FEATURE-BASED IDENTIFICATION OF URBAN ENDMEMBER SPECTRA USING HYPERSPECTRAL HYMAP DATA

Karl Segl, Mathias Bochow, Sigrid Roessner, Hermann Kaufmann[1] and Uta Heiden[2]

1. GeoForschungsZentrum Potsdam, Department 1 - Geodesy and Remote Sensing, Section 1.4 - Remote Sensing, Telegrafenberg, 14473 Potsdam, Germany, e-mail: segl@gfz-potsdam.de
2. German Remote Sensing Data Center, Dept. of Environment and Geoinformation, Oberpfaffenhofen, 82234 Wessling, Germany

ABSTRACT

Applications of hyperspectral remote sensing data within the urban environment are still rare, although there is a need for a time- and cost-efficient monitoring. Airborne hyperspectral sensors can help to improve this situation allowing more detailed and precise determination of urban surface cover types. For image classification or spectral sub-pixel analysis, definition of representative spectral endmembers is the most important but also most time-consuming operation. In order to simplify this procedure, unsupervised techniques were developed. Independent from their limitations to process the high spectral variability of urban materials, no material information is given and has to be determined by the operator.

In this paper a spectral identification procedure is presented which automatically detects and identifies representative endmember spectra from HyMap data. The tool benefits from the high spectral resolution which yields the potential for the identification of materials. The development of the tool consists of three steps. First, a comprehensive spectral library is built using several HyMap scenes. These data are necessary to make the tool more robust against effects resulting from atmospheric correction, image calibration, age of materials or illumination. The next step represents the definition of distinctive features based on the spectral information. This process is completely controlled by the computer selecting optimal features for an improved separation of materials. Since hundreds of features are defined an appropriate classification scheme is developed in the third step.

First results show that the tool is capable to automatically extract representative spectral endmembers from an unknown image scene. Especially roof materials showing a high spectral variability are well identified. All materials with a low spectral variability and characteristics are barely identified due to small spectral offsets caused by the atmospheric correction. However, the classifier is capable to learn new variations and thus will improve by adding more spectra from new image scenes in future.

Keywords: Endmember identification, spectral features, classification, spectral unmixing

INTRODUCTION

Image classification or spectral unmixing techniques are used to determine the surface materials and their fractions within the urban environment. Based on this information higher-order products such as degree of surface sealing, vegetation covers or composition of biotopes are derived. This makes it necessary to define spectral endmembers which is a time-consuming process especially for hyperspectral data. The use of endmember information from previous classified images is hardly possible, due to altering spectral characteristics caused by

different illumination, age and image preprocessing. The resulting variability also prevents the use of standard spectral libraries as often used within geological applications. Experienced users prefer the building of a so called “local spectral library” that contains only spectra from the image of interest. These spectra will yield the most accurate results as they include all scene-specific spectral characteristics.
In order to simplify the endmember selection procedure, techniques such as PPI (i), N-Finder (ii), ORASIS (iii), IEA (iv), and AMEE (v) were developed. These techniques try to select the purest pixels based on the analysis of spectral patterns within the image hypercube or in case of AMEE in combination with a spatial pattern analysis. Independent of their advantages and limitations they all represent unsupervised techniques and the resulting endmember spectra still have to be identified in an additional time-consuming process. Such techniques represent only a partial improvement.
In this investigation the effort is made to build a completely automatic endmember selection and identification procedure using hyperspectral image data. The focus lies on a first complete spectral exploitation of HyMap data in regard to an optimal differentiation of urban surface materials. The use of ancillary information e.g. high resolution spatial patterns (2/3D) or thermal data would greatly simplify this process but is not part of this study.

TEST SITES AND DATA

To determine the limitations of material identification it is necessary to know the spectral variability of urban materials. For this purpose spectra were collected from 6 HyMap scenes covering the cities of Dresden and Potsdam, Germany. The data were recorded during flight campaigns carried out by the DLR (Deutsches Zentrum für Luft und Raumfahrt) from 1999 to 2004. The pixel size varies between 3-6 meters. Atmospheric correction was performed by ATCOR 4 and an additional program and optimised using field spectra within an empirical line correction. More than 30.000 spectra were extracted as training data from five HyMap scenes while the sixth scene serves for the extraction of control spectra. The resulting spectral library includes material specific variations and data processing effects. The used materials are listed in Tab. 1.

