

Interactive Classification of Urban Areas Using Decision Trees

S. Abelen^{1,2}, H. Taubenböck¹, U. Stilla²

¹German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling-Oberpfaffenhofen, Germany

²Photogrammetry and Remote Sensing, Technische Universitaet Muenchen (TUM), Munich, Germany

Abstract— Conventionally decision trees are fixed structures for sequential classification, which are designed for certain regions and specific research questions. In heterogeneous and dynamic urban environments their broad application requires a continual change of their structure, which is time consuming and labor-intensive. This study focuses on the development of a user interface to facilitate the interactive adaption of decision trees. The platform of the user interface is composed of fixed feature sets which are equally applied to all scenes. They are selected on the basis of the Transformed Divergence. The features' thresholds are connected to controllers, which can be adapted by the user. For assessing the effectiveness of the user interface, its classification performance is compared to the one of a decision tree with fixed thresholds. By means of Landsat 7 imagery four land-cover classes are distinguished. Results show that in all analyzed test-sites the overall accuracy lies for adjusted thresholds above 80% and is by up to 33% higher than for fixed thresholds. Therefore the user interface proved to be more efficient in classifying a broad variety of scenes in urban environments than a decision tree with fixed thresholds.

I. INTRODUCTION

Triggered by the development of commercial software (eCognition, ERDAS) decision trees are increasingly used for land-cover classification with remotely sensed data. Several studies have also been conducted in urban areas using decision trees [1] [2] [3]. Decision trees are composed of hierarchically structured decisions, which have to be traced, when classifying each segment or pixel of an image. Conventionally decision trees are based on fixed decision rules and thresholds, which are often only feasible for one specific scene. If new knowledge was available or the target question changed, the decision tree would have to be redesigned by an expert [4]. Especially in urban environments where the spectral and spatial characteristics vary within and among cities [5], such classification algorithms are labor-intensive and require expert knowledge.

This study aims at the development of a user interface (UI), which allows non-experts to classify independent data sets of cities with only one decision tree platform. While the feature sets stay unchanged for different kinds of images, the threshold values can be interactively adapted by the user. The prototype is designed on the basis of the eCognition Developer 8 software (Architect Solution) for Landsat ETM+ data and distinguishes four land-cover classes: water, vegetation, (bare) soil and impervious surfaces. The paper will point out the

benefits and limitations of classifying a broad set of scenes via a UI and compare the accuracies achieved with those of a conventional decision tree with fixed thresholds.

II. METHODOLOGY

Fig. 2 shows the structure of the decision tree, which we implemented within the UI for land-cover classification [6]. After a multi-resolution segmentation, the classes are identified hierarchically, starting with classes of significant separability from other classes and ending with those of lower separability. The four-step approach allows the user to adjust the threshold values successively, lowering the complexity of threshold optimization. Remaining unclassified segments are iteratively assigned to the class of the neighboring segment, starting with the class of impervious surfaces and advancing back to the one of water.

For the design of the classification platform feature sets have to be identified which effectively distinguish a certain class in different scenes. We therefore analyzed the spectral patterns of ten different features on the basis of four Landsat ETM+ test-sites and their reference images. Eight of these features stem from the Landsat ETM+ sensor and are the blue, green, red, near infrared (NIR), short-wavelength infrared one (SWIR1), thermal infrared one (TIR low gain mode), thermal infrared two (TIR high gain mode) and the short-wavelength infrared two (SWIR2) band. The remaining two features are the vegetation indices NDVI (Normalized Difference Vegetation Index) [7] and SAVI (Soil Adjusted Vegetation Index) [8]. For the SAVI we chose an adjustment factor of $L = 0.5$, as proposed by [8].

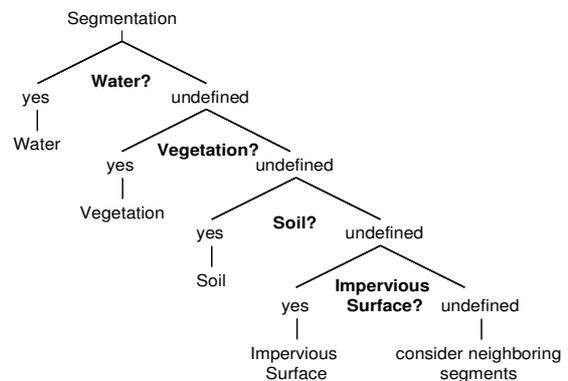


Figure 1. Step-wise structure of the classification algorithm [6].

