DESCRIBING BUILDINGS BY 3-DIMENSIONAL DETAILS FOUND IN AERIAL PHOTOGRAPHY

P. Meixner *, F. Leberl

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

KEY WORDS: Digital, Interpretation, Modelling, Visualization, Detection, Image, Three-dimensional

ABSTRACT:

A description of Real Properties is of interest in connection with Location-Based Services and urban resource management. The advent of Internet-maps and location aware Web-search inspires the development of such descriptions to be developed automatically and at very little incremental cost from aerial photography and its associated data products. Very important on each real property are its buildings. We describe how one can recognize and reconstruct buildings in 3 dimensions with the purpose of extracting the building size, its footprint, the number of floors, the roof shapes, the number of windows, the existence or absence of balconies. A key to success in this task is the availability of aerial photography at a greater overlap than has been customary in traditional photogrammetry, as well as a Ground Sampling Distance GSD exceeding the traditional values. We use images at a pixel size of 10 cm and with an overlap of 80% in the direction of flight and 60% across the flight direction. Such data support a robust determination of the number of floors and windows. Initial tests with data from the core of the City of Graz (Austria) produced an accuracy of 90% regarding the count of the number of floors and an accuracy of 87% regarding the detection of windows.

1. INTRODUCTION

Urban building models by computer vision have been a topic since the early 1990's (Gruber, 1997). Since 2006, this has evolved into a massive and systematic effort to map buildings in 3D to support a certain location-awareness in Internet-searches. While Google, Yahoo!, Ask and various regional search-providers all implemented 2D systems, Microsoft embarked on a 3D Internet mapping program (Leberl, 2007).
The US website www.zillow.com built an application on top of Microsoft's Internet mapping platform, then denoted as Virtual Earth, now Bing Maps, that attached a description and a value to each property in the USA. Both the description and the value are being taken from public records for property taxes, as shown in Figure 1. Adding the street-side view, one can obtain a rather complete assessment of a property's main characteristics, based on its 2D visualizations from the air and from the street level.

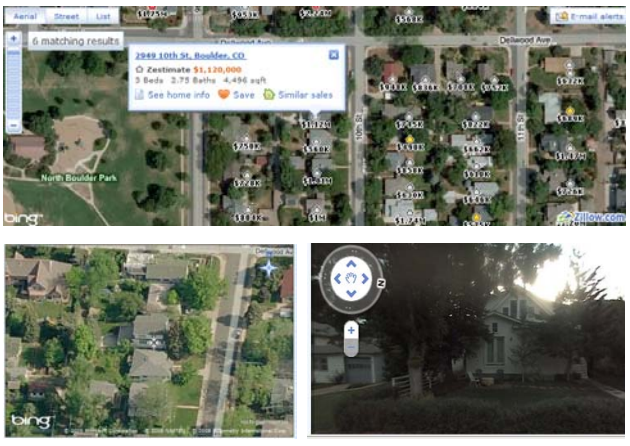

Figure 1. A property valuation website for North America is www.zillow.com. It associates public property tax records with a US$-value, built-out surface area and number of bathrooms to an orthophoto from the Microsoft Bing Maps website (top). The result is an easy access via a known address to the estimated value. In addition, presenting each property also on oblique Microsoft-images (below, left) and accessing the street-view data of Google (below, right) adds considerable visual information per property. However, the image information itself is not entering into the valuation nor description, and there is no searchable data being extracted from the images.

