Planar Roof Surface Segmentation using 3D Vision

Philipp Meixner
Institute of Computer Graphics & Vision
Graz University of Technology
A-8010 Graz, Austria
meixner@icg.tugraz.at

Franz Leberl
Institute of Computer Graphics & Vision
Graz University of Technology
A-8010 Graz, Austria
leberl@icg.tugraz.at

Mathieu Brédif
MATIS Laboratory
Institute Géographique Nationale
Université Paris-Est
94165 Saint-Mandé Cedex, France
Mathieu.bredif@ign.fr

ABSTRACT

Internet search has initially been a strong driving force for the rapid emergence of 3D building models of large urban areas. Additionally, many commercial and governmental initiatives have been started to develop urban 3D geographic information systems in a transition from the classical 2D- to the novel 3D-GIS. The modeling of building roofs is thus a relevant research topic. The focus has been on the use of aerial LiDAR point clouds (Light Detection And Ranging). However, recent progress in digital aerial cameras has rendered possible the acquisition of very dense point clouds from high overlap digital aerial imagery, and to use these point clouds jointly with the image information to generate 3D building models.

This paper presents a multi-step processing framework and work flow for the automatic segmentation of building roofs in densely built-up areas from high-resolution vertical aerial images. Details extruding from, or intruding into, a roof are being excluded so that each roof is being modeled by means of its planar segments and can then be classified as a specific roof type from a set of standard roof shapes. We show that the results from aerial photography compete well with LiDAR-results as reported by LiDAR researchers. Our experimental work employs a test area in Graz (Austria) with 186 buildings.

Categories and Subject Descriptors

I.4.8 Scene Analysis (Object reconstruction).

General Terms

Algorithms, Measurement, Experimentation, Verification.

Keywords

Semantic segmentation, roofs, roof shapes, aerial images, plane detection, smoothing

1. INTRODUCTION

The 2- dimensional GIS is rapidly morphing into a 3-dimensional model of the Earth and human habitat. Typical use of the resulting 3D GIS and its 3D virtual cities include urban and landscape planning, architectural design, tourist and leisure activities, environmental simulations, disaster preparedness, real estate transactions etc. Depending on the input data and the application, the required level of detail (LOD) may vary widely. This may range from the Lego-type parallelepipeds denoted as LOD-1 via models with roof shapes as LOD-2 to building models with detail such as windows and roof details as LOD-3 and on to LOD-4 to include the building interior [18]. In its most sophisticated form, the virtual city will model not only each building, but also each tree, street detail, bridge and water body is modeled in three dimensions, and many details will be included such as windows, doors, facade elements, sidewalks, manholes, parking meters, suspended wires, street signs etc., all as separate objects (see figure 1). LOD-2 to LOD-4 need a model of the roof shapes.

Figure 1. 3D city model, LOD2 (detail of Leiden, Netherlands) [Courtesy: Aerodata International Surveys, Belgium, http://www.aerodata-surveys.com/].

The state of the art methods of interpreting building roofs are based on airborne LiDARs (Light Detection And Ranging). They offer instant access to point clouds and an easy transition to a Digital Surface Model (DSM). The two fundamental LiDAR-approaches for 3D building reconstruction are model-driven or data-driven. Model-driven methods use a predefined list of roof forms as presented in figure 7 and match the data against those forms [1], [2]. Complex roofs typically are beyond such model-driven methods. Therefore data-driven roof reconstruction is often preferred [3], [4], [5], [6], [7], where a roof gets segmented into planar roof elements. Ideally, each data-plane describes exactly one roof plane. A recent paper [8] assesses the quality of reconstructed building models using LiDAR. It quotes a point density of 25 pts/m^2 which corresponds to a Ground Sampling Distance of 20cm. It quotes a total standard deviation for corner point identification of 10-15cm.

The progress in digital aerial cameras has made possible the acquisition of very dense point clouds at 100 pts/m² and up to 1000 pts/m², based on imaging with Ground Sampling Distances GSD between 10 cm and 3 cm. Those densities are desirable to model details of facades and roofs, street signs and suspended wires. The DSM from aerial photography no longer is being computed at point intervals of 10 to 20 pixels associated with traditional 2-image stereo overlaps. Current point clouds get acquired at a density of 1 point per pixel taking advantage of a 10-

image multi-view geometry [9], and thus to a concept of "super-resolution" [10]. Such dense DSM leads to well-defined horizontal building roof lines and is helpful in automating any roof-related analysis.

