

# GETTING THE ACT TOGETHER: SEGMENTATION-BASED LAND COVER CLASSIFICATION USING RAPIDEYE IMAGERY AND OPEN STREET MAP ANCILLARY DATA

L. Valozic<sup>a</sup>

<sup>a</sup> University of Zagreb, Faculty of Science, Department of Geography, Marulicev trg 19, 10000 Zagreb, Croatia - lvalozic@geog.pmf.hr

**KEY WORDS:** Land Cover Classification, OBIA, Rule-Based, GIS, RapidEye, Open Street Map

## ABSTRACT:

The research deals with land cover classification by means of segmentation-based image analysis and geoprocessing in GIS software. The goal of the research is the overall improvement of classification results obtained from rule-based land cover classification, by utilising ancillary vector data. The area of interest is the City of Zagreb in Croatia. This research is primarily based on 2015 RapidEye satellite imagery. The assumption that the mixture of urban, peri-urban, rural areas, that makes the territory of the City of Zagreb, is sufficiently interwoven with OpenStreetMap data was confirmed after examination of acquired shapefiles. Intent of the research was to find a solution for land cover classification of geographically heterogeneous region by reconciling the need for convenient and affordable spatial data. Preprocessing and postprocessing of spatial data, as well as the principle component analysis of the imagery, were performed in ESRI ArcGIS. The classification process was performed using Trimble eCognition Developer. Rule-based classification was employed on image segments that were created by the multiresolution segmentation algorithm. Threshold values for statistical features, vegetation indices, and values derived from PCA were set after careful examination of all input datasets. Overall classification accuracy of 91.5% was assessed by error matrix. Imagery of higher spatial resolution were used as reference data for the error matrix.

## 1. INTRODUCTION

### 1.1 Aim and Objective

The aim of this research is the successful land cover classification of urban and peri-urban areas by means of object-based image analysis of remotely sensed imagery. The evaluation criteria for the success of the land cover classification have been set as the overall accuracy of the classification, as well as the classification accuracies for the individual land cover classes, to be equal or more than 85 percent (Anderson, 1971).

The purpose of such classification tasks is the improvement of the spatial awareness of apparent geographic dichotomies of the urban versus rural, or artificial versus natural (Besussi et al, 2010).

The objective of this research work was the production of a land cover dataset by means of segmentation-based image analysis of multispectral satellite images with the support of ancillary GIS vector data layers.

Final land cover dataset created by classification consists of four classes: vegetation, impervious or built-up surfaces, bare soil, and water. Classification scheme is based on vegetation - impervious surface - soil (V-I-S) model for urban ecosystem analysis (Ridd, 1995).

### 1.2 Study Area

The study area is the administrative area of the City of Zagreb in the Republic of Croatia. This county-level administrative unit is the country's capital and the most populous city with almost 800 000 inhabitants. The City of Zagreb has an area of 641 sq km and is comprised of 70 urban, peri-urban, or even rural settlements. Even though this area has been characterised by a

considerable growth in population and built-up surfaces during the last several decades, significant parts of this administrative unit are either agricultural areas or inside the protected area of Medvednica Nature Park and are covered with forests. Most of the water surfaces in the City of Zagreb are highly regulated rivers and streams, or artificial lakes and ponds.

### 1.3 Data

The choice of suitable remote sensing data requires the settling of multiple criteria such as the quality and the quantity of data, as well as the financial costs. Two main data sources were used in the research: archived RapidEye's multispectral satellite imagery from 2015, and shapefiles of OpenStreetMap data. Former dataset comes from a commercial remote sensing vendor. RapidEye images have been chosen because of relatively convenient combination of scene tile size (25 x 25 km), spatial resolution (5 m) and spectral resolution (blue, green, red, red edge, near infrared). Five image scenes acquired for this work (3A) have already been geometrically and radiometrically corrected. Latter dataset, the free and easily obtainable OpenStreetMap data shapefiles, have been chosen because they have become a rich data source, and represent the dynamic crowdsourced knowledge of the area of interest (Rosina et al, 2016). More precisely, point, polyline, and polygon features that represent various urban objects, building footprints, and pedestrian, road, or rail networks, have been used as datasets that should help to classify the land cover of City of Zagreb. OpenStreetMap shapefiles have been acquired online from Geofabrik GmbH (Geofabrik, 2015). DigitalGlobe satellite imagery found in GoogleEarth application and digital colour orthophoto of City of Zagreb available in WMS Service (CDOF2012, 2016) were used as reference data for the accuracy assessment.

## 2. METHODOLOGY

### 2.1 Approach

This research was based on processing of raster images and vector datasets, by both remote sensing and GIS software in order to produce the final land cover dataset.

