame points, defined below, which is used to represent the frame content.

By using two different lists of frame points, the point similarity graph can be defined as follows.

Definition 2 (Point similarity graph – $G^{\delta, \lambda}(L_{f_{t_1}}, L_{f_{t_2}})$): Let $L_{f_{t_1}}$ and $L_{f_{t_2}}$ be two lists of frames. A point similarity graph $G^{\delta, \lambda}(L_{f_{t_1}}, L_{f_{t_2}}) = (N^{L_{f_{t_1}}} \cup N^{L_{f_{t_2}}}, E^{\delta, \lambda})$ is a bipartite graph. Each node $v^{L_{f_{t_1}}} \in N^{L_{f_{t_1}}}$ represents a point $P_1 \in L_{f_{t_1}}$ and each node $v^{L_{f_{t_2}}} \in N^{L_{f_{t_2}}}$ represents a point $P_2 \in L_{f_{t_2}}$. There is an edge $e \in E^{\delta, \lambda}$ between $v^{L_{f_{t_1}}}$ and $v^{L_{f_{t_2}}}$ if point similarity of associated points $P_1$ and $P_2$ are equal to 1, i.e.,

$$
E^{\delta, \lambda} = \{ (v^{L_{f_{t_1}}}, v^{L_{f_{t_2}}}) \mid v^{L_{f_{t_1}}} \in N^{L_{f_{t_1}}}, v^{L_{f_{t_2}}} \in N^{L_{f_{t_2}}}, \text{PS}(P_1, P_2, \delta) = 1 \}
$$

Figure 1 illustrates an example of the graph matching procedure. In Figure 1 (a), point similarity graph is obtained from visual rhythm with $\delta = 0$, i.e., two points are similar only if they have equal values.

Finally, for stating the dissimilarity measure proposed in [27], matching and maximum cardinality matching are defined as follows.
**Definition 3 (Matching – M_δ,λ):** Let G^δ,λ(L_{f_1}, L_{f_2}) be a point similarity graph between the frames \( f_1 \) and \( f_2 \) represented by their list of frame points \( L_{f_1} \) and \( L_{f_2} \). A subset \( M^δ,λ \subseteq E^δ,λ \) is a matching if any two edges in \( M^δ,λ \) are not adjacent.

**Definition 4 (Maximum cardinality matching – M^δ,λ):** Let \( M^δ,λ \) be a matching in a point similarity graph \( G^δ,λ(L_{f_1}, L_{f_2}) \). So, \( M^δ,λ \) is the maximum cardinality matching if there is no other matching \( M^Δ,λ \) in \( G^δ,λ(L_{f_1}, L_{f_2}) \) such that \(|M^Δ,λ| > |M^δ,λ|\).

Based on the size of maximum cardinality matching we can define max-matching dissimilarity measure as follows.

**Definition 5 (Max-matching dissimilarity measure – DIS^δ,λ):** Let \( M^δ,λ \) be a maximum cardinality matching in a point similarity graph \( G^δ,λ(L_{f_1}, L_{f_2}) \). So, the max-matching dissimilarity measure \( DIS^δ,λ(L_{f_1}, L_{f_2}) \) can be calculated as

\[
DIS^δ,λ(L_{f_1}, L_{f_2}) = 1 - \frac{|M^δ,λ|}{\max\{|L_{f_1}|, |L_{f_2}|\}}
\]

Two consecutive frames that are similar can be represented by only one. Thus, their lists of frame points are also expected to be similar, and consequently a high similarity score (computed using the size of the maximum cardinality matching) should be encountered. Equation 3 states that if two lists of frame points are not very similar, i.e., if it may have a transition or a new information between the associated frames, then \( DIS^δ,λ(L_{f_1}, L_{f_2}) \approx 1 \). The max-matching dissimilarity measure can be considered as a distance function. Thus, max-dissimilarity measure is calculated for each of those graphs and used to compute max-r dissimilarity measure as follows.

**Definition 6 (Max-r dissimilarity measure – rDIS^δ,λ):** Let \( f_k \) be frame at location \( k \), \( k \in [0, N-1] \) and \( L_{f_k} \) be the list of frame points associated with that frame. So, for a \( 2r+1 \) set of available frames centered at frame \( f_k \), i.e., for a frame interval \( f_{k-r}, f_{k-r+1}, \ldots, f_k, \ldots, f_{k+r-1}, f_{k+r} \), the max-r dissimilarity measure \( rDIS^δ,λ_k \) can be calculated as

\[
rDIS^δ,λ_k = \max_{i = k-r \ldots k+r} \left\{ DIS^δ,λ(L_{f_k}, L_{f_i}) \right\}
\]

### III. UNSUPERVISED HYPERSPECTRAL BAND SELECTION

In this section, we detail the proposed techniques for band selection. Firstly, we define the spectral rhythm and how to apply it in our band selection method. Secondly, the proposed unsupervised band selection method is described and detailed. Finally, we present the random pixel sampling approach to reduce the band selection processing time.
A. Spectral Rhythms

A spectral rhythm (SR) is an abstraction of the temporal visual rhythm introduced in II-A.

