UNSUPERVISED BAND SELECTION FOR HYPERSPECTRAL DATASETS BY DOUBLE GRAPH LAPLACIAN DIAGONALIZATION

Eduard Khachatrian [1]*, Saloua Chlaily* [1]*, Torbjørn Eltoft* [1]*, Paolo Gamba* [2]*, Andrea Marinoni* [1]

[1] Dept. of Physics and Technology, UiT The Arctic University of Norway, NO-9037 Tromsø, Norway. E-mail: {eduard.khachatrian, saloua.chlaily, torbjorn.eltoft, andrea.marinoni}@uit.no

[2] Telecommunications and Remote Sensing Lab., Dept. of Electrical, Computer, and Biomedical Engineering, University of Pavia, I-27100 Pavia, Italy. E-mail: paolo.gamba@unipv.it

ABSTRACT

The vast amount of spectral information provided by hyperspectral images can be useful for different applications. However, the presence of redundant bands will negatively affect application performance. Therefore, it is crucial to select a relevant subset that preserves the information of the original set. In this paper, we present an automatic and accurate band selection method based on Graph Laplacians. Unlike existing band selection methods, this method exploits two similarity measures simultaneously. Furthermore, it is performed on a superpixel level, so it allows us to preserve not only global but contemporaneously local particularities of original data. Experiments show the importance of measuring the relevance of the bands at local and global scales and the ability of the method to minimize intercorrelation among selected bands, hence improving the selection of the most informative spectral channels.

Index Terms— information theory, hyperspectral images, band selection

1. INTRODUCTION

Hyperspectral imagery provides large number of closely spaced narrow spectral bands in the visible and near-infrared portion of the electromagnetic spectrum with a high resolution, thus providing an enormous amount of information about the region of interest (ROI) [1]. However, not all the spectral bands are of actual importance, since some of the information can be redundant, corrupted, or unnecessary for a particular task [1, 2]. Besides, one of the main problems for the classification of hyperspectral images is the "curse of dimensionality" or "Hughes effect" [3] that makes the analysis challenging from the computational point of view. Therefore, dimensionality reduction methods, such as feature selection and extraction, are crucial steps in order to select relevant information [1].

This work is funded by Centre for Integrated Remote Sensing and Forecasting for Arctic Operations (CIRFA) and the Research Council of Norway (RCN Grant no. 237906).

Among the various dimensionality reduction methods, graph-based clustering methods play a key role [4, 2]. These methods represent hyperspectral data as a graph, where the spectral bands constitute the set of nodes and their connecting edges reflect their similarities. The connections among the nodes can be summarized by means of a graph Laplacian matrix, where each element identifies the degree of similarity (usually computed in terms of Euclidean distance) the nodes share [4, 2]. The dimensionality reduction is performed by determining similar nodes and selecting a representative attribute from each group using spectral clustering [2]. The clustering is fulfilled on a new representation of data generated using the eigenvectors of Laplacian.

In this paper, we introduce a band selection method, that relies on information theory metrics and on a representation based on graph Laplacians. Unlike existing graph-based clustering methods that are only using kernels as similarity measures, we are also considering the information content of the original data. Therefore, the similarity is quantified using two metrics simultaneously, which allows us to capture relevant information at different scales that improves the precision of the selection. The mutual information is performed globally and preserves the bands shared information [5]. The Gaussian kernel is applied locally and preserves the structure of the original data [4]. The superpixels refer to homogeneous regions of the image that share similar pixel information.

The remaining of this paper is organized as follows. The state-of-the-art is demonstrated in section 2. The method description is shown in section 3. The experimental results on hyperspectral datasets is illustrated in section 4. Finally, section 5 concludes the paper. For notational convenience, random scalars are denoted by lower case letters, *e.g.,* z. Random vectors are designated by bold lower case letters, *e.g.,* $\mathbf{z}$. Bold upper case letters refer to matrices, *e.g.,* $\mathbf{A}$.

