Urban land-cover mapping with hyperspectral data by means of data reduction techniques and machine learning classifiers

Luca Demarchi[1,2], Claude Cariou[3], Giorgio Licciardi[4], Frank Canters[1] and Jonathan Cheung-Wai Chan[5]

[1]*Department of Geography, Vrije Universiteit Brussel, Brussels, Belgium.*
[2]*European Commission, Joint Research Center, Water Resources Unit, Ispra, Italy.*
[3]*TSI2M - Institute of Electronics and Telecommunications, University of Rennes 1, France.*
[4]*Grenoble Images Parole Signal Automatique, Grenoble Institute of Technology, France.*
[5]*Research and Innovation Centre, Fondazione E. Mach 1, Italy. ldemarch@vub.ac.*

Abstract. While recently many papers have been published on methodological developments in dimensionality reduction, application oriented studies on the performance of these methods in combination with different types of classifiers are less numerous. In this study we evaluate the potential of two unsupervised data reduction techniques - the Autoassociative Neural Network (AANN) and the BandClust (BC) - for mapping of urban land cover at a high level of thematic detail, using an APEX 288-band hyperspectral dataset. Both methods were tested in combination with four state-of-the-art machine learning classifiers and assessed in terms of Kappa estimates. When the BC method is used in combination with a strong learner, classification accuracies are similar or higher than obtained with the full dataset, demonstrating the method's capability of preserving critical spectral information. On the other hand, important spectral information seems to be compromised or lost in the AANN data reduction process. This study also demonstrates the potential of the APEX sensor data for detailed mapping of land cover in spatially and spectrally complex urban areas.

Keywords. Dimensionality reduction, APEX data, urban land cover mapping, machine learning classifiers.

1. Introduction

The phenomenon of urbanization is characterizing the contemporary era, negatively affecting the Earth System at local, regional and global scales [1]. Urban land-cover mapping, which is required for improved environmental management, is a challenging task, due to the spatial and spectral heterogeneity of urban environments, consisting of a great variety of artificial and natural surface types [2]. Nowadays, airborne hyperspectral data prevail as probably the best choice for tackling urban land-cover mapping, since they fulfill both the high spectral and spatial resolution requirements of urban remote sensing applications [3]. Despite their high richness of information content, adjacent hyperspectral channels carry highly correlated information which may result in redundant data and possible noise. Dimensionality reduction methods are supposed to remove the redundancy in hyperspectral datasets, without losing the original information content that is required for further data analysis [4]. In the literature, several techniques are described for dimensionality reduction which can be split into two major groups. The transformation-based approaches project the original high-dimensional data into a lower-dimensional space producing a new reduced dataset representing the transformed initial information. A relatively novel non-linear transformation method which is reported to have a good computational efficiency, yielding high classification accuracies, uses an Auto-Associative Neural Network (AANN) [5] to generate a reduced set of representative features from the original data under analysis. The second group called feature-selection-based approaches is based on band selection, whose objective is to find a small subset of relevant data from the original information. Within this family, a novel unsupervised method called BandClust (BC) [6], is reported as being easy to implement and has shown to provide high classification accuracies when used in combination with Support Vector Machine classifiers.

While many studies in the literature focus on methodological development of data reduction techniques, less attention is paid to the assessment of the proposed methods in application oriented contexts. The objective of this study is to assess the efficiency of both AANN and BC in the framework of large-scale hyperspectral urban land cover mapping, where several land cover types with subtle spectral differences need to be distinguished in a spatially complex urban setting. A high spectral and spatial resolution dataset (APEX), composed of 288 bands between 400-2500 nm and acquired at 3m spatial resolution, was chosen as input. Reduced datasets, obtained by applying both dimensionality reduction approaches, were used to classify 22 urban land cover classes. Classification was done using four state-of-the-art machine learning classifiers: Random Forest (RF), Adaboost (ADB), the Multi-Layer Perceptron (MLP) and Support Vector Machines (SVM). Kappa values and overall accuracies were computed from each error matrix, and statistical tests were conducted to judge if classification results obtained with reduced input datasets are significantly different from results obtained with the original input data [7].

