INTERPRETATION OF 2D AND 3D BUILDING DETAILS
ON FACADES AND ROOFS

P. Meixner [a, *], F. Leberl [a], M. Brédif [b]

[a] Institute for Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria
(meixner, leberl)@icg.tugraz.at
[b] Université Paris-Est, Laboratoire MATIS, IGN, 94165 Saint-Mandé, Cedex
bredif@ign.fr

Working Groups I/2, III/1, III/4, III/5

KEY WORDS: Facades, roofs, semantic segmentation, aerial images, street side images, multi view

ABSTRACT:
Current Internet-inspired mapping data are in the form of street maps, orthophotos, 3D models or street-side images and serve to support mostly search and navigation. Yet the only mapping data that currently can really be searched are the street maps via their addresses and coordinates. The orthophotos, 3D models and street-side images represent predominantly "eye candy" with little added value to the Internet-user. We are interested in characterizing the elements of the urban space from imagery. In this paper we discuss the use of street side imagery and aerial imagery to develop descriptions of urban spaces, initially of building facades and roofs. We present methods (a) to segment facades using high-overlap street side facade images, (b) to map facades and facade details from vertical aerial images, and (c) to characterize roofs by their type and details, also from aerial photography. This paper describes a method of roof segmentation with the goal of assigning each roof to a specific architectural style. Questions of the use of the attic space, or the placement of solar panels, are of interest. It is of interest that roofs have recently been mapped using LiDAR point clouds. We demonstrate that aerial images are a useful and economical alternative to LiDAR for the characterization of building roofs, and that they also contain very valuable information about facades.

1. INTRODUCTION

Accurate and realistic 3-dimensional models of the urban human habitat are gaining importance for virtual tourism, city planning, internet search and many emerging municipal engineering tasks. They also represent location information for the evolving field of ambient intelligence. Internet search in Bing Maps or Google Earth is supported by 3D virtual cities worldwide. At this time these data are just used for visualization, but this is on the way to change. On the horizon are urban models that consist of semantically interpreted objects. In its most sophisticated form, each building, tree, street detail, bridge and water body is modeled in three dimensions, details such as windows, doors, facade elements, sidewalks, manholes, parking meters, suspended wires, street signs should exist as semantically identified objects.

Automatic mapping of facades and roofs in 3D is a fundamental element in building 3D virtual cities, but the tasks are surprisingly complex. We present in this paper several avenues of research we are pursuing to achieve automation from image sources, mostly in the form of aerial photography, but also in the form of street side images. For example, we have to define a building, its facades and roof with 3D detail extruding from or intruding into predominant planes. Such detail may be stair cases, balconies, awnings, dormers, chimneys, terraces, elevator shafts, air conditioning units, roof gardens and the likes.

Street side images are very important current sources of building information because such imagery is being created by both the vehicle-based industrial data collection systems as well as by Internet users in the form of Community Photo Collections. We therefore study facades using overlapping street level images. Recky et al. (2011) have shown that individual facades can be detected with a success rate of 97%. However, facades are also imaged in vertical aerial vertical image at the edge of the fields of view. Normally vertical aerial images are used for orthophotos and the mapping of roofscapes. The idea of using such data for the analysis of vertical walls may surprise. We demonstrate that the idea is valid, and that aerial vertical imagery is a good source to model building facades in 3D based on plane sweeps.

Mapping of roofs is also a fundamental element in characterizing buildings. The majority of research is based on LiDAR point clouds. We show that digital aerial images and point clouds extracted from them serve well for the characterization of building roofs. Roofs need to be modelled by their major planes and thus the architectural roof style, and by their 3D detail of chimneys, dormers, sky lights, terraces and such. An initial test area supports the conclusion that roof planes can be correctly mapped in 89% of all cases, and that the assignment of roofs to their roof type is successful at a rate of 82%.

---

* Corresponding author.

2. FACADE DETECTION USING STREET SIDE IMAGES

2.1 Approach

Street-side images need to get interpreted and facades need to get identified, each facade as a separate entity (Recky et. al. 2010, Hammoudi K. 2011). A street side will appear as a continuous agglomeration of connected buildings. At issue is the splitting of a building block into its individual buildings. The data source is a set of overlapping, thus redundant images taken from a moving vehicle carrying calibrated automated cameras.