Visual rhythms contain patterns or visual signatures of a video sequence. By a visual rhythm transformation, a video of dimensions $2D + t$, two spatial dimensions plus time, becomes a two dimensional image $(1D + t)$.

Likewise, a hyperspectral image can be seen as a sequence. In such case, time is replaced by frequency (or wavelength) leading to a $2D + \omega$ domain, where $\omega$ represents the frequency. Thus, we introduce the concept of spectral rhythm of a hyperspectral image $(2D + \omega)$, which transforms each spectral band into a vertical line and concatenates all bands creating a new image in the domain $1D + \omega$. Each line of a spectral rhythm image represents the evolution of the response of a spatial pixel on the frequency domain.

Formally, a spectral rhythm is a simplification of a hyperspectral image $S = \{b_\omega\}, \omega \in [1, \Omega]$, in domain $2D + \omega$, with $\Omega$ bands of dimensions $W_S \times H_S$, where each band represents a frequency range. Each band $b_\omega$ is transformed into a vertical line on an image $R$, in domain $1D + \omega$, such that,

$$R(\omega, z) = b_\omega(r_x \times z + a, r_y \times z + b), \omega \in [1, W_R], z \in [1, H_R],$$

where $W_R$ ($W_R = \Omega$) and $H_R$ are its width and height, respectively; $r_x$ and $r_y$ are the sampling rates along the horizontal and vertical directions; $a$ and $b$ are the horizontal and vertical offsets on each frame, respectively.

The values of $r_x, r_y, a$ and $b$ define the transformation applied into a $2D$ band $b_\omega$ to a $1D$ line of $R$. We used the following values for the band selection proposed in this work: $a = 0, b = 0, r_x = r$ and $r_y = q$, where $r$ and $q$ are the remainder and the quotient of the integer division: $z/W_S$.

B. Band Selection Scheme

Methods for band selection are based on metrics to select the most informative and distinctive bands in order to provide a subset that could be the most appropriate. Thus, our band selection scheme looks for bands which are most relatively dissimilar with respect to their neighborhood, providing a ranking of hyperspectral bands according to their representativeness. Moreover, our method can be subdivided into three steps: (i) computation of spectral rhythms; (ii) computation of max-$r$ dissimilarity measure $rDIS_k^{\delta,\lambda}$, and (iii) selection of salient bands.

Figure 2 shows these three steps as a scheme of the band selection process. After the computation of the spectral rhythm, which is illustrated in Figure 2 (a), a dissimilarity measure is computed among nearby
bands (Figure 2 (b)). In this work, we have used the bipartite graph matching, which is described in Section II for video sequences. However, instead of using visual rhythms, we use our new abstraction, the **spectral rhythm**. The outcome of this process is a one dimensional signal containing static and transition information according to respective neighbour bands, which we call **dissimilarity curve**. The selection of salient bands (Figure 2 (c)) is based on peaks of the dissimilarity curve. A peak of dissimilarity represents a salient band regarding its neighbors which may contain different information. Thus, instead of directly analyzing the dissimilarity values of the bands, the method identifies local maxima on the dissimilarity curve. This strategy allows the selection of the most representative bands in the set. It is interesting to note that other strategies could be implemented to extract good candidate bands from the dissimilarity curve, as we have proposed in [28].

In order to detail the proposed method of salient bands selection, we pursue with three definitions:

**Definition 7 (Forward difference operator [33]):** The Forward difference operator or discrete derivative operator $\Delta$ is defined for a function $g$:

$$\Delta g = x \mapsto g(x + 1) - g(x)$$
Definition 8 (Discrete Laplace operator [33]): The Discrete Laplace operator $\Delta^2$ is defined for a function $g$:

$$\Delta^2 g = x \mapsto g(x + 1) - 2g(x) + g(x - 1)$$

Definition 9 (Strict local maximum [34]): Let $g : S \rightarrow R$, $\omega_1 \in S$ be a strict local maximum of $g$, if there exists an $\varepsilon$-neighbourhood $N_\varepsilon(\omega_1)$ around $\omega_1$ such that $g(\omega_1) > g(\omega)$ for each $\omega \in N_\varepsilon(\omega_1) \cap S$.