2. EXISTING WORK

It is worth recalling that band selection represents a task in remote sensing data analysis that has been widely studied and

investigated in technical literature [1, 2]. Indeed, band selection methods can be categorized into five groups: ***ranking-based:*** measure the significance of each band by sorting them in terms of relevance using different metrics [1, 6]. Here we can highlight the covariance-based selection method (COVSEL) [7] and and the orthogonal subspace projection (OSP) [8, 1]; ***searching-based:*** convert band selection into an optimization problem of a given criterion function to form an optimal solution [1]. Among the methods in this family, we can mention ones based on linear projection (LP) used together with OSP [9], particle swarm optimization (PSO) that search for a band subset that optimizes the criterion function, typically in terms of minimum estimation abundance covariance (MEAC) or Jeffries–Matusita (JM) distance among different classes [10], minimum noise band selection (MNBS) [11]; ***clustering-based:*** grouping the bands into clusters and select representative ones from each cluster. Among this family, we can highlight discriminative kernel alignment (DKA) [12] and multigraph determinantal point process (MDPP) [13]; ***sparsity-based:*** emphasizing underlying structures within hyperspectral data by solving an optimization problem with sparsity constraints [1]. Here we can mention two methods, one with sparse non-negative matrix factorization (SNMF) that uses additional constraints, such as the thresholded Earth's mover distance (TEMD) [14], and sparse representation-based band selection (Spa-BS) [15] that does not require non-negative matrix factorization; ***hybrid scheme-based:*** integrate multiple operations mentioned above to obtain hyperspectral band selection. Here we can highlight a method where the ranking and eliminating searching schemes were combined to rank the bands according to minimum redundancy and maximum relevance (mRMR) along with sequential backward elimination to chose relevant bands [16].

3. METHOD

To preserve the particularity of every element in the observed scene, we implement the band selection on superpixels, i.e., regions showing homogeneous characteristics throughout the considered dataset. An image can be split into superpixels using different segmentation methods. In our work, we employ the Watershed segmentation algorithm [17].

We denote by $\mathbf{X} = (x_{lk}) \in \mathbb{R}^{L\times K}$ the matrix of spectral bands of the remotely sensed image $\mathbf{M}$, where K is the total number of bands, and L is the number of superpixels. We denote the k-th column of $\mathbf{X}$, that corresponds to the k-th band by $\mathbf{x}_{*k}$, so we can write $\mathbf{X} = [\mathbf{x}_{*1}, \ldots, \mathbf{x}_{*K}]$.

The main goal of band selection methods is to find, for a given superpixel l, the smallest subset of bands, $\{x_{l1}, \ldots, x_{lK}\}$, that preserves the structure and information content of the original set. To this aim, we build a fully connected graph $\mathcal{G}_l(\mathbb{V}_l, \mathbb{E}_l^{GK}, \mathbb{E}^{MI})$, where $\mathbb{V}_l$ denote the set of bands $\mathbb{V}_l = \{x_{l1}, \ldots, x_{lK}\}$, $\mathbb{E}_l^{GK}$ and $\mathbb{E}^{MI}$ are two set of edges that connect the nodes. The weight of the first edge is given by the Gaussian kernel (GK), while the weight of the second edge is quantified by the mutual information (MI). The Gaussian kernel between the i-th and k-th vertex is defined as follows [18],

$$w_{lij}^{GK} = \exp\left(-\frac{(x_{li} - x_{lj})^2}{2\sigma}\right), \tag{1}$$

where σ controls the width of the neighborhood in the graph, i.e., the number of connected vertices. The width of the neighborhood increases with σ. In this work, we set σ to 1. Gaussian kernel is applied at the local level selection, hence, the weights values are particular to each superpixel, i.e., $w_{pij}^{GK} \neq w_{qij}^{GK}$ for $p \neq q$.

The mutual information between the i-th and k-th vertex can be written as follows [19]:

$$\begin{aligned} w_{k_1,k_2}^{MI} &= I(\mathbf{x}_{*k_1}, \mathbf{x}_{*k_2}), \\ &= \sum_{i=1}^{L}\sum_{j=1}^{L} P(x_{ik_1}, x_{jk_2}) \log\left(\frac{P(x_{ik_1}, x_{jk_2})}{P(x_{ik_1})P(x_{jk_2})}\right), \end{aligned} \tag{2}$$

where $P(\mathbf{x}_i, \mathbf{x}_j)$ is the joint density function of $\mathbf{x}_i$ and $\mathbf{x}_j$, and $P(\mathbf{x}_i)$ and $P(\mathbf{x}_j)$ are the marginals. Mutual information quantifies the shared information between two random variables [19]. It is applied image-wise, therefore, the weights w_{k_1,k_2}^{MI} are equal for all superpixels. It is worth emphasizing that both similarity measures that are used as metrics are of equal importance since they maintain different information about the original dataset, GK preserves the local structure, while MI keeps the global information content. We refer to our method as GKMI.

