AUTOMATED MAPPING OF BUILDING FACADES BY MACHINE LEARNING

J. Höhle

Dept. of Planning, Aalborg University, Skibbrogade 3, DK-9000 Aalborg – jh@land.aau.dk

Commission III, WG III/2

KEY WORDS: Mapping, Classification, Automation, Point Cloud, Building, Learning, Decision Support, Extraction

ABSTRACT:

Facades of buildings contain various types of objects which have to be recorded for information systems. The article describes a solution for this task focussing on automated classification by means of machine learning techniques. Stereo pairs of oblique images are used to derive 3D point clouds of buildings. The planes of the buildings are automatically detected. The derived planes are supplemented with a regular grid of points for which the colour values are found in the images. For each grid point of the façade additional attributes are derived from image and object data. This 'intelligent' point cloud is analysed by a decision tree, which is derived from a small training set. The derived decision tree is then used to classify the complete point cloud. To each point of the regular façade grid a class is assigned and a façade plan is mapped by a colour palette representing the different objects. Some image processing methods are applied to improve the appearance of the interpreted façade plot and to extract additional information. The proposed method is tested on facades of a church. Accuracy measures were derived from 140 independent checkpoints, which were randomly selected. When selecting four classes ("window", "stone work", "painted wall", and "vegetation") the overall accuracy is assessed with 80% (95% Confidence Interval: 71% -88%). The user accuracy of class "stonework" was assessed with 90% (95% CI: 80%-97%). The proposed methodology has a high potential for automation and fast processing.

1. INTRODUCTION

Currently the restoration and renovation of building facades is an important task in Europe. Many old buildings are renovated in the old styles (cf. Figure 1). The planning of such work can take advantage of façade plots and information systems about the current and former state of the buildings. The recording of culture heritage is supported by UNESCO and government programs and takes place all over the world. Famous buildings and monuments are systematically photographed, mapped and recorded. Furthermore, 3D models with photo texture are today established for many cities. The recording of windows, doors, technical installations, and other objects will enrich these models. Automated and economic methods are required to record such objects and its features (position, dimensions, material, etc.). Oblique aerial images are playing a key role to realize efficient and economic mapping of building facades.

Figure 1. Façade restoration work in the German cities of Zittau (upper left), Görlitz (lower left), and Pirna (right)

New multi-camera systems are developed and oblique aerial imagery is nowadays systematically and regularly taken of larger towns. The availability of such imagery including its orientation data will make the economic mapping of facades possible. The generation of thematic maps of facades has to solve the geometric and semantic problems. In this contribution a solution to both problems will be presented. Other authors have contributed to this topic before. In (Yang and Förstner 2011) and (Delmerico, et al. 2013) the classification of façade objects is identified as an important task. Very different approaches have been investigated. These investigations comprise the use of various sensors, processing methods, and assessment procedures of the results. In (Höhle 2013) building facades are modelled from a point cloud derived from oblique imagery. Façade plots with photo texture are produced by projecting a dense grid into the original image and by collecting the intensities of three bands (red, green, blue) at the calculated image positions. The image quality of the facade plots was tested by edge analysis and proved to be as good as the original images. In (Yang, et al. 2012) the influence of various object features on the classification of building facades is investigated. The classification is based on images only. Eight object classes ("building", "car", "door", "pavement", "road", "sky", "vegetation", and "window") are selected and a set of 60 terrestrial images ("eTRIMS database") is investigated. These building façade images were randomly divided into training and test images. A randomized decision forest classifier was applied in order to assign a class to a segmented region. The assessment of the thematic accuracy was carried out by comparing the manually interpreted image with the classified image. It revealed accuracies of 76% ("vegetation"), 68% ("window"), and 60% ("building") for the three important classes. In (Meixner, et al. 2012) vertical aerial images are used to detect border lines between facades automatically. Their approach uses height profiles and repetitive patterns which are derived from large-format imagery. The achieved accuracy is quoted with 88%. Facades appear only in the corners of vertical images,

which have been taken with high overlaps (p=80%, q=60%) and a ground sampling distance (GSD) of 10 cm. The use of a multi-looking oblique view airborne laser scanning is investigated in (Tuttas and Stilla 2012). The obtained point cloud is not very dense (<10 points/m^2) and, therefore, the width and height of rectangular windows were reconstructed with a few decimetres only.