Before selecting salient bands, some operations on the dissimilarity curve are applied in order to filter out spurious information: (i) closing operation for removing small valleys [35]; (ii) white top hat operation [35] for computing morphological residues [36] which represent local variations; and finally (iii) thresholding for eliminating small variations.

The band selection proceeds by pointing out the bands that are represented by peaks on the function $g(\omega)$ where $g : S \rightarrow R$ is the smoothed unidimensional dissimilarity curve $rDIS^{\delta,\lambda}_k$ (for each frequency range $\omega \in S$, $g(\omega) \in R$ is the dissimilarity value computed from the spectral rhythm of the hyperspectral image for the band $b_{\omega}$).

Thus, a band $b_{\omega_1}$ is selected if $\omega_1$ is a strict local maximum of the function $g(\omega)$. In order to find the strict local maximum in a point $\omega_1$, two conditions must be truth:

1) First order condition: for $\omega_1 \in S$, $\Delta g(\omega_1) > 0$ and $\Delta g(\omega_1 - 1) < 0$ or $\Delta g(\omega_1) < 0$ and $\Delta g(\omega_1 - 1) > 0$. Which means that the discrete derivative crosses the zeros around the point $\omega_1$.

2) Second order condition: for $\omega_1 \in S$, $\Delta^2 g(\omega_1) < 0$. Which means that the discrete Laplace operator is negative at the point $\omega_1$.

The strict local maximum detection is achieved by finding the points $\omega_1$ in the function $g$ that made the two conditions truth. This process is performed iteratively, by selecting the most important bands until the algorithm finds the number of bands required, or all bands are ranked. Bands found at the same iteration and with the same dissimilarity value are sorted in ascending order. Thus, the iterative peak detection have as output a list with the spectral bands ranked by their importance on the selection.

C. Spectral Rhythm by Pixel Random Sampling

The full spectral rhythm representation contains a huge number of pixels computed from the hyperspectral image. Moreover, considering that the computation of a dissimilarity measure for sorting the bands is directly dependent on the size of such image, the methods are time consuming when a full representation is used. Thus, a simplification of this representation is useful for on-line applications in which the methods must be fast while preserving the classification accuracy. In this work, we propose
a strategy for reducing the size of spectral rhythm by taking a sample of the pixels from each band, instead of considering all pixels. A stratified random sampling of the spectral bands is used to select pixels samples in order to better estimate the original band [37].

Stratified random sampling allows to divide the population (the complete set of band pixels) of $N$ pixels into $P$ subpopulations, $N_1, ..., N_P$, where $N_1 \cup N_2 \cup ... \cup N_P = N$. Then, a simple random sampling is proceed for each subpopulation, where, all pixels from a subpopulation have the same probability to be chosen. The subpopulations $N_p, p \in [1, P]$ have the same size of $s$ pixels, being $s = N/P$, total number of pixels divided by the number of subpopulations. We define the subpopulations, $N_p$, as a nonoverlapping squared region of size $\sqrt{s} \times \sqrt{s}$ pixels in order to have an homogeneous spatial distribution to better represent a band content. By randomly choosing one representative pixel from each subpopulation containing $s$ pixels, we provide a reduction by a factor of $s$, passing from $N$ to $P$ pixels in each spectral band. The spectral rhythm for a random pixel sampling by a factor of $s$ will be denote by $s$-RPS-SR.

Figure 3 shows an example of a stratified random sampling process of one band, and the impact it causes to a column of the spectral rhythm for $s \in \{1, 4, 16, 64\}$, in which $s = 1$ means no simplification. It is important to note that for each spectral rhythm column, or spectral band, the pixels sampled are the same, which means that the samples are randomly chosen once and the selected pixel is used for all bands.

Fig. 3. Stratified random sampling example for computing $s$-RPS-SR-based for $s = \{1, 4, 16, 64\}$ in which 1-RPS-SR-based means no simplification.
D. Complexity Analysis

To show the efficiency of the proposed approach, we conducted a theoretical complexity analysis as follows.

Let us remember the three main steps of our method for band selection: (i) computation of spectral rhythms either without simplification or with a pixel random sampling; (ii) computation of max-$r$ dissimilarity measure $r\text{DIS}^{k,\lambda}$ for each band represented on the spectral rhythm; and (iii) selection of salient bands. The computation of max-$r$ dissimilarity measure is the most time consuming step since it is necessary to compute the lowest value of the maximum cardinality of the bipartite graph matching between the central representation of the hyperspectral image and all others from a specified window which corresponds to the highest dissimilarity score.