An initial segmentation divides the image into different contents like sky, cloud, roof, building, ground, vegetation, shadow and undecided. The segmentation was described by Recky et. al. (2010) and computes image patches using a watershed segmentation. Patches are consequently merged into larger segments depending on color and texture. A graph is then constructed where every image segment is a node and the edges define the type of relationship of the segments with one another. These relationships are examined using discriminative random fields (DRF). The spatial relations between segments represent context and permit one to differentiate between ground and sky or roof and facade. In a test area in Austria, the detection of building facades achieved a success rate of 94%.

The result of the previous step produces facade areas, not individual facades per building. In a next step repeated patterns in the images get associated with separate facades. The approach was introduced by Wendel et al. (2010) based on Wendel (2009). The method uses Harris corners as interest points. In a next step the color profile between every interest point and the 30 nearest neighbors is calculated. The color profiles are constructed using a 20-dimensional normalized descriptor for each of the three colors RGB, in total thus with 60-dimensions. A kd-tree method is then used for matching the descriptors. In a last step the repetitive patterns are located in a voting matrix.

In a next step the processing of the single facade is discussed in more detail. Due to the natural settings of objects in these images we assume that repetitive patterns occur along the horizontal direction and the separation of the facades occur in vertical direction. Therefore the lines between the matched interest points are projected into the horizontal axis constructing a match cost histogram. Then the facades are segmented by determining a separation area (area where one facade ends and the next begins). This is done by defining areas with a low likelihood as separation areas and areas with high likelihood as repetitive areas. To be able to determine the exact split between two facades in a last step they look for the global maximum in these areas.

By applying the pattern-based facade separation on image segments previously identified as facade space, the results improve over those achieved without the use of facade spaces (Recky et al., 2011). Figure 1 illustrates the result of the facade segmentation showing 4 test images.

2.2 Experimental Results

Tests were based on 9 separate building facades shown in 20-50 overlapping photos. The images are taken in a forward look so that the facades are shown under an oblique angle. This helps in

evaluating the influence of the perspective distortion. A detection rate was achieved of 97% (see Figure 1).

Figure 1: Four examples of automatically segmented street side images into individual building facades. Use is being made of overlapping street-side images. In this example, the camera is pointed forward (from Recky M. et al., 2011).

2.3 Discussion

The separation of buildings using their facades is an alternative to the separation of buildings by cadastral property data. Ideally, these two approaches produce the same result. However, buildings can extend over multiple properties, or one property can hold more than one building. Future work will have to address the comparison between the two alternatives and find means of resolving any discrepancies.

Future work will also have to deal with different architectural styles, non-plane facades, various imaging modalities and the problems with occlusions from vegetation. As we argue that aerial photography also offers facade information, the approaches thus far designed to work with street level images should be applied to aerial imagery as well.

3. FACADE DETECTION USING VERTICAL AERIAL IMAGES

3.1 Approach

Aerial vertical images are less affected by occlusions from vegetation and are available "freely" and at no cost of acquisition. The data will have been collected for traditional urban mapping anyway. Looking at facades in vertical aerial images offers therefore an added benefit. We combine aerial images with cadastral information. We have developed a framework for building characterization that is strictly built for aerial photography (Meixner et al., 2011). We start out by merging the aerial imagery with property boundaries to define each property as a separate entity for further analysis. The cadastral data may also contain preliminary information about a building footprint. In the next step the building footprints get refined vis-à-vis the mere cadastral prediction based on an image classification and on the definition of roof lines. 3D facade coordinates are computed from aerial image segments, the cadastral information and the DTM. This helps to determine the number of floors, the window locations (see figure 2) and offers candidates for attic and basement windows.
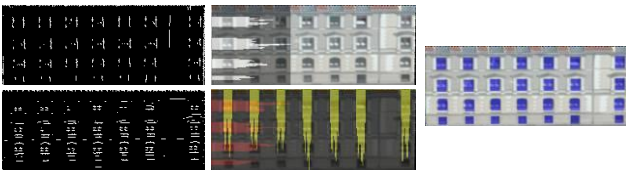
Figure 2: Processing steps for floor and window detection. (a) Horizontal and vertical edges, (b) maxima search in horizontal projection profile and overlay of the two profiles, (c) result of the window detection with highlighted window locations in blue. The count of floors and windows coincides with ground truth.

This method is well suited when a facade is generally plane, but fails with complex facades with extrusions like balconies, staircases and awnings. Figure 3 illustrates a failure.