Table 1: Surface categories and materials used for endmember detection.

roof materials	tiles (new), tiles (old), concrete, aluminum, zinc, copper, PVC, polyethylene, glass, plexiglass, bitumen bright/dark/red, tar-paper, schist, vegetation, gravel, facade, one still unknown material (other)
fully sealed materials	concrete, asphalt, tartan track, synthetics
partially sealed materials	cobblestone pavement, concrete, red /dark loose chipping trails
bare ground	sand, soil
water	river, pond, pool
vegetation	deciduous trees, coniferous trees, lawn, meadow, dry grass, field tilled, field untilled, fallow
shadow	falling on vegetation and non-vegetation

IDENTIFICATION APPROACH

Determination of material specific features

The spectra of the training library are used to build a classification system. First, material-specific features have to be defined that allow an optimal separation of materials. They will serve as the classification input and help to estimate the classifier's parameter set. Fig. 1 shows selected spectra of polyethylene. Their strong variability excludes the use of pure reflectance values which are used in most remote sensing studies. Therefore, numerical features will be used (e.g. ratios, absorption depth, coefficients of polynomial fit) that minimize the spectral variability for an optimal identification.

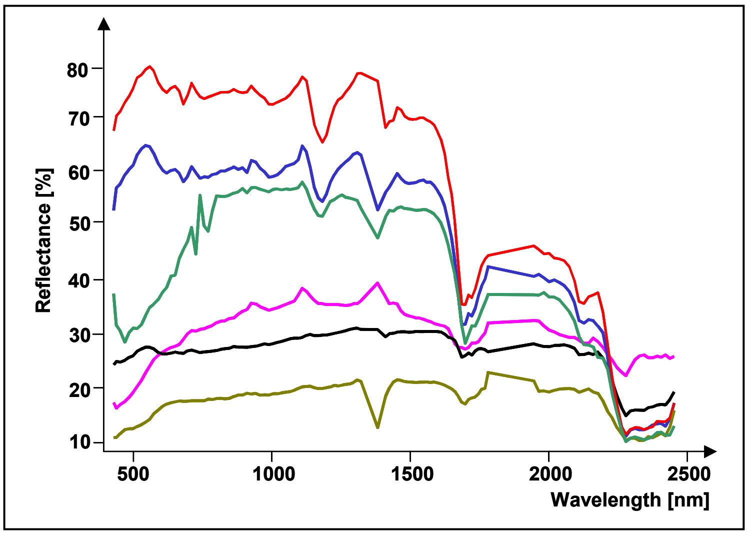

Figure 1: Selected spectra of polyethylene.

The problem is now to determine these features. Due to the complexity of this task this process is completely carried out by the computer. To exploit the full potential of the HyMap spectra more than 128.000 different features such as means, stdevs, ratios, areas, polynomial coefficients and RMS of line and curve regression, positions and depths of absorption and reflection maxima are calculated for each spectrum of the training data. Only neighbouring bands with a varying spectral position and number (2-126) are used for the calculation. During this process a preselection of significant features is already performed. Each feature is evaluated towards an optimal separation of two different materials. Only the best 20 results of each feature type and material combination are stored in a matrix structure.
The best 1000 features are finally selected for an optimal separation of all materials. An appropriate solution of this highly complex optimisation problem is achieved by applying the following iterative procedure: the first feature is selected that allows the best separation between two materials. The next feature should not correlate with the first one to improve the separability. This requirement can be fulfilled by the analysis of the confusion matrix classifying all training spectra based on the first feature. Tests showed that a classifier such as the parallelepiped classifier is well suited for this task due to a rough approximation of the classes in the feature space. This allows a further differentiation of best features that is not possible with high advanced classifiers. Such techniques evaluate much more features as absolutely perfect for separation. The material combination within the matrix that shows the lowest separability is identified. The corresponding feature with the best separation for this combination is selected as the second one, because it shows no correlation to the first one. Then a new confusion matrix is calculated based on the first and second feature and the third

feature is selected in the same way. This process is stopped until the maximum number of best features (user-defined) is reached. To exclude infinity loops a threshold is necessary to eliminate material combinations from the optimisation procedure that cannot be separated after 20 iterations.

Classification system

The classifier must be able to process a very high number of features with high accuracy. Another requirement is that the algorithm should be restrictive towards unknown spectral variations of materials and declare them as unknown. However, it should also allow the incorporation of new spectra that will improve the accuracy and enhance its applicability to new image scenes. This is very important since only 5 image scenes are used in this investigation containing a subset of possible spectral variations.
These requirements favour the use of a parallelepiped classifier. Although it shows a limited separability in low dimensional feature spaces, it greatly benefits from the high number of features that exclude most other techniques. Additionally, the use of feature values instead of pure reflectance values represents a problem for most classifiers due to their extreme different range. A high number of features improves the separability of most materials, but there still exist exceptions of spectra that are correctly classified as different material classes. To reduce this error, the classification scheme is extended (Fig. 2). All ambiguous pixels (e.g. U) are reclassified using two spectral subclasses of each material (A_1, A_2, B_1, and B_2) that were defined by a k-means clustering. If no distinct result is achieved, this process is continued using 4, 8, 16, and so on subclasses. The use of a high number of subclasses increases the accuracy of the classifier since the feature space is subdivided more precisely compared to the material's distribution. Few pixels that still fulfill the requirements of more than one class will be marked as ambiguous.