Solving the maximum cardinality matching on a bipartite graph could be done with $O(E\sqrt{V})$ operations [38], in which $V$ and $E$ represent the number of nodes and edges, respectively. The number of nodes is always equal to $2|L_{f_i}|$, where $2|L_{f_i}|$ is the number of points used to realize the matching. The number of edges depends on point similarity measures and threshold. It could be close to zero, but it also could be equal to $|L_{f_i}| \times \lambda$. In the worst scenario, $\lambda = |L_{f_i}|$, since all points are connected to all points in the consecutive list of points. So, the computational cost to compute the max-$r$ dissimilarity measure $r\text{DIS}^{k,\lambda}$ is given by $O(2r\lambda|L_{f_i}|\sqrt{2|L_{f_i}|})$ in which $2r$ means the number of point similarity graphs, $\lambda$ the neighbourhood in consecutive list of points to be analysed and $|L_{f_i}|$ the number of points. Therefore, the computational cost is directly related to the number of points $|L_{f_i}|$ to be matched that explains the low time in seconds (described later in Sec. V-C) for 64-RPS-SR-based with respect to the 1-RPS-SR-based, for example. Moreover, considering that in this work we fixed both $r$ and $\lambda$ to small values, the computational cost for computing the max-$r$ dissimilarity measure for each band can be re-written by $O(|L_{f_i}|\sqrt{2|L_{f_i}|})$ in which we ignore the $r$ and $\lambda$. Afterwards, a one-dimensional signal of size equal to the number of bands $\Omega$ is created. Thus, the computational cost of the proposed method for band selection is finally defined by $O(\Omega|L_{f_i}|\sqrt{2|L_{f_i}|})$.

IV. Experimental Setup

We carried out experiments to evaluate the effectiveness and efficiency of the proposed techniques. Experiments have been conducted in three real hyperspectral datasets which are detailed in Section IV-A. The experimental protocol consists in evaluating the band selection scheme by classifying the pixels of a hyperspectral image using, as features, the selected bands. For each dataset, the pixels are classified in land cover classes with a supervised classification method with a cross validation scheme.
of results considers the average accuracy and standard deviation in a 5-fold cross validation. We have organized the experiments in three main parts aiming at evaluating different aspects of the techniques proposed in this paper: (i) Band selection based on Spectral Rhythms analysis with random pixel sampling; (ii) Comparison with baselines; and (iii) Processing time.

For the former, we proceed the band selection with the proposed method and the classification scheme using several amounts of selected bands as features in order to build classification accuracy versus features set size curve. The spectral rhythm-based method was also tested with different sample rates, from 1 (no sampling) to 144, due to the different size of the datasets, different sample rates were chosen for each one. We chose sample rates that could satisfactorily reduce the processing time without degradation of the information, three sampling rates for each dataset were tested in order to verify the behavior of the method with the processing time reduction. According to these experiments, it will be possible to evaluate both the effectiveness and the high accuracy rates of our method with several random pixel sampling. Thanks to the random pixel sampling, the processing time is largely decreased in comparison with the method without this step.

Regarding the comparison, four methods of dimensionality reduction were used as baselines evaluation of efficiency and effectiveness. Those methods belong to different approaches, however, the same experimental protocol were used, allowing a fair comparison. Methods are detailed in Section IV-B. Finally, the time spent by our method to select bands were computed as well as the baseline’s to perform the same task.

The stochastic nature of the classification problem requires a careful analysis. By proceeding a cross-validation scheme and studying the values in average and standard deviation because of its random source, we converge our results to the reality and prevent outliers. Nonetheless, in order to compare two methods, it is still necessary to verify the equivalence between the results even if they have different values, which statistically means that is necessary to test if two random variables may came from the same distribution (same mean value) or not. [39] presents some of the most common statistical tests for supervised learning, one of them is the k-fold cross-validate paired t test. This test is based on statistical hypothesis Student’s t test, in which, under null hypothesis, the data follow a Student’s distribution with \( k - 1 \) degrees of freedom, where \( k \) is the number of samples. We used this test to compare the errors rate from our method with the baselines methods and with the sampling results, if the null hypothesis is rejected, the two method do not have the same error rate and they are not equivalent.
A. Datasets

The *Indian Pines dataset* is a real world dataset which was acquired by the AVIRIS sensor in 1992. This dataset is composed by 200 spectral bands of 145×145 pixels. The scene includes agriculture and forest areas and is divided in 16 land cover classes.

The *Salinas dataset* was also collected by AVIRIS sensor and comprises 204 spectral bands of size 217×512 pixels. There are 16 land cover classes of an agricultural area. Finally, the *Pavia University dataset* was collected by ROSIS sensor and has 103 spectral bands of size 340×610 pixels. The scene is divided in 9 regions from a University area.