This paper presents the automatic segmentation of building roofs in densely built-up areas from high-resolution vertical aerial images, it assigns to each roof a type, and it offers a measure of suitability for a roof to carry solar panels.

We describe a multi-step processing framework leading to a basic roof model. Tests of the proposed procedure are based on a test area in Graz (Austria) with 186 buildings We show that vertical aerial images provide us with a success assignment to a roof type in 88% of all roofs.

2. Approach

The basis for the proposed roof interpretation is a framework introduced in [11] and its improvement to deal with extruding masonry of a building [12]. This segments the images and photogrammetric point clouds into individual properties, building footprints and 3D point clouds associated with each building.

Once the building's point cloud has been isolated, roofs get segmented in 3 steps. First is a smoothing of the DSM with a total generalized variation method (TGV) [13]. Then all roof planes get extracted using an approach first introduced by [14] and employing so-called random sampling and conceptual representation. Finally the roof planes are input to a classification of roof types.

2.1 DSM Smoothing

Measured range data will be noisy so that an interpretation may fail. This topic is a very prominent one in photogrammetry and in literature there exist many algorithms that discuss this problem. The main requirement for such a smoothing algorithm is that it minimizes noise but preserves sharp edges and small structures. For this reason and to accelerate the plane detection we smooth the data using the "total generalized variation" TGV [13]. This method has the advantage that it can be efficiently implemented on the GPU and is therefore very fast.

The variational approach finds the solution of the model by the minimizer of an energy-functional that is usually composed of two terms. The regularization term copes with the a-priori assumption about the smoothness properties of the solution and the data term forces the solution to be similar to the input data. The second order variant of this functional is ideally suited for piece-wise affine surfaces and therefore well suited to buildings that can be approximated by piece-wise planar surfaces.

The data term is based on the robust Huber-L1 norm, the regularization term is based on TGV. The strength of the Huber-L1 norm is that it reduces stair-casing often found in other models and better reflects the noise model of real range images. By implementing the Huber norm to the data and regularization term the so-called "Huber model" is obtained.

$$\min_{u} \left\{ \alpha \int_{\Omega} |\nabla u|_{\varepsilon}\, dx + \sum_{l=1}^{K} \int_{\Omega} |u - f_l|_{\delta}\, dx \right\}$$

Ω … Image domain
K … observed range images
f_l … single observation
u … sought solution
α … smoothing factor ($\alpha \geq 0$)
ε, δ … Parameter of Huber norm

The main property of TGV this is that it allows to reconstruct piecewise polynomial functions of arbitrary order. It has the property to be convex, which means that a solution is globally optimal. For the approximation of buildings it is sufficient to use the TGV regularization of the second order since they can be approximated by piecewise planar surfaces. This method combines TGV regularization of second order with the Huber-L1 norm.

$$\min_{u,v} \left\{ \alpha_1 \int_{\Omega} |\nabla u - v|\, dx + \alpha_0 \int_{\Omega} |\varepsilon(v)|\, dx + \sum_{l=1}^{K} \int_{\Omega} |u - f_l|_{\delta}\, dx \right\}$$

Applying this method to a typical roof produces a smoothed result illustrated in Figure 2, taken from the test dataset in the city of Graz, Austria. In this image you can see that outliers and noise within the range data are minimized by maintaining the global structure of a roof.
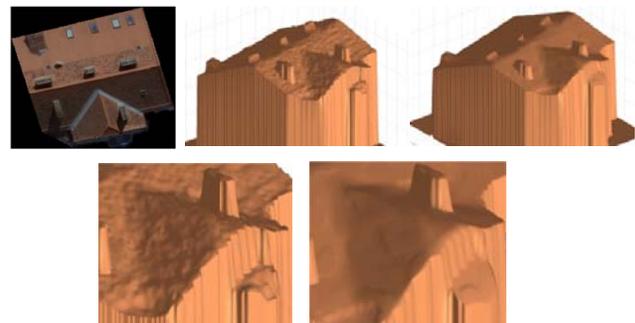

Figure 2: (a)Shows RGB image of single building of our test dataset (b) shows range image without using TGV model (c) shows range image using TGV model (d) shows detail of raw range data (e) shows detail of range data using TGV.

