

## CHARACTERIZING BUILDING FAÇADES FROM VERTICAL AERIAL IMAGES

P. Meixner, F. Leberl

Institute for Computer Graphics and Vision, Graz University of Technology, Inffeldgasse 16/II, Graz  
(meixner, leberl)@icg.tugraz.at

Commission III – WG III/4

**KEY WORDS:** Aerial images, Semantic image interpretation, Plane sweeping, 3D-Modelling, Facades, Window detection, Floor counting

### ABSTRACT:

Information is being extracted from vertical aerial photography and various data products for an efficient interpretation of terrain objects. Our focus lies on characterizing individual properties as accurately as possible, using aerial imagery. We want to determine the size of buildings, their number of floors, the number and size of windows, the existence of impervious surfaces, status of vegetation, roof shapes with chimneys and sky lights etc. To achieve robust results it is very important to incorporate all data that a set of images can offer. A key aspect therefore is the inclusion of the 3<sup>rd</sup> dimension when interpreting façade images and to deal with the different imaging angles when interpreting the details of building facades from vertical aerial photography. This paper first addresses the question which incidence angles are sufficiently large to get useful results. And secondly it describes a plane sweep algorithm to detect 3D façade objects such as eaves, balconies and porches. We expect to achieve an enhanced quality of the floor counts and window detection results. The topics are closely related since one first needs to understand which façade images have been taken under an angle that is too small, so that the facades excessively distorted. Second, the plane sweep algorithm needs images with as small a distortion as possible, and given a high degree of overlaps, one will need to prune-down the set of images actually used.

### INTRODUCTION

Can one analyze façade details of buildings from overlapping vertical, not oblique, aerial photography, and at what level of completeness and accuracy? Our research seeks to clarify those questions. The topic is relevant since the Internet has started to inspire an interest in modeling urban areas in 3D from vertical and oblique aerial photography, aerial LiDAR and from street side imagery and street side LiDAR. Vertical aerial photography is the workhorse to provide complete maps, orthophotos and dense 3D point clouds (Leberl, 2010). With the transition to digital sensing, image overlaps can increase without adding costs (Leberl & Gruber, 2005). This improves the accuracy, the automation opportunities and it reduces the occlusions. It is thus meaningful to extract from the aerial photography all the information about man-made objects such as buildings, and thus of their facades, from ubiquitous vertical aerial imagery. If successful, such façade information will be created at no added cost for new sensor data.

Our specific interest is the “value” of real properties (Meixner & Leberl, 2010a). That value is certainly based, among other factors, on a characterization of each property by as complete a set of numbers as can be extracted from sensor data. This includes the size of a lot, the building dimensions and building details, the impervious surfaces, the vegetation, the effect on a property by its neighbors such as shadows, etc.

In this paper we focus on façades, and we limit the sensor source data to easily available vertical aerial photography. We show that façades can indeed be mapped in 3D, provided we have aerial images at high overlaps and with small pixel sizes to support the modeling of human scale objects.

### THE PROCESSING FRAMEWORK

We start with geometric data from two sources: the aerial imagery and the cadastral information. Figure 1 shows an example for a 400 m x 400 m urban test area in the city of Graz

(Austria) and the suitable cadastral information of this area. We merge these two data sources to define each property as a separate entity for further analysis. The cadastral data also may contain preliminary information about a 2D building footprint.

In a next step we produce dense 3D point clouds associated with the aerial photography and extracted from it by means of a so-called dense matcher applied to the triangulated aerial photographs (Klaus, 2007), and we segment the data per property. This results in the areas occupied by one building as well as its height. The building footprints get refined vis-à-vis the cadastral prediction using image segmentation and classification to define roof lines.

The next step now is the determination of the façades (see Figure 4): For each façade we must find its intersection with the ground, thus its footprint. This is the basis for a definition of the façade in 3D by intersecting the footprint with the elevation data. We compute the corner points of each façade along the footprint and the roof line. These corners can then be projected into each image of a block of overlapping aerial photography. We can search in all aerial images for the best representation of the façade details; typically this will be the image with the maximum number of pixels for a given façade. Since there are multiple façade images, we prepare for a multi-view process.