The detection of objects in the facades is solved by classification where the attributes of the objects are analysed. Various methods are at disposal. Traditional methods like "maximum likelihood" are based on pixels of the images. Newer methods segment the images in regions and classify them thereafter. When point clouds are the starting point then machine learning may be used with advantage. This technology has successfully been applied in many fields, e.g., searching of information, speech recognition, robot driving (Mitchell 1997). There exist many different approaches in machine learning. One important method is the decision tree classification (Breiman, et al.). This approach is implemented in the software package "rpart" of the open source language and environment "R" (R Development Core Team 2013). It can, therefore, be easily and economically realized. In addition, functions for the assessment of the thematic accuracy are available in the R-package "survey".

The goals of this paper are to develop an efficient and highly automated method to detect objects in building facades and present the results in a thematic map together with other information on objects such as number, position, and dimensions. The paper deals first with the geometric tasks when deriving facades, describes the applied methods for the classification of the façade objects, the assessment of the thematic accuracy and of the cartographic refinement including the derivation of additional information. Tests with several facades of a church are carried out applying imagery of an oblique camera system. The obtained results are discussed and conclusions are given in the last sections.

2. GEOMETRIC TASKS AND PROBLEMS

In order to extract measures of the façade objects the oblique images have to be rectified. This can be done manually, e.g., by means of Adobes "Photoshop" using the function "correct perspective". A rectangle has to be present at the façade. A more accurate way is to use four points of known position at the façade and in the image. The eight coefficients of a perspective transformation are then calculated and applied. The rectification of a large amount of images and facades has to be carried out automatically or at least semi-automatically. The calibration data of the applied camera and the exterior orientation data of the images have to be known. From two or more images a point cloud is derived and the planes of facades have to be detected. The use of the Hough transformation is a solution for this task (Tarsha-Kurdi et al. 2007). The applied function is the equation of a plane (ρ=**n·x**) where ρ is the distance of the plane to the origin of the coordinate system, **n** the normal vector of the plane and the vector **x** represents a point lying in the plane. The normal vector is described by its three components (n_1=cosθ·cosφ, n_2=sinθ·cosφ, and n_3=-sinφ) and the point vector by three spatial coordinates (E, N, Z). All points of the point cloud use a set of plane angles (θ, φ) and calculate the distance (ρ) for each set. All points contained in a certain plane are accumulated in a voxel of the 3D parameter space. Figure 2 shows the result of a Hough transformation. The three axes of the cube are the parameters (θ, φ, ρ) of the Hough

transformation. The circles represent planes of buildings (roofs, facades). The points of the point cloud which are contained in such a plane are extracted and the plane can be modelled. Façade planes are normally vertical (φ=0°) and a value for the azimuth (θ) has to be determined only. An accurate value for this parameter may be calculated by means of regression.
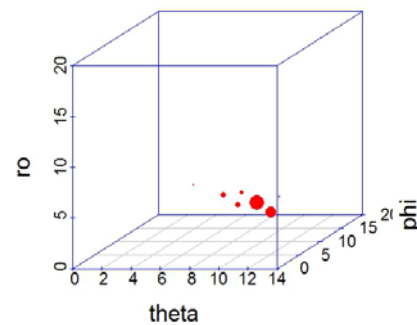

Figure 2. Parameter space of the Hough transformation

3. CLASSIFICATION

The objects of the façade to be determined (windows, doors, ledge, gutter, etc.) are usually specified by the user of the façade plans. The producer of the façade plans has to know which features characterize each object and where such data are available or how they can be derived. These attributes may be image-oriented and/or object-oriented. The proposed method uses the sum and the normalized difference of intensities of spectral bands as well as the height above ground and supplements these attributes to each point (cell) of the façade grid. A few training areas are collected in order to derive the decision tree. By means of the derived decision tree a class is assigned to each point of the façade grid. An annotated façade plan can then be plotted. The processing is semi-automatic. The borders of the training areas are manually digitized on top of the façade plot with photo texture. All points inside the training areas are automatically extracted and supplemented with attributes and a character representing the 'true' class.

4. ASSESSMENT

The assessment of the thematic accuracy of the annotated façade plot has to be based on statistical principles. An independent sample has to be taken. The reference values for cells are determined by observing the façade plot with photo texture. At the position of the check point a "true" value of the class is found. A confusion matrix can then be established from which the accuracy measures (overall accuracy, user accuracy, and kappa value) are derived. Also, their 95% confidence interval is important to know. The factors which influence the results (e.g., number of training points, selection of attributes, prior probability, number of classes, etc.) have to be studied in order to improve the results.