2.2 Plane Detection

The smoothed point clouds are now the input to plane detection. This applies to façades as well as to roofs. We use the "J-Linkage" method introduced by [14] that resembles the RANSAC method. The approach is described in [14] in detail, but we will present a short summary here. The process starts by random sampling to generate model hypotheses. The big difference to RANSAC is that minimal sets are constructed in a way that neighboring points are selected with higher probability. That means that if a point x_i has already been selected, that x_j has the following probability of being drawn:

$$P(x_j | x_i) = \begin{cases} \dfrac{1}{Z} exp - \dfrac{\|x_j - x_i\|}{\sigma^2} & if\ x_j \neq\ x_i \\ 0 & if\ x_j \neq\ x_i \end{cases}$$

Z … normalization constant
σ … heuristically chosen constant
x_i, x_j … single observations

RANSAC in contrast treats all points with the same weight that is in our case undesirable. Each surface point gets associated with the characteristic function of the set of random models that fit the point. Upon the creation of all hypotheses one computes for each point a preference set (set of hypothesis it prefers). Points that belong to the same structure have a similar preference set, meaning they are close in the conceptual space.

A tailored agglomerative clustering groups points belonging to the same model, where at each step the two clusters with the minimum pairwise distance are merged. This distance between two preference sets is calculated using the Jaccard distance:

$$d_j(A, B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}$$

where A and B are two sets. It measures the degree of overlap of the two sets and reaches from 0 (identical sets) to 1. Elements are linked together if their preference sets overlap. Outliers are represented as small clusters. Figure 3 shows the result of the plane detection for one building of our test dataset.
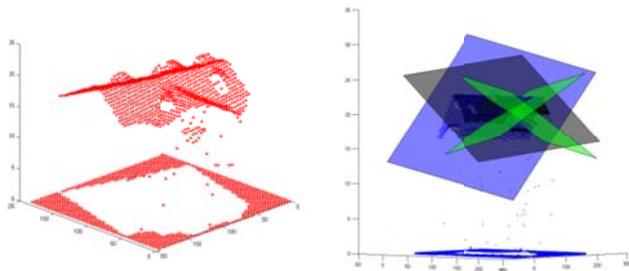

Figure 3: (left) Axonometric view of smoothed 3D point cloud of building roof (GSD 10cm) (points are highlighted in red); Dimensions: x,y-axis [in pixels], z-axis [in m] (right) 3D Point cloud of roof highlighted in blue; overlaid are the two major planes, marked in blue and the two smaller planes in green.

2.2.1 Elimination of Small Structures

The definition of segments of roof planes is followed by an elimination of smaller structures. These could for example be dormer windows or chimneys.

There are two main parameters that are important for this processing step: First, the size of the individual segments of planes must exceed a threshold for the segment to be retained. Additionally we need to consider the possibility that segments of planes are linked and really do represent parts of a larger and therefore relevant segment. And second the existence of height discontinuities at the borders of a segment. These discontinuities within a roof can be found primarily at dormer windows and chimneys. Figure 4 illustrates a building roof where some structures have been eliminated on the basis of the rules described here.

Figure 4: RGB image of building and building mask without smaller structures (downsampled by factor 4).

2.3 Roof Shapes

The roof planes are now available as an input into a classification of the roofs. The analysis is complex because building configurations can be rather complex. Therefore it makes sense to consider the building footprints, much as has been done in the modeling of facades in [11]. One can decompose a building footprint into so-called "cells" representing simple basic shapes, using a method introduced in [3]. Should an individual cell appear to still be complex, we review the 3D point cloud itself. The point cloud of a cell gets associated with different height classes using a method introduced in [15].

It is necessary to differentiate between simple and complex building footprints. The complexity of buildings and thus roofs may include urban L- and T-shapes (see figure 5). A simple footprint is of a building consisting of only four faces (see figure 7). Complexity gets increased if roof elements are connected, as shown in figure 9: a footprint is thus complex when a building has more than 4 faces and if it has connected roof elements. This is the case for L- ,T- and miscellaneous other buildings like the ones illustrated in figures 5 and 8. The complexity of the building is thus represented by the complexity of its footprint, and this must be determined. The big problem is that the used building classification (see figure 5b) is not very accurate because of misclassifications due to vegetation and shadows. Therefore to enhance our roof interpretation we have to determine the emerging masonry. We use a method to reconstruct building façades in 3D that was presented in [11]. This resulted in a description of the complex 3D façade and a refined building footprint. Figure 5 illustrates how the footprint gets represented by straight line segments.
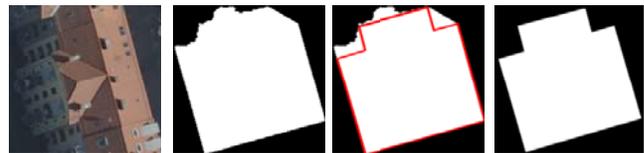

Figure 5: (a) building visible in vertical aerial image (b) segmentation result (c) in red modified building outline using 3D façade reconstruction (d) enhanced building footprint.