2. Study area and data

The study area covers the southern part of the city of Brussels, Belgium which is characterized by a strong heterogeneity of land cover types. An APEX flight mission campaign was undertaken on the 24th of September 2011. The APEX instrument is a joint Swiss/Belgian initiative, designed as a dual prism dispersive pushbroom imaging spectrometer which was programmed to acquire 288 spectral bands at 3m ground resolution: 148 bands in the VNIR wavelength and 140 in the SWIR range. Spectral bands in the ranges [1.349; 1.426] nm and [1.804; 1.949] nm were ill-conditioned after atmospheric correction and were therefore subsequently removed, leading to a final workable dataset consisting of 261 bands.

3. Methodology

3.1. Dimensionality reduction techniques.

3.1.1. Auto-Associative Neural Networks (AANN)

An AANN is a conventional feedforward neural network having sigmoidal activation functions in each node and trained by standard back-propagation or similar algorithms [8]. The network is trained to perform identity mapping, where the input X has to be equal to the output X. This means that if the training phase finds an acceptable solution, a good representation of the input must be embedded in the hidden bottleneck layer. Since there are fewer nodes in the bottleneck layer than the input/output, the bottleneck nodes encode the information obtained from the inputs for the subsequent layers to reconstruct the output. In other words, data compression caused by the network bottleneck forces hidden units to represent significant features in the data [5]. The AANN training can be performed in a fully automatic way and all pixels in the image can be used for this task. No independent target data are required and there is no need to have an a priori knowledge for the implementation of the learning phase. For a more detailed explanation of AANN, the reader is referred to [5].

3.1.2. BandClust (BC)

BandClust [6] is a novel data dimensionality reduction technique based on band splitting. The core idea of BC lies in recursively splitting an initial spectral band into several subbands. At a given level of recursion, an optimal splitting of the spectral range [□min, □max] is sought by finding the wavelength □opt such that the mutual information between the reflectance values in the two subbands [□min, □opt] and [□opt, □max] is minimized. The recursive band splitting is performed until each previously created subband cannot be further split. BC has several benefits: first, it is an unsupervised method; second, it automatically provides an estimate of the optimal number of bands; third, it preserves the physical meaning of the hyperspectral data; finally, its recursive nature makes it easy to implement. For a more detailed explanation of BC, the reader is referred to [6].

3.2. Machine Learning classifiers

Machine Learning techniques have been widely used for the analysis and classification of RS data. They are considered to be able to make full use of small, narrow-band features in hyperspectral datasets and at the same time be fast, robust, noise-tolerant, and adaptive. For this study we decided to focus on the one hand on Decision Tree algorithms (DT), which may be considered as “weak” learners, and on the other hand on Artificial Neural Networks and SVM, which are considered as “strong” learners. In order to boost the lower classification accuracies produced by DT [9], we selected two well-known ensemble methods: RF [10] and ADB [9]. Within the Artificial Neural Network family we selected the MLP, which can be seen as a modification of the standard linear perceptron able to distinguish data that are not linearly separable [11]. MLPs have demonstrated competitive performances with respect to other techniques. Also SVM was included in the study, given its ability to model complex, nonlinear class boundaries in high dimensional feature spaces through the concept of kernel functions and regularization [12]. SVMs are particularly appealing in the RS field because of the ability of handling few training data and of producing higher accuracy results compared to other methods.

The four chosen machine learning classifiers were run in default mode, with automated parameter tuning and using three different input scenarios: in *Scenario 1*, the original 261 APEX bands are used as input dataset. For *Scenario 2*, the transformed components resulting from the AANN method are used. Finally, for *Scenario 3* the subbands produced by the BC algorithm were used as input for the four classifiers.

3.3. Classification scheme and accuracy assessment

A classification scheme was defined based on the most important artificial and non-artificial land cover types present in the urban landscape of the study area (Table 1). The selected 22 land cover classes are characterized by high within-class spectral heterogeneity and by spectral similarities between some of the classes, emphasizing the complexity of the detailed land cover classification task faced in the present study. For each land-cover class, groundtruth polygons were manually collected with reference to in-situ knowledge. Each polygon was composed of maximum fifteen image pixels. Groundtruth polygons were separated randomly into independent “Training” and

"Validation" groups. A fixed number of 150 samples were randomly selected within the "Validation" group while for the "Training" group 250 randomly selected samples were retained.