What follows is a search for rows and columns of windows in the redundant photographic imagery. First of all, this serves to determine the number of floors. Second, we also are interested in the window locations themselves. And finally, we want to take a look at attics and basement windows to understand whether there is an attic or basement. Figure 2 summarizes the workflow towards a property characterization and represents the framework in which the effort is executed. Additionally the vegetation can be interpreted and roof details like eaves, skylights, chimneys or the type of attic can be detected.



Figure 1: Test area in Graz of 400 m x 400 m, with orthophoto (left) and cadastral map (right). We find here 221 buildings on 336 properties.

A more detailed description of this processing framework and the requirements that the source data must fulfill to successfully interpret the aerial imagery was presented by (Meixner & Leberl, 2010a).

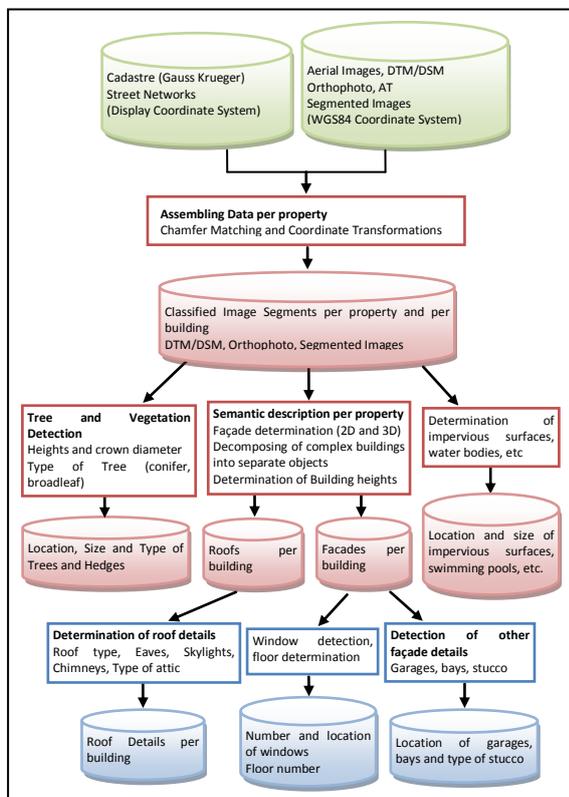


Figure 2: Diagram of the proposed work flow to characterize real properties from aerial images and associated cadastral data (from Meixner & Leberl, 2010a).

Figure 3 illustrates the floor counts and detected windows in each façade of one building. As one can easily determine, the automated floor count and the count of the windows is consistent with a visual inspection. The floor and window detection methods were discussed previously (Meixner, Leberl, 2010b).

However, this result is not always unambiguous. For that reason we proceed with the dependence of the result on the angle-of-view and the incidence angle under which a facade may appear. Furthermore we need to cope with the possibility that balconies, bay windows and other porches cause complications. The

reason why we speak about viewing angles is that we will introduce a plane sweeping approach that will benefit from keeping the number of images small, and thus only the most useful images should enter the analysis process.

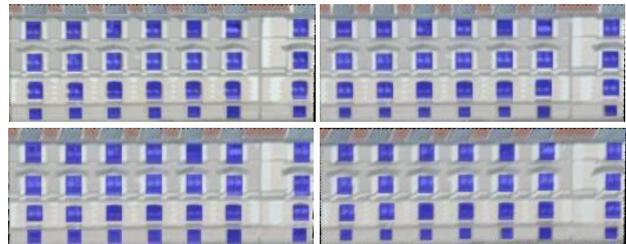


Figure 3: Four images of one façade of a building from 4 aerial images; it shows the result of the independent floor and window counts. The blue rectangles define the detected window positions.

