INTRODUCING CO-CLUSTERING FOR HYPERSPECTRAL IMAGE ANALYSIS

Emma Izquierdo-Verdiguier[1] *and Raúl Zurita-Milla*[1]

[1]Faculty of Geo-Information Science and Earth Observation (ITC). University of Twente, The Netherlands.
{e.izquierdoverdiguier,r.zurita-milla,}@utwente.nl, http://www.itc.nl/

ABSTRACT

This work introduces the use of *co-clustering* for hyperspectral image analysis. Co-clustering is able to simultaneously group samples (rows) and spectral bands (columns). This results in blocks, which do not only share spectral information (classical one way clustering) but also share sample information. Here, we propose using a co-clustering algorithm based on Information Theory -the optimal co-clustering is obtaining minimizing the loss of information between the original and the co-clustered images. A hyperspectral image (160000 samples and 40 bands) is used to illustrate this study. This image was clustered into 150 groups (50 groups of samples and 3 spectral groups). After that, blocks of the spectral groups was independently classified to assess the effectiveness of the co-clustering approach for hyperspectral band selection applications. Furthermore, the results were also compared with state-of-art methods based on morphological profiles, and the covariance matrix of the original hyperspectral image. Good results were achieved, showing the effectiveness of the Co-clustering approach for hyperspectral images in spatial-spectral classification and band selection applications.

***Index Terms*—** Co-clustering, hyperspectral images, k-means, Information Theory, band selection.

1. INTRODUCTION

Clustering methods are a fundamental tool in data analysis as they allow the exploration of complex datasets in an unsupervised fashion. Clustering groups the data according to their similarity. Different similarity criteria (or distance metrics) can be found in literature [1]. This is reflected in the vast amount of clustering literature. In this work, we focus on the use of information learning measures. In particular on the use of the *I-divergence*, a general case of the *Kullback-Leibler* divergence. Most clustering works are based on the use of Euclidean distance [2] and the use of Information Theory has, thus far, received less attention [3]. Clustering methods are often applied to assist with classification [4], regression [5] and feature extraction [6] tasks. For instance, clustering is used as a pre-processing step for Probabilistic Cluster Kernel applications [7]. Till now, most clustering applications are one way; this is, the similarity metric is computed along one dimension (samples or features). This can cause sub-optimal groupings because the clustering does not take into account the second dimension. *Co-clustering* algorithms fix this issue and consider both samples and features when performing the grouping of the data. In other words, co-clustering is able to find blocks of similar data in a matrix by simultaneously considering information along the rows and columns. Nowadays *Co-clustering* is increasingly used in several disciplines: seeks local pattern of genes expressions [8] or the relation between texts and documents [9]. However, to the best of our knowledge no co-clustering algorithms have been used for remote sensing applications despite the increasing interest in studying spectral and spatial relationships. This paper presents a first application of *Co-clustering* to analyze hyperspectral images.

The rest of the paper is outlined as follows. Section 2 reviews the co-cluster methodology. Section 3 presents the dataset used, the experimental results and its discussion. Finally, Section 4 concludes this paper.

2. I DIVERGENCE CO-CLUSTERING ALGORITHM

Co-clustering, also called Bi-clustering is a clustering method that is composed of generating block of rows and columns of the data taking into account the value of approximation error which is measure by a loss function. In the case of the Information Theory Co-Cluster (ITCC) the loss function is the *I-divergence* function [10]. The method consists of minimizing the distance between a matrix approximation ($\hat{X} \in \mathcal{R}^{m\times n}$) from the original data matrix ($X \in \mathcal{R}^{m\times n}$). The matrix approximation is obtained:

$$\hat{X} = \frac{E[X|\hat{U},\hat{V}] \times E[X|U] \times E[X|V]}{E[X|\hat{U}] \times E[X|\hat{V}]}, \tag{1}$$

where $E[\ ]$ is the expected value, U and V are the random initialization groups and $\hat{U}$ and $\hat{V}$ are the groups after co-clustering update. To obtain the approximation matrix, one must first row cluster clusters assign and after that assign the column clusters. Both steps are realized by means of decom-

Algorithm 1 Information Theory Co-clustering algorithm

Require: $X \in \mathcal{R}^{m \times n}$: data matrix, k: num. of row clusters, l: num. of column clusters.