At issue is the development of an ability to describe each property and each building automatically in the absence of detailed and publicly accessible property-tax records. Besides, even if such records exist, they typically will not contain certain details about a property's buildings. Therefore, an ability to describe buildings may be of interest in a broad range of tasks, typically related to the offerings of location-based services. Basing such a description on Internet-public data with its ortho-photos, but augmenting this with data products derived from aerial imagery, would appear to make this description largely a byproduct of aerial mapping, without added cost. Regarding the buildings of a property, major descriptive elements concern its number of floors, roof shape, number of windows, existence of a garage, of a basement or attic, of skylights and chimneys. These elements can be determined automatically, as we will demonstrate in this paper. However, a strictly 2-dimensional data set would be insufficient for the task. We do need 3D data since we approach the building as a 3D structure.
Our approach is based on data that have been created for regular mapping purposes, and we treat such data as input. Using these, we are building specific applications to extract building information. Using a demonstration data set from Graz (Austria) with 216 buildings on 321 parcels, we show that the detection of floors and windows from aerial photography is feasible at a detection rate regarding building floors of 90% and windows of 87%.

* Corresponding author

2. AERIAL PHOTOGRAPHY AND COMPUTED DATA PRODUCTS

Figure 2 is an orthophoto of a segment of the City of Graz and covering 400 m x 400 m. Such orthophotos are today being created from digital aerial photography using pixel sizes of perhaps 10 cm and image overlaps in the range of 80% forward and 60% sideward (Scholz and Gruber, 2009). A point on the ground will thus be imaged 10 times and the orthophoto will not have to have any occluded regions. Both a traditional orthophoto with relief displacements of vertical objects such as buildings and trees is a common product, and increasingly the true orthophoto is as well since the ability to avoid occlusions is essential in this case, and the novel high overlaps ensure that such occlusions get eliminated. However, in order to produce a true orthophoto at a good quality, one needs a Digital Surface Model DSM with well-defined building roof lines to avoid "ragged" building edges. A high-quality DSM requires a 3D capability at an accuracy level that is not needed for traditional orthophotos.

Figure 2. A 400 m x 400 m segment of an orthophoto of the urban core of the city of Graz (Austria). The pixel size is at 10 cm. The orthophoto is of the type "true"; therefore the facades are not visible.

Associated data are computed from the aerial images. They consist firstly of the results of the aerial triangulation with their pose information per image. Given the high overlaps among images and the digital format, the accuracies of the pose and attitude are higher than those of the traditional two image stereo image blocks on film. The demonstration data set in Graz is produced at an accuracy of 10cm on the ground.

Secondly, we have available the DSM plus its filtered Bald Earth DTM (regular rasters). It may be remarkable that the DSM is computed at an interval of the elevation postings at only 2 pixels. Traditional photogrammetry had postulated a distance between elevation postings as a multiple of the height accuracy. That horizontal spacing was recommended to be in the range of perhaps 20 times the elevation error. If one were to assume an elevation error of ± 1 pixel, then the postings were to be spaced 20 pixels apart. However, these recommendations were based on 2-image stereo. This is now changing to a 10-image multi-view geometry (Hartley, Zisserman, 2000), and thus to a concept of "super-resolution", as if the pixel sizes were in effect much smaller than they actually are. The result is a much denser DSM than was ever computed previously (Klaus, 2007). This leads to well-defined horizontal edge information such as building roof lines. This approach also is very competitive with the direct elevation measurements from aerial laser scanners (see Figure 3).

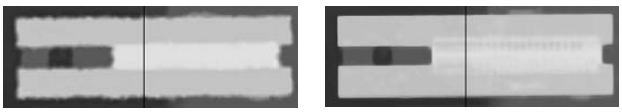

Figure 3. Comparing a building outline obtained from high-overlap digital aerial photography (right) using 8 cm pixels, with the result from an aerial LIDAR measurement (left) using 40 cm postings. This result has been obtained from the Vaihingen test near Stuttgart under supervision by the University of Stuttgart (Cramer and Haala, 2009). This example had been developed in a separate project (Leberl et al, in print).

The third type of derived information is the image classification into roofs, grassy areas, vegetation, water bodies and circulation spaces such as roads, parking spaces, driveways and other impervious surfaces.