## ANGLES-OF-VIEW AND FAÇADE INTERPRETATION

### Angles and Distortions

The angle of view of a high resolution digital aerial image of a facade affects the floor and window detections. To assess the effect we have performed experiments in a test area of the city of Graz with a dimension of 400m\*400m with a Ground Sampling Distance of 10 cm and image overlaps in the range of 80% forward and 60% sideward (see Figure 1). The images were taken with the digital aerial camera UltraCamX with a lateral angle of view of 55 degrees.

This means that in the lateral left and right borders of the aerial image the façades can be examined under a maximum viewing angle of 27.5 degrees. The newest UltraCam model is the UltraCam XP-WA with a lateral angle-of-view at 70 degrees. Façades will thus be seen under maximum angles of 35 degrees. This angle approaches those available from current oblique cameras where angles are in the range of 45-60 degrees (see Prandi, 2008).

The angle of view is one of the key aspects when interpreting façades using aerial images. Depending on the angle under which a façade is photographed the analysis can deliver useless data. The steeper the incidence angle is, the more distorted the pixels will appear in the plane of the façade. Table 1 illustrates the pixel dimensions which nominally are square at 10 cm in the horizontal plane, but which in fact will appear as rectangles on the façade.

Angle (deg)	0	5	10	15	20	25	30
Pixel vertical [cm]	∞	114	57	37	27	21	17

Table 1: Conversion of the incidence angle into a vertical pixel dimension. Pixel size is nominally 10 cm.

The high overlaps of the images ensures that each given façade is likely to be imaged near the edge of an image and thus at a large incidence angle in the range of 20° or more.

### Façade Experiments

Of the total of 221 buildings in the test area in Fig. 1, 104 buildings with 233 façades were selected for an evaluation. The selection was to include complete façades not excessively obscured by vegetation. Proceeding façade by façade, all images get collected that indeed do show that façade. Then those images were eliminated that present the façade under an

incidence angle smaller than a defined minimum, set at 5 degrees. Even after the elimination of all compromised images, we still find on average that the façades are shown in 4 images. We find thus 870 façade images for the 233 façades, and those are taken from a total of 20 aerial photographs covering the area of Figure 1.

We are interested in a floor count and find that in 694 out of 870 façade images, the number of floors has been correctly determined, resulting in an accuracy of about 80%. Separately we have a correct window count in 573 out of 870 façade images, representing an accuracy of about 65%. Table 2 illustrates the dependency of the floor count on the incidence angle. Table 3 presents the rate of the correct window detection.

Angle [deg]	< 5	5-10	10-15	15-20	20-25	> 25
Floor detection	0	$\frac{7}{21}$	$\frac{79}{103}$	$\frac{191}{221}$	$\frac{255}{279}$	$\frac{228}{246}$
Percentage rate	0%	33%	77%	86%	91%	93%

Table 2: Floor detection rate

Angle [deg]	< 5	5-10	10-15	15-20	20-25	> 25
Window detection	0	$\frac{6}{21}$	$\frac{69}{103}$	$\frac{174}{221}$	$\frac{233}{279}$	$\frac{212}{246}$
Percentage rate	0%	29%	67%	79%	83%	86%

Table 3: Window detection rate

One can see from Table 2 that the floor detection, even at a compromised incidence angle of only 15 degrees, is correct in 86% of the façades. One may conclude that when the incidence angle gets as small as 10 degrees, the floor detection is no longer meaningful. Similarly, the detection of windows is correct in 80% of the cases at incidence angles of  $\geq 15$  degrees.

Tables 2 and 3 contain results computed from single images of single façades. By a fusion of the analysis per façade from more than a single image, one can increase the rate of success. Additionally, one can support the floor count by the introduction of constraints, for example for the height of windows/floors or their spacing. With the consideration of these constraints, the success of floor counts increases from 80% to 93%.

If one imposes onto the window detection a minimum and maximum dimension of a window and a minimum distance between windows, that accuracy can as well be improved and our experiments show the success increasing from 65% to 88%.

**Discussion**

Of course, these experimental accuracies may differ from test area to test area. The current numbers all concern the Graz test. As we proceed with our work, we will increase the diversity of test areas. This is expected to address both urban and suburban cases as well as various types of urban settings.

