URBAN OBJECT EXTRACTION FROM DIGITAL SURFACE MODEL AND DIGITAL AERIAL IMAGES

D. Grigillo [a, *], U. Kanjir [b]

[a] Faculty of Civil and Geodetic Engineering, University of Ljubljana, Jamova cesta 2, Ljubljana, Slovenia – dejan.grigillo@fgg.uni-lj.si
[b] Institute of Anthropological and Spatial Studies, ZRC SAZU, Novi trg 2, Ljubljana, Slovenia – ursa.kanjir@zrc-sazu.si

Commission III, WG III/4

KEY WORDS: Digital surface model, Building extraction, Object-based image analysis, Object classification, Segmentation

ABSTRACT:

The paper describes two different methods for extraction of two types of urban objects from lidar digital surface model (DSM) and digital aerial images. Within the preprocessing digital terrain model (DTM) and orthoimages for three test areas were generated from aerial images using automatic photogrammetric methods. Automatic building extraction was done using DSM and multispectral orthoimages. First, initial building mask was created from the normalized digital surface model (nDSM), then vegetation was eliminated from the building mask using multispectral orthoimages. The final building mask was produced employing several morphological operations and buildings were vectorised using Hough transform. Automatic extraction of other green urban features (trees and natural ground) started from orthoimages using iterative object-based classification. This method required careful selection of segmentation parameters; in addition to basic spectral bands also information from nDSM was included. After the segmentation of images the segments were classified based on their attributes (spatial, spectral, geometrical, texture) using rule set classificator. First iteration focused on visible (i.e. unshaded) urban features, and second iteration on objects in deep shade. Results from both iterations were merged into appropriate classes. Evaluation of the final results (completeness, correctness and quality) was carried out on a per-area level and on a per-object level by ISPRS Commission III, WG III/4.

1. INTRODUCTION

1.1 Motivation and aims

Urban systems are very complex and are composed of a large number of spatially heterogeneous components. By their very nature, such systems require advanced methods and algorithms in order to obtain results closer to automatic extraction. The automated extraction of urban objects from data acquired by airborne sensors has been an important topic of research in photogrammetry for at least two decades. In March 2011 ISPRS Commission III, WG III/4 provided test data to the participants of the ISPRS Test Project on Urban Classification and 3D Building Reconstruction in order to evaluate techniques for the extraction of various urban object classes. The aim of this project was to analyse state-of-the-art data sets which were used to obtain urban objects with selected methods and algorithms. In attempt to provide results closely matching to the reference data two approaches were implemented: a method for automatic building extraction and an object based-classification based on rule set classifier for automatic vegetation extraction.

1.2 Overview

The paper presents two different methods used to obtain buildings and urban vegetation. For building extraction we combined lidar DSM and multispectral photogrammetric imagery. Images were used to automatically produce DTM employing image matching techniques and morphological filtering to remove the objects that do not belong to the ground. Additionally orthoimages were produced using photogrammetric methods. By subtracting lidar DSM and photogrammetric DTM a nDSM was calculated. Building extraction procedure included the production of several building masks by employing nDSM threshold, vegetation removal based on orthoimages and several morphologic operations to produce building outlines. Final vectorisation of buildings was performed using Hough transform.

On the other hand automatic extraction of vegetation was based on the object-based image analysis (OBIA). OBIA was developed to bridge the gap between the increasing amount of detailed geospatial data and complex feature recognition problems (Blaschke, 2010). Thoroughly selected segmentation parameters yielded segments that were classified into most resembling class (vegetation, shadows) using selected rule set. Procedure was repeated for visible vegetation and vegetation under shadows separately. Obtained results were later divided into high vegetation (trees) and lower vegetation (natural ground), based on height attribute of segments derived from nDSM data.

