

Fast Segmentation of Hyperspectral Images by Combining Textural and Spectral Information

UWE KNAUER¹, DAVID KILIAS¹ & UDO SEIFFERT¹

Abstract: In this paper different approaches for dimensionality reduction of hyperspectral images are compared in order to apply a texture based segmentation algorithm which efficiently operates in low dimensions. Detection of Powdery Mildew infection is used as a real-world test case for the proposed approach. For this application, an adaptive Principal Component Analysis based projection preserves the most relevant spectral information within the combined process of spatial-spectral feature extraction.

1 Introduction

A rich and meaningful feature space is one of the key components of a successful classification in remote sensing applications. In hyperspectral remote sensing a rich but highly correlated feature space (given by the typically large number of observed wavelengths) favors pixel-wise segmentation based on spectral data. However, also the texture of surfaces should be considered as an additional source of information leading to a spectral-spatial classification scheme.

We adapt a recently proposed fast segmentation algorithm for RGB images to the segmentation of hyperspectral images. First, certain discriminating spectral bands are identified and second a vast number of spatial features per pixel are computed within the selected bands. A major benefit of the approach is the use of integral images and compatible basic features for fast computation of a much higher number of derived features. The classification step is based on modified Random Forest classifiers which use the integral images as a cache memory for fast lookup of feature values, if and only if they are required in the decision process. This use of features on demand results in a speed advantage compared to other classification methods such as support vector machines which have to incorporate all dimensions of the feature space into the decision.

We present and compare three approaches of different complexity for the selection of discriminating spectral bands. The first is dimensionality reduction by Principal Component Analysis, which is the most computational expensive method, because projection of all spectra is required. The second and simplest approach is the selection of commonly used default bands. Adaptively selecting spectral bands from images with high spectral resolution requires preprocessing based on training data, but does not require any additional processing time for segmentation of images.

We benchmark the implementation of the algorithm for different parameterizations of the classifier (number of trees, number of used features) on hyperspectral images of Powdery Mildew infection.

Segmentation results are compared to pixel-wise classification based on spectral information only to address the accuracy-vs-speed-tradeoff between reduction of spectral information (lowering the dimension) and incorporation of textural information (increasing the dimension).

¹ Fraunhofer IFF, Sandtorstr. 22, 39106 Magdeburg; E-Mail: [knauer, kilias, seiffert]@iff.fraunhofer.de

2 Related Work

In XIA (2014) Rotation Forests has been proposed as a method for classification of hyperspectral data. This approach extends Random Forests by introducing an individual transformation of the input feature space for each decision tree of the ensemble. This contributes to the diversity of the ensemble decisions and the authors report a significant gain in classification accuracy.

In VEGANZONES (2014) the coupling of partition trees with pruning based on spectral-unmixing techniques has been proposed. The authors show that the reconstruction error of hyperspectral images can be reduced and meaningful segmentations of hyperspectral remote-sensing data are obtained.

In REN (2014) spectral and spatial features are extracted in parallel. Then, a so-called hybrid feature vector is created and used for training of a Random Forest classifier. Finally, results are improved by imposing a label constraint which is based on majority voting.

FASSNACHT (2014) provides a survey on the performance of different feature selection approaches and classifiers for tree species classification from hyperspectral data obtained at different locations and with different sensors. The authors conclude that the selection of 15-20 bands provides the best classification results. They found, that the location of the selected bands strongly depends on the classification method used, which is conform to the general interdependence between classification method and selected features. However, best classification results for all forest datasets have been obtained with Minimum Noise Fraction Transformation (MNF) and selection of the first 10-20 principal components of MNF as input features for classification.

The article of CAMPS-VALLS (2014) reviews recent developments in hyperspectral image classification. The authors present a statistical learning theory (SLT) based framework for analysis of hyperspectral data. They highlight the ability of SLT to identify relevant feature subspaces to enable the application of more efficient algorithms. The review categorizes existing spatial-spectral classification approaches into spatial filters extraction, spatial-spectral segmentation, and advanced spatial-spectral classification.

In AMINI (2014) the segmentation performance of a conventional Random Forest classifier is improved by a semi-supervised sampling approach for classifier training. In an initial procedure a supervised classifier is trained with a labeled dataset. Next, the whole hyperspectral image is classified, class probabilities are computed, and crisp labels are assigned to all pixels. Finally, a new classifier is trained with manually and automatically labeled data.