After the determination of the building footprints in a next step we have to split the footprint into mostly quadrilateral regions to ease a further interpretation. A method presented in [3] divides a building footprint into mostly quadrilateral shaped polygons in accordance with figure 6. The decomposition considers the requirement that the number of cells be a minimum. This can be done by defining an adequate subset that still reflects the characteristics of the building.
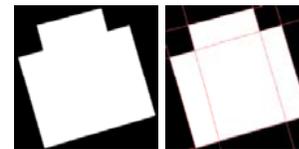

Figure 6: Decomposition of building footprint from figure into quadrilateral cells.

The cells now determine the roof shape of the building. We are particularly interested the standard roof shapes illustrated in figure 7 and table 1, and their combinations when roofs get connected as illustrated in figure 9.
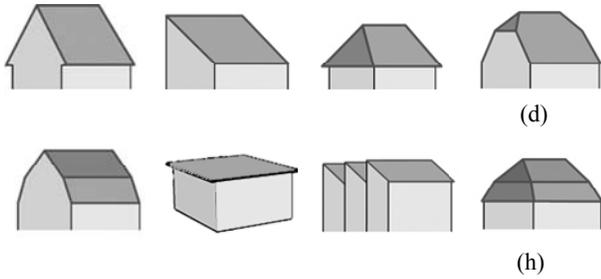
(d)

(h)

Figure 7: Different roof shapes (a) gable roof (b) shed roof (c) hip roof building (d) half hipped roof (e) mansard roof (f) flat roof (g) saw-tooth roof (h) mansard hipped roof.

A fully automatic classification of roofs into the standard shapes uses a concept introduced in [14]. It works with "height classes" associated with the 3D point cloud (see figure 8): in our case we rely on a lower, a middle and an upper class.
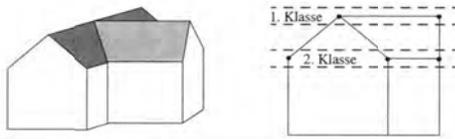

Figure 8: Classification of a building roof into different height classes [15].

This idea is very crucial when dealing with different kinds of hip roofs and mansard roofs. In contrast to the method used in [15] we employ both the roof pitch and the building footprint. Therefore all segments of roof planes are assigned to their appropriate height class. The number of points belonging to each of the three classes is counted per plane. Points in the upper and lower class are of special interest. The ratio between the numbers in these two classes is relevant for the assessment of the roof shapes. This is especially effective with hip roofs. The classification uses the following features of a roof:

- Number of planes
- Roof pitch
- Orientation of normal vector
- Points belonging to lower and upper class and their ratio
- Adjacency matrix (planes)

One important parameter is the orientation of the normal vector. This is important to determine the orientation of two adjacent roof planes. The result of this calculation is a function of the position of two roofs, it is 0 if two planes stand normal to each other, 1 if two planes face in a similar direction and -1 if they look in opposite directions. The use of these parameters and the decomposition of a roof into quadrilateral cells results in a classification of each roof into the major roof shapes of table 1.

Table 1. Different roof shapes

In addition to the principal roof types of table 1, the approach also works with the connecting roof shapes of figure 9.
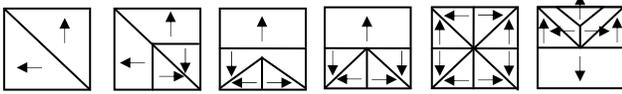

Figure 9. Examples for possible connecting roof shapes

One significant improvement of the proposed method over [3] is the ability to deal with non-symmetric and more general roof shapes.

3. Experiments

The test area covers 400m x 400m near the core of the city of Graz with 186 different buildings. The vertical aerial photography was taken with a GSD of 10 cm and 80% forward and 60% sideward overlaps, using the large format digital aerial camera UltraCam-X.