5. CARTOGRAPHIC ENHANCEMENT

The façade plot may have some deficiencies. There can be misclassifications or gaps. The outlines of man-made objects like windows or doors may not be straight or orthogonal as they are in reality. Many of these defects can be corrected. One solution to these problems is the conversion of the raw façade plot into an image and applying image processing methods such

as filtering, segmentation, or object manipulation. Furthermore, information about the classes (number, position, dimensions) may be extracted from the 'imaged' and annotated façade plan. The need for automated processing as well as economic considerations will decide about these efforts.

6. TESTS WITH FACADES OF A CHURCH

The general description of the automated mapping of facades will now be supplemented by some tests. The object is an old church in Widnau, Switzerland, which has recently been renovated. The church has several facades which are photographed by a medium-format aerial camera system (Leica RCD30 Oblique) which consists of four oblique cameras and one nadir camera. Details of the different steps of the applied approach are given in the following sections. The main emphasis is given to the automated mapping by means of machine learning and to the assessment of the thematic accuracy.

6.1 Description of the object

The church has several tall and bright facades, dark roofs and a squared big tower. It is surrounded by trees, bushes and other buildings. The main façade (cf. Figure 3) has three facades which are parallel to the main axis of the church. The facades contain windows, doors, vegetation, painted walls, and stonework. Smaller objects are ledges, gutters, and lightning rods.

Figure 3. Oblique image of the church (forward view)

6.2 Image acquisition

Oblique and nadir images were taken from 670 m above ground by means of the RCD30 Oblique Penta camera system. The lenses of the five cameras have a focal length of 50 mm, the sensors a pixel size of 6 µm. The oblique images have a nadir distance of 35°, their GSD at facades varies between 15 and 30 cm. The oblique images are taken simultaneously to both sides and forward/backward. The images to the sides overlap 68% (front) and 82% (rear). The flight lines are 430 m apart and the exposure stations form a regular pattern. Stereo pairs are, therefore, possible in flight and across flight direction. The orientation data of the oblique images were derived by an aerotriangulation using a few control points. The quality of the images has been tested using edges within images and deriving the point spread function. The calculated factors for effective GSD were σ=1.06 only.

6.3 Derivation of the point cloud

Point clouds were derived by means of the program "Match-T DSM" of the Trimble Company. A dense grid of spatial points is derived in the reference system. The points are not regularly distributed in the facades. The facades were therefore first modelled and then supplemented with a regular grid of points. The parameters of the façade were derived from the result of a Hough transformation and a subsequent regression. The spacing of the grid was selected with Δ=0.1 m. The points of the façade are supplemented with photo texture using "back-projection". Figure 4 depicts the derived plot. Each point (cell) of this point cloud has coordinates in the reference system (Easting, Northing, elevation), the image coordinate system (col, row), and intensities in the three channels (I_{red}, I_{green}, I_{blue}). Other attributes are derived and added. These are the height above ground (dZ), the sum of the three intensities (Srgb), and the normalized difference index (Pndvi). The latter attribute is derived by Pndvi = (I_{red}-I_{blue}) / (I_{red}+I_{blue}).
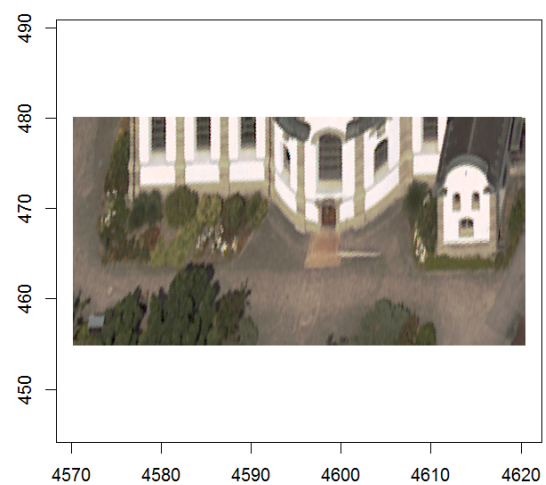

Figure 4. Plot of the point cloud with photo texture in the façade coordinate system (yZ). The measures are in meters.