**CONSIDERING THE COMPLEXITY OF FAÇADES**

The above characterization of facades was based on the assumption that a façade is planar and thus essentially has two dimensions (x and y). There are cases where this 2-dimensional approach to detect windows and floors will fail. While problems will be caused by vegetation and by one building covering up

another, our interest is in coping with balconies, bay windows, attached staircases and elevator shafts. These all do represent a deviation of a façade from a planar object.

Figure 4 illustrates the footprint of a façade and shows how façade details extend into the third dimension. When the emphasis is on fast throughput and a simple approach, the third dimension gets suppressed, as seen in Figure 4c.

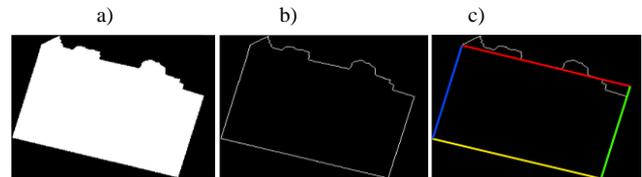


Figure 4, The classification layer “building” is based on color and texture. (a) shows the binary layer, (b) its contour and finally in (c) the simplified geometric figure of the façade footprint.

Problems will exist if parts of the façade lie in different planes. Figure 5a is the rectified image of a façade with balconies and awnings. A search for “floors” and “windows” in Figures 5b and c fails.

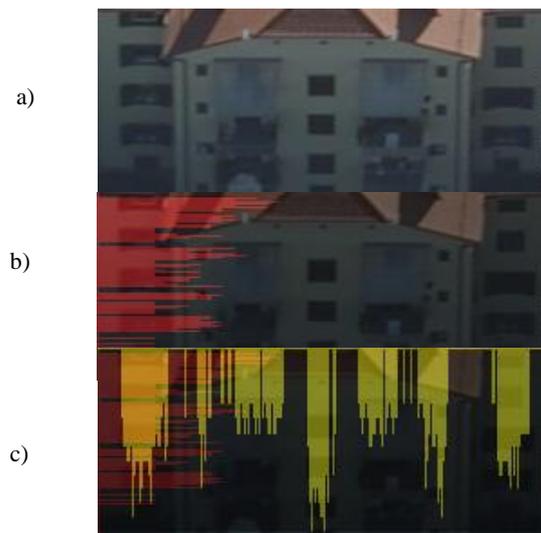


Figure 5: a) rectified façade image. b) the floor detection remains ambiguous; c) the window detection finds vertical structures but fails in the horizontal segment of the automated approach presented in Meixner & Leberl (2010a).

A possible elimination of these problems could be a splitting of the façades into multiple façade fragments. However, for our experimental data set of 104 buildings with 233 facades this method would yield a quadruple number of facades, and each image would only show a small element. One will have to cope with ambiguities because of the smallness of the façade elements.

A more promising solution is the explicit consideration of the 3<sup>rd</sup> dimension. We want to use the so-called *plane sweeping* method with its advantage that one no longer needs to assume a single vertical plane per façade. One starts from an approximate location of a façade. The method produces a 3D depth map for a façade. We employ a method that has been described by Zebedin et al. (2006). It supports the definition of architectural elements that stick out from a façade.

### USE OF THE 3<sup>RD</sup> DIMENSION

The plane sweep produces a depth map. The algorithm consists three distinct steps:

- the 2D space is iteratively traversed by planes that are parallel with the main façade plane;
- an image is being warped;
- a multi-view correlation is performed.

#### Plane sweep

The plane sweep approach is a well established method in computer vision for an image based reconstruction from multiple views. This is in contrast to traditional computational 2-image stereo methods. A 2D space is defined by multiple planes that lie parallel to the “key-plane” (see Figure 6).

A *key-plane* is the approximate façade-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.