Ensure: $R \in \{0,1\}^{m \times k}$: Row cluster membership matrix $C \in \{0,1\}^{n \times l}$: Column cluster membership matrix.

1-R and C matrix $\leftarrow$ Randomly initialization

2-Calculate $d_r = (\hat{X}_{const}) * \hat{X}'_{variable} - (X) * log(\hat{X}'_{variable})$ using eq. 3

3-Update row clusters: $R^* = \arg\min d_r$

4-Calculate $d_c = (\hat{X}_{const}) * \hat{X}'_{variable} - (X) * log(\hat{X}'_{variable})$ using eq. 4

5-Update column clusters: $C^* = \arg\min d_c$

6-Update data matrix eq. 1

posing the reconstructed matrix in a product of matrix:

$$\hat{X} = \hat{X}_{const} \times \hat{X}_{variable}, \tag{2}$$

where $\hat{X}_{const}$ is a constant factor to update rows (columns) and $\hat{X}_{variable}$ depends on grouping of the rows (columns), where in the row case:

$$\hat{X}_{const} = \frac{E[X|U] \times E[X|V]}{E[X|\hat{U}]}, \hat{X}_{variable} = \frac{E[X|\hat{U}V]}{E[X|\hat{V}]} \tag{3}$$

whereas column case:

$$\hat{X}_{const} = \frac{E[X|U] \times E[X|V]}{E[X|\hat{V}]}, \hat{X}_{variable} = \frac{E[X|U\hat{V}]}{E[X|\hat{U}]} \tag{4}$$

The application of this methods requires the calculation of the joint probability distribution between the rows and columns of the input matrix or that the data is treated as a co-occurrence matrix. The latter case is assumed here by building a matrix whose rows represent pixels and whose columns are the spectral bands. This matrix is co-clustered using not only the information of the block to update the co-clusters, but also the information of the complete rows/columns of the block.For more information of the ITCC algorithm, see [10] and Algorithm 1.

3. EXPERIMENTAL RESULTS

3.1. Data

The image selected to apply the *co-clustering* was acquired by the DAIS7915 sensor which has four spectrometers used 80 spectral bands over the city of Pavia (Italy). Taking into account the previous studies of the image [11] we have used 40 collected by two of four spectrometers. One of them used 32 bands whose spectral range is between $[496, 1035]$ nm with a spectral resolution of 17 nm and the second used 8 bands in the range $1539 - 1756$ nm with a spectral resolution of 27 nm.

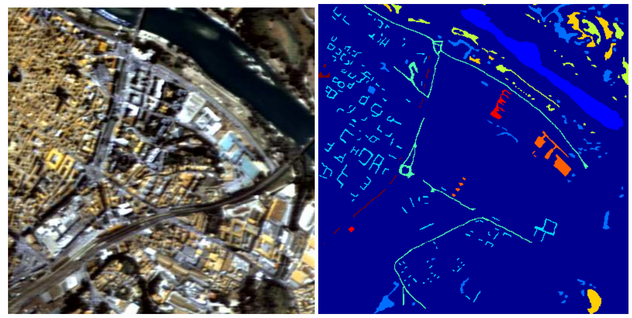

Fig. 1. RGB composite (left) and ground truth (right) of Pavia image.

The Pavia image has 5 meters of spatial resolution and contains 9 urban classes: water, trees, asphalt, parking, bitumen, roofs, meadows, soil and shadows. Figure 1 shows the Pavia RGB composite and the ground truth map.