In ABDEL-RAHMAN (2014) use of SVM and Random Forest for classification of pine trees is reported. An important aspect is the utilization of Random Forest variable importance to identify the most relevant wavelength bands. Importance is based on out of bag error and measures the average loss of accuracy when a single variable is not used.

Classification of tree species in boreal forests was investigated in DALPONTE (2013). Experiments include dimensionality reduction of hyperspectral data. The authors conclude that identifying the most relevant wavelength bands prior to classification yields similar results to classification based on the complete spectral data. These findings indicate that feature reduction is possible without significant loss of accuracy.

3 Proposed Algorithm

3.1 Dimensionality reduction

The proposed segmentation algorithm consists of three major processing steps. The first step is dimensionality reduction. The hyperspectral image with typically more than 100 channels is converted into a 3-channel image. On one hand, compatibility with an existing image segmentation framework is an important reason for the dimensionality reduction. On the other hand, the computational costs of pixel-wise classification of unreduced hyperspectral images or a channel-wise processing makes a number of approaches not suitable for applications where real-time segmentation of the images is required.

Hence, the most important requirement of dimensionality reduction is to maintain as much of the discriminative power of a high spectral resolution image as needed for a given segmentation task. So far, we implemented and tested the following options:

- Canonical band selection (inspired by human perception or existing applications),
- Relevance based band selection based on importance histograms,
- Synthesis of orthogonal bands based on Principal Component Analysis (PCA),
- Target class specific synthesis based on adapted data sampling before PCA.

For canonical band selection the same image bands as used by the software PARGE are selected. For VNIR cameras such as NEO Hypspx VNIR 1600 or SPECIM aisaEagle, the red-channel of the resulting RGB-image is mapped to the 651 nm band, the green-channel to 549 nm, and the blue-channel to 440 nm. Another option for canonical band selection is close infrared (CIR), where the 3 channels are mapped to the 811 nm, 640 nm, and 498 nm bands, respectively. In the short-wave infrared, the following mapping is used: $\{R,G,B\} \rightarrow \{1,081 \text{ nm}, 1,652 \text{ nm}, 2,253 \text{ nm}\}$.

The relevance based band selection is based on supervised pixel-wise classification of spectral information with Random Forest classifiers. During the induction of a decision tree, many local optimizations take place for optimal feature selection. Hence for each classification, the visited tree nodes are checked for which feature (band) was used to create a histogram of band importance. Finally, the three highest ranked bands are selected.

Principal Component Analysis is used to derive a new orthogonal base of the original feature space. The resulting bands are linear combinations of all original bands. Using PCA for dimensionality reduction has a long history, because most of the original data can be reconstructed using only a few projections. Random subsets of spectra are used to calculate the projection matrices. For target class specific PCA the input spectra are sampled from predefined pixels only. Optionally, a radiometric calibration is applied before dimensionality reduction.

3.2 Spatial features

The use of integral images is a crucial component for fast segmentation. Integral images allow extraction of spatial features at different scales in a constant amount of time. Mean intensity, standard deviation of intensity values, and homogeneity within arbitrary upright rectangular image regions are used as base features. Features are extracted at different scales and for each

image channel separately. As a default and for all presented tests, a 225-dimensional feature vector per pixel is used (25 different regions * 3 base features * 3 channels).

3.3 Coupled classification and feature extraction

Fast pixel-wise segmentation based on the proposed high-dimensional feature space requires a tight coupling of classification and feature extraction. The common processing scheme, where all features are extracted before the resulting feature vector is processed, is not suitable.

Instead, the definitions of all features are linked to corresponding nodes of the decision trees. This allows for computation of the feature values on demand. The precomputed integral images serve as a feature cache memory, since any feature value can be looked up in constant time by accessing only the four corner points of the rectangular region of interest.

This ‘features on demand’ approach is a modification of the idea of a cascade of Haarlike features proposed in VIOLA (2001). It has been successfully applied to fusion of neural network outputs (KNAUER 2014a) as well as for variable selection in multiple classifier fusion for reduction of the computational load (KNAUER 2014b). Its application for a pixel-wise operating image segmentation framework for 3-channel images has been proposed in KNAUER (2014c) and is here extended to differently preprocessed hyperspectral images.

