Bochow, M., Segl, K., Kaufmann, H. (2008): Comparison of multi- and hyperspectral remote sensing data for use in comprehensive urban biotope mapping.
IEEE International Geoscience & Remote Sensing Symposium - IGARSS (Boston 2008)

COMPARISON OF MULTI- AND HYPERSPECTRAL REMOTE SENSING DATA FOR USE IN COMPREHENSIVE URBAN BIOTOPE MAPPING

Mathias Bochow, Karl Segl and Hermann Kaufmann

Helmholtz Centre Potsdam – German Research Centre for Geosciences (GFZ Potsdam),
Section 1.4 Remote Sensing, Telegrafenberg, A 17, 14473 Potsdam, Germany;
phone (+49)331-288-1198, fax (+49)331-288-1192, {mbochow}{segl}{charly}@gfz-potsdam.de

ABSTRACT

We classified 922 urban biotopes from 11 different biotope types in a 50.6 km^2 study area in Berlin, Germany. As input advanced data products were derived from hyperspectral and simulated multispectral data. Urban surface materials were derived from the hyperspectral data by classification and linear spectral unmixing. Multispectral data was classified using four different per-pixel and object-oriented classifiers. The results show that our developed method for biotope classification works well with hyperspectral and with multispectral input data yielding comparable overall accuracies of 88.1 and 91.3 percent.

Index Terms— urban biotope mapping, remote sensing, hyperspectral, multispectral, automation, spatial metrics

1. INTRODUCTION

Urban biotope maps are of high importance for ecological urban planning and well established in Germany. Most of the bigger cities accomplished initial mappings of urban biotopes, which can be thought of as city structure types categorized by their ecological value and functions, over the last 20 years using a detailed and complex mapping key [1, 2]. In common practice, area-wide mapping and monitoring of urban biotopes is based on visual interpretation of color-infrared aerial photographs and field investigations. Since this procedure is very time-consuming and costly it often cannot keep up with the pace of urban development. Against this background, there are astonishingly few investigations dealing with the automation of these mappings by an automated analysis of remote sensing data although such automation, even of parts of the mapping procedure, holds the potential of being faster, reproducible and objective.

This paper presents a novel method for the automatic identification of urban biotope types based on remote sensing and GIS data. This method is the essential part of an automatic update system for existing urban biotope maps [3]. Whereas in former investigations [4, 5] we used hyperspectral images as input data the focus of this paper lies on the needed spectral resolution of the optical input data by a comparison of airborne hyperspectral (HyMap) and high resolution multispectral data (simulated IKONOS).

The remainder of this paper is organized as follows: Section 2 describes the basic input data. Section 3 deals with the retrieval of advanced data products from hyperspectral (3.1) and multispectral (3.2) input data needed for biotope classification. In section 4 the overall approach of urban biotope classification is briefly introduced. Section 5 and 6 show and discuss the achieved results and give an outlook on future work.

2. DATA

The developed method for classifying urban biotopes needs three types of basic input data: (1) Optical high resolution remote sensing data, (2) a high resolution digital surface model (DSM) and (3) the existing biotope map. The analysis was carried out on hyperspectral HyMap data (126 spectral bands, 3.5 m spatial resolution) and a DSM (1 m spatial resolution) generated from images of the HRSC-AX camera by a multi-image matching technique [6, 7]. Both data types were acquired in 2005 over a test site in the city of Berlin, Germany.

A normalized digital surface model (nDSM) that stores the heights of objects over ground surface was derived from the DSM in two steps: First, a digital terrain model (DTM) was calculated by an advanced filtering technique. Then the DTM was subtracted from the DSM. The resulting nDSM was resampled to the resolution of the HyMap data and serve as input for feature calculation (section 4). At last, a mask image of high objects was created by applying a height threshold of 1.8 meters to the nDSM.