B. Baselines

In order to assess the performance of the proposed *spectral rhythm*-based method, a comparison study is made with four methods in the state-of-the-art: (i) a feature transformation based on PCA method; (ii) a ranking based on entropy method; (iii) a clustering based on mutual information method; and (iv) a clustering based on Kullback-Leibler divergence. Next, we summarize these methods pointing out their main limitations.

1) **Principal Component Analysis (PCA):** PCA is a transformation for dimension reduction based on eigen decomposition of the data covariance matrix, such a way that the orthogonal components have the highest variance as possible. PCA is widely used in hyperspectral simplification [40] and it can reduce \( N \) band images into a set of \( M \) features, \( M < N \), in an effort to keep as much as information. This method may find a good feature when the classes of the data are located along the largest eigenvector [3].

Although PCA may produce good results, [23] proved that it is not adequate for hyperspectral images reduction because their class distribution. Besides, methods based on feature transformation are subject to loss of physical meaning and interpretation.

2) **Entropy:** In [18] and [19], a band selection method based on entropy is described. The entropy of a random variable \( A \) defined on a sample space \( \mathcal{A} \) with probability function \( p : \mathcal{A} \rightarrow [0, 1] \) is defined by:

\[
H(A) = - \sum_{A \in \mathcal{A}} p(A) \log p(A)
\]

The authors analyze each spectral band as a random variable with sample space defined by the range of values of the pixels, normally the gray level scale (from 0 to 255). The probability function \( p \) is calculated empirically by the frequency of each value on each spectral band (the histogram of the band).
The main property of the entropy-based method is to select bands with highest entropy since they normally have a large amount of information in the considered data. Entropy methods are usually very simple, easy to compute and computational costless if compared to most of the other methods that require complex calculations between each pair of bands. However, the entropy method was considered not consistent with the requirement of image classification [16]. Because of entropy calculation considers only one random variable, band, without any reference, [16] pointed that the amount of information measured by the entropy may not guarantee the information content useful in the classification.

3) Mutual Information (MI) clustering: Martinez et al. [12] use MI measures among hyperspectral bands as a computation of statistical distances.

MI is a measure of the statistical dependency between two random variables. Given two random variables $A$ and $B$ with sample spaces $\mathcal{A}$ and $\mathcal{B}$, marginal probabilities $p_A : \mathcal{A} \to [0, 1]$ and $p_B : \mathcal{B} \to [0, 1]$ and joint probability $p_{AB} : \mathcal{A}, \mathcal{B} \to [0, 1]$, the MI is defined by:

$$I(A, B) = \sum_{A \in \mathcal{A}} \sum_{B \in \mathcal{B}} p(A, B) \log \frac{p(A, B)}{p(A)p(B)}$$

And a distance measure based on MI may be calculated as follows:

$$D_{MI}(A, B) = \left(1 - \sqrt{\frac{2I(A, B)}{H(A) + H(B)}}\right)^2$$

where $H(A)$ and $H(B)$ are the marginal entropies of $A$ and $B$. Similarly to the entropy-based approach, each spectral band is considered as a random variable and $D_{MI}(A, B)$ is calculated for all $A$ and $B$ being bands from the hyperspectral image and $A \neq B$. Thus, a matrix $N \times N$ of $D_{MI}(..,..)$ is built, where $N$ is the number of spectral bands. From this matrix a clustering algorithm based on Ward’s linkage method is applied to divide the set in subsets of similar bands. The representative band from each cluster is selected as the band with the lowest average distance from the other bands in the cluster. Which means that, on average, the band with a high correlation with regard to the other bands in the cluster is selected.

The main drawback of this method is the processing time. For each pair of bands, it is necessary to compute the mutual information distance $D_{MI}$. Considering the high dimensionality of the bands set and the resolution of each band, it may became expensive.

4) Kullback Leibler (KL) clustering: Martinez et al. [12] also proposed a cluster approach based on KL divergence, which may be used to compute the distance between two probabilities distributions. Similarly to MI approach, a clustering algorithm (Ward’s linkage) is used to group the similar bands and
representatives from each cluster are chosen, however, by considering KL divergence based distance as follows.

\[ D_{KL}(A, B) = \sum_{A \in A} p(A) \log \frac{p(A)}{p(B)} + \sum_{B \in B} p(B) \log \frac{p(B)}{p(A)} \]

Being \( A \) and \( B \) two bands representing two random variables with sample spaces \( A \) and \( B \), marginal probabilities \( p_A : A \rightarrow [0, 1] \) and \( p_B : B \rightarrow [0, 1] \).

Same problems with processing time encountered on the previous method may be highlighted here.