3.4 Boosting of classifiers

Changes of lighting conditions, different perspectives, and object variability lead to variation of the underlying feature distributions. Therefore, in our image segmentation framework a second classification layer is added to improve the results. While Random Forests are typical examples of bagging classifiers, boosting is another beneficial technique to iteratively adapt classifiers to difficult cases. Since the training of the initial Random Forest classifier is based on a subset of randomly chosen pixels, the second classification layer is trained by sampling based on the errors made by segmentation of the complete images of the training dataset.

4 Dataset

For presentation and evaluation of the proposed segmentation algorithm, a dataset from a recent hyperspectral imaging campaign has been selected. The underlying problem is to establish a reliable detection of powdery mildew disease for quality inspection in wine production.

The dataset consists of recordings of infected and healthy berries. In addition, images from two sides of 30 complete bunches have been recorded. These bunches have been selected to represent three levels of infection with powdery mildew: healthy control, light infection, and heavily infected. Hyperspectral images have been recorded at the Wine Institute of Adelaide University, Waite Campus. NEO Hypspx VNIR 1600 and SWIR 320m-e have been used for simultaneous acquisition of visible, near-infrared, and shortwave-infrared bands between 400 nm and 2,500 nm. Objects have been moved beneath these line scanning cameras by a translation stage to create a hyperspectral image. Acquisition was performed under controlled illumination in a darkened room.

5 Experiments and Results

First, classification performance based on spectral information is investigated. These results provide a baseline for the performance of the proposed approach which operates in a drastically reduced spectral dimension. For classifier training we manually selected a number of pixels, where powdery mildew infection is clearly visible. 10-fold cross-validation is used to predict the classification accuracy for separating these pure pixels. Tab. 1 lists the results.

Tab. 1: Average accuracy of pixel-wise classification for different feature spaces

| # of samples | Spectral | PCA | Adaptive PCA | Adaptive PCA 2 | Selection | RGB | CIR |
|--------------|----------------------------------|--------------------|----------------------------------|--------------------|----------------------------------|--------------------|-------------------|
| 1,000 | 0.984 +/- 0.001 | 0.590 +/- 0.033 | 0.960 +/- 0.015 | 0.570 +/- 0.038 | 0.770 +/- 0.047 | 0.770 +/- 0.036 | 0.710 +/- 0,04 |

To overcome limits of classification accuracy after dimensionality reduction of the input image data, a number of additional spatial features are introduced (see section 3.2). While increasing dimensionality after reduction of the spectral dimension may look confusing, it offers several benefits. First, using texture features can add a different source of information and second, memory demands are slightly reduced due to the use of only a few integral images compared to the number of original image bands. Third, depending on the particularly used method, dimensionality reduction can offer insights into how to design an adapted imaging device (e.g. by identifying relevant bands). Adaptive PCA and Adaptive PCA2 refer to two different results of the same dimensionality reduction algorithm, which provided very different results due to random sampling. Tab. 2 lists the results of 10-fold cross-validation for the spatial-spectral dataset. Results indicate that the chosen spatial features can compensate the loss of information by dimensionality reduction.

Tab. 2: Average accuracy of spatial-spectral classification compared to spectral classification

| # of samples | Spectral | PCA | Adaptive PCA | Adaptive PCA 2 | Selection | RGB | CIR |
|--------------|----------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| 100 | 0.950 +/- 0.050 | 0.920 +/- 0.040 | 0.985 +/- 0.032 | 0.975 +/- 0.030 | 0.965 +/- 0.050 | 0.950 +/- 0.035 | 0.960 +/- 0.037 |
| 1,000 | 0.984 +/- 0.001 | 0.967 +/- 0.001 | 0.995 +/- 0.005 | 0.984 +/- 0.006 | 0.993 +/- 0.006 | 0.984 +/- 0.006 | 0.989 +/- 0.010 |
| 10,000 | 0.996 +/- 0.001 | 0.994 +/- 0.001 | 1.000 +/- 0.000 | 0.998 +/- 0.001 | 1.000 +/- 0.001 | 0.998 +/- 0.001 | 0.999 +/- 0.000 |

Next, the most relevant features within the spectral-spatial representation are searched. A selection scheme similar to relevance based band selection in section 3.1 is used, creating a compact representation of both spectral and spatial features. Tab. 3 lists the results.