3. IDENTIFYING BUILDINGS WITHIN INDIVIDUAL REAL PROPERTIES

3.1 What is a "Building"?

A central task exists to identify "buildings". The definition of a "building" is less obvious than it may initially seem. The imagery needs to be related to parcel maps in the form of cadastral records. Figure 4 presents a cadastral map segment and superimposes it over the orthophoto. The first observation concerns the geometric relationship: the visual data from the imagery are not in complete agreement with the cadastral parcels and a geometric change is needed to achieve a optimum match. The second observation concerns the fact that buildings as seen in aerial imagery cut across property boundaries because they may be attached to one another in dense urban situations.

Figure 4. A cadastral vector data set is superimposed onto the true orthophoto for a segment of the Graz demo site. Note the small discrepancies between the data along property boundary lines manifesting themselves as visual feature in the imagery.

What then is a "building"? In our context, this is a structure of sufficient size within a parcel. Therefore what may be experienced as a single building in aerial photography will be represented by a collection of buildings, each defined by its own parcel. The inverse may also exist, where multiple buildings are defined on a single parcel. This fact leads to a third issue, namely a need to separate smaller structures such as garages or sheds from a building properly.

A fourth topic addresses complex building shapes with many facades. For analysis purposes it would be desirable to have

buildings with only 4 facades. An approach to cope with the complex building shapes may consist of separating an individual building into its parts so that fairly basic building shapes are then be achieved, in analogy to separating the concatenated buildings of urban landscapes along parcel boundaries. In the demo area of Figure 2 we count 216 buildings. Of these, 139 are with a simple rectangular footprint, and at least 2 viewable facades. We find that occlusions form vegetation prevent one often from actually being able to have multiple facades per building available for redundant analyses.
To deal with the second through the fourth issues, we first need to identify the data per parcel.

3.2 Matching Cadastral Parcels with the Orthophoto

In a separate paper we have presented a solution to the problem of mismatches between cadastral and image data (Meixner & Leberl, 2010). Such mismatches can be the result of the different histories of the cadastral data and their focus on 2D local information. We do not allow for a local deformation of the cadastral data. Instead, the cadastral maps are treated as rigid 2D entities where changes are only permitted in rotation and scale. We apply the widely available method of chamfer matching to conflate the vector-type parcel data with the raster-type Orthophoto. Details about the chamfer-matching, the handling of roof overhangs and mismatches between the cadastre and the DSM are illustrated in Meixner and Leberl (2010). This is applicable if sufficient image information is available to define the parcel boundaries by natural features. Major parcel-vector matches with imagery are along street outlines and where fences exist. In our demonstration data set in Graz, we have shown that the initial mismatches in the range of ±7 pixels could be reduced to ±3 pixels.

3.3 Data per Property

Once the orthophoto and cadastral parcel match, one can proceed to cut all data sets along parcel boundaries. Figure 5 illustrates the result for a single property with its DSM, its image classification and its multiple individual overlapping image segments. This example represents a case with no special complexity since there is a single simple building shape with four facades. One complexity is caused by occlusions due to vegetation.

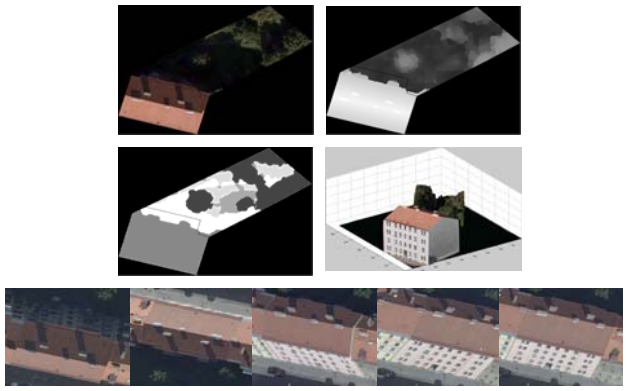

Figure 5. A sample parcel with its data from 10 overlapping aerial photographs, consisting (a) of a True Orthophoto (b) the DSM, (c) the classification layers and (d) a selection of some individual aerial image segments. Also shown is (e) a perspective view of the DSM and the aerial imagery.