C. Classification protocol

We have used a supervised classifier scheme in order to apply the selected features and be able to compare the accuracy provided by each dimensionality reduction method. The Support Vector Machines (SVM) classifier was chosen because it is considered a reference method in supervised classification field and it has been widely employed in remote sensing tasks. We used a RBF kernel and a multi-class support based on one-against-one scheme. The two parameters of the method, \( C \) and \( \gamma \), were chosen by proceeding an exhaustive search in a list of 225 parameter combinations. To assure statistical validity of the classification results a 5-fold cross validation scheme was used. In each fold run, the parameters were optimized according to the above exhaustive search.

V. Results and Discussions

This section contain the results and discussion of the experiments conducted to evaluate: (i) the capability of the proposed method in selecting bands and (ii) the effectiveness of the proposed sampling procedure in order to reduce processing time. Hereafter, our method will be denoted by \( s \)-RPS-SR-based which means that the method is based on spectral rhythm using a random pixel sampling by a factor of \( s \).

This section is summarized as follow. Firstly, the accuracy results of the proposed method and the pixel sampling procedure are exposed. Secondly, our results are also compared with four state-of-the-art methods. And finally, in order to allow full analysis of performance, we present the processing time of each method.

A. Spectral Rhythm-based band selection

The band selection process provides a rank of the bands by their information, which means that the outcome of our method scheme (Figure 2) is a list of bands ranked by their importance on representing
the group. Thus, we used different sampling rates to compute the *spectral rhythms* and the dissimilarity curves from each dataset. An example of a dissimilarity curve obtained for each dataset is presented in Figure 4. For computing these curves, we have used the following parameters for India Pines dataset $\lambda = 5$, $\delta = 2$ and $r = 2$. For Salinas and Pavia University, the parameters are $\lambda = 5$, $\delta = 3$ and $r = 3$.

![Indian Pines dissimilarity curve](image)

![Salinas dissimilarity curve](image)

![Pavia University dissimilarity curve](image)

(a) Indian Pines  
(b) Salinas  
(c) Pavia University

Fig. 4. Dissimilarity curves for the three datasets using 1-RPS-SR-based method. The horizontal axis represents each band of the dataset and the vertical represents the amount of dissimilarity of this band. The red color intensity represents the top ranked bands.

From the dissimilarity curve, the most informative bands are selected by identifying the strict local maxima on each curve, Figure 4 red color represents the bands selected in a rank. Thus, bands are sorted and ranked by using the complete *spectral rhythm* and the sampled ones. Figure 5 shows the ranking of bands for each dataset. We have used a gray level color range to represent the importance of the bands in the ranking. The white bands represents high importance and black less important ones. This kind of representation allows the visualization of the selected bands by using $s$-RPS-SR-based method for $s = \{1, 4, 16, 64\}$ for Indian Pines dataset and $s = \{1, 16, 64, 144\}$ for Salinas and Pavia University datasets. It is important to note that when $s = 1$ there is no random pixel sampling, *i.e.*, the *spectral rhythm* contains the whole data. The datasets do not have the same random pixel sampling factors, $s$, due to a size constraint of the Indian Pines dataset, which is very smaller than the others, we should not apply a random pixel sampling by a factor of 144.

By analyzing Figure 5, one can see that 1-RPS-SR-based method achieved very similar results when compared to the results obtained by the random pixel sampling by a factor greater than one. Several ranges of similar gray level appear in the same range of bands.

1) *Classification results:* A curve of accuracy versus the amount of used bands is shown in Figure 6 for each dataset. The main goal is to decrease the size of the spectral rhythm without performance loss as
Fig. 5. **Indian Pines**: bands from 0 to 199, colored by their rank score for \(s\)-RPS-SR-based method for \(s = \{1, 4, 16, 64\}\). **Salinas**: bands from 0 to 203, colored by their rank score for \(s\)-RPS-SR-based method for \(s = \{1, 16, 64, 144\}\). **Pavia University**: bands from 0 to 102, colored by their rank score for \(s\)-RPS-SR-based method for \(s = \{1, 16, 64, 144\}\). Bands with more importance have larger values (white) and the less important bands have smaller values (black).

much as possible. For Salinas and Pavia University datasets, it is illustrated the accuracy results for the \(s\)-RPS-SR-based method for \(s = \{1, 16, 64, 144\}\) and \(s = \{1, 4, 16, 64\}\) for Indian Pines dataset. Observe that we try to compare the method without random pixel sampling to other three sampling rates to well illustrate the trend curve for the accuracy rates.