Tab. 3: Classification accuracy for different numbers of used spatial features

| Transformation | Number of selected features | | | | | | | | |
|----------------|-----------------------------|--------------|--------------|--------------|-------|-------|-------|-------|-------|
| | 3 | 5 | 10 | 15 | 20 | 25 | 30 | 35 | 40 |
| Adaptive PCA 2 | 0.500 | 0.870 | 0.940 | 0.980 | 0.970 | 0.990 | 0.996 | 0.996 | 0.997 |
| PCA | 0.873 | 0.944 | 0.975 | 0.994 | 0.994 | 0.997 | 0.997 | 0.997 | 0.997 |
| RGB | 0.919 | 0.932 | 0.985 | 0.996 | 0.995 | 0.999 | 0.998 | 0.997 | 0.997 |
| CIR | 0.910 | 0.965 | 0.998 | 0.999 | 0.999 | 0.999 | 0.998 | 0.999 | 0.998 |
| Adaptive PCA | 0.84 | 0.978 | 0.998 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| Selection | 0.972 | 0.980 | 0.992 | 0.990 | 0.990 | 1.000 | 0.999 | 1.000 | 1.000 |

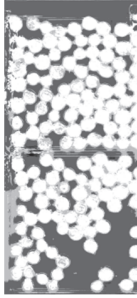
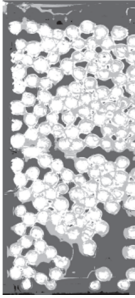
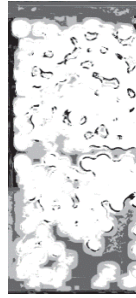
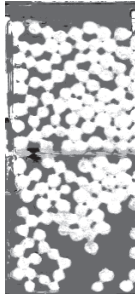

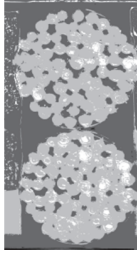
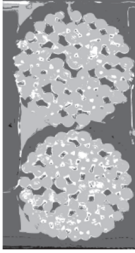
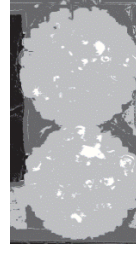
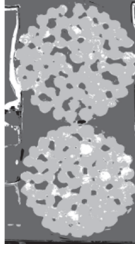
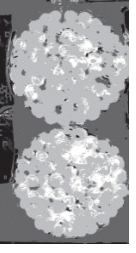
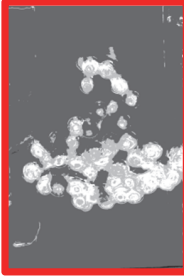
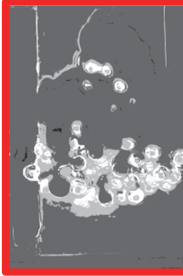


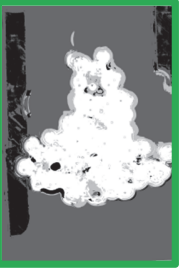
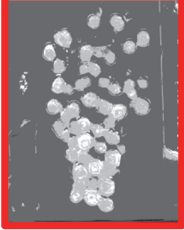
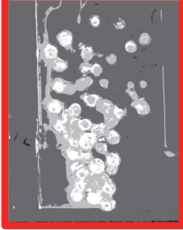

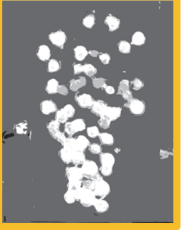
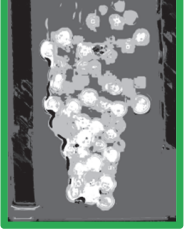




The results show that an adaptive band selections as well as an adaptive PCA preserve more of the spectral information. Hence, a lower number of spatial features is required to achieve the same classification accuracy. The presented results (Tab. 1 - 3) have been obtained with small Random Forests of 10 trees. In contrast to the default values of (Breiman, 2001) a sparse set of 10 random features is used at each tree node during classifier training. For completeness, results for different ensemble sizes are given in Tab. 4. For the given problem, only minor improvements can be achieved by increasing the ensemble size.

Tab. 4: Classification accuracy for different ensemble sizes

| # Trees | RGB | Selection | Adaptive PCA |
|---------|-----------------|-----------------|-----------------|
| 1 | 0.948 +/- 0.009 | 0.977 +/- 0.009 | 0.988 +/- 0.008 |
| 3 | 0.973 +/- 0.010 | 0.991 +/- 0.005 | 0.995 +/- 0.005 |
| 5 | 0.982 +/- 0.011 | 0.992 +/- 0.006 | 0.996 +/- 0.004 |
| 11 | 0.984 +/- 0.009 | 0.995 +/- 0.005 | 0.996 +/- 0.004 |
| 51 | 0.984 +/- 0.006 | 0.994 +/- 0.004 | 0.997 +/- 0.005 |

So far, training and testing datasets were based on random selection of predefined pixels acquired under optimal conditions (single berries, no bunches). Analysis of bunches is more difficult, because berries may be partially occluded and cast shadows. Tab. 5 shows a selection of segmentation results including segmentation of bunch images.