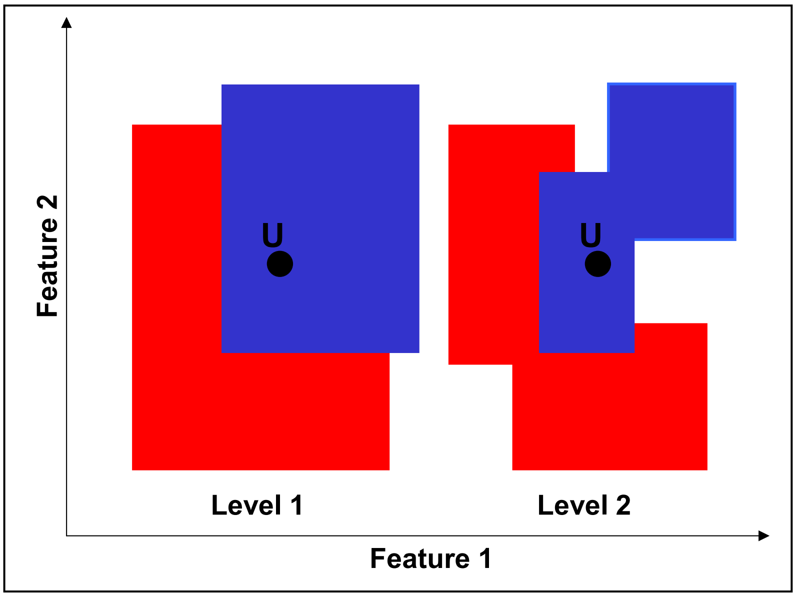

Figure 2: Modified parallelepiped classification scheme using subclusters on different levels.

Object analysis and hyperspectral clustering

This classification system is used to identify pixels in an unknown image scene. The use of all classified pixels as endmembers for further image processing is not possible due to their high number. In a first step only objects larger than a material specific minimum size (3-30 pixels) are considered as significant and thus accurate. Smaller objects probably include a higher fraction of mixed pixels. The position of the detected materials is used to compute representative endmember spectra from the HyMap scene. Cluster algorithms can be used to identify important sub-categories in the multidimensional scatter plot of each material. In this approach, a cluster technique was used which minimizes the number of meaningful subclasses and optimizes gaussian distributions (vi). The algorithm uses a multivariate test to check the normality of the single subclusters. It also directs further cluster splittings in the high dimensional feature space. The maximum number of subclusters varies between 5-20 depending on the spectral variability of the material. It also has the advantage of minimizing the influence of sparely miss-identified pixels.

RESULTS AND OUTLOOK

The new endmember identification tool was tested using another HyMap image covering a subset of Dresden, Germany (Fig. 4). Within this procedure the materials of about 6 % of the image pixels were identified. For evaluation purpose an accuracy assessment was performed.