6.4 Derivation of decision tree

The decision tree is derived from training areas. The borders of a few areas representing a class are digitized on top of the façade plan with photo texture. The points (cells) within the borders are then extracted and supplemented with their attributes (Easting, Northing, elevation, two image coordinates, height above ground, intensity in the red, green, blue band, sum of intensities in three colour bands, and the normalized difference index). The decision tree is then computed by the R-function "rpart". The probability that a cell will be one of the classes can be set as "prior" value in this function. This value is chosen according to the estimated size of the class area within the façade (P_{window}=0.5, $P_{stonework}$=0.2, $P_{vegetation}$=0.1, $P_{painted\ wall}$=0.2). The calculated decision tree is depicted in Figure 5. Weights are not applied even though the number of training points differ (n_{window}=459, $n_{stonework}$=354, $n_{vegetation}$=118, $n_{painted\ wall}$=430). There are three intermediate nodes and four end nodes. The three attributes representing the intensities in the colour bands (r,g,b) were not considered in the calculation of the decision tree. Cells are assigned to class "window", e.g., when the sum of intensities (Srgb) is less than 384.5 and the height above ground higher or equal 3.35 m. Other tests incorporated also other attributes (e.g., intensities of each channel). The function "rpart" applies also pruning which simplifies the decision tree.
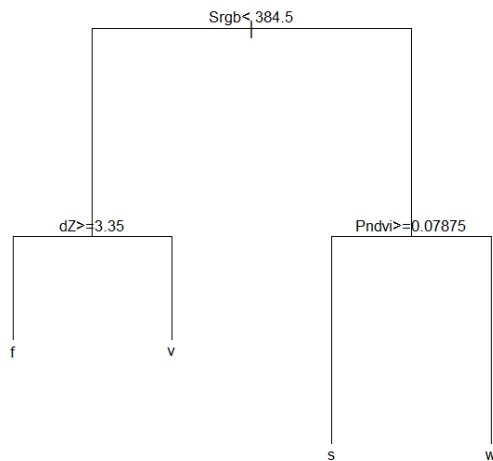
Figure 5. Decision tree derived from training areas. The nodes are thresholds of the class attributes (dZ=height above ground [m], Srgb=sum of intensities in three colour bands, Pndvi= normalized difference index). The end nodes of the tree are depicted with the names of the classes ("window"=f, "vegetation"=v, "stonework"=s, "painted wall"=w).

6.5 Generation of the annotated façade plot

The annotated façade plot is generated using the decision tree derived before. For each cell of the point cloud a class is assigned. The result is plotted using a symbol and a colour for each class (cf. Figure 6).
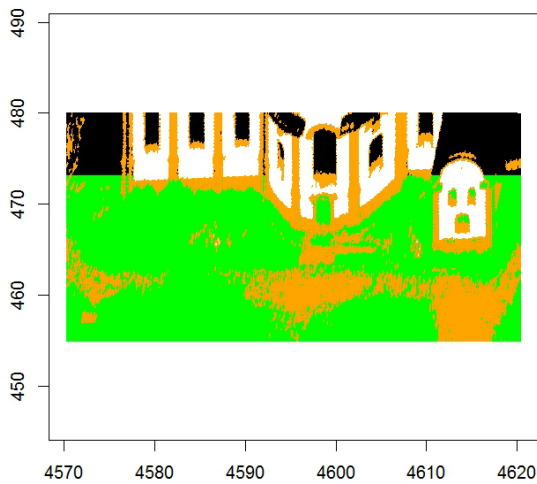

Figure 6. Generated interpreted façade plot with four classes ("vegetation"=green, "stonework"=orange, "painted wall"=white, "window"=black) and coordinate axes with units in meters

The rearmost façade is classified with the correct height above ground. The other two facades have to use modified ground heights and need to be classified in addition. The assessment of the thematic accuracy has, therefore, to be carried out for each of the three façades separately. When evaluating the rearmost façade, the four classes can clearly be distinguished. The edges are not very straight and the areas have some gaps. The plot is, therefore, a raw result which may be cartographically enhanced (cf. section 6.7). The coordinate axes allow for measuring object

dimensions, e.g. the width and height of the windows or the elevation of the lower edge of the painted wall. The number of cells belonging to one class can be counted and used for weighting in the assessment of the thematic accuracy (cf. section 6.6).

