Empirical Evaluation of Pattern Recognition Techniques in Remote Sensing

Research Proposal for EuroSDR

Wolfgang Förstner, Björn Waske
Department of Photogrammetry
University Bonn

We apply fort he support of an empirical study on the power of today's methods for interpreting remote sensing data. The benchmark is meant to be performed by the University Bonn and in conjunction with the ISPRS Intercomission Working Group *'Pattern Recognition in Remote Sensing'*.

Motivation

The last decade in remote sensing is characterized by an increased **availability of**

- *Sensor data:* This especially refers to higher in spatial, temporal and spectral resolution and the increased availability of data from different, often complementary sensors and sources.

- *Software for classification available in commercial software.* Not only classical software packages for remote sensing as ENVI/IDL or ERDAS Imagine contain a large number of classifiers but also image processing tools in Photoshop and other publically available packages contain techniques from pattern recognition, e. g. to segment images into for- and background. Besides this different publically available packages from the field of machine learning, pattern recognition and data mining are available, which often do not enable a direct processing of image data.

- *Research in pattern recognition and machine learning.* In the areas of pattern recognition, computer vision and machine learning the relevant of image data has increased in the last years. Triggers are requirements in medical image processing, image sequences for surveillance, and the provision of geo-coded image data by Google and Microsoft.

The development in classification of remote sensing imagery over the last decades was to a great extent driven by the development of enhanced remote sensing systems and the availability of more diverse remote sensing data sources. Contrary to many common methods that are based on simple techniques, such recent developments usually require more sophisticated methods. Consequently a shift from simple approaches to more powerful machine learning algorithms for image classification has taken place. This development corresponds with the parallel increase in computing power that enables the handling of huge and diverse remote sensing data sets as well as more rigorous performance requirements, such as speed and accuracy.

In contrast to these positive developments one can observe a **lack of**

- *Comparative studies of methods.* The success in an area can best be documented by comparative studies on the performance of competing methods. Such comparisons are limited in the remote sensing area, for which many reasons can be found, e. g. the locality of the investigation area, the specificy of the research goals, the embedding of remote sensing as an data provision tool for a geo- or agro-research related theme. As a consequence one observers the main hindrance for benchmarking:

- *The lack of publically available sensor data with ground truth.* Evaluating the performance of classification and interpretation methods requires an adequate amount of sensor data with ground truth. In spite each publication containing quantitative performance measures is based on sensor data with ground truth, a comparison in general cannot be achieved. One reason is: these data are proprietary, thus are not public. Another reason is: these data are insufficient for comparing different methods. Partially the amount of ground truth data is not large enough to achieve significant results. Partially the ground truth data are sparse: this prevents testing methods, which require – at least partially - dense ground truth for training.

- *Transparency of quality of methods.* As a result of this situation a user of pattern recognition software in general cannot evaluate the useful of a certain method for his/her task, as there is not only one method which is superior to all others in all contexts, e. g. resolution or number and type of classes. This hinders pattern to exploit the power of currently available recognition methods in practice.

Application for support of an empirical study

We apply for support of an empirical study with the following *goals*

- Only an *empirical study* can provide adequate information about the usefulness of existing methods with respect to certain tasks relevant to users. Of course, due to the complexity of natural phenomena and the variety of mathematical, physical and phenomenological models we only can address the problem of classification performance, not the usefulness of the classification results for a specific application.

- We want to compare *commercial and research software*. There are quite some commercial software packages for the evaluation of remote sensing data available, which are regularly used in practice. Partially they reflect the progress of current research. However, methods developed in the research community of machine learning and pattern recognition may not have been applied to remote sensing data, for the reasons mentioned above or for other reasons, such as performance or different interest of the research group. A comparison of the state of the art classification methods with commercial software would be of great interest to both, companies as well as researchers. A comparison of state of the art classification methods with algorithms that are publically available and widely used in the field of remote sensing (e.g., commercial software packages), would be of great interest to companies, researchers as well as users.

- The final goal is a *categorization of methods w. r. t. data and tasks*. This certainly will be of utmost relevance, but at the same time will be qualitative.

The *means* to perform such an empirical study would be the following

- The members of the test, especially of EuroSRD, would be asked to provide sensor data with ground truth. It is decisive to have qualitatively and quantitatively good reference data in order to evaluate the performance. But also the adequateness of the sensor data w. r. t. the addressed classes is of utmost importance: The identification of the classes based on the sensor data should be good, neither to easy nor too difficult, in order to differentiate the various classification schemes. Ideally the benchmark data sets comprise various remote sensing data sets from regions with different environmental settings.

- The partners from EuroSDR and ISPRS are asked to provide the data in a common format or in a format which allows bringing them into a common format. The data, both image and ground truth data must be made available to the research community. There may be a formal procedure to make the distribution of the data transparent to the provider of the data.

- Although benchmark data sets (i.e., remote sensing imagery + ground truth data) would be available, a comparison of different studies might be limited, due to different and inconsistent accuracy assessments. Thus, a central evaluation is planned to be performed at the University of Bonn, which allows a consistent assessment of the provided classification results.

We have the following time schedule, referring to the actual start of the project (PI=principal investigator, PA=participant, D=data provision, C=classification):

[0=start]	Preparation of call: specification of data, ground truth, evaluation scheme (PI)
[+2=02]	Call for participation for data provision data provision (PI)
[+6=08]	Provision of sensor data with ground truth (PA-D)
[+2=10]	Evaluation of submitted data (PI)
[10]	***Decision on continuation***, Call for participation on classification (PI)
[+2=12]	Homogenisation and distribution via web interface (PI)
[+6=18]	Provision of classification results using web-interface (PA-C)
[+2=20]	Result of analysis, distribution (PI)
[+2=22]	Workshop, interactive test (training+classification) (PI, PA-C)
[+2=24]	Final report (PI)

The call for participation will make the usage of the data and the evaluation procedures transparent, such that both the providers of data as well as the participants of the benchmark will know in advance how their data and results will be evaluated.

We presume, the annotated data to fulfil the following requirements:

1. approximately 8-20 land cover classes, if possible with class hierarchy,
2. large enough annotated data (at least 500 pixels per class),
3. rights to publish the data internationally without charge.