Tab. 5: Segmentation results

| | RGB | PCA | Adaptive PCA | Selection | Adaptive PCA 2 |
|--|---|--|--|---|---|
| Hold-out testing Healthy |  |  |  |  |  |
| Hold-out testing Powdery mildew |  |  |  |  |  |
| Independent samples Healthy bunch #7B |  |  |  |  |  |
| Independent samples Infected bunch #24A |  |  |  |  |  |
| Rating | Poor generalization from training set | Poor generalization from training set | Best approximation of infection level | Specific, but reduced sensitivity | Best with new samples, high sensitivity, good specificity |
| |  clutter |  background |  infected |  healthy | |

To highlight the performance of spatial-spectral segmentation based on Adaptive PCA 2, the segmentation performance for a low level infected image is shown in Fig. 1.

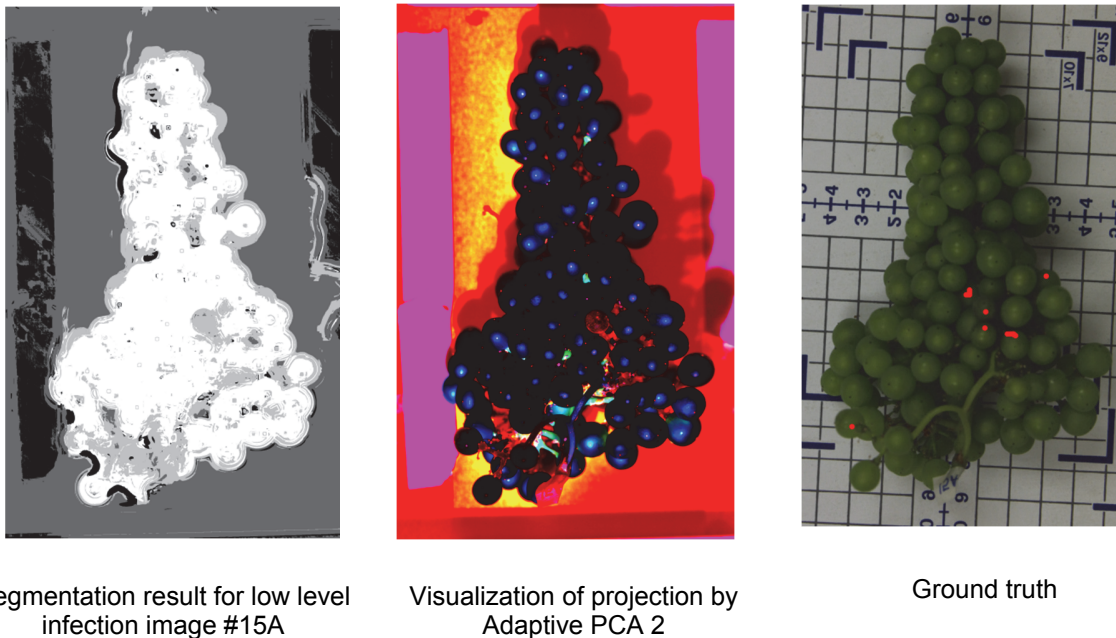


Fig. 1: Segmentation performance for low level infected material

The visualization of the PCA projection shows that infection with Powdery Mildew can be detected. However, automatic segmentation is distracted by image background components which do not belong to berries and have not been explicitly modeled. Hence, segmentation based on the proposed transformation can be easily further improved by an adapted training dataset.

6 Conclusions

A recently proposed image segmentation algorithm for fast processing of 3-channel images has been successfully applied to segmentation of hyperspectral images. Different dimensionality reduction methods have been implemented and tested with a real-world problem, namely the detection of Powdery Mildew infection of wine.

It was shown, that adaptive PCA provides the best segmentation results of bunch images among the used methods. Also, this projection yields stable results under changes of illumination and of perspective which are introduced with the imaging of complete bunches instead of berries. However, using a different random subset of samples for deriving the projection matrix of adaptive PCA can result in a very different projection and less convincing results. Hence, investigation of proper sampling strategies for different segmentation tasks is needed.