Land-cover types	
Manmade surfaces	**01.** Metal roofing bright
	02. Metal roofing dark
	03. Slate roofing
	04. Asphalt
	05. Concrete
	06. Tiles
	07. Cobblestones
	08. Gravel
	09. Red clay roof tiles
	10. Red pavement tiles
	11. Red clay (tennis courts)
	12. Red synthetic surfaces
	13. Artificial turf
Vegetation	**14.** Grassland maintained
	15. Grassland un-maintained
	16. Cropland
	17. Mixed forest
Bare soil	**18.** Bare soil field
	19. Sand
Water	**20.** Water
Shadows	**21.** Building shadow
	22. Tree shadow

Table 1. Land cover scheme definition.

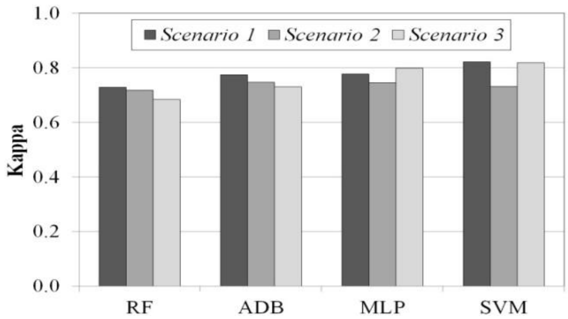

Fig.1. Kappa values calculated for the three input scenarios using the four classifiers.

	Sc.1 vs. *Sc.2*	*Sc.1* vs. *Sc.3*	*Sc.2* vs. *Sc.3*
RF	1.02	**3.83**	**2.81**
ADB	**2.62**	**4.04**	1.42
MLP	**3.00**	**2.09**	**5.09**
SVM	**8.54**	0.23	**8.31**

Table 2. Z-statistic values for pairwise comparison of accuracies obtained for different scenarios using a particular classifier. Statistically significant differences (Z-values greater than 1.96) are highlighted in bold.

Accuracy assessment was done based on Kappa analysis, the Kappa value (K) providing an estimate of how well the remotely sensed classification agrees with the reference data [7]. To determine if two independent K values are significantly different from one another and therefore to compare the performances of the classifiers for different input scenarios, a test of significance was performed based on:

$$Z = \frac{|K_1 - K_2|}{\sqrt{var(K_1) + var(K_2)}} \tag{1}$$

where K_1 and K_2 denote the estimates of Kappa for the two error matrices being compared. Given the null hypothesis H_0:$(K_1$-$K_2)=0$, and the alternative H_1:$(K_1$-$K_2) \neq 0$, H_0 is rejected if $Z \geq 1.96$ with a confidence level of 95% [7].

For each type of classifier applied in this study, Z statistic values were calculated for the comparison of *Scenario 1* (original data) with *Scenario 2* (AANN) and *Scenario 3* (BC), and for the pairwise comparison of *Scenario 2* with *Scenario 3*, in order to assess the impact of using one of the two data reduction methods on the performance of the classifier.

4. Results and discussion

For *Scenario 2*, the AANN method was run on the APEX dataset, resulting in a bottleneck layer composed of 21 nonlinear components. Applying the BC algorithm on the APEX dataset (*Scenario 3*), resulted in the selection of ten splitting bands, yielding eleven subbands which, according to the BC algorithm, capture the essential information content present in the original dataset.

The K values obtained for each classifier in the three input band scenarios are plotted in Fig. 1, while Table 2 summarizes the Z values resulting after comparing the results. Two distinct trends in performance are noticeable, one for the DT classifiers (RF and ADB) and one for the "strong" classifiers (MLP and SVM). For both RF and ADB, *Scenario 1* produces the highest accuracies (K=0.729 for RF and K=0.774 for ADB). Of the two data reduction methods tested, AANN produces the best results. Only with RF though, the accuracy obtained with AANN (K=0.717) is comparable with the accuracy obtained in *Scenario 1*. Furthermore, for the DT classifiers BC (*Scenario 3*) produces the lowest accuracies (K=0.684 for RF and K=0.730 for ADB) which are significantly lower than the accuracies obtained with the entire dataset.