Figure 3: To the left is a rectified facade image with a depth structure, to the right a failed count of windows. The 3D structure needs to get considered.

To be able to deal with those facades we have to reconstruct them in 3D so that we can separate a facade in planar segments. This is achieved with the so-called plane sweeping method along the proposal by Zach (2007), with its advantage that one no longer needs to assume a single vertical plane per facade but also complex facades with awnings, bay windows, staircases and balconies can be analyzed. The plane sweep operates with multiple planes that lie parallel to a key-plane. A *key-plane* is the approximate facade-plane. Additional planes are set parallel to the key-plane about one pixel apart (in our test area, this is at 10 cm) in both directions from the key-plane (see figure 4).
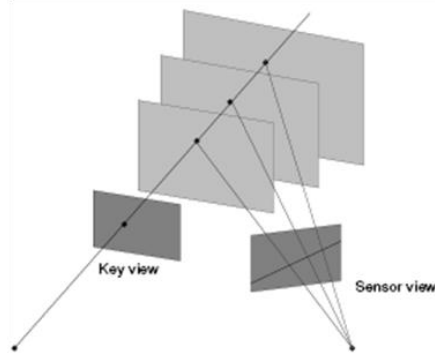

Figure 4: Plane sweeping principle. The homography between the facade's reference plane and the sensor view varies for different depths. (Zach, 2007).

If the plane at a certain depth passes exactly through parts of the object's surface to be reconstructed, a match will exist between the relevant parts of the new sensor view and the key view, the match being computed as a correlation. The sensor images are warped onto the current 3D key plane using the projective transformation.

After projecting a sensor image onto the current plane hypothesis, a correlation score for the current sensor view is calculated. The final correlation score of the current plane hypothesis is achieved by integrating all overlapping sensor views. For the accumulation of the single image correlation

scores a simple additive blending operation is used. We repeat this process for all parallel planes and all corresponding images. The results of this calculation are n_k matching probabilities for every pixel x(i,j) of a facade for all n facade planes. Figure 5 illustrates the result of this correlation for 4 different planes.

In a next step we determine the depth map of a facade using a total generalized variation TGV multi labelling approach proposed by Pock et al. (2008). Figure 6 shows the resulting raw 3D point cloud and the 3D point cloud overlaid with RGB photo texture.

Figure 5: Correlation coefficients calculated for 4 different planes visualized as binary images (white areas have the largest correlation values).
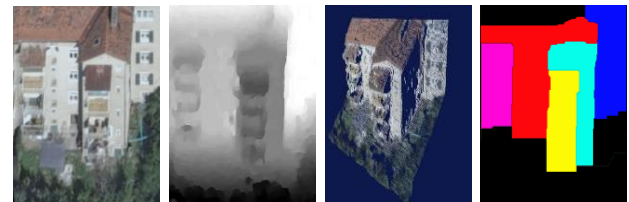

Figure 6: Reconstructed building facades using plane sweeping. (a) key view of facade (b) raw 3D point cloud (c) 3D point cloud overlaid with RGB information (d) segmentation of that point cloud into facade areas belonging to specific vertical planes.

The method produces a 3D point cloud that can now be used to determine if a facade is planar or complex depending on how many points of the 3D point cloud lie within a certain range of a regression plane. One now has to analyze the 3D points with the goal of segmenting the facade into its planar sub-facades. The problem is one of detecting planes in 3D point clouds for which various solutions exist.

The 3d point cloud is projected into the horizontal xy-plane and will present a footprint of the facade. We thereby reduce the point cloud from 3D to 2.5D, because we just use the most common depth values for every facade.

The projection is by column and starts by eliminating all outliers for every column and by searching for local maxima in each individual column. Then these values are weighted according to their appearance in the column. In a next step we look for maxima in the xy- plane (footprint of facade, see figure 7). These maxima in the xy-plane are the major facade planes of one facade. These resulting multiple planes do represent sub-facades ready for the detection of floors and windows.

The xy-plane contains the footprint. It is the basis for dealing with facade details and masonry of a building. We want to detect balconies. This is possible using the lower weights of the areas that are probably balconies, as shown in figure 7a (green highlighted profiles). Figure 8 illustrates a complex building facade, the footprint of the building and the effect of the elimination of the balconies and roof overhangs.

