ENSEMBLE METHODS FOR SPECTRAL-SPATIAL CLASSIFICATION OF URBAN HYPERSPECTRAL DATA

Jón Atli Benediktsson*, Björn Waske* and Xinlu Wang* **

(*) Faculty of Electrical and Computer Engineering
University of Iceland
107 Reykjavik, Iceland

(**) College of Computer Science and Technology,
Jilin University, 2699 Qianjin Street, Changchun
130012, P.R. China

ABSTRACT

In this paper the use of an ensemble of classifiers or multiple classifiers for a spectral-spatial classification of urban hyperspectral data is addressed. Whereas usually a single classifier is used to determine the class memberships, in many experiments the classification accuracy was improved by using an ensemble of classifiers. The ensemble classifier concept is based on a combination method that makes use of the benefits of each (weak) classifier. In doing so, each individual classifiers support each other in making a decision. Boosting, bagging and random forests are perhaps the widest used multi classification strategies. These methods are manipulating the input data (training samples and image bands) during the individual training processes. After all the individual classifier outputs are combined to create the final results. Contrary approaches that are based on the statistical consensus theory, handle each data of the sources separately, and use all the training data only once.

The classification of urban remotely sensed data can be challenging, because urban land cover classes are often spectrally heterogeneous and surface materials from different classes might have similar spectral properties. Purely spectral classifications can be limited under these circumstances and result in relatively low classification accuracy. However, additional spatial features often help overcoming such spectral limitations. Methods based on mathematical morphology (i.e., morphological profiles) for preprocessing of the hyperspectral data were introduced during the last years. In several studies it was demonstrated that a combination of the derived information with the original imagery improves the classification accuracies. However, such approaches generate large data sets, which have to be handling by efficient classifiers and/or feature reduction methods. In the study presented here the performance of classifier ensembles is investigated in detail. Moreover the so-called variable importance, which can be provided by random forest classifier system, is used to reduce the dimensionality. This method is based on a raking of the features, which depends on the impact of each feature on the overall accuracy.

A ROSIS-03 (Reflective Optics System Imaging Spectrometer) data set acquired over the city of Pavia is used for our experiments. The flight over the city of Pavia, Italy, was operated by the Deutschen Zentrum fur Luft- und Raumfahrt (DLR, the German Aerospace Agency) in the framework of the HySens project, managed and sponsored by the European Union. According to specifications, the number of bands of the ROSIS-03 sensor is 115 with a spectral coverage ranging from 0.43 to 0.86µm. The spatial resolution is 1.3m per pixel.

Various morphological profiles are generated, using different sizes and shapes for the structural element. The RF variable importance is used to determine impact of the bands of the overall accuracy. Beside the Random Forests different classifier approaches were applied to the data sets. Overall the obtained results using the different ensemble approaches are excellent and outperform traditional single classifier algorithms on the whole data set in terms of accuracies. These results clearly demonstrate that multiple classifier systems are well suited for the integration of spatial information into the classification process of remote sensing data.