6.6 Assessment of the thematic accuracy

Several accuracy measures were derived from checkpoints which were randomly selected. The reference values are determined by observing the façade plot with photo texture at a calculated random position. This true value is compared with the value of the classification. The results are summarized in the confusion matrix (cf. Table 1). Only the accuracy of the rearmost façade is assessed. When selecting four classes ("window", "stonework", "painted walls", and vegetation) the overall accuracy is 80% (95% CI: 71% -88%). The "survey weighted kappa" value is 0.73 (95% CI: 0.62-0.84). The weight of each class is based on the ratio between the number of cells and the number of checkpoints. The cells and the check points have to be part of the façade to be assessed. The user accuracy of class "stonework" is 90% (95% CI: 80%-97%). The other classes are less accurate (cf. Table 2).

reference / classification	f	s	v	w	Σ
f	**11**	2	0	0	13
s	3	**46**	0	2	51
v	1	7	**20**	0	28
w	0	12	0	**36**	48
Σ	15	67	20	38	140

Table 1. Confusion matrix (f=window, s=stonework, v=vegetation, w=painted wall)

class	user accuracy [%]	95% CI [%]
f	85	54-98
s	90	80-97
v	71	52-87
w	75	61-86

Table 2. User accuracy for each class (f=window, s=stonework, v=vegetation, w=painted wall)

The number of training points has influence on the results (cf. Table 3).

test #	training points #	check points #	overall accuracy (CI) [%]	survey weighted kappa (CI)
1	1361	140	80 (71-88)	0.73 (0.62-0.84)
2	128	51	78 (61-90)	0.69 (0.50-0.89)

Table 3. Results of the rearmost façade

When 128 training points are selected only, the thematic accuracy drops to 78%. Important for reliable results is the number of the check points. This can be noticed at the size of the confidence interval. When the number of check points is reduced from 140 to 51 points, the confidence interval becomes wider. The confidence intervals of the survey weighted kappa value will also proof this (0.62-0.84 versus 0.50-0.89). Its values (0.73 and 0.69 respectively) indicate both a moderate result. Improvements may be achieved by means of additional relevant attributes. The increase of the number of classes may

also give some improvements. Other objects, e.g., gutter and ledge, are also part of the image and may be of interest to the user of the façade plot. In the current tests these objects will belong to one of the four classes and thereby contribute to errors. The total area of these two objects is small. When these two objects are treated as classes, the overall accuracy of the classification may become slightly better.

6.7 Cartographic enhancement and derivation of new attributes

In order to improve the cartographic quality of the ("raw") façade plot, techniques from image processing are used. The R-package "EBImage" provides useful functions for this purpose (Pau, et al. 2013). In the following, the applied operations are mentioned and the functions of the package "EBImage" are added in brackets. In order to make the objects more distinct and to close gaps, the used operations (functions) are:

1. Dilation ('dilate') and erosion ('erode')
2. Filtering ('makeBrush') and thresholding ('thresh')
3. Filling of holes ('fillHull')

Misclassification is corrected by the following manipulations.

4. Labelling of objects (bwlabel)
5. Compute features (computeFeatures)
6. Removal of objects (rmObjects)

The removal of objects uses attributes like the size of area or the coordinates of the centre of mass. The objects have to be well separated and their attributes have to be descriptive for the selected class. The dZ-values are calculated for each façade separately, because the distance of the three facades to the camera differs. Furthermore, the image of each class is processed separately and the enhanced façade plot is composed of four images representing one class. The result of these enhancements for all three facades is depicted in Figure 7.

Figure 7. Enhanced façade plots. All three facades are corrected for the position of the plane (distance to the camera) and enhanced

Other information about the objects can be added, e.g., position and dimensions. Table 4 displays the derived information on three windows. The measures are related to the object space.

# of window	y [m]	Z [m]	dZ [m]	area [m^2]
1	4580.2	478.3	7.1	8.9
2	4584.8	478.4	7.2	8.4
3	4589.4	478.5	7.3	8.2

Table 4. Derived information on the windows of the rearmost façade (y, Z=façade coordinates of centre, dZ=height of centre above ground)