The HyMap data was atmospherically corrected using an in-house developed hybrid method (ACUM algorithm). It employs the radiative transfer models of MODTRAN in order to calculate at-surface reflectance taking a DSM as supplementary data. The geometric correction was done with an in-house developed parametric geocoding approach utilizing the in-flight recorded exterior orientation parameters, a DSM and ground control points. The resulting root mean square (RMS) error of less than one pixel allows an accurate overlay of the input data.

Multispectral data was simulated by spectrally degrading the HyMap data using the spectral response function of the IKONOS sensor. This ensures an accurate comparison of the results from both data types.

The areas of analyzed biotopes were taken from the official biotope map of Berlin with the state of June 2006. 11 biotope types, namely detached house development, block development, perimeter block development, apartment block development, mid-rise dwellings development, high-rise building development, industrial areas, traffic areas, allotment gardens, sport areas and ornamental lawns were selected for the study. These are very common types in German cities representing about 50% of the test site which has an extent of 50.6 km^2.

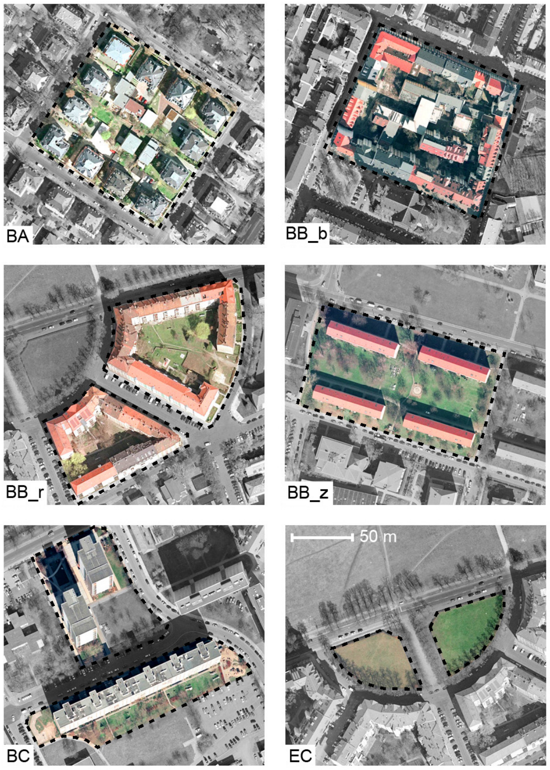

Figure 1. Example biotopes in aerial images for 6 of the 11 biotope types: (BA) detached house development, (BB_b) block development, (BB_r) perimeter block develop., (BB_z) mid-rise dwellings develop., (BC) high-rise building develop., (EC) ornamental lawns

3. RETRIEVAL OF ADVANCED INPUT DATA FOR BIOTOPE CLASSIFICATION

It is the main aspect of this investigation to examine the feasibility of multispectral data for use with the developed method for biotope classification. In our study with hyperspectral data presented in [4] we used five types of advanced input data for the method: (1) The areas of the biotopes from an existing biotope map, (2) the nDSM, (3) unmixing layers of surface cover types, (4) unmixing layers of thematic class groups and (5) an image segmentation of the unmixing layers. 3-5 are explained in section 3.1. During the investigations it turned out that the best results could be obtained using only the thematic main classes instead of the detailed surface cover types. Models that were trained with features (explained in section 4) calculated on these thematic main classes were significantly more stable when applying them to an unknown dataset than models that were trained with features calculated on the individual surface cover types. From this finding the main question of this investigation arose whether the thematic main classes could be extracted from a classification of multispectral data. The following subsections explain the processing of hyperspectral and multispectral data to retrieve appropriate advanced data products for the biotope classification.