For MLP and SVM a distinct trend is noticed: the highest accuracies are obtained with both *Scenario 1* and *Scenario 3*, while *Scenario 2* produces significantly lower accuracies (Table 2). In the case of MLP, the reduced BC dataset is even capable of outperforming the original dataset (K=0.799 for *Scenario 3* and K=0.777 for *Scenario 1*), although the Z value of 2.09 obtained is very close to the critical value of 1.96. In the case of SVM, testing of the null hypothesis shows that the accuracies obtained with *Scenario 1* and *Scenario 3* are statistically equal (Z=0.23, K=0.821 for *Scenario 1* and K=0.818 for *Scenario 3*). In other words, when the BC method is used in combination with the two "stronger" classifiers (MLP and SVM), it succeeds in substantially reducing the dimensionality of the hyperspectral dataset without decreasing classification accuracy. The AANN method, on the contrary, is not able to achieve satisfactory results with MLP and SVM.

These results demonstrate that the effectiveness of the data reduction methods is dependent on the type of classifier used. In fact, most classifiers do not seem able to achieve a satisfactory level of accuracy using the 21 transformed AANN components. Probably, the original spectral information is not well preserved, and critical information has been compromised or distorted through the transformation procedure. On the other hand, the eleven subbands resulting from applying the BC algorithm on the original dataset, which preserve the physical meaning of the hyperspectral data, lead to good performances for MLP and SVM. The results obtained seem to confirm the efficiency of the BC method when used in combination with SVM, as already indicated in [6] and on the other side, highlight the potential of SVM of fully exploiting the richness of information present in the APEX hyperspectral dataset.

Fig. 1 shows an extract of the land-cover map produced by SVM for *Scenario 1* (Fig. 5(b)) and *Scenario 3* (Fig. 5(c)) respectively. A comparison of the spatial distribution of the 22 land-cover classes in both maps confirms the similarity of the classification result obtained with and without data reduction.

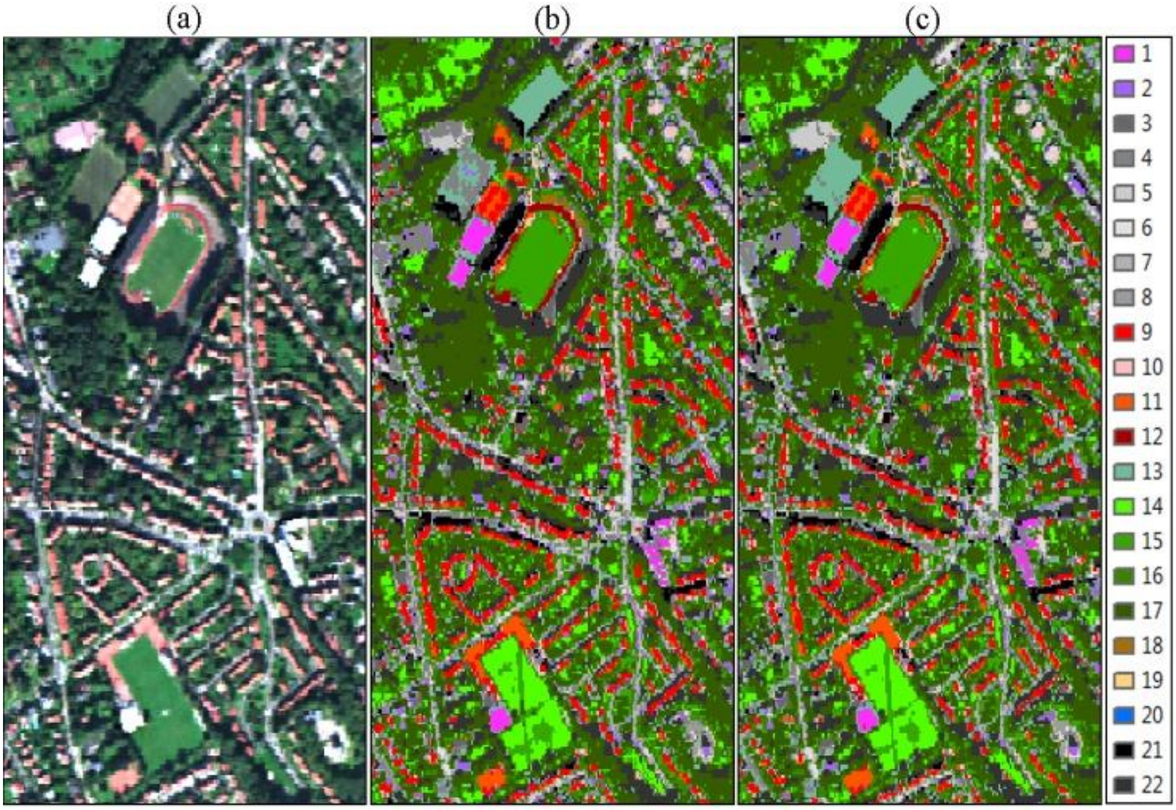

Figura 1. True color composite (a), SVM classification for *Scenario 1* (b), and SVM classification for *Scenario 3* (c) for an extract of the study area (for class numbering, see Table 1).