1.3 References to related work

A large number of authors use the combination of lidar data and photogrammetric imagery for building extraction. Techniques integrating lidar data and imagery can be divided into two groups (Awrangjeb et al., 2010): techniques which use the lidar data as the primary cue for building detection and employ the imagery only to remove vegetation, and integration techniques,

* Corresponding author.

which use both lidar data and imagery as primary cues to delineate building outlines. Our method belongs to the first group. Authors of the first group use the full information of the lidar point cloud (Rottensteiner et al., 2005) or interpolate point cloud into a grid format (Vu et al., 2009). Demir and Baltsavias (2010) used DSM/DTM comparison in combination with normalized differential vegetation index (NDVI) analysis for building detection. The vectorisation of buildings by Hough transform was proposed by several authors (e.g. Lee et al., 2003, Cha et al., 2006, Koc San and Turker, 2010). The similar approach employing Radon transform is described in Grigillo et al. (2011).

Research of the automatic tree detection and delineation from digital imagery dates back to the mid- 1980s. On the other hand lidar data emerged as source for vegetation analysis only in recent years, however only few studies are focusing on a detailed identification of urban vegetation. Ardila et al. (2012) successfully extracted tree crown objects with OBIA at multiple segmentation scales in urban areas using VHR satellite images, since object-based classification has been argued as the most appropriate method to obtain information from urban remote sensing applications. Ardila et al. (2011) used Markov random field based super resolution mapping approach for crown detection from VHR satellite data. Walton et al. (2010) used a combination of aerial and satellite imagery to estimate the canopy cover in urban areas, whereas Van der Sande (2010) combined optical imagery with topographic data to extract urban trees. Đurić (2011) used object-oriented approach on VHR satellite imagery with the additional lidar data for estimation of urban tree crowns in urban park. Not many studies have been found on extraction of only natural grounds, but there are plenty of researches where this land use was extracted as part of whole urban scene.

2. METHODS

2.1 Study area and data description

The data set was captured over Vaihingen in Germany. Area consists of three test areas: Area 1 (Inner City), Area 2 (High Riser) and Area 3 (Residential Area). Each test area contains different urban structure, showing different degrees of shape complexity and urban structure. The test data consist of digital aerial images with orientation parameters and Airborne Laserscanner (ALS) data. Images are pan-sharpened colour infrared images with ground resolution of 8 cm and radiometric resolution of 11 bits. The mean point density of the ALS point cloud is 4 points per m^2. In addition to the original ALS point cloud, a DSM with a grid width of 25 cm was provided. For both presented methods we used only digital aerial images and DSM as input data.

2.2 Preprocessing

For all three Vaihingen data sets the DTM with 1 m grid cell was produced from digital aerial images and interpolated to 0.25 m grid cell. The orthoimages with spatial resolution of 0.25 m were created using DTM and aerial images. DTM and orthoimages were generated with the Socet Set software. The NGATE software module enables the automatic creation of DSM employing image matching and various morphological operations for removing objects that do not belong to the relief to produce DTM. By calculating the difference between DSM and DTM a nDSM was also generated for all three test areas.

The paper describes results based on DTM produced from aerial images by photogrammetric procedures. This DTM contains some errors due to incorrect image matching or inefficient morphological filtering of high objects. Consequently we were not able to detect some smaller buildings from the derived nDSM. Some of those buildings could have been detected if we would have used lidar-based DTM for nDSM generation instead, which we prepared from ALS point cloud using software LAStools. The comparison of the results is given in the Results and discussion section.

2.3 Automatic building extraction

By applying the 1.5 m threshold to nDSM a nDSM high objects mask was produced. Apart from buildings the mask also included other objects higher than 1.5 m. Such objects were mainly vegetation. By removing vegetation from high objects mask the initial building mask was created. Vegetation was extracted from orthoimages with the object-based methodology as described in section 2.4. Faster approach for the production of vegetation mask would be calculation of the NDVI from red (R) and infrared (IR) band of the orthoimage (1):

$$\text{NDVI} = \frac{IR - R}{IR + R} \tag{1}$$

We used NDVI to remove vegetation from nDSM high objects mask only in Area 3. For this area the 0.17 NDVI threshold was applied to produce the vegetation mask.