Feature selection involves feature prioritization, excluding irrelevant information, and feature decorrelation, removing redundant information [9]. For feature prioritization we calculated the Transformed Divergence (TD) for all potential feature sets of each class. The TD is a statistical distance measurement and describes the difference of spectral signature-derived probability distributions for different classes [10]. It originates from the Divergence D_{F1_F2} between two classes A and B based on Feature 1 and Feature 2 [11]:

$$D_{F1_F2} = 0.5 \operatorname{tr} \left[(V_A - V_B)(V_B^{-1} - V_A^{-1}) \right] + 0.5 \operatorname{tr} \left[(V_A^{-1} - V_B^{-1})(M_A - M_B)(M_A - M_B)^T \right] \quad (1)$$

where V_A is the covariance matrix of Feature 1 and Feature 2 for Class A, V_B is the covariance matrix between Feature 1 and Feature 2 for Class B, M_A is the mean vector of class A, M_B is the mean vector of class B and $\operatorname{tr} [\]$ is the trace of the matrix. Higher values indicate better separability than lower values. If more than two features are applied, the covariance matrix and the mean vectors increase in their dimension. In the case of the UI, we set the maximum number of features to three, assuming that an adjustment of more than three features per class is time-consuming and complex for the user.

To take saturation effects into account, the Divergence was altered to the TD [11], which we used for the feature selection of this study:

$$TD_{F1_F2} = 2000 \left[1 - \exp \left(\frac{-D_{F1_F2}}{8} \right) \right] \quad (2)$$

The TD scales the distance values of the Divergence so that no values higher than 2000 are reached.

If more than two classes are to be distinguished with the same feature set, the TD values of different class combinations are averaged [4]. For example if one wanted to distinguish class A from class B and class C, then the TD value for class A (TD_A) would be the average of the TD value to distinguish class A from class B (TD_{A_B}) and the TD value to distinguish class A from class C (TD_{A_C}). For the UI we also took the hierarchical structure of the decision tree into account and introduced a system of weighting. Thereby the distance values of previously classified classes loose in significance by a factor of three:

$$TD_{_wat} = \frac{3 TD_{wat_veg} + 3 TD_{wat_soil} + 3 TD_{wat_imp}}{9} \quad (3)$$

$$TD_{_veg} = \frac{TD_{veg_wat} + 3 TD_{veg_soil} + 3 TD_{veg_imp}}{7} \quad (4)$$

$$TD_{_soil} = \frac{TD_{soi_wat} + TD_{soi_veg} + 3 TD_{soi_imp}}{5} \quad (5)$$

$$TD_{_imp} = \frac{TD_{imp_wat} + TD_{imp_veg} + TD_{imp_soi}}{3} \quad (6)$$

For feature decorrelation we calculated the Pearson product momentum correlation coefficient for each feature pair within a class-specific feature set. If two features correlated in all possible scene pairs by more than 99%, we removed the feature with the lower average TD value from the set.

On the basis of the feature selection process we chose one feature set for each class to be implemented in the UI. Afterwards we assigned to each feature upper and/or lower controllers, which we adjusted to each individual test site, yielding four site-specific threshold sets. A conventional, fixed threshold set was simulated by applying the average of the site-specific thresholds to all test-sites. Furthermore we combined unrelated test-sites and threshold sets (e.g. applying the threshold set of Test-site 1 to Test-site 2). For the evaluation of the classification performance we derived the user's, producer's and overall accuracy [4] [12]. To test the UI in terms of an independent scene, we finally applied it to a fifth Landsat test-site, which had not been involved in the design of the classification platform.

III. STUDY AREA AND DATA SETS

To reflect spectral similarities and differences between urban Landsat ETM+ images we selected four test sites of $15 \times 15 \text{ km}^2$ (Fig. 1). Three of them are located in Mexico City, Mexico. The first one shows the city center (S1 Jan 02). The second and third one are located at the northern urban fringe of the city and only differ in their time point of acquisition (S2 Jan 02 and S2 Apr 02). The fourth test site shows a suburban area of Istanbul, Turkey (S3 Mar 02). The fifth one (150×450 pixels), which was not integrated in the feature selection process, is located at the urban fringe of Cairo, Egypt.

For each test-site we created manually a reference image with the land cover type of each segment. As S2 Jan 02 and S2 Apr 02 only differ in the date of acquisition, we generated only one common reference image. Table 1 highlights the appearance of each land cover type in the four test-sites.