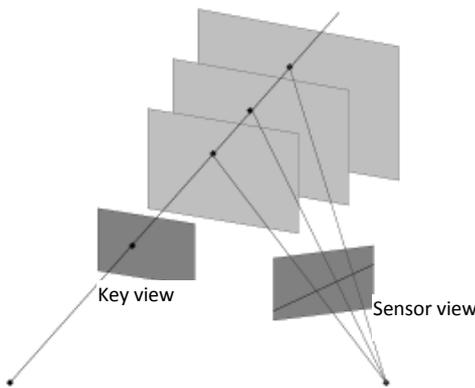


Figure 6: Plane sweeping principle. The homography between the façade’s reference plane and the sensor view varies for different depths.

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.

#### Image Warping

The sensor images are warped onto the current 3D key plane using the projective transformation. This appropriate  $H$  is obtained from

$$R = R_2 * R_1^T$$

$$t = t_2 - R_2 R_1^T t_1$$

$$H = K(R - t n^T / d) K^{-1}$$

$K$  ... intrinsic Matrix of the camera

$(R|t)$ ... relative pose of the sensor view

Details on epipolar geometry and the mathematics behind these equations are described in (Hartley & Zisserman, 2004).

#### Image Correlation

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. The

calculation of the correlation coefficient  $r$  for window-based local matching of two images  $X$  and  $Y$  is

$$r = \frac{\sum_{i \in W} (X_i - \bar{X}) * (Y_i - \bar{Y})}{\sqrt{\sum_{i \in W} (X_i - \bar{X})^2 * \sum_{i \in W} (Y_i - \bar{Y})^2}}$$

This formula is invariant under affine linear changes of luminance between images. To receive robust results for the determination of the correlation coefficient neighboring pixels are added up. This can be done for a neighborhood of 3x3 pixels. Figure 7 shows the result of the image correlation for 4 different planes.

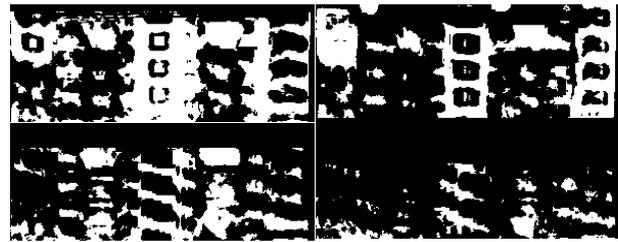


Figure 7: Correlation coefficients calculated for 4 different planes visualized as a binary image (white areas have strongest correlation)

For the determination of the best correlation coefficients we use a winner-takes-all (WTA) approach. With the WTA method first the median of all images of one plane will be determined and stored in a matrix  $M_k$ . In a next step, all pixels  $i,j$  from the previously determined matrices  $M$  are compared with each other and the maxima get searched. These results can directly be converted to a 3D depth map and visualized. For a more detailed description of this method (see Zach, 2007).

### EXPERIMENTAL RESULTS

In our experimental data we have produced parallel planes with a distance between each other of 1 pixel (10cm) to gain as much information about the structure of a façade as possible. That means that the depth map can have a maximum accuracy of 10cm if it would be possible to eliminate noise in the depth map. Figure 8 shows a result of this calculation, where we tried to find the major planes of a façade.

For the determination of the major planes of one façade we were using the result shown in Figure 8 and calculated column wise the most likely planes. The calculation was done by determining the mean values for every column and comparing this value with the neighboring columns.

Figure 9 shows the final footprint of the façade from Figure 5 and the 4 detected major planes. Figure 10 illustrates the resulting façade image split into planes according to the results from Figure 9.

As one can see in Figure 11, the splitting of one facade in smaller parts can improve the floor and window detection rate (see Figure 5). In Fig 11a), b), d), f) and g) all windows and floors are properly detected. The same results were achieved for other complex facades, where the floor and window detection rate could be improved by the use of plane sweeping. We were repeating these for other façades and received almost everywhere the same results.