In further processing we deal with three different masks: the nDSM high object mask, initial building mask and the mask for the building area test. Additionally we used building outlines image. The elaboration of mask for the building area test and the building outlines image is in detail discussed in Grigillo et al. (2011) within section 3.3. The mask for the building area test is made with a series of morphological operations from the initial building mask. Morphological operations were used to remove irregularities from the initial building mask. Irregularities are the consequence of different objects, which are higher than 1.5 m and do not represent buildings (fences, vehicles), noise in nDSM, errors in the calculation of the vegetation etc. The mask for the building area test contains only buildings in the test area.

For the building extraction the building outlines image was also included in the procedure. The nDSM high objects mask, the mask for the building area test and serial of morphological operations were employed to produce the nDSM building mask, where one object in the mask represented one building or several buildings, if these touched each other. Finally, the building outlines were produced from the nDSM building mask.

Since the building outlines image was in raster format, the contours were not presented with straight lines as expected for buildings and angles between outlines were not rectangular. For this reason, the final building outlines were vectorised using Hough transform, separately for each object within the building outlines image. The full scenario of vectorisation procedure with Hough transform included the following steps:

- Orientation of the main building axis is determined by straight line which included the largest number of pixels that belonged to the building outline (Figure 1 (a)).

- Detection of all straight lines parallel to the main building axis that described individual building in the building outlines image. The detection of straight lines went on until at least four pixels were included within the line or until at least two parallel straight lines were detected. Straight lines were detected with the mutual distance of 0.5 m.
- Since we assumed that buildings have rectangular sides (which however is not true for some of the buildings in Area 1), in the same way also all rectangular straight lines describing building outline were detected. In this case the threshold when to stop straight line detection along the axis perpendicular to the main building axis was slightly lowered (until 3 pixels were in the line or at least two rectangular straight lines were detected). Figure 1 (b) shows parallel and perpendicular lines in the building outline.
- Intersections among rectangular straight lines were detected and out of them rectangles were made (Figure 1 (c)).
- All obtained rectangles do not necessarily correspond to buildings (Figure 1 (d) shows constructed rectangles on the mask for the building area test). To detect false positives we used the following criteria: if the ratio between the surface area obtained by the intersection of the rectangle with the mask for the building area test and the surface area calculated from the rectangle's corners was larger than 0.5, the rectangle was retained, otherwise it was discarded (Figure 1 (e)). When dealing with building interiors, the surface area test was carried out by negating the mask for the building area test.
- Vectorised building was obtained using the outline of retained rectangles. Figure 1 (f) shows the vectorised building on the DSM.

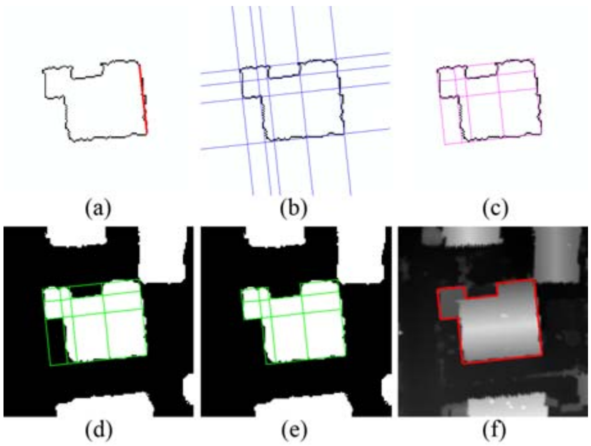


Figure 1. Building vectorisation procedure.

Building vectorisation procedure is shown in Figure 1. Figures 1 (a), (b) and (c) are shown as negatives.

2.4 Automatic vegetation extraction

Object-based classification was performed with iterative object based image analysis implemented on orthoimages with three spectral bands (IR, R, G); as the fourth band a nDSM layer was added. The resulting combined image (3+1) gave best results to locate and delineate trees and other green areas, marked as natural ground. All three test sites contain several tree species, implying large variation of distance between trees, crown sizes and surrounding elements, resulting in challenging and realistic cases for mainly deciduous tree detection.

The data has been processed with ENVI EX Feature Extraction module that includes object-based approach implemented in two steps. In the first step the segments are created based on the spectral signature of the different image parts (segmentation), while in the second step the segments are analysed and classified into the classes (classification).