5. Conclusions

In this work we assessed the performance of two relatively novel unsupervised data reduction techniques - the Autoassociative Neural Network approach (AANN) and BandClust (BC) - in the context of high-resolution urban land-cover mapping using APEX hyperspectral data. Classification in this study was done using four state-of-the-art machine learning methods: Random Forest (RF), AdaBoost (ADB), the Multiple Layer Perceptron (MLP), and support vector machines (SVM).

Results show that BC is able to retain or slightly improve the level of classification accuracy obtained with the full dataset when used in combination with a "strong" learner (SVM, MLP). On the other hand, for AANN accuracies are reduced for three of the four classification methods used. The results of this study seem to indicate that the BC algorithm succeeds well in preserving critical spectral information needed for successfully distinguishing spectrally similar urban land-cover classes. More work with other hyperspectral datasets is needed though to examine whether the results obtained in this study can be generalized to other urban settings.

This research also demonstrates the potential of APEX data for producing high-resolution, thematically detailed urban land-cover maps, representing common urban surface types that are characterized by subtle spectral differences and high within-class spectral heterogeneity. Best classification results in this study are obtained with the SVM classifier, reaffirming the effectiveness of this classifier in modeling complex, nonlinear class boundaries in high dimensional feature spaces, as reported in numerous other studies. Combining the BC band selection method with SVM seems to be an efficient approach for reducing the high data volume of hyperspectral data covering urban areas, without compromising on classification accuracy.

References

[1] Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X. and Briggs, J.M., (2008). *Global Change and the Ecology of Cities*. Science, 319, (5864), 756-760.

[2] Herold, M., Roberts, D.A., Gardner, M.E., and Dennison, P.E. (2004). *Spectrometry for urban area remote sensing – development and analysis of a spectral library from 350 to 2400 nm*. Remote Sens. Environ., 91, (3-4), 304-319.

[3] Gamba, P., Dell' Acqua, F., Lisini, G. and Trianni, G. (2007). *Improved VHR Urban Area Mapping Exploiting Object Boundaries*. IEEE Trans. Geosci. Remote Sens, 45, 2676-2682.

[4] Bruce, L.M., Koger, C.H. and Li, J. (2002). *Dimensionality Reduction of Hyperspectral Data Using Discrete Wavelet Transform Feature Extraction*. IEEE Trans. Geosci. Remote Sens, 40, 2331-2338.

[5] Licciardi, G.A. and Del Frate, F. (2012). *Pixel Unmixing in Hyperspectral Data by Means of Neural Networks*. IEEE Trans. Geosci. Remote Sens, 49, 4163-4172.

[6] Cariou, C., Chehdi, K. and Le Moan, S. (2011). *BandClust: An Unsupervised Band Reduction Method for Hyperspectral Remote Sensing*. IEEE Trans. Geosci. Remote Sens, 8, 565-569.

[7] Congalton, R.G. and Green, K. (2009). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC Press.

[8] Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford Univ. Press, London, UK.

[9] Chan, J.C.-W. and Paelinckx, D. (2008). *Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery*. Remote Sens. Environ., 112, 2999-3011.

[10] Breiman, L. (1996). *Bagging Predictors*. Machine Learning, 24, 123-140.

[11] Benediktsson, K.A., Swain, P.H. and Ersoy, O.K. (1990). *Neural Network Approaches Versus Statistical Methods In Classification Of Multisource Remote Sensing Data*. IEEE Trans. Geosci. Remote Sens, 28, 540-552.

[12] Mountrakis, G., Im, J. and Ogole, C. (2011). *Support vector machines in remote sensing: A review*. ISPRS Journal of Photogrammetry and Remote Sensing, 66, 247-259.