Figure 7: Plane detection for a complex facade; (a) key-view of a facade and projection in xy-plane (red lines are strongly weighted (>66% of points lie in this plane); (b) determined major facade planes (segmented areas with balconies are marked in red)
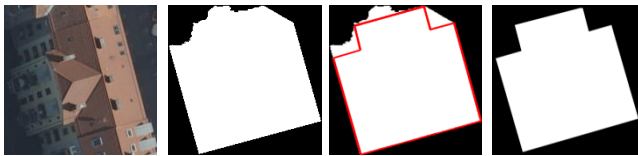

Figure 8: (a) building visible in vertical aerial image (b) segmentation result (c) in red modified building outline using 3D facade reconstruction (d) enhanced building footprint.

3.2 Experimental Results

For the evaluation of the 3D facades we have performed experiments in a test area of the city of Graz with a dimension of 400m x 400m with a Ground Sampling Distance of 10 cm and image overlaps in the range of 80% forward and 60% sideward. We have randomly selected 131 buildings with a total of 230 facades and have performed our plane detection method. Table 1 summarizes the results of this calculation depending on the number of aerial images used for the reconstruction. We achieve a detection rate of 70% for complex facades and a detection rate of 80% for planar facades.

	1 image	2 images		3 images		4 images		5 images		total	
images	5	68		77		52		28		230	
complex	0 %	64%	18/28	63%	15/24	78%	18/23	82%	9/11	**70%**	61/87
planar	0 %	75%	30/40	74%	39/53	86%	25/29	94%	16/17	**80%**	114/143
total	0 %	71%	48/68	70%	54/77	83%	43/52	90%	25/28	**76%**	175/230

Table 1: Evaluation of the facade plane detection using the column-wise approach

The main reasons why the plane detection fails are the combination of aerial images from two different flight strips for the computation and an inability to deal with dissimilarities in images taken with vastly different viewing perspectives. This is the case in about 20 facades of our data set. The results are noisy matching results that influence the outcome of the depthmap. One solution would be the reliance on imagery from just a single flight line. This would imply a very dense arrangement of flight lines and thus an increase of flying costs. At issue therefore would be further research and innovation in dealing with dissimilar facade images to obtain point clouds and matches.

By using these sub-facades for floor and window detection we achieve a detection rate for floors of 87% and for windows of 80%. Without the 3D reconstruction none of the complex facades could be interpreted correctly.

3.3 Discussion

We show that facades are being imaged usefully in vertical aerial photography. We also show that facades cannot be modelled as planes since there often is significant 3D structure. Use of the 3rd dimension for the interpretation of building facades is feasible with aerial photography and strongly improves the results. We achieve success rates of 87% for floor detection and 80% for window detection for facades that fail completely when a plane is assumed to be applicable.

We also show that it is possible to determine the extruding masonry of a building by eliminating balconies and roof overhangs. There are several avenues for improvements of the detail extraction from complex facades, and improvements of our understanding how well this works. First are data experiments in the form of a study with vertical aerial images with different GSD and different overlaps. Second is the ability of automatically recognizing occlusions and then responding to occluded facades by taking advantage of the overlapping images and their multitude of viewing angles. Innovations in window recognition will then become relevant, as will be site-dependent approaches to architectural styles.

4. CHARACTERIZATION OF BUILDING ROOFS USING AERIAL IMAGES

4.1 General Approach

Roofscapes offer similar complexities as facades, with predominant planes and multiple structures extruding from, sometimes also intruding into these planes. Nowadays, the preferred data source is LiDAR, and most of the recent literature on roof analysis is LiDAR-related. However, digital aerial photography is available and can be used. At issue is the segmentation of extended roofscapes into individual roofs, the measurement of the predominant roof planes and then the mapping of the 3D details. We build an approach within an overall framework introduced by Meixner et al. (2010). It employs vertical aerial images in order to characterize real properties.