3.2. Methodology

We applied ITCC method to Pavia image using 50 row clusters and 5 column clusters. The spectral clusters were independently used to obtain classification maps by means of K-means clustering. We compared the classification results obtained by co-clustering with spectral spatial classification methods based on morphological profiles. For this, we extracted the Extended Morphological Profile (EMP) of the three first principal components of the original hyperspectral data [12]. Next, we applied K-means clustering to original data (Spectral), the EMP and the combination of the both (EMP-spectral). The accuracy of these classifications was assessed with the available ground truth (7456 samples). We also compared the co-cluster result with the covariance matrix. Note that the covariance matrix provides the variance (diagonal elements) or covariance (off-diagonal elements) for the bands of the data, i.e. a dispersion measure between bands and as such, the covariance matrix shows the relation between bands.

3.3. Results and discussion

Figure 2[left] shows the data matrix ordered by the groups obtained by the co-clustering using $k = 50$ and $l = 5$. In the column case, first group contains the Visual and a fraction of Near Infrared ($1539 - 1756$ nm) ranges whereas the other Near infrared fraction ($717 - 1035$ nm) has been divided in two groups (Fig. 2[right]). Noteworthy, the first group joins the spectral ranges that are distant in terms of *Euclidian distance* but that, nonetheless have a similar information content. Also important to notice that the co-clustering yielded 3 instead of 5 spectral groups. This is because the algorithm self-tunes the number of groups according to the information available in the image.

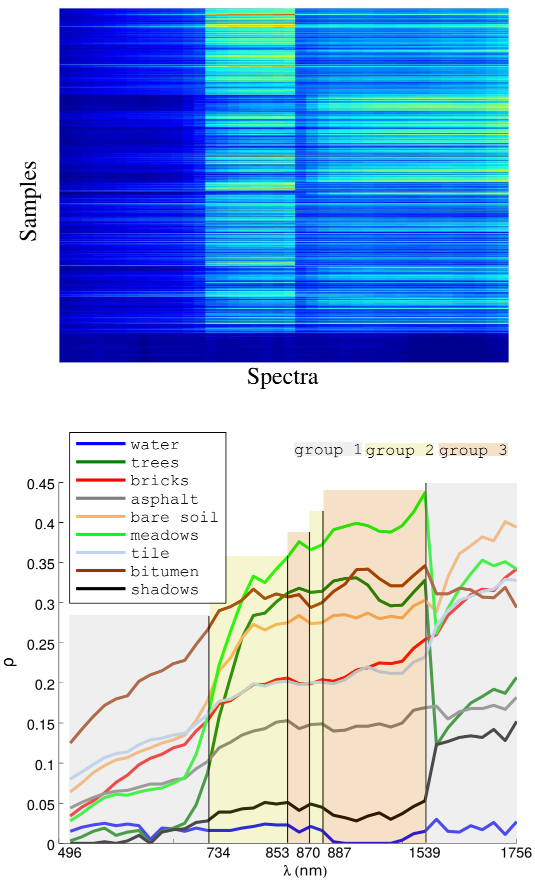

Fig. 2. Co-clustering results of Pavia image with $k = 50$ and $l = 5$ (top) and the spectrum of 9 classes of Pavia image along the column groups obtained by the co-cluster (bottom).

Figure 3 show the covariance matrix of the original data in which it is possible distinguish 3 groups as in the co-clustering case. The covariance matrix, in contrast to co-clustering, ignores the spatial information of the original data generating different groupings. Note that the groups generated by the covariance matrix are those expected by means of Euclidean distance. This is due to the relation between covariance and Euclidean distance.