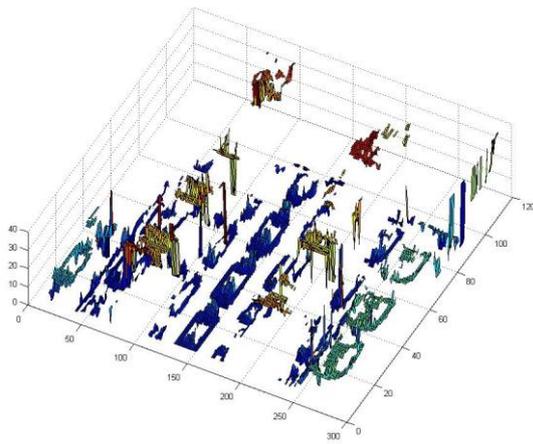


Figure 8: Simplified depth map of the façade image.

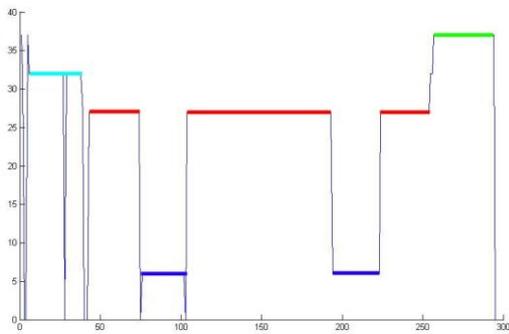


Figure 9: Footprint of the façade from Figure 5a. We find that 4 major planes of the façade exist, using different colors to mark the different façade planes.

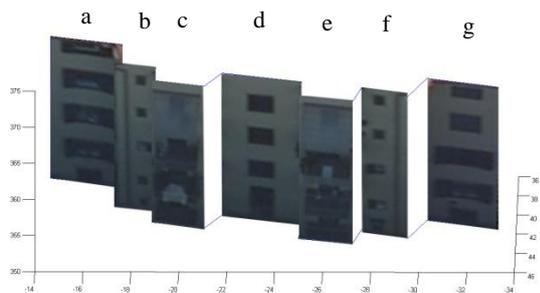


Figure 10: The 4 planes of a façade in 3D, with photo texture.

Using these major façade planes we can also improve our building classification. It is not possible to determine the emerging masonry using aerial images because image segmentation cannot deal with roof overhangs. Moreover there are also misclassifications within the segmentation image that have to be removed for a robust characterization of real properties. Using plane sweeping we can calculate the major façade planes and through that we can calculate the emerging masonry by projecting the footprint of a façade on the footprint of the building as one can see in Figure 11.

### CONCLUSION

We are describing approaches to count floors and windows of buildings using vertical aerial photography and images of those

façades at rather steep look angles. First test results have shown that for façades with very little architectural detail, aerial photography will deliver a floor count with an accuracy of 93%, and a window count with 88% accuracy, using test data from Graz (Austria).

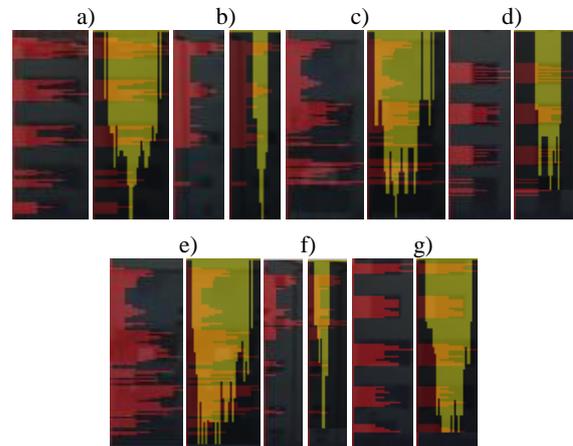


Figure 11: Floor and window detection separately in each of the 7 segments of the single façade in Fig. 5, with each segment shown in Fig. 10.