3.1. FROM HYPERSPECTRAL HYMAP DATA

The derivation of advanced data products from hyperspectral data is only briefly described here because it was developed in former studies [4, 5]. Surface fraction layers were produced by a classification and unmixing of HyMap data comprising 54 man-made and natural surface cover types (25 roof materials, 16 other man-made materials, 5 vegetation types, 2 soil types, 4 water types and 2 types of shadow). We used a processing chain that has been developed at the Helmholtz Centre Potsdam – German Research Centre for Geosciences (GFZ Potsdam) over the past years. It has reached a high degree of automation consisting of (1) a feature-based endmember identification approach [8, 9], followed by (2) a maximum likelihood classification to detect spectrally pure pixels and (3) an iterative neighborhood-oriented linear spectral unmixing procedure [10, 11]. As a new developed feature the classification and unmixing was supplemented with the mask of high objects to improve the identification of objects standing on ground surface (e.g. buildings and trees).

Based on the unmixing result thirteen additional layers were generated which contain the thematic main classes vegetation, trees, soils, roofs, metal roofs, tile roofs, flat roofs, industrial roofs, traffic areas, sport areas, shadow, water and courtyard. These layers were built by summing up the surface fractions of the respective classes per pixel.

Further advanced input data for the feature computation were obtained from a segmentation of the surface fraction layers of classes and thematic main classes. This was done by thresholding the fraction layers with 0.51 (i.e. more than 51% of the pixel have to belong to the class) and by clumping adjacent pixels of a class (or thematic main class, respectively) to a segment. The segments of the thematic main classes are saved in individual segmentation layers since overlapping segments are possible due to several classes that belong to more than one thematic main class (e.g. deciduous trees is in the thematic main class trees and vegetation). The segments of all individual classes are stored together in one layer.

3.2. FROM MULTISPECTRAL IKONOS DATA

We compared the performance of four different classifiers for the classification of the IKONOS data: Support Vector Machines (SVM) [12] and the Maximum Likelihood (ML) classifier are both pixel-oriented classifiers. SVMs are applied with One-Against-All (OAA) and One-Against-One (OAO) strategy. The ECHO classifier [13] is an extension of the ML classifier with a simple preceding image segmentation into homogeneous image objects. The new software ENVI Fx 4.4 [14] performs a more advanced image segmentation based on a region growing algorithm with a subsequent kNN classification. All classifications were supplemented with the mask of high objects to obviate confusion between roofing materials and pavement materials.

To examine the thematic content of multispectral data we started with all 54 surface cover types for the classification (not possible in ENVI Fx; at present a max. of 47 classes is supported). As expected all multispectral classifications show inferior results to the hyperspectral classification assessed by the overall accuracy with hold-out strategy (ECHO: 68.7%, SVM-OAA: 74.5%, SVM-OAO: 75.7%, ML: 79.8%; hyperspectral: 86.2%). Based on a separability measurement with Jeffries-Matusita distance and the confusion matrices of the classifications we discovered groups of spectrally similar classes (Table 1, A). These classes were merged into spectral class groups in the image and in the test pixel sets only if they belong to the same thematic main class. Thereafter, the overall accuracies improved to ECHO: 86.2%, SVM-OAA: 83.1%, SVM-OAO: 84.1%, ML: 84.6%. The remaining confusion with multispectral classifiers is mainly caused by the classes named in Table 1, B. In addition, a simple majority vote was done with these 4 classification results to eliminate specific problems of the individual classifiers. With ENVI Fx and the merged 43 classes only 55.6% could be achieved and thus was not considered in subsequent analyses. Here, problems occurred with the segmentation due to low contrast materials and shadow effects (Fig. 2).

Table 1. Groups of spectrally similar classes that could be merged (A.) and not merged (B.)

A	Red clay tile & auburn clay tile & red concrete tile Aluminum & PVC & Plexiglas (all used for roofing) Dark & bright bitumen roof sheeting Steel with protective coating & glass River & pond 3 different types of concrete pavements Lawn & meadow
B	Coniferous trees & shadow on vegetation & greenhouse Lake & shadow on non-vegetation Deciduous trees & green roofs

Figure 2. A building with low contrast to it's surrounding (HyMap, middle image). The borders of the ENVI Fx segmentation (right) do not meet the building's border yielding to false classification compared to the ML result (left image; building roof: red class).