To benchmark the classification and band selection skills of our spatial-spectral co-clustering algorithm, we compared our results with that obtained by the EMP and its combination with the spectral data. We applied a K-means clustering using 9 clusters to the original data matrix (spectral case), EMP, the combination of EMP and spectral data, and the data matrix using the bands join by the column cluster results (Fig. 2[right]). The classification maps are shown in Figure 4 along with the average of Overall accuracy (OA) and kappa index value of 10 realizations for all experiments. To assess the cluster method [7], we took into account the label data available of the Pavia image. In this experiment, the proposed method improves not only the accuracy but also the kappa value index obtained by the previous methods (spectral, EMP and EMP-spectral) and the others groups of bands obtained by co-clustering method. The results suggest that the co-clustering method and the distance based on Information Theory constitute a alternative to standard spatial-spectral combinations methods and distance measures based on Euclidean distances.

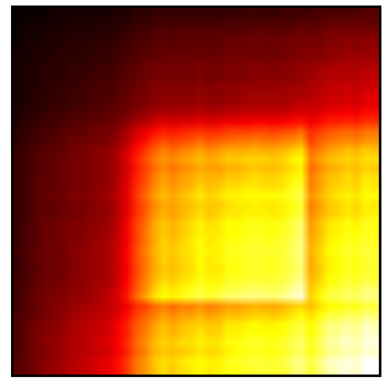

Fig. 3. Covariance matrix from the Pavia image.

4. CONCLUSIONS

This paper presents the use of co-clustering methods to simultaneously exploit the spectral and spatial information present in hyperspectral images. The two main objectives of this work were to study: 1) the possibilities of co-clustering to evaluate informative blocks with the best spatial-spectral information; 2) whether the spectral clusters obtained by co-clustering are a good option to reduce the dimensionality (by means of band selection). Good results were obtained in the classification task by means of K-means clustering, where the combination of Visible and the high Near Infrared wavelengths obtained by co-clustering method performs better than previous spectral-spatial combination methods based on morphological profiles. This suggests that our information theory based co-clustering is a good tool to select bands for classification tasks in hyperspectral images.

5. REFERENCES

[1] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. PAMI-1, no. 2, pp. 224–227, April 1979.

[2] K. K. Singh, M. J. Nigam, K. Pal, and A. Mehrotra, "A fuzzy kohonen local information c-means clustering for remote sensing imagery," *IETE Technical Review*, vol. 31, no. 1, pp. 75–81, 2014.