Ground truth of the roof shapes was collected by hand. Table 2 summarizes the types of roofs. Since some roofs fall outside the standard types, a "miscellaneous" category with 24 entries was required. An additional classification was into "simple" and "complex" buildings, and the latter was further classified into L-buildings (corners), T-buildings and buildings with U- and other irregular shapes. The test area has 125 simple and 61 complex buildings. Of the complex buildings 25 are L-and 17 are T-buildings, and 19 fall into the miscellaneous category.

Table 2. Ground truth: types of roof shapes for the 186-building test area

Flat roof	Shed roof	Gable roof	Hip roof	Half Hipped Roof	Mansard roof	Mansard Hipped roof	Sawtooth roof	Non-allocatable roofs
16	14	121	2	7	0	0	2	24

The smoothing of the DSM not only eliminates outliers, but increases the throughput of plane detections by a factor two. Plane detection requires a focus on the major planes, and thus a meaningful threshold for the acceptable minimal size for plane segments. This threshold is calculated for every building depending on the size of the building footprint and the size of the single plane segments.

The 186 buildings consist of a total of 614 major roof planes. Of these, our approach detected 567 planes. Table 3 presents the detection rates at 92%. A major limitation for the plane detection presents itself when the roofs are curved or very fragmented, thus when there are no planes. The test area has four such buildings.

Down-sampling of the DSM point clouds by a factor 4 is acceptable to first suppress minuscule plane segments and to secondly accelerate the throughput. This down-sampling eliminates small structures in advance. Nonetheless even with this down-sampling not all small structures can be eliminated especially larger dormer windows. Before the roof type can be assigned one needs to eliminate the remaining small structures caused by bigger dormers and chimneys and other extrusions or intrusions. Success in defining and removing such small structures was at 83%, identifying 612 of the 738 smaller roof structures.

Table 3. Different roof shapes

	roof planes	smaller structures
total	614	738
detected	567	612
Detection rate [%]	92	83

The assignment of a standard roof type concerns 162 of the 186 buildings, since 24 fall outside the standards. If we could detect these non allocatable buildings automatically we achieve a detection rate of 88%, thus 145 buildings out of 162 were correctly classified (see table 4). If we include the 24 non-allocatable buildings we achieve 78%.

While the flat roof would seem to be the easiest type to identify, only 88% were correct. This translates to an error in 2 of 16 buildings, and it turns out that these 2 flat roofs have gardens. The shed roofs were identified at a rate of 86% since 2 buildings were incorrectly classified due to large extrusions in the form of large dormer windows. Gable roofs were detected at a 91% success rate. Problems again occur with large extrusions. All hip roofs were correctly detected. However, the half-hipped roofs only were correctly classified in 43% of the cases. The problem with this category derives from the two small plane segments in the upper height class. The current approach eliminates those plane segments during plane detection

The test area did not include any mansard and mansard hipped roofs. We therefore processed 10 additional buildings from another data set (Annecy, France) for each of these two types. All 5 mansard roofs were detected. Errors occurred with the mansard hipped roofs in 2 of the 5 cases. The plane segments in such roofs are at very similar pitches so that plane detection merges plane segments when they should be kept separate.

Table 4. Evaluation of roof shape detection;

Roof shape	Flat roof	Shed roof	Gable roof	Hip roof	Half Hipped Roof	Mansard roof	Mansard Hipped roof	Saw-tooth roof	Total
Total number	16	14	121	2	7	0	0	2	**162**
Detected roof shapes	14	12	112	2	3	0	0	2	**145**
Detection rate (%)	88	86	91	100	43	-	-	100	**78**

4. Conclusion

We propose a method for automatically mapping roofs and classifying them into architecturally accepted standard types, based on traditional digital large format color aerial photography. This method relies on point clouds at 25 pts/m^2 extracted from highly overlapping vertical aerial imagery, and on an image classification using color and texture to delineate building outlines and footprints. Experimental work in a Graz-test area with 186 buildings with 614 roof planes, results in correct roof planes in 92% of the cases. Small roof structures do confuse the analysis and must therefore be detected and eliminated. This is successful in 83% of all the test cases. Roof types get classified correctly at a rate of 88%. LiDAR literature quotes its success with roof type assignments at xxx%. Limitations exist with complex roof shapes that include large dormers, or with curved roofs. However, those

difficulties are not specific to aerial photography and also present themselves with LiDAR data. We may thus conclude that aerial photography produces results at least as good as those from LiDAR point clouds.