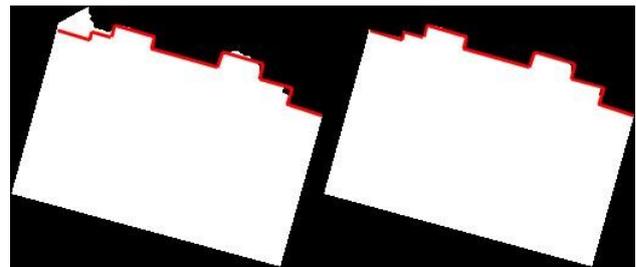


Figure 11: Building contour of one single building. Left: before determination of the major façade planes. Right: after determination of the major façade planes.

In cases where architectural structure exists in 3D, one will have to employ a 3D approach based on sufficiently accurate depth maps for a detailed analysis. We explain our plane sweeping method and illustrate it with the help of a single example. Initially, using just a 2D method, the window and floor count fails completely. However, using a 3D approach, we can successfully separate the façade into separate planes, one each for the various architectural details such as extrusions for stair cases.

Encouraged by the improvement of the result in one façade, we now want to continue with more experimental work using many façades and buildings in one test site, and also using multiple test areas with different architectural styles. It will be relevant to consider coastal resort environments, historical small towns, alpine terrains and industrial zones.

Continuing work will also have to address improvements of the new method that are needed to avoid ambiguities. We have observed in some experiments not shown nor discussed here that problems will occur when the assignment of a pixel to a certain plane of the depth map is incorrect. This may be due to similar correlation coefficients of one pixel for different planes, and a resulting ambiguity to which plane this pixel is to get assigned. The resulting 3D data set can be very noisy. We

believe that this is a weakness of the method caused by the Winner-Takes-All WTA-approach that we need to eliminate.

Given that we have multiple matches per pixel from the many image overlaps, a remedy would involve consideration of neighborhoods as the match results get fused, for example by means of the so-called Belief Propagation method. (e.g. Weiss and Freeman, 200). This is an approximation technique for global optimization on graphs, which is based on passing messages on the arcs of the underlying graph structure. The algorithm iteratively refines the estimated probabilities of the hypotheses within the graph structure by updating the probability weighting of neighboring nodes (see Zach 2007).

#### REFERENCES

Hartley R., Zisserman A. (2004) *Multiple View Geometry in Computer Vision*. Second Edition. Cambridge University Press, March 2004, pp. 219-243.

Klaus A. (2007) *Object Reconstruction from Image Sequences*. Dissertation, Graz University of Technology, 2007.

Leberl F., M. Gruber (2005) *Ultracam-D: Understanding Some Noteworthy Capabilities* Proceedings, Proceedings of the Photogrammetric Week 2005, Stuttgart University; Wichmann-Verlag Heidelberg.

Leberl F. (2010) *Human Habitat Data in 3D for the Internet*. Ranchordas, A., Madeiras Pereira, J., Araújo, H.J., Tavares, J.M.R.S. (Eds.); Springer Communications in Computer and Information Science (CCIS) Volume 68, Computer Vision, Imaging and Computer Graphics: Theory and Applications · International Joint Conference, VISIGRAPP 2009, Lisboa, Portugal, February 5-8, 2009. Revised Selected Papers, pp. 3-17.

Meixner P. Leberl F. (2010a) *From Aerial Images to a Description of Real Properties: A Framework*. Proceedings of VISAPP International Conference on Computer Vision and Theory and Applications, Angers 2010.

Meixner P. Leberl F. (2010b) *Describing Buildings by 3-dimensional Details found in Aerial Photography*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, in print Vienna, Austria.

Prandi F. (2008) *Lidar and Pictometry Images Integrated Use for 3D Model Generation*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVII (part B2) Beijing, China, pp. 763-770.

Weiss, Y. and Freeman, W. T. (2001). *On the optimality of solutions of the max-product belief propagation algorithm in arbitrary graphs*. IEEE Transactions on Information Theory, pp. 723–735.

Zach C. (2007) *High-Performance Modeling from Multiple Views using Graphics Hardware*. Dissertation, Graz University of Technology, 2007

Zebedin L., Klaus A., Gruber B., Karner K. (2006) *Façade Reconstruction from Aerial Images by Multi-view Plane Sweeping*. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. XXXVI (part 3), Bonn, Germany.