After pre-processing the data to segment them into point clouds per individual property, and classification of roof areas, we move on to the characterization of the single building roofs. The interpretation of building roofs consists of 3 major processing steps:
- DSM Smoothing
- Plane Detection
- Roof Segmentation

4.2 DSM Smoothing

Photogrammetrically measured elevation data (range data) are noisy at the pixel to sub-pixel level and therefore may not easily be interpreted. For this reason and to accelerate the plane detection we smooth the elevation or range data using total generalized variation TGV using the approach developed by Pock et al. (2011). The result is illustrated in figure 9.
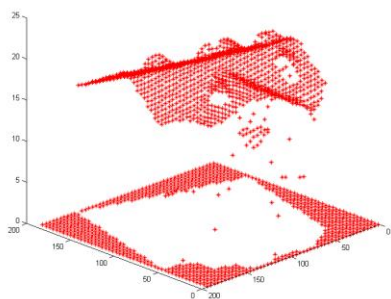
Figure 9: Smoothed point cloud of building roof (GSD 10cm); Dimensions: 200 x 200 pixels

4.3 Plane Detection

Roof planes are found using the "J-Linkage" method introduced by Toldo et al. (2008) that resembles the RANSAC method. It starts by random sampling where model hypotheses are generated. The essential difference to RANSAC is that minimal sets are constructed in a way that neighboring points are selected with higher probability. RANSAC treats all points the same. After all hypotheses are created, a preference set (set of hypothesis it prefers) is created for each point. Points that belong to the same structure have a similar preference set, meaning they are close in the conceptual space. To find the models "J-Linkage" uses an agglomerative clustering procedure, where at each step the two clusters with the minimum pairwise distance are merged. This distance reaches from 0 (identical sets) to 1 and just elements are linked together whose preference sets overlap. Figure 10 illustrates the result of this plane detection.
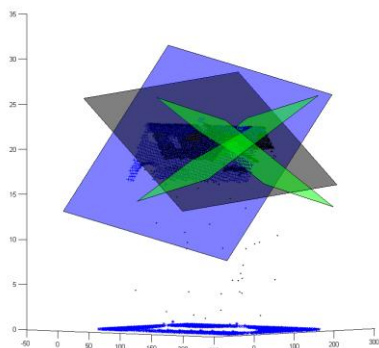

Figure 10: Point cloud from figure 9 highlighted in blue; the two major roof planes are marked in blue and black and the two smaller planes in green.

4.4 Roof Segmentation

After we have detected all roof planes we have to label them. Therefore we segment the roofs into 3 different classes, namely major roof planes, minor roof planes and superstructures. Using the major roof planes we obtain information about the style of the roof.

We determine if a plane segment is linked to any other segment to build a "region". Then the size of the resulting region or assembly of plane segments gets considered. Depending on the size of the regions with respect to the overall size of the roof we assign each region to an appropriate category. A refinement step serves to look at smaller plane segments associated perhaps

with superstructures. Depending on their height values with respect to the neighbouring pixels these smaller regions are classified as part of a smaller plane (regions without height continuities at the borders), superstructures or are eliminated entirely. To achieve meaningful results we differentiate of course between height discontinuities at the edges of the roof and within the roof by using the information about a building from the building classification.

Of particular interest in our case are superstructures because they give us information about the use of the roof and building. We divide the superstructures into three groups: dormer windows, chimneys and other structures. For the example of chimneys all 4 edges have height discontinuities and that the maximum height is not lower than the height of the roof's ridge. By contrast the dormer windows have height discontinuities on at last three edges. Moreover their area is much larger than the area of the chimneys and the geometric form is more "quadratic" than elongated. Chimneys have usually smaller and narrower forms (< 0.5m width). Figure 11 illustrates the segmented roofscapes for two buildings.
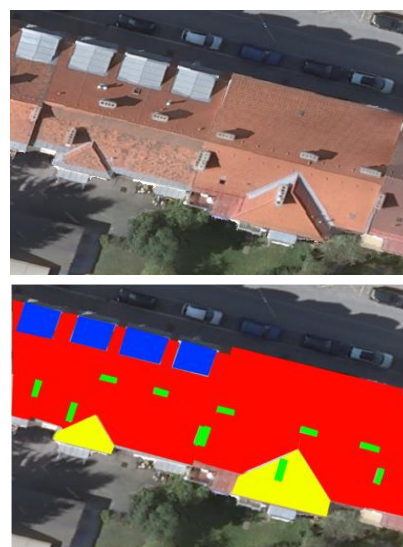

Figure 10: Extract of an orthophoto of our test data set and overlaid Segmentation of roofscape into different roof categories (red: major planes; yellow: minor planes; green: chimneys; blue: dormer windows). [Source: Detail of the Graz test dataset]

4.5 Experiments

The Graz test site has 186 different buildings. A random selection of 20 buildings from this dataset serves as the basis for a segmentation experiment.