While we propose to continue with the development of roof analysis work based on aerial photography, the experimental work has revealed that some weaknesses exist in the proposed method. These weaknesses from confusions with large dormers, or roof gardens, or small differences in pitch angels of mansard hipped roofs will be addressed. In addition work is needed with a wide range of buildings in diverse areas of the World where architectural styles are different from those in the initial test area Graz. An initial look at French urban areas immediately shows the strong use that is being made there of curved roofs. Snow-free coastal resort environments may have various types of flat roofs, historical small towns, alpine towns, urban cores with skyscrapers and the urban fringe with its industrial zones will offer different challenges.

We expect that a focus will be needed next on the analysis of the detailed extrusions, sometimes intrusions, on roofs. These concern chimneys, dormers, sky lights, terraces etc. Success with those details will reflect back on an improvement of the roof analysis and type assignments.

5. REFERENCES

[1] Maas, H.-G. and Vosselman, G. 1999. Two Algorithms for Extracting Building Models from Raw Laser Altimetry Data. *ISPRS Journal of Photogrammetry and Remote Sensing* 54 (2-3), 153-163.

[2] Haala, N., Brenner, C. and Anders, K.-H. 1998. Generation of 3D City Models from Digital Surface Models and GIS. *ISPRS Workshop on 3D Reconstruction and Modelling of Topographic Objects*, Stuttgart, Germany, 68-75.

[3] Kada, M. and McKinley, L. 2009. 3D building reconstruction from lidar based on a cell decomposition approach. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38, Part 3/W4, 47–52.

[4] Dorninger, P. and Pfeifer, N. 2008. A comprehensive automated 3D approach for building extraction, reconstruction, and regularization from airborne laser scanning point clouds. *Sensors* 8, (11), 7323–7343.

[5] Milde, J. and Brenner, C. 2009. Graph-based modeling of building roofs. *AGILE Conference on GIScience*, Hannover, Germany (on CD-ROM).

[6] Rottensteiner, F. 2010. Roof plane segmentation by combining multiple images and point clouds. In *Proceedings of Photogrammetric Computer Vision and Image Analysis Conference* 38, Part 3A, Paris, France, 245-250.

[7] Jochem, A., Höfle, B., Rutzinger, M. and Pfeifer, N. 2009. Automatic Roof Plane Detection and Analysis in Airborne Lidar Point Clouds for Solar Potential Assessment. *Sensors* 9, 5241-5262.

[8] Elberink S. O. and Vosselman G. 2011. Quality analysis on 3D building models reconstructed from airborne laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, Issue 2, 157-165, ISSN 0924-2716.

[9] Hartley, R., Zisserman, A. 2004. *Multiple View Geometry in Computer Vision.* Second Edition. Cambridge University Press, 219-243.

[10] Klaus A. 1997. *Object Reconstruction from Image Sequences.* Doctoral Thesis, Graz University of Technology.

[11] Meixner, P. and Leberl, F. 2010. From Aerial Images to a Description of Real Properties: A Framework. In *Proceedings of the Int'l. Conference on Computer Vision Theory and Applications,* Angers, France, 283-291.

[12] Meixner, P. and Leberl, F. 2010. Characterizing Building Façades From Vertical Aerial Images. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38, Part 3B, 98-103.

[13] Pock, T., Zebedin, L. and Bischof, H. 2008. TGV-fusion. In *Rainbow of Computer Science*. Springer-Verlag. To appear.

[14] Toldo, R. and Fusiello, A. 2008. Robust Multiple Structures Estimation with J-linkage. In *Proceedings of the European Conference on Computer Vision*, Part 1, 537-547, ISBN: 978-3-540-88681-5

[15] Dan, H. 1996. *Rekonstruktion generischer Gebaeudemodelle aus Punktwolken und deren Abbildungskorrekturen in Orthobildern.* Doctoral and Habilitation Theses, ETH Zuerich. doi:10.3929/ethz-a-001693020.

[16] Kassner, R., Koppe, W., Schüttenberg, T. and Bareth, G. 2008. Analysis of the solar potential of roofs by using official lidar data. In *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 37, Part B4. Vienna, Austria, 399–403.

[17] Kolbe T.H., C. Nagel, A. Stadler (2009) *CityGML – OGC Standard for Photogrammetry*? Proc. of the Photogrammetric Week 2009, Wichmann Verlag, pp. 265-278.