Although many image segmentation methods have been developed, there is a strong need for new and more sophisticated segmentation methods to produce more accurate segmentation results for urban object identification on a fine scale (Li et al., 2011). Segmentation can be especially problematic in areas with low contrast or where different appearance does not imply different meaning. In this case the outcomes are represented as wrongly delineated image objects (Kanjir et al., 2010). Object features of type vegetation were examined at multiple segmentation scales which yielded different shapes of image-objects. After the segmentation different attributes (spatial, spectral, geometrical, texture) were calculated for each segment. NDVI (see Equation 1) was assigned to each segment.

After image was segmented segments were classified by creating a rule set according to suitable segment attributes. Texture, shape and contextual features are keys to the identification of trees in urban scenes (Ardila et al., 2012). Due to the complexity of the scene and several factors limiting vegetation detection, we detected vegetation objects using the following processing steps: 1.) classify 'visible' vegetation (i.e. vegetation within unshaded areas) and shadows, 2.) create mask of shadows, 3.) detect vegetation under shadows, 4.) merge vegetation results from visible vegetation and vegetation under shadows, 5.) divide vegetation results into trees and natural ground (high and low vegetation). Figure 2 presents a diagram of the workflow described in this section.

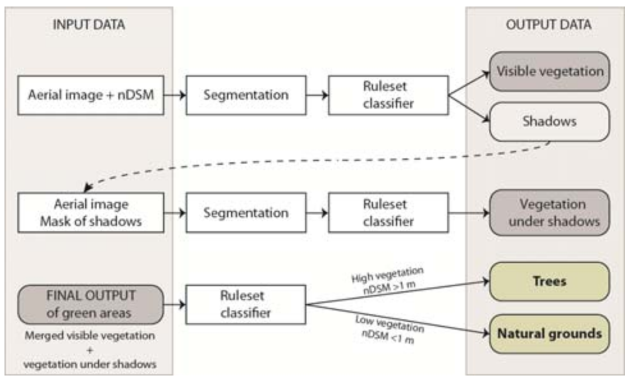


Figure 2. Workflow for urban vegetation extraction with object-based approach.

Objects of visible vegetation were detected based on NDVI which is high on vegetated areas. Simultaneously, we detected also class "shadows"; these dark areas may also contain vegetation, however in the shades the contrast between the image objects of adjacent land covers was smaller, therefore shaded vegetation areas could not be detected exclusively from NDVI. The rule set of attributes used in this procedure that correspond to class shadows is listed in Table 1. The threshold value for each attribute was identified through visual interpretation.

Class	Attribute	Value
Visible vegetation	NDVI	>0.15
Shadows	nDSM R band value area NDVI	<2.5 <800 >10 <0.15
Vegetation under shadows	R band value G band value NDVI area	<725 >260 >0 >1.5

Table 1. Rule set of attributes for three given classes.

Later, shadowed areas were masked out from the whole image scene and analysed in the same way as visible vegetation. Shadows produced by high objects (high buildings or high vegetation) are elemental in urban vegetation analysis. They represent a great challenge in remote sensing and symbolize a factor that considerably influences the results. Although land use under shadows has low spectral separability, vegetation under shadows still appears to have higher NDVI than other land use classes under shadows. Further promising characteristic we successfully utilize is that objects that are assigned to vegetation under shadows class have specific geometrical and spectral properties (Table 1).

The set of all vegetation objects was created by merging segments of obtained visible vegetation and vegetation under shadows. Last step was to distinguish trees from the natural grounds. It was based on nDSM value of segments – values > 1 m were assigned to trees and values < 1 m were assigned to natural grounds.

3. RESULTS AND DISCUSSION

The results were evaluated by the organizers based on reference data. The reference data for Vaihingen were generated by photogrammetric plotting. All objects were evaluated by a comparison of label images and provide completeness, correctness and quality of the results both on a per-area level and on a per-object level. The detailed description of evaluation techniques is given in Rutzinger et al. (2009).