The proposed approach provides an easy and computationally inexpensive way to derive a number of meaningful spatial features for hyperspectral image segmentation. Calculation of such features in the original image space is either time- or memory-consuming, e.g. as integral images

for all bands are needed. Hence, dimensionality reduction based on band selection or projection has proven a valuable tool for fast processing of hyperspectral images.

7 Acknowledgements

The authors acknowledge the supply of the biological material by The University of Adelaide, School of Agriculture, Food & Wine. Hyperspectral image acquisition and labelling of infected berries was done in close collaboration with Tijana Petrovic and Timothy Zanker.

This work was partly supported by a grant of the German Federal Ministry of Education and Research (BMBF) under contract number 01DR14027A.

8 References

- ABDEL-RAHMAN, E. M., MUTANGA, O., ADAM, E. & ISMAIL, R., 2014: Detecting Sirex noctilio grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing* **88**, S. 48-59.
- AMINI, S., HOMAYOUNI, S. & SAFARI, A., 2014: Semi-Supervised Classification of Hyperspectral Image using Random Forest Algorithm. *IEEE International Geoscience and Remote Sensing Symposium, Québec, Canada*, S. 2866-2869.
- BREIMAN, L., 2001: Random Forests. *Machine Learning* **45** (1), S. 5-32.
- CAMPS-VALLS, G., TUIA, D., BRUZZONE, L. & BENEDICTSSON, J. A., 2014: Advances in Hyperspectral Image Classification. *IEEE Signal Processing Magazine* (**31**) 1, S. 45-54.
- DALPONTE, M., ØRKA, H. O., GOBACKEN, T., GIANELLE, D. & NÆSSET, E., 2013: Tree Species Classification in Boreal Forests With Hyperspectral Data. *IEEE Transactions on Geoscience and Remote Sensing* **51** (5), S. 2632-2645.
- FASSNACHT, F., NEUMANN, C., FÖRSTER, M., BUDDENBAUM, H., GHOSH, A., CLASEN, A., JOSHI, P. K. & KOCH, B., 2014: Comparison of Feature Reduction Algorithms for Classifying Tree Species With Hyperspectral Data on Three Central European Test Sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **7** (6), S. 2547-2561.
- KNAUER, U., BACKHAUS, A. & SEIFFERT, U., 2014a: Beyond Standard Metrics - On the Selection and Combination of Distance Metrics for an Improved Classification of Hyperspectral Data. *Workshop on Self Organizing Maps, Mittweida, Germany*, S. 167-177.
- KNAUER, U., BACKHAUS, A. & SEIFFERT, U., 2014b: Fusion Trees for Fast and Accurate Classification of Hyperspectral Data with Ensembles of Gamma-divergence based RBF Networks. *Neural Computing and Applications, Springer*, S. 1-10.
- KNAUER, U. & SEIFFERT, U., 2014c: Fast Image Segmentation with Boosted Random Forests, Integral Images, and Features on Demand. *IEEE Symposia Series on Computational Intelligence, Orlando, USA*, S. 1-6.
- REN, Y., ZHANG, Y., WEI, W. & LI, L., 2014: A spectral-spatial hyperspectral data classification approach using random forest with label constraints. *IEEE Workshop on Electronics, Computers and Applications, Ottawa, Canada*, S. 344-347.

- VEGANZONES, M. A., TOCHON, G., DALLA-MURA, M., PLAZA, A. J. & CHANUSSOT, J., 2014: Hyperspectral Image Segmentation Using a New Spectral Unmixing-Based Binary Partition Tree Representation, *IEEE Transactions on Image Processing* **23** (8), S. 3574-3589.
- VIOLA, P. & JONES, M., 2001: Rapid object detection using a boosted cascade of simple features, *IEEE Conference on Computer Vision and Pattern Recognition*, Kauai, USA, S. 511-518.
- WEI, L., LI, S., ZHANG, M., WU, Y., SU, S. & JI, R., 2013: Spectral-spatial classification of hyperspectral imagery based on Random Forests. *International Conference on Internet Multimedia Computing and Service*, Xiamen, China, pp. 163-168.
- XIA, J., DU, P., HE, X. & CHANUSSOT, J., 2014: Hyperspectral Remote Sensing Image Classification Based on Rotation Forest, *IEEE Geoscience and Remote Sensing Letters*, **(11)** 1, pp. 239-243.