4. BUILDING FOOTPRINTS

4.1 Computing Footprints

We have two information sources for building footprints. One is from the image classification of roofs. The other is from the vertical elements of the DSM. The classification typically is based on color and texture, not however, on the 3D information of the DSM. Therefore the two information sources are independent.

(a) Using the Classification Layer “Roofs”
Figure 6 illustrates the classification layer “roof” for a building and its contour in the form of contour pixels. The selection of contour pixels in the binary building layer is trivial. The conversion of the raster- into a vector-format follows a standard procedure according to Douglas D. and Peucker T. (1973). The result consists of straight line segments. Knowledge that this is a building footprint will enter at this point by replacing the line segments by a rectangular shape from a library of such shapes. The match between the line segments and the geometric figure is achieved via a best fit between the geometric shape and the line segments. The measure of fit consists of 4 lines.

Figure 6. The classification layer “building” is based on color and texture. (a) shows the binary layer, (b) its contour in raster and finally in (c) the geometric figure of the footprint.

(b) Using the Vertical Elements of the DSM
A computational pass through the difference between DSM and DTM of a parcel will result in height -postings representing vertical objects.

- Loading the height-postings Hij of the parcel for all rows i and columns j;
- Calculation of the first and second derivative Hij’ and Hij’’ of the height data Hij in each line and column;
- Locating the maximum 1st and 2nd derivatives Hmax’ and Hmax’’ in each line and column, delivering candidate footprint postings;
- For a neighborhood around each candidate footprint location, determine the associated height H of a structure;
- Decide on valid footprint positions from the verticality of the DSM expressed by the values of H’, the curvature expressed by H” and building object expressed by the height H.

The positions of candidate footprint pixels are now in the raster format. We again convert this to line segments as in alternative (a) above. The information now can be fed into the computation of a geometric figure of the building footprint as previously described. This geometric figure is the resulting “building mask”. Other vertical objects may be trees and those also will produce candidate footprint pixels. However, there will not exist straight line segments to replace those pixels and therefore these footprint pixels will get deleted.

4.2 Attaching Heights to the Footprints

The use of the DTM in defining footprints produces, as a by-product, an estimate of an elevation value for each candidate footprint. While this has been computed for candidate positions where a footprint location is possible, this now needs to be converted to a set of elevation values along the path of the footprint. For this purpose the geometric figure of the footprint is placed into the DSM and the elevation profiles get interpolated along the straight lines of the footprint:

For each straight line of the footprint repeat the following process:

- Define positions i,j along the straight footprint line at equal intervals;
- Determine the XY-pixel –locations perpendicular to the line at positions i,j;
- From the short elevation profiles along the pixel locations XY, determine the base height and the top height associated with that footprint element, and thus the elevation difference.

The result of this procedure is a set of elevation profiles along the footprints.

4.3 Buildings Cutting Across Parcel Boundaries

With the elevation values along the footprints, we have the 3D outlines of the buildings. At issue is the situation along a parcel boundary where there may be a valid building footprint, or the building is attached to a structure on the adjacent parcel and the footprint is merely virtual.
To determine whether the footprint is virtual or real, we revisit the elevation data. Along a footprint at the edge of a parcel, one defines a small mask of perhaps 20*20 pixels. If one is dealing with a real footprint, then half of the elevation values should be zero. If the footprint is virtual, then a majority of the elevation values will be large. We select a threshold of 2/3 of all values to be large to determine that the footprint is virtual.

4.4 Small Structures versus Buildings

With elevation profiles along the footprints, we also have the means to separate actual residential housing from detached garages. The latter will have a small surface area of 50 m2 or less and not exceed a height value of 2.5 m.