Figure 3 shows the overlay of the automatic building extraction results in Area 3 and DSM. Table 2 shows the evaluation of automatic building extraction for all three Vaihingen test areas respectively. Completeness (Comp.) is the percentage of entities in the reference that were detected. Correctness (Corr.) indicates how well the detected entities match the reference and is closely linked to the false alarm rate. The Quality of the results provides a compound performance metric that balances completeness and correctness. The last column in Table 2 (Area3*) shows the evaluation results for building extraction based on the nDSM produced using lidar DTM.

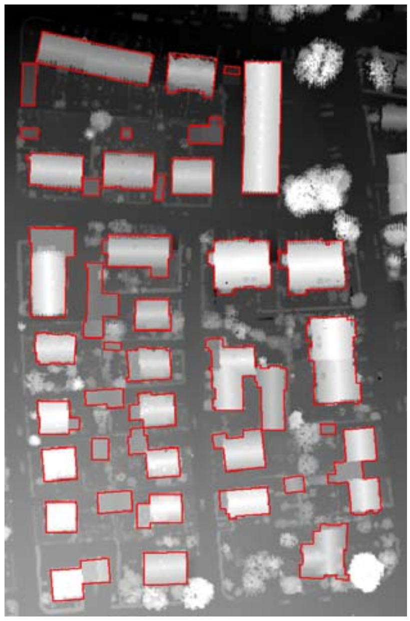
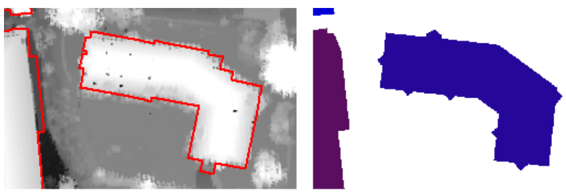


Figure 3. Results of the automatic building extraction.

Evaluation method		Area 1	Area 2	Area3	Area3*
Per-area	Comp.	87.5	93.8	89.7	94.4
	Corr.	96.2	95.4	97.0	95.4
	Quality	84.6	89.7	87.3	90.3
Per-object	Comp.	81.1	85.7	76.8	82.1
	Corr.	100.0	100.0	100.0	100.0
	Quality	81.1	85.7	76.8	82.1
Per-object (balanced by area)	Comp.	98.8	99.7	97.1	98.7
	Corr.	100.0	100.0	100.0	100.0
	Quality	98.8	99.7	97.1	98.7

Table 2. Evaluation of automatic building extraction.

Per-area evaluation method actually means pixel based evaluation, where the raster representation of the detection results and the reference are compared. Building extraction results are influenced by some errors, e.g. use of orthoimages for vegetation removal, created using photogrammetric DTM where tall objects are not included, slightly miscalculated orientation of some buildings in the area from building outlines and the assumption that all the buildings have rectangular shape (which is not the case especially in Area 1; see Figure 4).



(a) (b)

Figure 4. Example of low quality building extraction due to assumption that building has rectangular shape (a) compared to the reference image (b).

In per-object evaluation, an object is considered to be a true positive if a certain minimum percentage of its area is covered by objects in the reference. Per-object correctness shows that all extracted buildings actually represent buildings. Lower completeness shows that some buildings were not detected by the procedure. All of these buildings had small surface area or were low. The nDSM generated using lidar DTM enabled us to detect some of the smaller buildings, as can be seen from the last column in Table 2. Consequently, the completeness and the quality are enhanced, not only for per-object evaluation but also for per-area evaluation.

Tables 3 and 4 show the detection rates of trees and natural ground in all three areas. Object-based method for vegetation identification described in Section 2.4 gave better results than solely NDVI calculations since it enabled extraction of vegetation also under shadowed areas. Figure 5 shows the results of tree and natural ground extraction in Area 3.

Evaluation method		Area 1	Area 2	Area3
Per-area	Comp.	59.3	88.9	76.7
	Corr.	61.8	59.2	58.7
	Quality	43.4	55.1	49.8
Per-object	Comp.	63.8	79.0	70.3
	Corr.	47.2	55.2	39.1
	Quality	37.2	48.1	33.6
Per-object	Comp.	71.9	95.8	90.8
(balanced	Corr.	71.1	72.5	69.9
by area)	Quality	55.6	70.2	65.3

Table 3. Evaluation of automatic tree extraction.

