Photogrammetrie • Fernerkundung • Geoinformation 2/2008, S. 65–81, 14 figs.

Integrated LIDAR and Image Processing for the Modelling of Building Facades

SUSANNE BECKER & NORBERT HAALA, Stuttgart

Keywords: Three-dimensional, point cloud, LIDAR, orientation, modelling, facade interpretation

Summary: The acquisition of 3D city and landscape models has been a topic of major interest within the photogrammetric community and a number of algorithms became available both for the automatic and semiautomatic data collection. These 3D models are mainly used for visualisation to generate realistic scenes and animations e. g. for planning and navigation purposes. Since area covering urban models are usually extracted from airborne data, a refinement is required especially if realistic visualisations from pedestrian viewpoints are aspired. Of particular interest are algorithms that allow for a fully automatic facade reconstruction process at different levels of detail. For this purpose, in our approach terrestrial LIDAR data as well as facade imagery is used to increase the quality and amount of detail for the respective 3D building models. The a priori information which is provided by the models from airborne data collection is integrated both for the automatic georeferencing of the terrestrial data and the subsequent geometric refinement. After alignment of the terrestrial data, window structures are determined approximately from the LIDAR point clouds and then further refined by 3D edges which are extracted from the overlapping facade images. Our modelling process applies a 3D object representation by cell decomposition, which can be used efficiently for building reconstruction at different scales.

Zusammenfassung: *Integrierte LIDAR- und Bildprozessierung für die Modellierung von Gebäudefassaden.* Die Erfassung von 3D-Stadt- und Landschaftsmodellen ist zu einer wichtigen Aufgabe der Photogrammetrie geworden. Mittlerweile existiert eine Vielzahl von Algorithmen sowohl für die automatische als auch für die halbautomatische Datenerfassung. Genutzt werden die 3D-Modelle hauptsächlich zu Visualisierungszwecken, um beispielsweise für Planungs- und Navigationsanwendungen realistische Szenen und Animationen zu erzeugen. Üblicherweise erfolgt die flächendeckende Bereitstellung von Stadtmodellen auf der Grundlage von luftgestützten Sensoren. Eine Verfeinerung dieser Modelle ist daher vor allem dann notwendig, wenn wirklichkeitsnahe Visualisierungen aus der Perspektive eines Fußgängers angestrebt werden. Von Interesse sind hierbei insbesondere Algorithmen, die eine vollautomatische Rekonstruktion von unterschiedlich detaillierten Fassaden erlauben. Unser Ansatz verwendet terrestrische LIDAR-Daten und Fassadenbilder, um die Qualität und den Detailgrad der 3D-Gebäude zu erhöhen. Die bestehenden 3D-Modelle liefern dabei eine Vorinformation, die sowohl in die Georeferenzierung der terrestrischen Daten als auch in die nachfolgende geometrische Verfeinerung integriert werden kann. Nach der Koregistrierung aller Datensätze werden zunächst Fensterstrukturen aus den LIDAR-Punktwolken näherungsweise extrahiert. Diese werden anschließend durch zusätzliche 3D-Kanten, welche aus den Fassadenbildern abgeleitet werden, weiter verfeinert. Innerhalb des Modellierungsprozesses werden die 3D-Objekte mittels Zellzerlegung repräsentiert, welche für die Gebäuderekonstruktion in verschiedenen Maßstäben effizient genutzt werden kann.

1432-8364/08/2008/065 $ 4.25

1 Introduction

The area covering collection of urban models is usually based on the evaluation of aerial data like stereo images or LIDAR. By these means, 3D building representations become available which are sufficient for applications like simulations and visualisations at small or medium scale. However, for large scale applications like the generation of very realistic visualisations from pedestrian viewpoints, the quality and amount of detail for urban models from aerial data has to be improved. As an example, due to the viewpoint restrictions of airborne platforms, detailed information for the facades of the buildings frequently is not available. To improve the visual appearance of the buildings, terrestrial images are often mapped against the facades. However, this substitution of geometric modelling by real world imagery is only feasible to a certain degree. For instance, protrusions at balconies and ledges, or indentations at windows will disturb the visual impression of oblique views. Thus, geometric refinement is still necessary for a number of applications.