The preprocessing and postprocessing tasks of the datasets were performed in ESRI ArcGIS 10.3. and they include: geometry checks and geometry repairs of OSM polygon shapefiles, selection of features by attributes, overlays of data layers that originate from different sources, and principle component analysis of multispectral imagery. Additionally, GIS software was used for spatial and attribute data management operations, and visualizations, from initial previews to the final representations of the results of the research.

After the geoprocessing in GIS software, spatial data was imported in Trimble's eCognition Developer 9, software that offers a myriad of tools for image segmentation and classification. Imported raster datasets included original RapidEye five-band images and five raster image layers created by PCA. They were joined by OpenStreetMap shapefiles as thematic layers.

Image segmentation was performed by using multiresolution segmentation algorithm (Benz et al, 2004; Trimble, 2014). Original image bands and thematic layers for OSM points, roads, railways, and buildings, were used for the segmentation process. Additional weight (2) was set for the NIR and red edge bands. Scale parameter was set to 75, shape criterion weight to 0.1, and compactness criterion to 0.5.

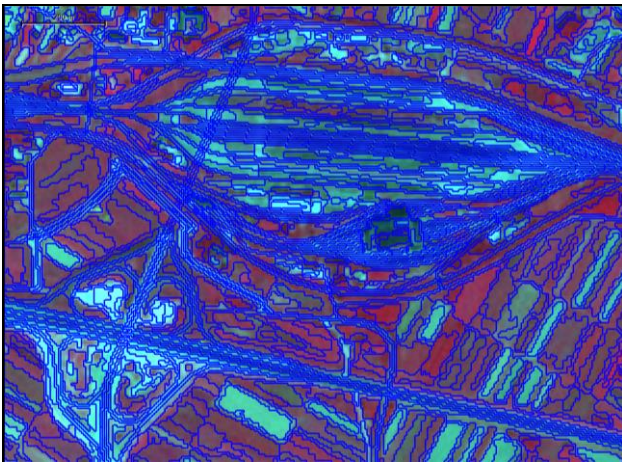


Figure 1. Example of image segments produced by segmentation algorithm executed on RapidEye satellite image and OpenStreetMap data, © 2016 Planet Labs Netherlands BV, © 2016 Geofabrik GmbH, © 2016 Trimble Germany GmbH

The rule-based classification (Zhu, 2009) of image segments in the eCognition environment is the foundational part of this research. A step-wise process tree that assigns image segments to different classes according to determined threshold values has been developed. Careful examination of features and a priori knowledge of the area of interest were crucial for defining of the threshold settings.

Thematic layers imported from OSM shapefiles were used immediately for the classification of impervious surfaces (built-up areas) because they already represented buildings and transport networks. Points and polylines that did not represent artificial surfaces were removed from the dataset in advance by actions in GIS software. Decisions about removal of certain

features were made based on inspection of attribute data and a priori knowledge of the City of Zagreb area.

Most of the other object features that were used for classification are statistical data (mean, SD) on image segments for individual image layers (original image bands and PCA bands), arithmetic features such as ratios between couple or more image bands, HSI transformation, and vegetation or water indices such as SAVI (Huete, 1988), NDWI (McFeeters, 1996), BNDVI (Yang et al, 2004).

Executed process tree resulted with a crisp classification of image segments into land cover classes with binary membership values assigned to them.

### 2.2 Results and Discussion

The final results were vector datasets of four land cover classes for the area of City of Zagreb. Accuracy assessment was performed by error matrix (Story, Congalton, 1986). Reference and classification data samples were extracted on locations that were obtained by stratified random sampling. Colour digital orthophotos of Zagreb area and imagery from Google Earth were used as reference data sources for the purpose of accuracy assessment. As indicated in the Table 1., most of the classification errors occurred between classes of impervious surfaces and bare soil.

%	Impervious Surface	Bare Soil	Vegetation	Water
Producer's Accuracy	88.5	83.3	95	100
User's Accuracy	90	85.7	95	92
Overall Accuracy	91.5			