From the resulting classification the same data products are derived as from the hyperspectral unmixing result. The pseudo-unmixing layers contain only the values 0 or 100. Based on the IKONOS classification with 43 merged spectral classes it was possible to generate all 13 thematic main classes. Only high confused classes were not included. The pseudo-unmixing layers of classes and thematic main classes were also segmented as described in section 3.1.

4. BIOTOPE CLASSIFICATION

In this section we describe the use of the advanced input data from both sources, multi- and hyperspectral data, to classify urban biotopes. Therefore, in [5] we developed a new approach employing an individual fuzzy logic model for the identification of every biotope type, i.e. each model is able to recognize biotopes of it's own type and to reject biotopes of other types. In [4] the build-up process of these models was automated.

The nDSM, the (pseudo-)unmixing results of the optical input data, the segmentation images and the rasterized areas of the biotopes are the basis for the calculation of numerical feature. Each biotope type model consists of quantitative features that describe the composition of the biotopes of different surface cover types and their arrangement within the biotope. The features [15] were developed to numerically capture the characteristics of the biotope types.

Applied to a biotope every biotope type model (Fig. 3) calculates a similarity value expressing the similarity of the biotope to the corresponding biotope type of the model. Finally, the biotope is classified with the biotope type whose model calculated the highest similarity value.

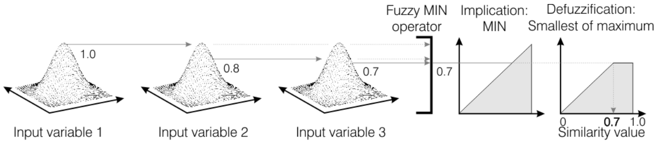

Figure 3. Schematic model design of the biotope type models. See [4] for details.

5. RESULTS AND DISCUSSION

We classified 922 biotopes from 11 different biotope types in the study area of Berlin using hyperspectral and simulated multispectral input data. The classification accuracy was assessed with a confusion matrix using cross validation with stratified random sampling and tenfold repartitioning. Based on the hyperspectral data the overall accuracy was 88.1%. Based on multispectral data and the surface cover type classifications with SVM-OAA, ML, SVM-OAO and ECHO as well as the majority vote of the four, comparable accuracies of 86.3%, 87.3%, 87.3%, 88.7% and 91.3% could be achieved. The majority vote outperforms the individual classifications slightly because specific problems of the individual classifiers could be eliminated.

6. CONCLUSIONS AND OUTLOOK

In this investigation we showed that high resolution multispectral satellite data is a suitable input for the derivation of advanced data products needed for urban biotope classification. Compared to the spectral unmixing of hyperspectral data several urban surface materials could not be separated correctly with multispectral data but appeared to be not relevant for the classification of the 11 selected biotope types.

The use of multispectral satellite data can be a great step towards operability and global applicability of a method for updating urban biotope maps of which the developed method for classifying urban biotopes is the essential part. The method will be tested on real satellite data in the next step. Furthermore, the focus of future work will lie on the implementation of vegetation-dominated biotope types.

ACKNOWLEDGEMENTS

This work was made possible by several flight campaigns carried out by the Deutsches Zentrum für Luft- und Raumfahrt (DLR) Oberpfaffenhofen, Germany. The Senate Department for Urban Development of Berlin provided the biotope map. We further thank the DLR, Institute of Planetary Research for providing the HRSC derived DSM and CREASO GmbH for a test license of ENVI Fx. Many thanks go to the people of the Geomatics Lab of Humboldt University of Berlin for providing the HyMap data and a wonderful free software called ImageSVM.