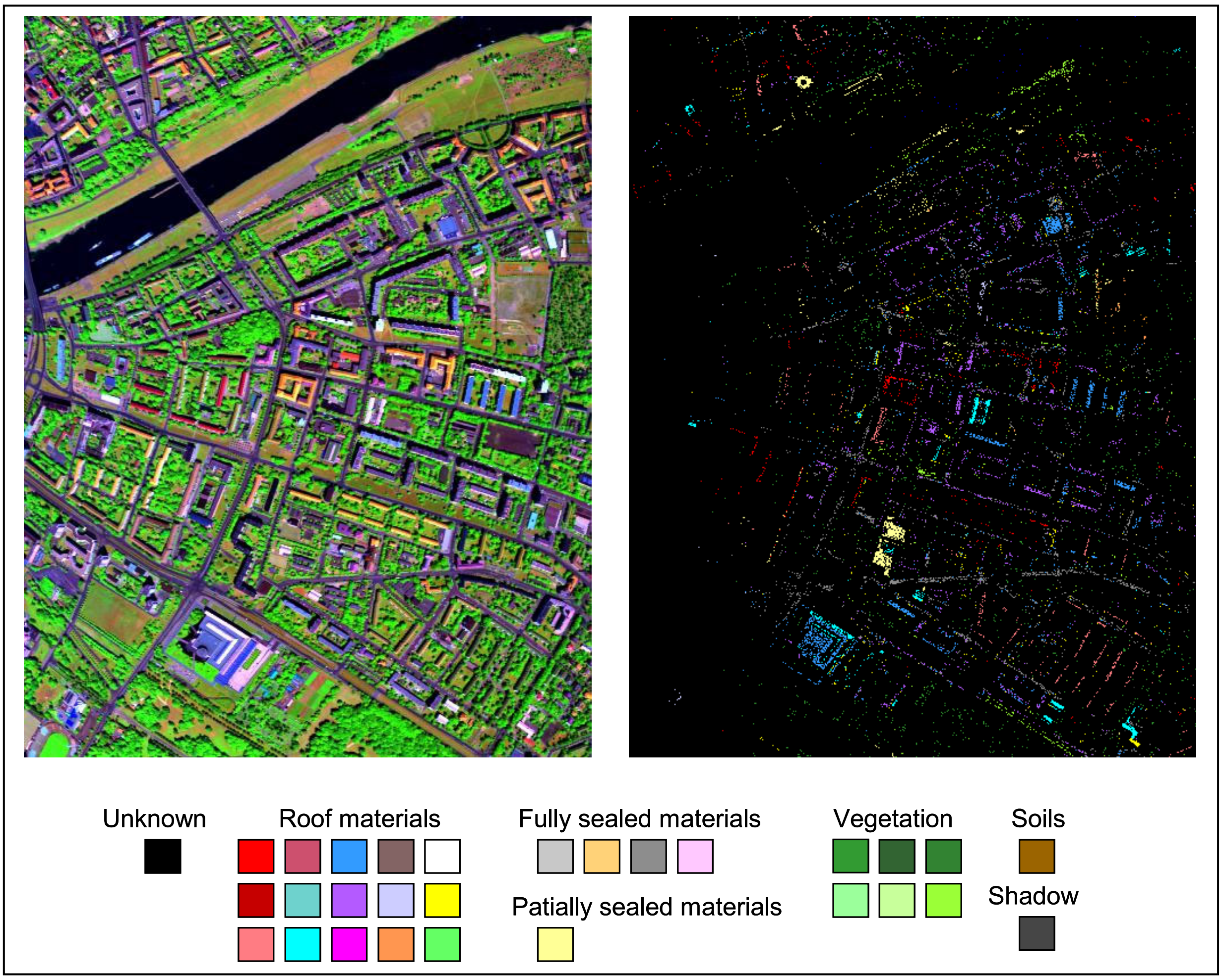

Figure 3: Dresden: Hymap RGB (bands 109/25/3) and identified pixels.

Control regions were marked in the image for all materials and used to calculate a confusion matrix. The resulting errors of omission and commission were calculated for these materials. Tab. 2 contains the means of all materials.

Table 2: Accuracy assessment of endmember selection.

Average error	Training data	Training and control data
Omission Commission	94.8 % 1.9 %	0.0 % 0.0 %

The errors of omission and commission cannot be interpreted in the same way as for a supervised image classification due to different objectives. Within a classification the most likely material is selected based on high accuracy training data. The objective of the presented identification procedure is to extract such training data. These data approximate only a representative cross-section of the image characteristics. This means that a high omission can be partially tolerated as long as suitable spectra are identified. However, a low commission error is more important describing how many identified spectra are miss-identified.
The omission of 94.8 % means that only 5.2 % of the control pixels were identified (producer accuracy). The best results (10 - 50 %) could be achieved for materials that include a high degree of spectral variation such as most roofing materials (tiles, metals, and synthetics). Other materials are barely identified (< 10 %). This problem can be explained by smaller spectral offsets induced by the atmospheric correction. This new variations still exceed the variations known by the classifier so far. A good example represents the water classes with 0.1 % identified pixels. In contrast, these offsets show a minor impact on the identification of spectrally variable materials. To improve the identification accuracy in future, the system has to consider new variations, because the used 5 HyMap scenes still reflect only an excerpt of possible spectral variations. In a second test the control and training data were added and used to define a new identification system. The resulting omission and commission error of 0.0 % proofs the capability of the system to adapt.
For the more important commission error an average value of only 1.9 % could be achieved representing an excellent result. The maximum values (3 - 25 %) can be found within roofing tar, shadow, bright bitumen and red loose chipping trails that possess similar constituents as other materials.