4.5 Complex Buildings

The split of a complex building into simpler building elements has been discussed by Zebedin et al. (2008) and implemented in a workflow to replace a dense point cloud by simple building geometries.
There exist three measures of complexity for a building. One is the geometric figure of the building's footprint. One may restrict the complexity to be for 4 façades only. The second is the elevation profile along the footprint. One may determine a measure of the building symmetry for the elevations along the footprint: if facades get associated with different building heights, one may have reason to break the building into its parts. The third is the number of local maxima in the elevations of the roof: the roof shape is defined by the elevation values inside the footprint figure. By computing local maxima for those elevations, one will have the means to determine a separation of the building into building elements, each with a separate roof. Zebedin et al. (2008) evaluates the height differences between manual and automatic reconstruction of a building for a test data set of Manhattan (1973 buildings). It shows that 67.51% of the pixels have a height difference smaller than 0.5m, 72.85% differ by less than 1m and 86.91% are within 2m. Details on this method are described in Zebedin et al. (2008).

5. FACADES

The interest is in describing floors and windows, and for this purpose one needs to identify the façades. These are available along the building mask's straight segments, and the elevation profile associated with that line segment. Independent of the actual shape of the façade and where it touches the roof, and how the ground slopes, one can for simplicity define a quadrilateral in 3D space by computing a façade height from the DSM profile. The footprint will define one edge of the quadrilateral in 3D by computing a slope from the DSM values. The end points of the straight line segment define the two opposing vertical edges of the quadrilateral. The DSM-values along the roof line will be replaced by the 4th segment.
Figure 7 illustrates the façade quadrilaterals for the simple building, together with the image texture of one of the aerial photographs covering those facades.

Figure 7. Façades of one building, with the computed quadrilateral for each of the facades. Note that the replacement of the elevation profiles along the building footprints by a straight line serve to obtain a simple façade figure in 3D.

6. FLOORS AND WINDOWS

6.1 Image Texture per Façade

The definition of the façade quadrilaterals produces 4 façade corner points in 3D object coordinates. These must be projected into each of the aerial images to associate image content to each façade. Typically, many aerial images will show the texture of each façade. Figure 8 is an example for one of the separate facades of the building in Figure 7. The projection is based on the pose values of each image from the aerial triangulation.

Figure 8. Of one single façade of the building in Figure 7 one will obtain multiple aerial image segments. These have been rectified into a façade coordinate system. From an aerial image block showing for each object point typically 10 images, not all will contain useful data for a specific vertical façade. Selected here are the 4 best, where "best" is defined as the largest area of a façade quadrilateral in the projection into an image.

6.2 Floors

From the building's appearance, floors get defined by windows. In turn, windows form a defining structure in describing a façade's detail. A procedure for finding a floor count has been developed using the following steps.

For each façade i of a building j, repeat:
Import all n image segments showing this façade i.

- For each image segment repeat:
- Transform the segment into the façade coordinate system.
- Apply a contrast enhancement.
- Apply the Prewitt edge detection horizontally.
- Apply the Prewitt edge detection vertically.
- Convert the maximum horizontal and vertical edge values into a binary format.
- Create for each image row, and image column, a summation of all pixel values, resulting in a vertical and horizontal edge profile.
- From the summation, remove outliers, normalize the values and remove low values as "noise".
- Determine the number of maxima of the sums of vertical gradients and use this as the number of floors.
- Perform a verification by eliminating floors that do not have the proper vertical spacing (minimum distance between floors); and removal values from along the edges of the image texture inside the façade quadrilateral.

This approach will result in data as illustrated in Figure 9.

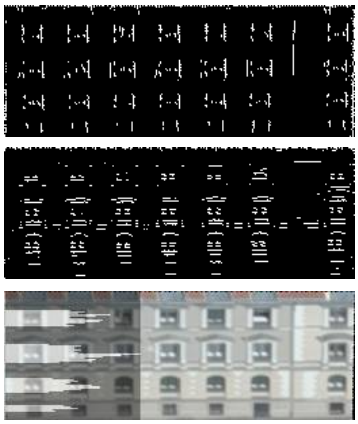

Figure 9. Binary Prewitt edges in (a) are vertical, in (b) horizontal. The sums of edge values are shown in (c) as a count of the number of floors.