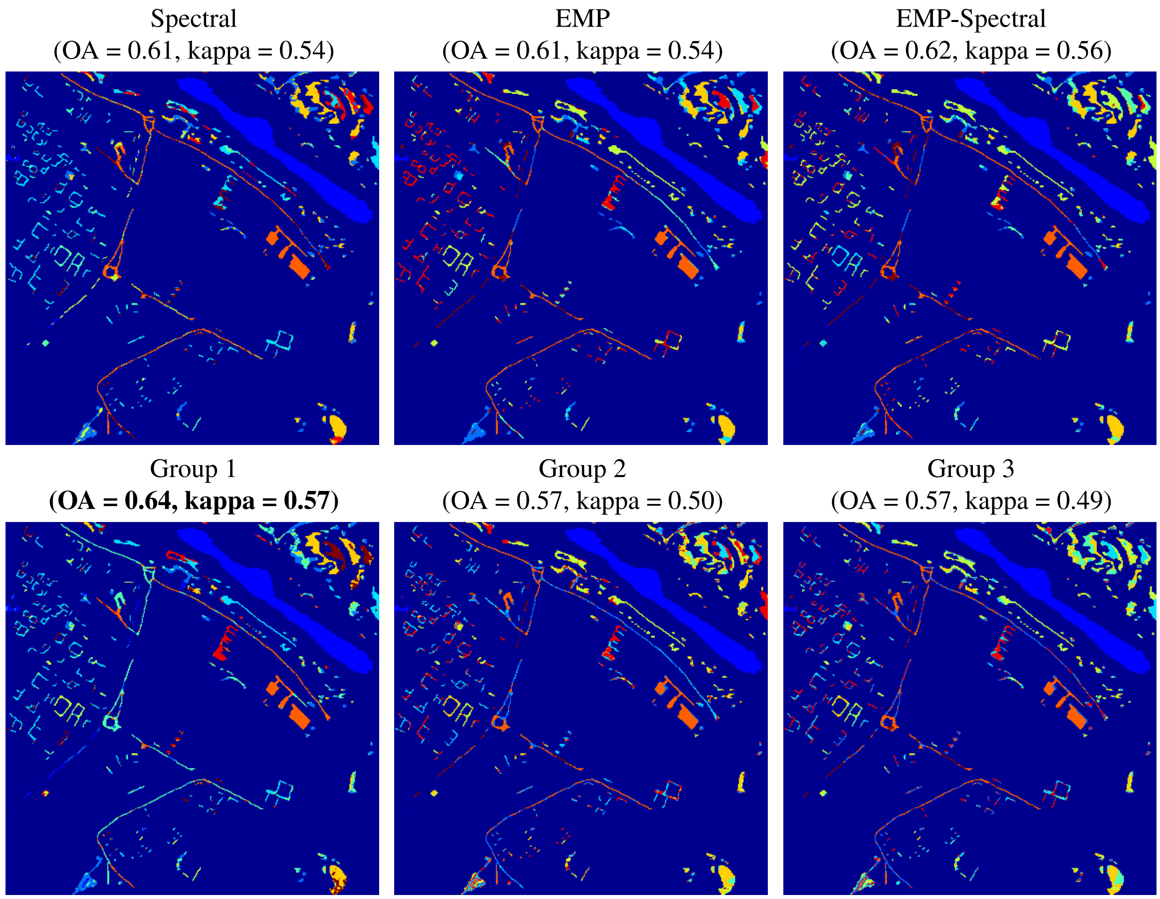

Fig. 4. Classification maps along with overall accuracy (OA) and Kappa index (kappa) for the *Pavia* image: Spectral, EMP, and Spectral and EMP methods (Top). Bands assigned to groups 1, 2 and 3 by the co-clustering method (bottom).

[3] N Slonim, G. S. Atwal, G. Tkaĉik, and W. Bialek, "Information-based clustering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 51, pp. 18297–18302, 2005.

[4] X. Yu, H. He, D. Hu, and W. Zhou, "Land cover classification of remote sensing imagery based on interval-valued data fuzzy c-means algorithm," *Science China Earth Sciences*, vol. 57, no. 6, pp. 1306–1313, 2014.

[5] B.P. Shukla, C.M. Kishtawal, and P.K. Pal, "Prediction of satellite image sequence for weather nowcasting using cluster-based spatiotemporal regression," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 7, pp. 4155–4160, July 2014.

[6] Q. Lu, X. Huang, and L. Zhang, "A novel clustering-based feature representation for the classification of hyperspectral imagery," *Remote Sensing*, vol. 6, no. 6, pp. 5732–5753, 2014.

[7] E. Izquierdo-Verdiguier, R. Jenssen, L. Gómez-Chova, and G. Camps-Valls, "Spectral clustering with the probabilistic cluster kernel," *Neurocomp*, vol. 149, Part C, pp. 1299 – 1304, 2015.

[8] A. Thévenin, L. Ein-Dor, M. Ozery-Flato, and R. Shamir, "Functional gene groups are concentrated within chromosomes, among chromosomes and in the nuclear space of the human genome," *Nucleic acids research*, p. gku667, 2014.

[9] I. S. Dhillon, S. Mallela, and D. S. Modha, "Information-theoretic co-clustering," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 89–98.

[10] A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. S Modha, "A generalized maximum entropy approach to bregman co-clustering and matrix approximation," *Journal of Machine Learning Research*, vol. 8, pp. 1919–1986, August 2007.

[11] C. Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*, John Wiley & Sons, 2013.

[12] M. Fauvel, J.A. Benediktsson, J. Chanussot, and J.R. Sveinsson, "Spectral and spatial classification of hyperspectral data using svms and morphological profiles," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 46, no. 11, pp. 3804–3814, Nov 2008.