Figure 2. (a) Overview on the selected test-sites of Landsat 7 and exemplarily (b) test-site S2 Jan 02 in RGB+NIR and (c) test-site S3 Mar 02 in RGB+NIR.

TABLE I. DISTRIBUTIONS OF SEGMENTS FOR EACH TEST SITE

| Test-site | Water | Vegetation | Soil | Imp. Surf. | Total Nr. |
|-----------|-------|------------|------|------------|-----------|
| S1 Jan 02 | 0.2% | 8.2% | 3.5% | 88.1% | 11204 |
| S2 Jan 02 | 0.5% | 41.6% | 1% | 56.9% | 11879 |
| S3 Mar 02 | 6.5% | 40.8% | 0.7% | 52% | 7805 |

IV. RESULTS

Fig. 3 shows in class-specific plots for each test-site the features with the four highest TD values. Higher ranks indicate a higher suitability of the feature to distinguish the class from others. For the class of water, the features indicated differ strongly among the test-sites. For the remaining classes at least one feature can be found which is in all test-sites among the four highest ranks. The highest ranked features take on average over different test-sites the following TD values: water: 1688, impervious surfaces: 1211, vegetation: 1102, soil: 747.

Based on the feature selection process the feature sets of Table 2 were implemented in the UI. The TD values of those features differ at maximum by 10% from the highest distance value reached by a site-specific feature. The lowest TD for water lies (with respect to all test-sites) at 1782, for vegetation at 1426, for soil at 1229, and for impervious surfaces at 1438. Due to feature decorrelation the features of one class correlate at most by 99%.

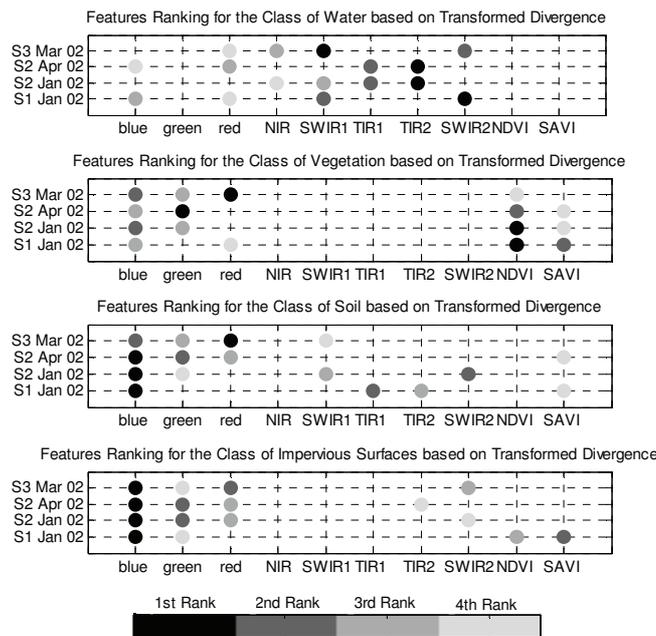


Figure 3. Features with the four highest distance values in different test-sites.

TABLE II. SELECTED FEATURE SETS FOR THE USER INTERFACE

| Class | Final Feature Sets | | |
|---------------|--------------------|-----------|-----------|
| | Feature 1 | Feature 2 | Feature 3 |
| Water | SWIR2 | SAVI | - |
| Vegetation | blue | NDVI | - |
| Soil | blue | red | NDVI |
| Imp. Surfaces | blue | red | SAVI |

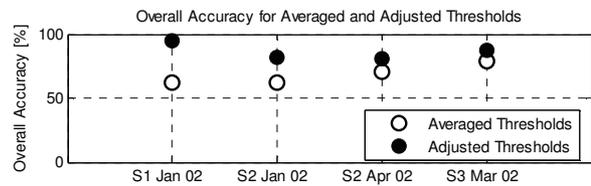


Figure 4. Overall accuracies for fixed and adjusted thresholds.

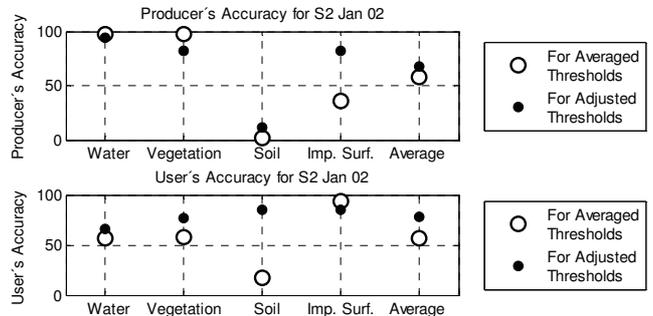


Figure 5. User's and producer's accuracy for S2 Jan 02.