[1] Sukopp, H., and S. Weiler, "Biotope Mapping and Nature Conservation Strategies in Urban Areas of the Federal Republic of Germany," *Landscape and Urban Planning*, 15: pp. 39-58, 1988.

[2] Arbeitsgruppe "Methodik der Biotopkartierung im besiedelten Bereich", W. Schulte, H. Sukopp, and P. Werner, (eds.), "Flächendeckende Biotopkartierung im besiedelten Bereich als Grundlage einer am Naturschutz orientierten Planung," *Natur und Landschaft*, 68(10): pp. 491-526, 1993.

[3] Bochow, M., K. Segl, and H. Kaufmann, "An update system for urban biotope maps based on hyperspectral remote sensing data," in Proc. of *5th EARSeL SIG IS Workshop on Imaging Spectroscopy*, Bruges, Begium, 2007.

[4] Bochow, M., K. Segl, and H. Kaufmann, "Automating the build-up process of feature-based fuzzy logic models for the identification of urban biotopes from hyperspectral remote sensing data," in Proc. of *URBAN / URS 2007 Joint Conference*, Paris, France, 2007.

[5] Bochow, M., K. Segl, and H. Kaufmann, "Modelling of urban biotope types from hyperspectral imagery using a fuzzy logic approach," in Proc. of *2nd Workshop of the EARSeL Special Interest Group on Remote Sensing of Land use & Land Cover*, Bonn, Germany, 2006.

[6] Wewel, F., and F. Scholten, "Erstellung von digitalen Oberflächenmodellen mit dem Mehrzeilen-Stereoscanner HRSC-A," *Photogrammetrie - Fernerkundung - Geoinformation*, 8/2000: pp. 57-66, 2000.

[7] Scholten, F., S. Sujew, F. Wewel, J. Flohrer, R. Jaumann, F. Lehmann, R. Pischel, and G. Neukum, "The high resolution stereo camera (HRSC) - digital 3D-image acquisition, photogrammetric processing and data evaluation," in Proc. of *Proc. of the ISPRS Joint Workshop "Sensors and Mapping from Space"*, Hannover, Institut für Photogrammetrie und Ingenieurvermessungen, Universität Hannover, 1999.

[8] Heiden, U., *Analyse hyperspektraler Flugzeugscannerdaten zur ökologischen Charakterisierung städtischer Biotope*, Dissertation, Fakultät VII - Architektur Umwelt Gesellschaft, TU Berlin, p. 136, 2004.

[9] Segl, K., M. Bochow, S. Roessner, H. Kaufmann, and U. Heiden, "Feature-based identification of urban endmember spectra using hyperspectral HyMap data," in Proc. of *1st workshop of the EARSeL Special Interest Group Urban Remote Sensing*, Berlin, 2006.

[10] Roessner, S., K. Segl, U. Heiden, and Kaufmann, "Automated differentiation of urban surface based on airborne hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, 39(7): pp. 1523-1532, 2001.

[11] Segl, K., S. Roessner, U. Heiden, and H. Kaufmann, "Fusion of spectral and shape features for identification of urban surface cover types using reflective and thermal hyperspectral data," *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1-2): pp. 99-112, 2003.

[12] Vapnik, V.N., *Statistical Learning Theory*, Wiley, New York, 736 pp., 1998.

[13] Kettig, R.L., and D.A. Landgrebe, "Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects," *IEEE Transactions on Geoscience Electronics*, GE-14(1): pp. 19-26, 1976.

[14] ITT-VIS, *Introducing the ENVI Feature Extraction Module*, website, last visit: 20.06.2008, Available from: http://www.ittvis.com/envi/featureextraction/index.asp

[15] Bochow, M., K. Segl, and H. Kaufmann, "Potential of hyperspectral remote sensing data for the automated mapping and monitoring of urban biotopes," in Proc. of *1st Workshop of the EARSeL Special Interest Group Urban Remote Sensing*, Berlin, Germany, 2006.