Table 1. Accuracy assessment results

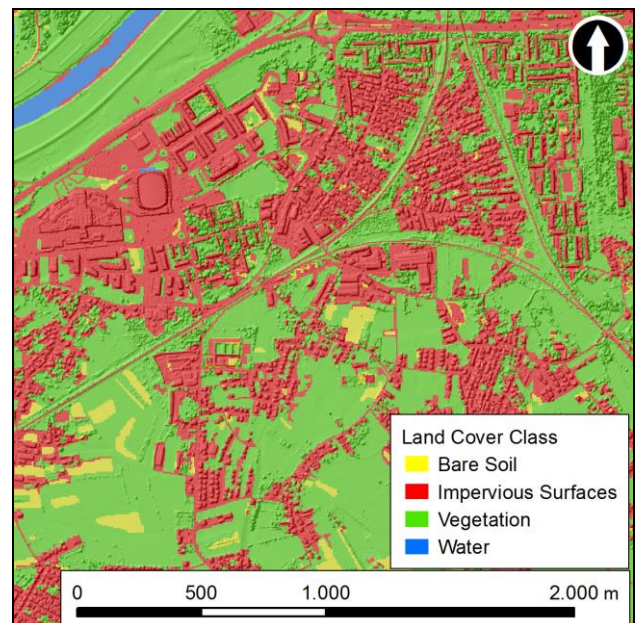


Figure 2. Part of a land cover thematic map of the City of Zagreb produced in the research

Visual inspection of the classification results (Figure 2.) has revealed the main advantage of OpenStreetMap shapefiles' usage as ancillary data for land cover classification. Relatively small or narrow anthropogenic objects that do not come to light

because of spatial resolution or are secluded by e.g. tree canopy, are clearly mapped because of vector datasets and methods used for their processing.

### 3. CONCLUSIONS

On the basis of the accuracy assessment, the land cover classification performed for this research work can be regarded quite successful. The imperfections of both remote sensing raster data and GIS vector data have been settled by their combined use.

Obtained land cover data can be used for several purposes. Primarily, it is used in spatial analyses and thematic mapping in the discipline of urban geography. Such data can help the study of relationships between built-up or artificial surfaces, and vegetation or natural (quasi-natural) land covers. In an environment where naturally occurring bare surfaces are rare, data on bare soil land cover can indicate environmental transition on the observed locations. In combination with population data, even land cover data based on simple classification schemes as one employed in this research, improve the mapping of population distribution and density.

Even though development of rule-based classification requires relatively large amount of time and a certain a priori knowledge of the area of interest, once created rulesets can be executed rapidly and used on similar datasets (similar environments or geographical areas) with just some fine-tuning of threshold values.

### ACKNOWLEDGEMENTS

This research is part of CRORURIS – interdisciplinary project that is focused on the study of changes in the rural areas of Croatia and aims to develop framework for identification and geographical differentiation of predominant trends and key uncertainties as well as their projection by using statistical modelling and Delphi method, and to construct alternative future scenarios and relate them to the context of rural Europe.

### REFERENCES

- Anderson, J. R., 1971. Land-Use Classification Schemes, *Photogrammetric Engineering* 37(4), pp. 379-387.
- Benz, U. C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information, *ISPRS Journal of Photogrammetry & Remote Sensing*, 58, pp. 239– 258.
- Besussi, E., Chin, N., Batty, M., Longley, P. A., 2010. The Structure and Form of Urban Settlements. In: *Remote Sensing of Urban and Suburban Areas, Remote Sensing and Digital Image Processing 10* (ed. Rashed, T., Jürgens, C.), Springer, Dordrecht, Heidelberg, London, New York, pp. 13-31.
- Huete, A. R., 1988. A soil-adjusted vegetation index (SAVI), *Remote sensing of environment*, 25 (3), pp. 295-309.
- McFeeters, S. K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features, *International journal of remote sensing*, 17 (7), pp. 1425-1432.
- Ridd, M. K., 1995. Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities, *International Journal of Remote Sensing*, 16 (12), pp. 2165-2185.
- Rosina, K., Hurbánek, P., Cebecauer, M., 2016. Using OpenStreetMap to improve population grids in Europe, *Cartography and Geographic Information Science*, 43 (5).
- Story, M., Congalton, R. G., 1986. Accuracy assessment - A user's perspective, *Photogrammetric Engineering and remote sensing*, 52 (3), pp. 397-399.
- Trimble Documentation, 2014. *Trimble eCognition Developer 9 User Guide*, Trimble Germany GmbH.
- Yang, C., Everitt, J. H., Bradford, J. M., & Murden, D., 2004. Airborne hyperspectral imagery and yield monitor data for mapping cotton yield variability, *Precision Agriculture*, 5 (5), pp. 445-461.
- Zagrebačka infrastruktura prostornih podataka, Gradski ured za strategijsko planiranje i razvoj grada, "Digitalni ortofoto 2012 (CDOF2012)", Zagreb, Croatia [https://geoportal.zagreb.hr/Public/Ortofoto2012\\_Public/MapServer/WMSServer](https://geoportal.zagreb.hr/Public/Ortofoto2012_Public/MapServer/WMSServer) (13 Jun. 2016).
- Zhu, A. X., 2008. Rule based mapping. In: *The handbook of geographic information science* (ed. Fotheringham, A. S., Wilson, J. P.), Blackwell, Malden, Oxford, Carlton, pp. 273-291.