HAL
open science

Unsupervised processing of hyperspectral images

Benoît Jaillais, Karima Meghar, Ludivine Nus, Sebastian Miron, David Brie, Saïd Moussaoui

To cite this version:

Benoît Jaillais, Karima Meghar, Ludivine Nus, Sebastian Miron, David Brie, et al.. Unsupervised processing of hyperspectral images. CHIMIOMETRIE XIX, Jan 2018, Paris France. hal-01774088

HAL Id: hal-01774088

https://hal.science/hal-01774088

Submitted on 23 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Unsupervised processing of hyperspectral images

B. Jaillais[1] K. Meghar[1] L. Nus[2] S. Miron[2] D. Brie[2] S Moussaoui[3]

[1] StatSC Unité de Statistique, Sensométrie, Chimiométrie, F44300 Nantes. Benoit.Jaillais@inra.fr

[2] Université de Lorraine, CRAN, UMR 7039, F54506 Vandœuvre-lès-Nancy. ludivine.nus@univ-lorraine.fr, sebastian.miron@univ-lorraine.fr, david.brie@univ-lorraine.fr

[3] Ecole Centrale Nantes, LS2N, F-44321 Nantes, said.moussaoui@ec-nantes.fr

Keywords: Wood sorting, Near infrared hyperspectral imaging (NIR-HIS), k-means, on-line NMF.

1 Introduction

This work is a part of the CNRS project "ALOHA: Analyse en Ligne de dOnnées Hyperspectrales pour l'industrie Agroalimentaire" and of the ANR-OPTIFIN (Agence Nationale de la Recherche-OPTImisation des FINitions). The aim of these projects is to develop analytical tools adapted to the high throughput online analysis of samples by acquisition and processing of hyperspectral images. One output of the ALOHA and ANR OPTIFIN projects consists in the development of sequential algorithms for the deconvolution and on-the-fly unmixing of hyperspectral data. The main goal is to be able to predict and classify the quality of wood pieces renderings.

However, the first step is to compare results obtained by usual chemometrics such as Principal Components Analysis (PCA) and K-means clustering analysis to *on-line* Non-negative Matrix Factorization (NMF). In this preliminary work, these algorithms were applied to NIR hyperspectral images of wood.

2 Material and methods

The hyperspectral images of wood samples were acquired by a pushbroom imager in a near infrared spectral range between 1000 and 2500 nm. Each hyperspectral image obtained has the dimensions *288*384*1200* where *288* characterize the number of wavelengths and 384*1200 the spatial dimensions. The acquisition is performed in a sequential "slice by slice" manner by moving the imager above the object. One "slice" corresponds to a 288*384 matrix.

Principal Components Analysis (PCA) and K-means clustering are two unsupervised algorithms which are usually applied on matrices to gather individuals according to the distance between their spectra. By dealing with hyperspectral image, it is possible to build a model with a selected number of pixels and then to project all the pixels of an image, leading to a false-colors image [1]. In case of a series of images, models were built with a selection of the different images and applied to the entire collection of images, which makes it possible to compare images.

On-line NMF is an adaptive version of the classical NMF [2]. For a non-negative matrix $\boldsymbol{X}$, the NMF consists in finding two non-negative matrices $\mathbf{S}$ and $\mathbf{A}$ such that $\mathbf{X} \approx \mathbf{SA}$. In hyperspectral imagery, the columns of $\mathbf{X}$ represent the data samples recorded at different wavelengths. $\mathbf{S}$ is a matrix containing the normalized source spectra (the endmembers) and $\mathbf{A}$ is a matrix containing on its columns the mixing coefficients (abundances) for the recorded samples. The goal of *on-line* NMF methods is to sequentially update in real-time the sources and the abundances for each new acquired sample. We propose an on-line source separation algorithm that comply with the acquisition rate of the pushbroom imager, inspired by the Incremental Non-negative Matrix

Factorization algorithm (INMF) [3]. We adapted the method to the pushbroom data acquisition scheme and integrated a minimum volume constraint on the sources to restrict the domain of feasible solutions [4].

The images were pre-processed by smoothing and standard normal variate before applying PCA, k-means, and NMF algorithms with Matlab®. The number of groups for the k-means was fixed to five and the number of spectral signatures to extract in the *on-line* NMF algorithm was set to three.

3 Results and discussion

A selection of ten samples with different proportions of sapwood was analyzed. Figure 1 shows the results obtained by PCA (left part) and k-means methods (right part) and for each of the analyzed faces. Figure 2 shows one of the abundance maps obtained by on-line NMF with minimum volume constraint on the sources, following the treatment of the hyperspectral image, for each of the analyzed faces.

All the three methods are able to distinguish different tissues of wood which is not possible by a visual inspection. Despite the difference between the methodological concepts underlying each of the three methods, the results are highly similar between all the three, which validates the approach.

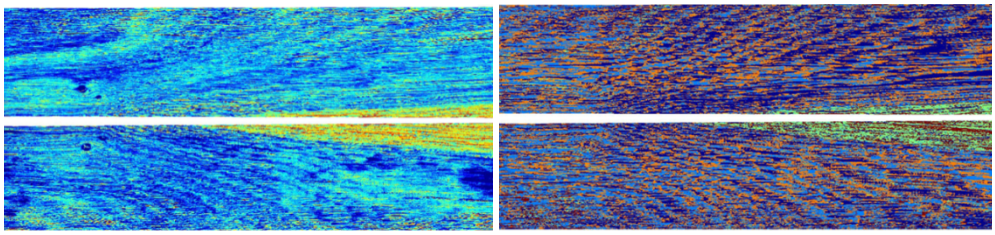

Figure 1: PC1 scores images (left) and k-means images (right) for both faces of the same sample.

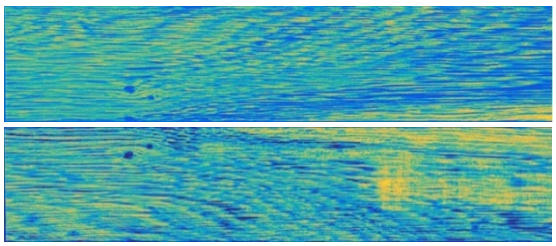

Figure 2: Abundance maps for both faces of the same sample.

4 Conclusion

Combination of near infrared hyperspectral imaging with chemometrics makes it possible to highlight the different tissues of wood, according to their spectral signatures. PC1 Scores, k-means images and abundance maps of on-line NMF demonstrated that the use of unsupervised algorithms is relevant for the purpose of sorting samples. Moreover, the *on-line* NMF approach allows a real-time classification directly on the production line.

5 References

[1] Amigo JM, Babamoradi H, Elcoroaristizabal S. Hyperspectral image analysis. A tutorial. *Anal Chim Acta*. 2015 896:34-51. doi: 10.1016/j.aca.2015.09.030.

[2] D. D. Lee, H. S. Seung. Algorithms for non-negative matrix factorization. In Advances in neural information processing systems, 556–562, 2001.

[3] S. S. Bucak, B. Gunsel. Incremental subspace learning via non-negative matrix factorization. Pattern recognition, 42(5):788–797, 2009.

[4] L. Nus, S. Miron, D. Brie. On-line source separation of hyperspectral image by Non-negative Matrix Factorization, SSP 2017 (to be submitted).