To select the relevant subset of bands based on both measures, we partition the graph $\mathcal{G}_l$ into subgraphs, such that two vertices of the same subgraph have strong connections via both links. Two vertices from different subgraphs show at least one weak connection, either GK or MI. To this aim, we perform graph clustering by means of graph Laplacians [4]. We associate two graph Laplacians to the graph $\mathcal{G}_l$, one based on Gaussian kernel metric, $\mathbf{L}_l^{GK}$, while the other one is based on mutual information, $\mathbf{L}^{MI}$.

$$\mathbf{L}_l^{GK} = \mathbf{I} - \mathbf{D}_l^{GK^{-1/2}} \mathbf{W}_l^{GK} \mathbf{D}_l^{GK^{-1/2}}, \tag{3}$$

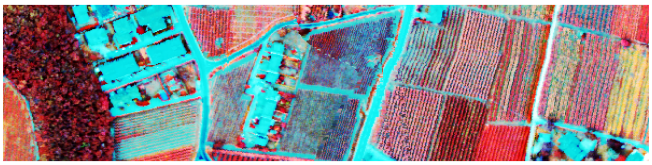
$$\mathbf{L}^{MI} = \mathbf{I} - \mathbf{D}^{MI^{-1/2}} \mathbf{W}^{MI} \mathbf{D}^{MI^{-1/2}} \tag{4}$$

where $\mathbf{I}$ is the identity matrix. $\mathbf{W}_l^{GK} = (w_{lij}^{GK})$ and $\mathbf{W}^{MI} = (w_{ij}^{MI})$ are the adjacency matrices of the graph $\mathcal{G}_l$, and $\mathbf{D}_l^{GK} = \text{diag}(\sum_{i\neq j} w_{lij}^{GK})$, and $\mathbf{D}^{MI} = \text{diag}(\sum_{i\neq j} w_{ij}^{MI})$ are their corresponding degree matrices, respectively.

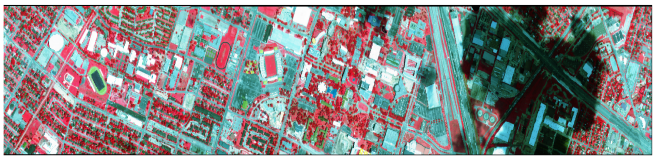
We embed the original set of bands into a lower-dimensional manifold using the *joint* null eigenvectors of $\mathbf{L}_l^{GK}$ and $\mathbf{L}^{MI}$, that are defined such as,

$$\mathbf{L}_l^{GK} = \mathbf{V}_l \mathbf{\Lambda}_l^{GK} \mathbf{V}_l^T, \tag{5}$$

$$\mathbf{L}^{MI} = \mathbf{V}_l \mathbf{\Lambda}_l^{MI} \mathbf{V}_l^T, \tag{6}$$



(a)



(b)

Fig. 1: False-color composite representation for both datsets: Trento RGB (a), Houston RGB (b).

where $\mathbf{\Lambda}_l^{GK} = \text{diag}(\lambda_{l1}^{GK}, ..., \lambda_{lK}^{GK})$, $\mathbf{\Lambda}_l^{MI} = \text{diag}(\lambda_{l1}^{MI}, ..., \lambda_{lK}^{MI})$ are diagonal matrices of the corresponding eignevalues, and $\mathbf{V}_l$ is the matrix of eigenvectors. $\mathbf{V}_l$ can be determined using joint diagonalization (JD) algorithms, which minimize a criteria of diagonality of $\mathbf{V}_l^T \mathbf{L}_l^{GK} \mathbf{V}_l$ and $\mathbf{V}_l^T \mathbf{L}^{MI} \mathbf{V}_l$. Different diagonalization constraints and distances can be used leading to a multitude of algorithms. In this work, we perform the joint diagonalization using the Quasi-Newton algorithm [20].

We stack the first N common eigenvectors of $\mathbf{L}_l^{GK}$ and $\mathbf{L}^{MI}$, $\mathbf{u}_{li}$ $(i = 1, \ldots, N)$, into one matrix $\mathbf{U}_l = [\mathbf{u}_{l1}, \ldots, \mathbf{u}_{lN}]$, where N is the number of bands to be selected. After the original set of attributes is embedded into a lower-dimensional manifold using the joint null eigenvectors of the Laplacian matrices [4], the partition of the embedding can be done using a clustering method: in this work, we chose a k-means [5] clustering algorithm. As a final step, we select the closest band to each cluster centroid as a representative one and form the set of relevant spectral bands that can be used for further analysis.

4. EXPERIMENTS AND RESULTS

The following section demonstrates the datasets description as well as experimental analysis and results, namely intercorrelation comparison with existing band selection algorithms.

Datasets: The first dataset was acquired from an agricultural area in the south part of the city of Trento, Italy, by the AISA Eagle sensor with 1 m spatial resolution and includes 63 bands ranging from 0.40 to 0.99 μm, where the spectral resolution is 9.2 nm.