A floor count can be applied to each of a set of overlapping façade images. If there were a discrepancy in the result, some logic would have to get applied to resolve the ambiguity.

6.3 Windows

Window detection has been of some interest in recent years. Algorithms like "boosting" have been applied by Nguyen et al. (2007) to detect cars and windows in aerial images. Cech and Sara (2007) have developed a window detection based on a library of window shapes. Lee and Nevatia (2004) have based their approach on edge images. These approaches have been subjected to only limited experimental analysis, but are generally reported to find windows in a rather robust manner.

Given our floor counts, we are reusing the intermediate Prewitt edges to also find the windows. An approach that simply "intersects" the locations along the pixel rows and columns with the maximum edge sums will work if all windows are regularly arranged. While this is often the case, it is not always true. Therefore Lee and Nevatia (2004) have proposed a variation of the approach.
To refine the locations of the windows a one dimensional search for the four sides of a window is performed. For every line of a window hypothesized lines are generated by moving the lines to its perpendicular direction. The refined positions of the windows are determined where the hypothesized line has the best score for the window boundary. For a more detailed description of the used algorithm read Lee and Nevatia (2004).
The big advantage of this method is that one can also use images with lower resolution, and that not only rectangular windows but almost all window designs can be automatically detected rather quickly without training the program in advance.
The window count is applicable in each image segment of a given façade, separately. Or one might want to merge the edge data sets and apply a single window detection to the sum of all edges. Initial tests have shown that the window count is a rather robust method that delivers no discrepancies between the separate images of one façade in the examples chosen thus far.
A comparison of the various different methods for window detection should be performed and will be the subject of ongoing work.

6.4 Multiple Facades per Building

The redundancy not only applies to the image coverage per façade from the high overlaps of aerial photography. We also find that we have multiple measures for the number of floors from multiple facades. These must be consistent with one another. It is possible that a building has different floor counts on a sloping terrain. Since the "bald Earth" as well as the slope of a building footprint are known, they must enter into the floor count.
Figure 7 presented facades of one building. Figure 10 illustrates the floor counts and detected windows in each façade of that one building. As one can easily determine, the automated floor count and the count of the windows is consistent with a visual inspection.

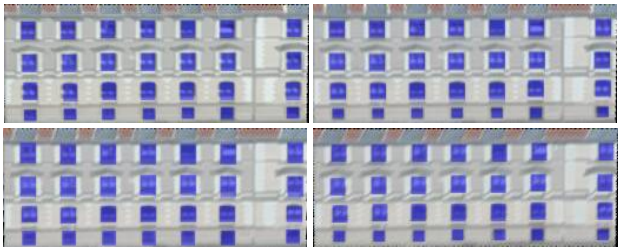

Figure 10. Four facades of one building from Fig. 7 lead to independent floor counts and window counts. It hss to be noted that the floor counts and the number of windows coincide with the visual inspection.

We have extended this exercise to a selection of 150 properties in the Graz demo data set. In those properties we have identified 102 buildings with a total of 225 facades. The total number of floors was 387, the number of all windows was 2646. Running the approach through this data set results in the following:

Success rate of Building detection: 100%, all 102 building were found.

Success rate of Floor detection: 90% of the 387 floor were correctly counted.

Success rate of Window detection: 87.1% of the 2646 windows were correctly counted.

7. CONCULSION: TOWARDS AN EXTENSIVE PROPERTY DESCRIPTION

The search for a description of individual buildings per property is but an element in a larger effort. The development of as detailed a description of real properties will have the buildings as the most important element, but other features of a property are also in need of a description. One will want to consider the land, the vegetation, the impervious surfaces, even the interaction between properties casting shadows or affecting privacy. And one will also be interested in the traffic, distances to businesses or public transportation etc. A full system for property descriptions will involve business addresses, traffic information, street network information, as well as sun angles.