In the horizontal axis, the amount of features from few to all bands. In the vertical axis, the average accuracy in the classification. The error bar in each plot represents two times the standard deviation value calculated by the 5-fold experiment.
The results exposed in Figure 6 show that the proposed $s$-RPS-SR-based method achieves good precision accuracy with a small set of bands. By using 5% or 10% of the number of bands, the classification accuracy is very close to the accuracy obtained by using all bands. It means that we succeed in selecting bands that can represent the entire set without degradation of information.

By comparing the results for different random pixel sampling, it is possible to see that results are very similar in most part of the cases. The sampling process proceeds by randomly choosing the sample set. Because of that, to avoid outliers, we computed 5 times the accuracy of the sampling. We observed that the results were statistically equivalent in all tests.

In order to verify statistical equivalence between different random pixel sampling strategies when compared to 1-RPS-SR-based method, we conduct the Student's $t$ test as described in Section IV. The outcome of the test is a value that may came from a Student distribution with a probability called $p$-value. A threshold for the $p$-value should be chosen to reject or to accept the null hypothesis. Figure 7 shows the $p$-values obtained in the $t$ test for each dataset, each sample rate and each amount of bands. A gray color level represents the probability from 0.0 to 1.0. Red points highlight the probabilities smaller than 2.5%, which means that the distributions are not equivalent.

By analyzing these results, we note that results are considered equivalent for most part of the instances. However, in a few cases, with $p$-value smaller than 2.5%, the results are considered not equivalent according to the test. Accordingly to [39], the cross-validation $t$-test is very powerful in the sense of being capable of detect statistical difference when a statistical difference exists. However, this test may present a considerable probability in incorrectly detecting difference when no difference exists. That means that for all cases where the method did not detect a statistical difference, it was correct, and it may be wrong on the contrary. Based on the argument above and on the fact that the statistical difference were detected in just few cases, the simplification for computing spectral rhythm according to a random pixel sampling by a factor greater than 1, successfully reduces the processing time without loss in terms of classification accuracy.

B. Baseline comparison

Figure 8 shows the accuracy reached for the four baselines and the proposed method using the higher random pixel sampling. In other words, for Salinas and Pavia University, we apply 144-RPS-SR-based to the classification task. For Indian Pines, we apply 64-RPS-SR-based to the same task.

Several number of features set size are represented in the graphics. The statistical paired test between the baselines method and the proposed method are shown in Figure 9.
By analyzing each method separately or together (Figure 9 and 8), it is possible to point out some observations. Concerning a few number of features: (1) the proposed $s$-RPS-SR-based method produced the best accuracies; (2) For Indian Pines dataset, with 5, 10 and 15 bands selected, the proposed method presented the highest accuracy value. T-student test show that is equivalence with MI method using 5 features, while better than the other three. For 10 features, the proposed method was equivalent to PCA, while they were better than the other three, and finally, it was the best method for 15 features; (3) In Salinas dataset, considering 5, 10 and 15 features, the proposed method was equivalent to PCA and MI method whereas better than Entropy and KL (equivalent to Entropy for 10 features); and (4) In Pavia University dataset we have achieved higher results, equivalent with MI method and Entropy using 10 features.

Concerning a medium number of features (from 10% to 70% of the maximum number) PCA developed better or equivalent results than the proposed method for Pavia University and Indian Pines dataset and equivalent results for Salinas dataset.

Generally, with more than 70% of the band all methods have similar classification accuracy.

In conclusion, the proposed method sucessfully proceed to effectively and efficiently select a small number of bands. As discussed before, our method $s$-RPS-SR-based is, at least, as good as the compared methods in terms of accuracies but it is faster than the compared one.

C. Processing time comparison

Although all methods exposed in this paper have dimensionality reduction proposals, they work differently from each other, making the comparison difficult in terms of time processing. Concerning our method and the entropy-based one, the outcome is a rank of bands and the decision of how many bands will be selected can be performed *a posteriori* without any extra cost. In other words, the method is performed once and the processing time does not depend on the number of selected bands. In contrast, PCA and MI method have different processing times which depends on the size of the feature set used.

Despite of this, we attempt to compare processing time from each method. We have computed the time spent to select 10 bands by using the same computational resources. We report the average value after five runs for each method. Table I shows the time spent for ranking the bands by using the proposed method and the baseline methods.