The second dataset was acquired over the University of Houston campus and the neighboring urban area by the Compact Airborne Spectrographic Imager (CASI) with 2.5 m spatial resolution and was distributed for the 2013 IEEE GRSS Data Fusion Contest [21]. The hyperspectral dataset includes 144 spectral bands ranging from 0.38 to 1.05 μm.

Figure 1(a) and 1(b) illustrates the false-color composite representations of Trento and Houston hyperspectral datasets.

Intercorrelation Analysis: There are various criteria that can be defined to evaluate the relevance of selected bands, such as classification accuracy, error rate, precision, recall, F-measure, clustering accuracy [22]. However, the above-mentioned criteria are used for specific applications, for instance, classification, target detection, unmixing whilst band selection is applied as a preprocessing step. Nevertheless, the general aim of the band selection is to remove spectral redundancy and reduce computational costs, while preserving relevant information about the ROI, independently from the task under exam. Therefore, we decided to employ the Pearson intercorrelation factor over the selected bands as a common criterion independent of the subsequent application that could be used, so to directly evaluate the performance of the different band selection methods [3, 19]. Accordingly, in order to investigate the performance of the proposed scheme when dealing with different hyperspectral datasets, we focused our attention on exploring the ability of our method to extract for each sample a subset of spectral bands that would minimize the intercorrelation among bands, i.e., reduce the redundancy in the spectral channels and minimize their linear dependency. This step is a required operation for most of the hyperspectral remote sensing processing and analysis, as it leverages the problem of "curse of dimensionality", and helps in improving the efficiency of the investigation of hyperspectral records [1].

We report in Table 1 the Pearson intercorrelation factor obtained on the subset of bands selected by each scheme introduced in Section 2 [3, 19]. This factor measures the linear relationship between the variables, which makes it suitable for estimation of the relevance of selected bands. To evaluate intercorrelation we are extracting mean, μ, and variance, σ^2 from a cross-correlation matrix of selected bands. 100 experiments have been conducted for each method over both datasets. Various methods presented for comparison use an automatic selection of N, except GKMI where this parameter is fixed. However, N does not vary considerably among different methods, which does not make it a significant factor in our case. It is worth noting that some of the methods presented for comparison are relying on classic spectral clustering, based on k-means [23] and affinity propagation [12], and graph theory-based methods [13].

It is indeed worth recalling that a low value of intercorrelation is associated with a reduction of the redundancy and linear dependency between bands, hence signifying higher relevance of the considered bands. Thus, it is possible to appreciate that the method we use in this work, GKMI, is actually outperforming the other algorithms for hyperspectral band selection, as it is apparently able to maximize the significance of the selected spectral contributions with respect to the other schemes. It is also worth noting that the methods among the traditional schemes for hyperspectral band selection showing intercorrelation performances closer to GKMI are DKA and MDPP, which are intrinsically weighting the separability of the bands by means of a global coherence con-

Table 1: Performance comparison for band selection. Mean, μ, and variance, σ^2, of the intercorrelation between the selected bands of Trento and Houston hyperspectral datasets. N refers to the number of selected bands for which the intercorrelation is obtained.

	Trento			Houston		
Method	N	μ	σ^2	N	μ	σ^2
Original	63	0.36	0.33	144	0.78	0.04
COVSEL [7]	25	0.29	0.02	24	0.72	0.03
OSP [8]	21	0.32	0.01	23	0.74	0.01
LP [9]	22	0.31	0.01	22	0.74	0.02
PSO-MEAC [10]	22	0.27	0.01	22	0.70	0.03
MNBS [11]	24	0.29	0.03	23	0.71	0.02
K-means [23]	26	0.33	0.01	27	0.72	0.02
DKA [12]	25	0.23	0.01	25	0.62	0.01
MDPP [13]	24	0.24	0.02	24	0.61	0.01
SNMF-TEMD [14]	23	0.30	0.02	23	0.73	0.04
Spa-BS [15]	22	0.26	0.02	25	0.70	0.04
mRMR [16]	23	0.28	0.03	23	0.68	0.02
GKMI	20	**0.10**	0.08	20	**0.55**	0.01

straint (DKA), and implementing a spectral clustering to discover the subgraph structures in the high-dimensional band space (MDPP) [1, 12, 13]. These properties further emphasize the importance of measuring the relevance of the bands at local and global scales. Moreover, it highlights that conducting the investigation by considering Gaussian kernel and mutual information metrics simultaneously - one of the main contributions of this work - represents a key factor in identifying the most significant bands associated with each pixel in the given dataset.