The threshold adjustment for the chosen feature sets gave the overall accuracies shown in Fig. 4. The graph shows that the overall accuracy is always higher for thresholds which have been optimized for the target area and lower for averaged thresholds which are uniformly applied to all test-sites. The highest difference amounts to 33% for S1 Jan 02. While for adjusted thresholds the overall accuracy always lies above 80% for averaged thresholds it can be as low as 62% (S2 Jan 02).

The results for the user's and the producer's accuracy are shown exemplarily for S2 Jan 02 in Fig. 5. Both accuracies are generally higher for adjusted thresholds. However, for the class of impervious surfaces averaged thresholds lead to a higher user's (by 8%) but lower producer's accuracy (by 47%). Also for vegetation averaged thresholds lead to a higher producer's accuracy. For the class of soil with the least segments in the image the producer's accuracy is in both cases very low (<15%). The problem of identifying classes of low occurrence could also be observed in other test-sites where for example only few segments of water exist (e.g. S1 Jan 02, Table 1).

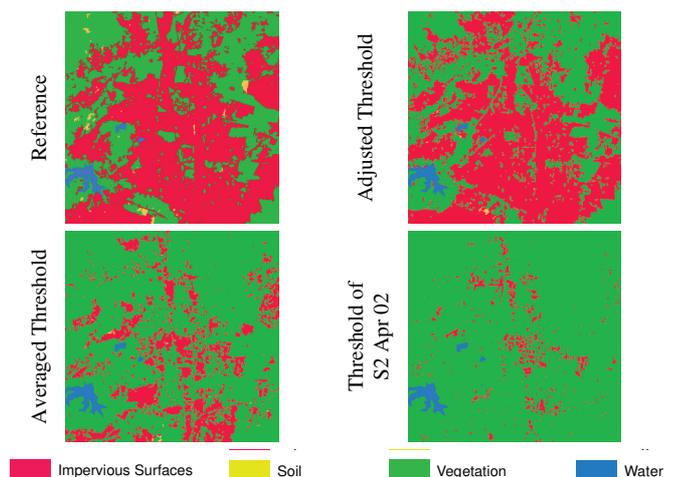


Figure 6. Various classification results for S2 Jan 02.

Fig. 6 shows that the higher producer's accuracy for averaged thresholds in the case of vegetation is based on over-classification (lower left). The resulting under-classification of impervious surfaces is an explanation for the class's high user's but low producer's accuracy for averaged thresholds. Taking the threshold set, optimized for the same area but a different time point (three months later), a similar outcome can be observed (lower right). The threshold set produced for S2 Apr 02 again over-classifies the vegetation cover, when applied to S2 Jan 02. Using this set for its descend-test-site (S2 Apr 02) this trend is not observed. Further combinations show that the sun elevation angle at the time point of acquisition has most influence on the threshold setting. For example in S2 Jan 02 and S2 Apr 02 the sun elevation angles differ between 39.2° (January) and 61.6° (April).

For the independent test-site of Cairo we reached an overall accuracy of 75% within 30 minutes of threshold adaptation. With another 10 minutes of manual classification the overall accuracy rose to 86%. The manual classification was specifically beneficial for the class of soil, where the producer's accuracy rose from 36% to 73%. Taubenböck et al. applied the UI to Landsat data of Mexico City and three further Mexican cities: Guadalajara, Monterrey and Puebla. They thereby reached on average an overall accuracy of 94% [13].

V. DISCUSSION

In the analyzed test-sites the class of water reaches higher separability values than any other land-cover type. At the same time no feature can be found which is among the four highest ranks in all test-sites. Based on this observation we find two main guidelines to be effective in identifying a suitable feature set for the implementation in the UI. Firstly the TD value, telling how much the probability distributions differ between classes, should be maximized. Secondly the difference of the distance values for the most suitable site-specific feature and the uniformly applied feature should be minimized.