	Major planes	Smaller planes	Dormer windows	Chimneys	Other structures
Total planes	43	12	32	84	3
Detected planes	42	9	28	68	3
Detection rate [%]	98	75	88	80	100

Table 2: Detection rate for segmented roof structures

Major planes are being detected with a 98% success, 75% success was scored with smaller planes, 88% of the planes for the dormer windows and 80% of those for the chimneys were

correctly found. The overall success rate over all planes was at 87%. Misclassifications occur in complex roof structures like roof terraces or non-planar roof structures, thus in curved surfaces.

4.6 Discussion

We have presented in this section a method to segment building roofs into different roof categories: major plane, smaller planes, superstructures, and to interpret these in association with types of roofs, dormer windows, chimneys and other structures. We show that we can segment a roof with an accuracy of 87%. The method builds on point clouds and classifications of buildings from overlapping aerial photography with a GSD at 10 cm, and producing thus a point density of 25 pixels/m^2. This data source differs from the current predominant sensor for point clouds in the form of the airborne LiDAR.

Current results do encourage continued development of roof analysis work based on aerial photography. However, the experiments have shown that the proposed method should be improved. Particular difficulties occur when rather large dormers or roof gardens exist. Additionally, we need to extend the experimental effort to include different architectural styles and building uses to include coastal resort environments, historical small towns, alpine terrains, urban cores with skyscrapers and industrial zones.

5. CONCLUSION

GeoVirtual Environments include the advent of 3D virtual cities in the form of 3D models of the urban human habitat. The Internet as an initial inspiration to quickly develop 3D city models has put the focus in the visual appeal of the result, not on the ability to use the building data for any analysis.

We argue that this needs to change and that images need to serve to characterize the real properties, that building details be part of the data base and can be searched. Buildings should be found based not only on an address, but also on the number of floors, the size of facades, the number of windows and the architectural style of a building.

Aerial photography is a work horse for urban mapping and exists for all urban spaces. It contains information about facades and roofs that needs to get extracted. Initial work succeeds in finding relevant information with accuracies in the range of 88% and more. Aerial and street side imagery needs to be used cooperatively to overcome the limitations of each data source, such as occlusions in street side data or poor facade texture in aerial data.

LiDAR has not been addressed in this contribution, although LiDAR is a contender in any urban geo-sensing and is in everybody's mind when 3D point clouds are at issue. We did show that imagery is a valid and useful source of geometric 3D information of building facades and roofs, and by this we want to highlight that digital high overlap imagery produces point clouds for successful roof as well as facade analysis.

References

Hammoudi, K., Dornaika, F., Soheilian, B., Paparoditis, N.. A Pipeline for Modeling Urban Street Facades from Terrestrial Laser and Image Data. *in ICVSS International Computer Vision Summer School*, Sicily, Italy, to appear.

Meixner, P., Leberl, F. 2011. 3-Dimensional Building Details from Aerial Photography for Internet Maps. *Remote Sening.* 3, no. 4, pp. 721-751.

Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D. 2008. A Convex Formulation of Continuous Multi-Label Problems. In: *Proceedings of the European Conference on Computer Vision* ECCV, pp. 792-805

Pock, T., Zebedin, L., Bischof, H. 2008 TGV - Fusion. In: *Rainbow of Computer Science*. Springer-Verlag. To appear.

Recky, M., Wendel, A., Leberl, F. 2011. Facade Segmentation in Multi-View Scenario. In: *Proceedings International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission* [3DIMPVT 2011], Hangzhou, PR China. Published by IEEE Computer Society, pp. 358-365.

Recky, M., Leberl, F. 2009. Semantic Segmentation of Street-Side Images. In: *Proceedings of the Annual OAGM Workshop*. Austrian Computer Society in OCG, pp. 271–282.

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

Wendel, A, Donoser, M., Bischof, H. 2010. Unsupervised facade segmentation using repetitive patterns. In: *Proceedings of the 32nd Annual Symposium of the German Association for Pattern Recognition* (DAGM'10), Springer LNCS 6376, pp. 51-60.

Wendel A. 2009. *Facade Segmentation from Streetside Image.* Master's Thesis, Graz University of Technology, 72 p.

Zach, C. (2007) *High Performance Modelling from Multiple Views using Graphics Hardware.* Doctoral Thesis, Graz University of Technology, 153 p.