In the current contribution we have focused on basic descriptions of buildings. This involves the definition of a building on a property, even if two buildings are connected along a property line. It deals with complex buildings having many facades and a complex roof-scape. From the outside, thus from aerial imagery, one can count the floors and windows, and identify the window areas on a façade for further analysis. At this stage of research we are beginning with the experimental evaluation of the various approaches. We will have to cope with occlusions from vegetation, with ambiguities regarding garages and sheds, the difficulties arising from an inability of matching parcel maps with aerial imagery, and with ambiguities from basement and attic windows.

Initial results are encouraging. Using 150 properties with 102 buildings having 387 facades and 2646 windows, 90% of all floors and 87.1% of all windows were found automatically. The result addresses, however, a specific situation in a mature core area of Graz (Austria). Reasons for misclassifications regarding floors and windows result from inaccuracies of the DTM, occlusions from vegetation and other buildings, partial shadows on the facades, very complex facades and steep camera angles. All these reasons for misclassifications have to be analyzed very carefully. Fore that the building interpretation has to be repeated by increasing the sample data in one city, and then by looking at vastly different environments such as a coastal resort environments, historical small towns, alpine terrains and industrial zones.

REFERENCES

Cech J., R. Šara , 2007. Windowpane detection based on maximum a-posteriori labeling. Technical Report TR-CMP-2007-10, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic.

Cramer, M. & Haala, N., 2009. *DGPF project: Evaluation of digital photogrammetric aerial based imaging systems - overview and results from the pilot centre.* Published at ISPRS Workshop High-Resolution Earth Imaging for Geospatial Information, Hannover, Germany, June 2 - 5. Digitally published on CD, 8 pages.

Douglas D., T. Peucker, 1973. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature, *The Canadian Cartographer* pp. 112-122.

Gruber M., 1997. *„Ein System zur umfassenden Erstellung und Nutzung dreidimensionaler Stadtmodelle“*, Dissertation, Graz University of Technology,2007.

Hartley R, A. Zisserman, 2000. Multiple View Geometry for Computer Vision. *Cambridge University Press*, 1st Edition.

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

Leberl F., 2007. Die automatische Photogrammtrie für das Microsoft Virtual Earth *Internationale Geodätische Woche Obergurgl.* Chesi/Weinold(Hrsg.), Wichmann-Heidelberg-Publishers, pp. 200-208

Leberl F., A. Irschara, T. Pock, P. Meixner, M. Gruber, S. Scholz, A. Wiechert (in print) Point Clouds from Laser Scanning Versus 3d Vision. *Photogrammetric Engineering and Remote Sensing.*

Lee S.C., R. Nevatia, 2004. Extraction and Integration of Window in a 3D Building Model from Ground View Images. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* CVP’04

Meixner P, F. Leberl, 2010. From Aerial Images to a Description of Real Properties: A Framework. Manuscript submitted for publication, Graz University of Technology – Institute of Computer Graphics and Vision, Graz.

Nguyen T., H. Grabner, B. Gruber, H. Bischof, 2007. On-line Boosting for Car Detection from Aerial Images. *Proceedings of the IEEE International Conference on Research, Innovation and Vision for the Future* (RIVF'07), pages 87-95.

Scholz S., M. Gruber, 2009. Radiometric and Geometric Quality Aspects of the Large Format Aerial Camera UltraCam Xp. *Proceedings of the ISPRS*, Hannover Workshop 2009 on High-Resolution Earth Imaging for Geospatial Information, XXXVIII-1-4-7/W5, ISSN 1682-1777

Zebedin L., Bauer J. Karner K., Bischof H., 2008. Fusion of Feature- and Area-Based Information for Urban Buildings Modeling from Aerial Imagery. *Proceedings of the ECCV* 2008, Marseille, France, pages 873-886