It is possible to see that the proposed random pixel sampling reduces drastically the processing time of the method making it efficient when compared to the others methods. By analyzing the results, with the proposed sampling strategies, our method was the second fastest method. The fastest method is entropy-
TABLE I
Processing time relation for the three datasets and the methods s-RPS-SR-based, Entropy, MI, PCA, and KL

<table>
<thead>
<tr>
<th></th>
<th>Indian Pines</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-RPS-SR</td>
<td>4-RPS-SR</td>
<td>16-RPS-SR</td>
<td>64-RPS-SR</td>
<td>Entropy</td>
<td>MI</td>
</tr>
<tr>
<td></td>
<td>1054s</td>
<td>66s</td>
<td>5s</td>
<td>0.9s</td>
<td>0.8s</td>
<td>810s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Salinas</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-RPS-SR</td>
<td>16-RPS-SR</td>
<td>64-RPS-SR</td>
<td>144-RPS-SR</td>
<td>Entropy</td>
<td>MI</td>
</tr>
<tr>
<td></td>
<td>60025s</td>
<td>245s</td>
<td>20.7s</td>
<td>6.7s</td>
<td>1.52s</td>
<td>3918s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pavia University</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-RPS-SR</td>
<td>16-RPS-SR</td>
<td>64-RPS-SR</td>
<td>144-RPS-SR</td>
<td>Entropy</td>
<td>MI</td>
</tr>
<tr>
<td></td>
<td>91777s</td>
<td>377s</td>
<td>29.5s</td>
<td>7.5s</td>
<td>1.31s</td>
<td>2617s</td>
</tr>
</tbody>
</table>

based one, but it achieves the worst accuracy results. PCA-based is the third in the processing time rank. MI-based method and KL-based method were very expensive in terms of processing time. The construction of the distance matrix provides the high processing time of the method since it is computed the distance between all pair of bands.

VI. CONCLUSIONS

In this paper, we propose a new unsupervised band selection method based on the dissimilarity among neighboring bands by exploiting an intermediary representation called spectral rhythm. We also propose a pixel sampling strategy to improve the efficiency of our method without significant reduction on the quality of selected bands. We can point out the following conclusions on our research: Spectral Rhythm (SR) representation is a suitable way to compute bands dissimilarities; the proposed SR-based band selection strategy is as effective as state-of-the-art hyperspectral dimensionality reduction techniques; the proposed Random Pixel Sampling is able to speed up the band selection process keeping high accuracy rates on classification tasks.

As further work, we intend to improve the spectral rhythm representation by using super pixel approaches and investigate other pixel sampling strategies.
ACKNOWLEDGMENT

The authors would like to thank Rodrigo Y. M. Nakamura for the contributions. Also, CAPES and CNPq/INCT-InWeb.

REFERENCES


**Lilian Chaves B. dos Santos** received the M.Sc. degree from École Supérieure d’Ingénieurs en Électronique et Électrotechnique, France. B.Sc degree in Electrical Engineering from Universidade Federal de Minas Gerais, Brazil.

**Silvio Jamil F. Guimarães** received his B.Sc., M.Sc. and D.Sc. degrees in Computer Science, from Universidade Federal de Viçosa (UFV), Universidade Estadual de Campinas (UNICAMP) and Universidade Federal de Minas Gerais (UFMG), in 1997, 1999 and 2003, respectively. He is the founder and header of the Audio-Visual Information Processing Laboratory. His research interests include digital image and video processing and computer vision applications, hierarchical information analysis, multimedia information systems, and content based information (image and video) retrieval.

**Jefersson A. dos Santos** received his B.Sc., M.Sc. and D.Sc. degrees in Computer Science from Universidade Estadual do Mato Grosso do Sul (UEMS), Brazil, University of Campinas (Unicamp), Brazil and University of Cergy-Pontoise, France/University of Campinas (Unicamp), Brazil, in 2006, 2009 and 2013, respectively. He is currently a professor in the Department of Computer Science at the Universidade Federal de Minas Gerais (UFMG), Brazil. His research interests include image processing, remote sensing, machine learning, and content-based multimedia information retrieval.
Fig. 6. Average accuracy of the classification per number of features by using features selected by $s$-RPS-SR-based method. For Salinas and Pavia University datasets, it is illustrated the accuracy results for the $s$-RPS-SR-based method for $s = \{1, 16, 64, 144\}$. For Indian Pines dataset, it is illustrated the accuracy rates for the $s$-RPS-SR-based method for $s = \{1, 4, 16, 64\}$.
Fig. 7. Matrix of p-values of paired t-test in the three datasets. Horizontal axis have the amount of bands selected and vertical axis have different random pixel sampling when compared to 1-RPS-SR-based. Red points indicate that the p-value is smaller than 2.5% and the null hypothesis is rejected.
Fig. 8. Average accuracy of the classification per number of features by using features selected by the proposed spectral rhythm method and the four baselines.
Fig. 9. Matrix of $p$-values of paired t-test in the three datasets. Horizontal axis have the amount of bands selected and vertical axis have the four baselines that were tested with the s-RPS-SR-based method. Red points indicate that the p-value is smaller than 2.5% and the null hypothesis is rejected.