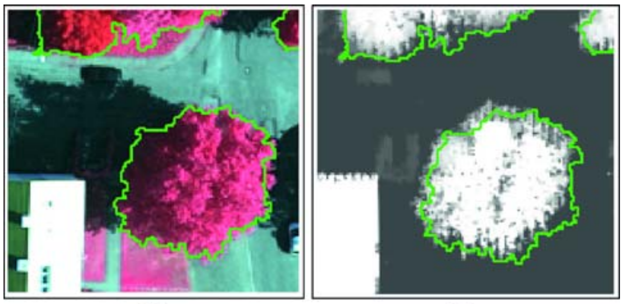
Evaluation method		Area 1	Area 2	Area3
Per-area	Comp.	67.5	68.7	83.4
	Corr.	65.9	85.2	60.6
	Quality	50.0	61.3	54.0
Per-object	Comp.	73.7	36.8	88.0
	Corr.	26.3	17.6	46.4
	Quality	24.1	13.5	43.7
Per-object	Comp.	95.8	97.5	98.8
(balanced	Corr.	78.2	94.0	72.1
by area)	Quality	75.6	91.8	71.4

Table 4. Evaluation of automatic extraction of natural ground.

Compared to building extraction this methodology gave poorer results. This is mainly due to: 1.) low spectral separability between vegetation and other land use (typically sealed areas), especially when under shadows, 2.) extraction of trees from the orthoimages produced from DTM (see Figure 6), 3.) substantial difference between reference geometrical shapes (tree crowns were simplified into correct circles) and obtained results (variety of irregular pixel shapes). There was no geometrical delineation improvement performed in our methodology (see Figure 7).

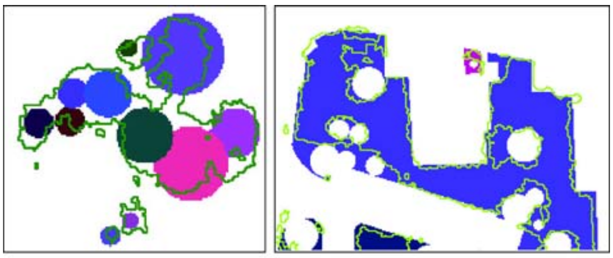
False negatives were concentrated mostly in places with trees or higher natural grounds. This shows the necessity of improving rule sets for vegetation distinction. Nevertheless, general quality of vegetation mask was still high enough to obtain qualitative building extraction.

Figure 5. Extracted trees (green) and natural ground (light green).



(a) (b)

Figure 6. Extracted tree from orthoimage (a) shown on DSM (b). The example shows the radial image displacement of orthoimage produced from photogrammetric DTM.



(a) (b)

Figure 7. Different geometrical shapes of reference data and obtained results for trees (a) and natural ground (b).

Regarding the vegetation results in urban areas it can also be seen that remote sensing analysis overestimate vegetation and underestimate land use that is located under trees (usually impervious area).

4. CONCLUSION

This paper has demonstrated two methods for obtaining two types of urban structure: buildings and vegetation. On all three test areas we have shown that objects of both types can be separated from the rest urban scenery with high accuracy. Extracted buildings were obtained from nDSM, generated from lidar DSM and photogrammetric DTM. Multispectral images were used to remove vegetation. Building outlines were produced using morphologic operations and vectorised using Hough transform. Average per-pixel completeness was 90.3%. Results could be improved if we would use DTM produced from lidar data since the automatically generated DTM from aerial images contains some errors due to incorrect image matching or inefficient morphological filtering of high objects. Consequently we were not able to detect some smaller buildings from the derived nDSM. Per-pixel quality was the lowest in Area 1 where not all buildings were rectangular. For the future work we should consider expanding the method of building delineation to allow lines in arbitrary directions.