5. CONCLUSIONS

In this paper, we employed an unsupervised GKMI band selection method that applies two different similarity measures and performed on a superpixel level.

The experimental results obtained when analyzing the considered hyperspectral datasets show the ability of GKMI to minimize the intercorrelation between the selected bands and therefore remove the redundant information, which is crucial for band selection methods. Moreover, the GKMI method shows the advantage of applying two similarity measures simultaneously, which is not used in classical graph-based clustering methods. It is especially important that the use of two metrics allows preserving more information about the original data, therefore selecting the most relevant bands subset.

Future works will be focused on the use of GKMI as a preprocessing step for hyperspectral image unmixing and classification, so to enhance the information extraction in hyperspectral data analysis. In this respect, the proposed method will be enhanced in order to take into account variability and sparsity of the data, hence improving the adaptivity of the analysis for various test cases.

6. REFERENCES

[1] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 7, no. 2, pp. 118–139, 2019.

[2] V. Kumar, J. Hahn, and A. M. Zoubir, "Band selection for hyperspectral images based on self-tuning spectral clustering," in *21st European Signal Processing Conference (EUSIPCO 2013)*, 2013, pp. 1–5.

[3] D Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing (Wiley Series in Remote Sensing and Image Processing)*, Wiley: Hoboken, USA, 2005.

[4] U. Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[5] S. Theodoridis and K. Koutroumbas, *Pattern Recognition, Fourth Edition*, Academic Press, Inc., Orlando, FL, USA, 4th edition, 2008.

[6] P. Bajcsy and P. Groves, "Methodology for hyperspectral band selection," *Photogramm. Eng. Remote Sens.*, vol. 70, pp. 793–802, 2004.

[7] J.-H. Kim, J. Kim, Y. Yang, S. Kim, and H. S. Kim, "Covariance-based band selection and its application to near-real-time hyperspectral target detection," *Opt. Eng.*, vol. 56, 2017.

[8] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, 1994.

[9] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, 2008.

[10] H. Su, Q. Du, G. Chen, and P. Du, "Optimized hyperspectral band selection using particle swarm optimization," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2659–2670, 2014.

[11] K. Sun, X. Geng, L. Ji, and Y. Lu, "A new band selection method for hyperspectral image based on data quality," *IEEE J. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2697–2703, 2014.

[12] J. Feng, L. Jiao, T. Sun, H. Liu, and X. Zhang, "Multiple kernel learning based on discriminative kernel clustering for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6516–6530, 2016.

[13] Y. Yuan, X. Zheng, and X. Lu, "Discovering diverse subset for unsupervised hyperspectral band selection," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 51–64, 2017.

[14] W. Sun, W. Li, J. Li, and Y. M. Lai, "Band selection using sparse nonnegative matrix factorization with the thresholded earth's mover distance for hyperspectral imagery classification," *Earth Sci. Informat.*, vol. 8, no. 4, pp. 907–918, 2015.

[15] S. Li and H. Qi, "Sparse representation based band selection for hyperspectral images," in *Proc. 18th IEEE Int. Conf. Image Processing (ICIP)*, Brussels, Belgium, Sept. 2011.

[16] Y. Jiang and C. Li, "mrmr-based feature selection for classification of cotton foreign matter using hyperspectral imaging," *Comput. Electron. Agriculture*, vol. 119, pp. 191–200, 2015.

[17] S. Beucher, "The Watershed Transformation Applied to Image Segmentation," *Proceedings of the 10th Pfefferkorn Conference on Signal and Image Processing in Microscopy and Microanalysis*, pp. 299–314, 1992.

[18] B.M. Romeny, *Front-End Vision and Multi-Scale Image Analysis: Multi-Scale Computer Vision Theory and Applications, Written in Mathematica*, Springer Publishing Company, Incorporated, 1st edition, 2009.

[19] T. M. Cover and J.A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*, Wiley-Interscience, USA, 2006.

[20] P. Ablin, J. Cardoso, and A. Gramfort, "Beyond pham's algorithm for joint diagonalization," *CoRR*, vol. abs/1811.11433, 2018.

[21] "2013 ieee grss data fusion contest. online. http://www.grss-ieee.org/community/technical-committees/data-fusion/2013-ieee-grss-data-fusion-contest/.," .

[22] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artificial Intelligence Review*, 2019.

[23] M. Ahmad, D. I. U. Haq, Q. Mushtaq, and M. Sohaib, "new statistical approach for band clustering and band selection using k-means clustering," *IACSIT Int. J. Eng. Technol.*, vol. 3, no. 6, pp. 606–614, 2011.