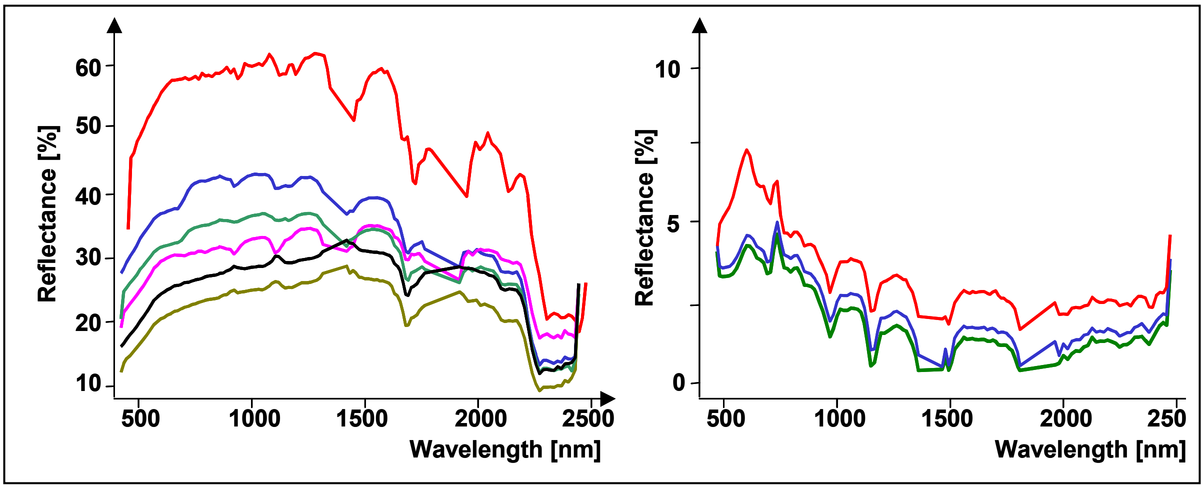

Figure 4: Selected endmember spectra representing polyethylene (left) and water (right).

Based on the identified spectra representative endmembers were calculated using the cluster algorithm. Fig 4. shows exemplary selected endmember spectra representing polyethylene and water. Polyethylene is a highly variable material and several typical spectral variations were extracted. Water is highly problematic, but the extracted endmembers allow an overall good classification of water pixels. Altogether 190 spectra were identified. Such a high number of endmembers can be used e.g. in spectral unmixing techniques as described in vii and viii. Missing endmembers have to be determined by e.g. a standard classification using a threshold defining classes with a high likelihood. A first test shows that about 79 % of the image pixels were well classified. The remaining unknown pixels, forming larger areas, represent candidates for new endmember spectra. These already promising results will be improved including additional HyMap scenes in future. Further improvements can also be expected by the incorporation of thermal data as provided by the new airborne sensor ARES (ix).

ACKNOWLEDGMENTS

This work was made possible by several flight campaigns carried out by the Deutsches Zentrum für Luft- und Raumfahrt Oberpfaffenhofen, Germany.

REFERENCES

i Boardman J W, Kruse F A & Green R O, 1995.Mapping target signatures via partial unmixing of AVIRIS data: in summaries, Fifth JPL Airborn Earth Science Workshop, JPL Publication 95-1,1, 23-26

ii Winter M E, 1999. N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data, Proc. SPIE Imaging Spectrometry, V, 266-275

iii Bowles J, Palmadesso P J, Antoniades J A, Baumback M M & Rickard L J, 1995. Use of filter vectors in hyperspectral data analysis, Proc. SPIE Infrared Spaceborne Remote Sensing, III, 148-157

iv Staenz K, Szeredi T & Schwarz J, 1998. ISDAS - A system for processing/analyzing hyperspectral data, Canadian Journal of Remote Sensing, 24, 99-113

v Plaza M, Martinez P, Peres R & Plaza J, 2002. Spatial/spectral endmember extraction by multidimensional morphological operations.

vi Segl K, 1996. Integration von Form- und Spektralmerkmalen durch künstliche neuronale Netze bei der Satellitenbildklassifizierung.Deutsche Geodätische Kommission, Reihe C, Nr. 468,ISBN 3-7696-9508-9, München

vii Roessner S., Segl K, Heiden U & Kaufmann H, 2001. Automated differentiation of urban surface based on airborne hyperspectral imagery, IEEE TGARS, 39/7, 1523–1532

viii Segl K, Roessner S, Heiden U, & Kaufmann H, 2003. 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, 99-112

ix Richter R, Müller A, Habermeyer M, Dech S, Segl K & Kaufmann H, 2005. Spectral and radiometric requirements for the airborne thermal imaging spectrometer ARES, International Journal of Remote Sensing, 26/15,3149-3162