Adjusting the thresholds for the selected feature sets to each test-site, the classification accuracy increases by up to 33% and lies for all test-sites with adjusted thresholds above 80%. Therefore two objectives for the functioning of the UI are met. Firstly the UI is more efficient in classifying a variety of scenes than a fixed decision tree. This is especially true when the sun angles differ between scenes. Secondly the overall accuracy achieved does not fluctuate significantly between scenes. Therefore the UI can be applied to diverse scenes, differing in location, state of urbanization and time point of acquisition.

The limitations of the UI become apparent through the poor users' or producers' accuracies for classes of low occurrence. In the case of soil, the poor accuracies are based on the class's low TD values. For the class of water the TD values are high, but its spectral pattern overlaps with the one of the frequently occurring class of vegetation. As consequence the shared part of the feature space is assigned to the class of vegetation with the larger amount of segments or pixels to keep the overall accuracy high. Consequently two main limitations to the UI can be identified. Firstly difficulties of distinguishing classes with similar spectral signatures cannot be overcome. Secondly classes of low occurrence are likely to be masked by broadly

occurring classes to keep the overall accuracy high. Manual classification is specifically effective for the identification of less frequently occurring classes. This was shown in the fourth test-site of Cairo, where manual classification of 10 minutes raised the producer's accuracy by 37% for the class of soil.

VI. CONCLUSION AND OUTLOOK

The study has shown that for the classification of a broad spectrum of scenes, decision trees with adjustable thresholds yield higher overall, producer's, and user's accuracies than conventional ones, lowering over- and under-classification. Especially for scenes that have been acquired at different sun-elevation angles the improvement in classification accuracy is significant. Therefore the UI is expected to be advantageous for multi-temporal and multi-spatial image analysis.

The classification outcome of the UI depends on the user's skills to adjust the platform settings to each scene. Therefore in the future features should be developed, which simplify the threshold optimization e.g. by providing default values.

VII. ACKNOWLEDGMENT

The authors thank RapidEye for providing reference data of Mexico City within the project "Urbane-Struktur-Analyse mit RapidEye, USARE" at DLR-DFD. Further thanks go to the team of Achim Roth at DLR-DFD for their broad support.

REFERENCES

- [1] D. Lu and Q. Weng, "Spectral mixture analysis of the urban landscape in Indianapolis with Landsat ETM+ imagery," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 9, pp. 1053–1062, 2004.
- [2] S. McCauley and S. J. Goetz, "Mapping residential density patterns using multi-temporal Landsat data and a decision-tree classifier," *International Journal of Remote Sensing*, vol. 25, no. 6, pp. 1077–1094, 2004.
- [3] H. Taubenböck, T. Esch, M. Wurm, A. Roth, and S. Dech, "Object-based feature extraction using high spatial resolution satellite data of urban areas," *Journal of Spatial Science*, vol. 55, no. 1, pp. 117–133, 2010.
- [4] J. Jensen, *Introductory digital image processing: a remote sensing perspective*, 3. ed. Upper Saddle River N.J.: Prentice Hall, 2005.
- [5] Q. Weng and D. Lu, "Subpixel analysis of urban landscapes," in *Urban remote sensing*. CRC Press, 2007.
- [6] H. Taubenböck, "Vulnerabilitätsabschätzung der erdbebengefährdeten Megacity Istanbul mit Methoden der Fernerkundung," Bayerische Julius-Maximilian Universität, 2008.
- [7] J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, "Monitoring vegetation systems in the great plains with ERTS," in *Third ERTS Symposium, NASA SP-351 I*, vol. 351, pp. 309–317, 1973.
- [8] A. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sensing of Environment*, vol. 25, no. 3, pp. 295–309, 1988.
- [9] M. Dash and H. Liu, "Feature selection for classification," *Intelligent data analysis*, vol. 1, no. 3, pp. 131–156, 1997.
- [10] C. Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. Springer, 2003.
- [11] P. Mausel, W. J. Kramber, and J. K. Lee, "Optimum band selection for supervised classification of multispectral data," *Photogrammetric Engineering and Remote Sensing*, vol. 56, no. 1, pp. 55–60, 1990.
- [12] P. K. Varshney and M. K. Arora, *Advanced image processing techniques for remotely sensed hyperspectral data*. Springer, 2004.
- [13] H. Taubenböck, M. Klotz, A. Felbier, M. Wegmann, and R. Ludwig, "Spatio-temporal cross-city comparison using multi-sensoral remote sensing for Mexican cities," *Joint Urban Remote Sensing Event 2011* (to appear in the same proceedings).