Specific characteristics of urban areas proved to be very challenging for semi-automatic image identification of trees and natural ground. Proposed methodology yielded irregular shapes; only such are found in nature and also on optical data. This was the main reason of poor quality assessment comparing to reference data, where trees were presented as circles. Results were overestimating vegetated areas and underestimating impervious areas. For the future work we propose contextual identification of vegetation. Contextual rules can model the occurrence of natural ground and trees in urban areas, since context (information that can be used to characterize the situation of an entity) is an essential element for feature recognition (Ardila et al., 2012). Also more detailed separation between high and low vegetation is suggested.

5. REFERENCES

Ardila, J.P., Bijker, W., Tolpekin, V., A., Stein, A., 2012. Context-sensitive extraction of tree crown objects in urban areas using VHR. *International Journal of Applied Earth Observation and Geoinformation*, 15, pp. 57-69.

Awrangjeb, M., Ravanbakhsh, M., Fraser, C. S., 2010. Automatic building detection using LIDAR data and multispectral imagery. In: *Digital Image Computing: Techniques and Applications,* Sydney, Australia, pp. 45-51.

Blaschke, T., 2010. Object based image analysis for rempte sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65, pp. 2-16.

Cha, J., Cofer, R. H., Kozaitis, S. P.,2006. Extended hough transform for linear feature detection. *Pattern Recognition*, 39, pp. 1034-1043.

Cramer, M., 2010. Evaluierung digitaler photogrammetrischer luftbilkamerasysteme. http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html (9 Jan. 2012)

Demir, N., Baltsavias, E., 2010. Combination of image and lidar data for building and tree exraction. In: *IAPRS*,Saint-Mandé, France, Vol. XXXVIII, Part 3B, pp. 131-136.

Đurić, N., 2011. Object oriented classification for tree species identification and detection of Japanese knotweed. *Graduation thesis – university studies.*

Grigillo, D., Kosmatin Fras, M. and Petrovič, D., 2011. Automatic extraction and building change detection from digital surface model and multispectral orthophoto. *Geodetski vestnik*, 55(1), pp. 28-45.

Kanjir, U., Veljanovski, T., Marsetič, A., Oštir, K., 2010. Application of object based approach to heterogeneous land cover/use. In: Addink, E.A. (ed.). *GEOBIA 2010 : Geographic Object-Based Image Analysis, Ghent, Belgium, 29 June - 2 July*. Ghent, 2010. http://geobia.ugent.be/proceedings/html/papers.html.

Koc San, D., Turker, M., 2010. Building extraction from high resolution satellite images using Hough transform. *The International Archives of Photogrammetry, Remote Sensing and Spatial Science* 2010, 38 (8), pp. 1063-1068.

Lee, D. S., Shan, J., Bethel, J., 2003. Class-guided building extraction from IKONOS imagery. *Photogrammetric Engineering and Remote Sensing*, 69, pp. 143-150.

Li, P, Guo, J., Song, B., Xiao, X, 2011. A multilevel hierarchical image segmentation method for urban impervious surface mapping using very high resolution imagery. *IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing*, 4(1), pp. 103-116.

Rottensteiner, F., Trinder, J., Clode, S., Kubik, K., 2005. Using the Dempster-Shafer method for the fusion of LIDAR data and multi-spectral images for building detection. *Information Fusion*, 6(4), pp. 283-300.

Rutzinger, M., Rottensteiner, F., Pfeifer, N., 2009. A comparison of evaluation techniques for building extraction from airborne laser scanning. *IEEE Journal on Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1), pp. 11-20.

Van der Sande, C.J., 2010. Automatic object recognition and change detection of urban trees. *The international Archives of Photogrammmetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII-4/C7.

Vu, T., Yamazaki, F., Matsuoka, M., 2009. Multi-scale solution for building extraction from LiDAR and image data. *International Journal of Applied Earth Observation and Geoinformation*, 11(4), pp. 281-289.

Walton, J.T, Nowak, D.J., Greenfield, E.J., 2010. Assessing urban forest canopy cover using airborne or satellite imagery. *Arboriculture & Urban Forestry,* 34 (6), pp. 334-340.

6. ACKNOWLEDGEMENTS

The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010).