7. RESULTS AND DISCUSSION

The proposed method applies machine learning in the classification of point clouds which are derived from oblique images. The obtained results for the thematic accuracy of objects within the building facades are promising. When selecting four classes ("window", "stonework", "vegetation", and "painted wall"), the overall accuracy is assessed with 80%. Cells of the class "stonework" are classified with 90% and of the class "window" with 85% accuracy. The "survey weighted kappa" value is 0.73. Improvements seem possible. Imagery of multi-spectral cameras (including a near-infrared band) will improve the accuracy of the class "vegetation". The selection of training areas needs some manual work. The interpreted façade plot may be cartographically enhanced by applying image processing methods. The Hough transformation and regression were used to extract the azimuth of the façade from the point cloud derived from stereo imagery. The proposed method has a potential for automation and fast processing. The tests have to be extended to other facades containing different types of objects. The applied approach is relatively simple but effective. The decision tree classification is based on the cells of the point cloud and not on the pixels of the image. Object dimensions (height above ground) are used as attributes in addition. This requires a precise orientation of the images and a precise modelling of the facades. More objects may be necessary to be detected and to be mapped. Additional objects in the used façades are "lightning rod", "ledge", "gutter", "ground", and "sky". The misclassifications could then be reduced. Further information on the objects of the facades (position, dimensions) is extracted using image analysis techniques.

8. OTHER BUILDING ENVIRONMENTS AND INNOVATIONS

The proposed methodology has been tested on facades of a church. Other buildings have other facades and other objects within the façades. Buildings may be much higher and the objects within the facades may be more numerous and complex. The objects of interest may be situated in narrow streets or are surrounded by vegetation. It is difficult to generalize to other building environments from this example. The use of aerial oblique imagery is attractive from the economic point of view, especially when imagery already exists. The oblique view from the air is superior to the view from the ground. The automated interpretation of façade objects requires also attributes which characterize these objects. These attributes can be derived from the imagery or be inherent in the object. The connection of image data with object data is, therefore, an important feature of the proposed approach. The machine learning method can also be more sophisticated than the applied decision tree classification. Other innovations are in the aerial oblique cameras. The latest model of the RCD30 Oblique camera system has some new features. The image sensor has smaller pixels resulting in 80 Megapixel (MP) images. All five cameras may be equipped with an 80 mm lens and the oblique images may also have a near infra-red channel. The imagery can then be taken from higher altitudes and "true" NDVI values can be derived. The RCD30 oblique camera system is very compact and may be installed in an unmanned aerial vehicle as well. A new multi-camera oblique system, the UltraCam Osprey, has recently been announced by the photogrammetry division of Microsoft. Their oblique cameras are equipped with a 120 mm lens and a 60 MP sensor. These new oblique camera systems have been built to meet the increasing market for oblique photography.

9. CONCLUSIONS

The proposed methodology for automatic mapping of facades by oblique imagery has been successfully tested with facades of a church. The selected four classes and a few attributes derived from the object **and** image space enabled an overall accuracy of 80% by means of a simple decision tree classification. New camera systems may improve the thematic accuracy and enable a universal application in various building environments.

ACKNOWLEDGEMENTS

REFERENCES

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. *Classification and regression trees*. CRC press.

Delmerico, J.A., David, P., Corso, J.J., 2013. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision. *Image and Vision Computing* 31 (11), pp. 841-852.

Höhle, J., 2013. Oblique aerial images and their use in cultural heritage documentation. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information, Sciences,* vol. XL-5/W2, pp. 349-354.

Meixner, P., Wendel, A., Bischof, H., Leberl, F., 2012. Building facade separation in vertical aerial images. *Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 239-243.

Mitchell, T.M., 1997. *Machine Learning*. McGraw-Hill Education.

Pau, G., Sklyar, O., Huber, W., 2014. Introduction to EBImage - an image processing and analysis toolkit for R. http://www.bioconductor.org/packages/release/bioc/html/EBImage.html (July 18, 2014).

R Development Core Team, 2013. R: A language and environment for statistical computing. http://www.r-project.org (July 18, 2014).

Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., 2007. Hough-transform and extended ransac algorithms for automatic detection of 3d building roof planes from lidar data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Systems,* vol. XXXVI, part 3/W52, pp. 407-412.

Tuttas, S., Stilla, U., 2012. Reconstruction of rectangular windows in multi-looking oblique view ALS data. *Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. I-3, pp. 317-322.

Yang, M. J. and Förstner, W., 2011. Regionwise classification of building facade images, In: *PIA 2011*, LNCS 6952, Stilla, U. (Ed.), Springer, Berlin/Heidelberg, pp. 209-220.

Yang, M.J., Förstner, W., Chai, D., 2012. Feature evaluation for building facade images-an empirical study. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Systems*, vol. XXXIX-B3, pp. 513-518.