Within this paper, the geometric modelling of building facades is demonstrated exemplarily for the extraction of window objects. For this purpose, both terrestrial laser scanning and image measurement are used. Images provide high resolution information and can be collected at little effort. On the other hand, densely sampled point clouds from terrestrial laser scanning usually show depth displacements between elements like windows and walls and can therefore efficiently support their segmentation. Thus, an image based approach like it is for example presented by (Mayer & Reznik 2006) considerably profits from the additional availability of terrestrial laser scanning. An integrated collection of images and 3D point clouds is feasible by mobile systems, where a laser scanner and a camera are mounted on a car. Such a system was for example applied by (Früh & Zakhor 2003) to generate textured meshes for visual representation of building facades. In contrast, we use standard equipment consisting of a digital camera and a terrestrial laser scanner. In our scenario, the camera is not directly integrated with the laser scanning system. Independent camera stations allow for a flexible data collection, however, the automatic co-registration of the collected range and image data sets which is required as first processing step can become labour intensive. Additionally, complex areas like urban environments can only be completely covered by terrestrial laser scanning if data is collected from different viewpoints. Standard approaches align these scans from different viewpoints based on tie and control point information, which has to be measured at specially designed targets.

In contrast, our approach aims on the fully automatic georeferencing for the collected LIDAR and image data. As discussed in Section 2, this data preparation is facilitated by closely integrating existing building models. First, the LIDAR data is georeferenced by matching the point clouds from terrestrial laser scanning against the corresponding faces of the given 3D building model. In a second step, the reflectivity images, which are also provided during laser scanning, are used to link the camera images against the point cloud by a suitable feature extraction and matching process. Existing building models as they are provided from airborne measurements are also used during the modelling process, which is described in Section 3. This allows for a geometric enrichment of building facades while avoiding a data collection from scratch. In contrast to approaches using grammar based facade descriptions, which are more likely to focus on semantic modelling and interpretation (Brenner & Ripperda 2006, Alegre & Dallaert 2004), we aim at a more data driven approach. Within our process, cell decomposition is used in order to integrate window objects to the existing coarse building model. This type of modelling simplifies the problems to generate topologically correct boundary representations while fulfilling geometric constraints such as meeting surfaces, parallelism and rectangularity. Within our two-step reconstruction ap-

proach, the windows and doors are modelled from the terrestrial LIDAR data, while the window frames are further refined by photogrammetric analysis of the images in a subsequent step.

2 Data preparation and alignment

We applied our reconstruction approach to a data set which was acquired on the Schillerplatz, Stuttgart. It consists of a 3D point cloud obtained by a terrestrial laser scanner (Leica HDS 3000), image data from a calibrated camera (NIKON D2x Lens NIKKOR 20 mm, producing 4288 by 2848 pixel images) and a 3D city model from the City Surveying Office of Stuttgart. The Leica HDS 3000 laser scanner is based on a pulsed laser operating at a wavelength of 532 nm. Besides range measurements also the backscattered energy of the emitted signal is recorded for each sample point. These intensity values will be used within the georeferencing of the image data, which will be discussed in Section 2.2.2. In contrast to mobile mapping systems like the StreetMapper (Kremer & Hunter 2007) featuring high quality GPS/IMU systems for direct georeferencing, we limit ourselves to a low-cost system, which only allows for a coarse alignment of the different laser scans. However, as discussed in Section 2.1, a fully automatic registration and precise georeferencing can be realised by matching the collected 3D point clouds against the corresponding faces of the given 3D building model from the existing city model (Böhm & Haala 2005).

After georeferencing the laser scans, the independently captured images have to be aligned to these 3D point clouds by an automatic co-registration. In general, the automatic orientation of terrestrial images considerably benefits from the recent availability of feature operators, which are almost invariant against perspective distortions. One example is the affine invariant key point operator proposed by (Lowe 2004). It extracts points and suitable descriptions for the following matching, which is based on histograms of gradient directions. By these means, robust automatic tie point measurement is feasible even for larger baselines. As described in Section 2.2, in our application, this operator is used to align both the terrestrial images and the terrestrial LIDAR data. For this purpose, the feature extraction and matching is realised based on the reflectivity images that are usually provided from the laser scanner.

2.1 *Georeferencing of LIDAR data*

During collection of the 3D point clouds, a low-cost GPS and a digital compass were mounted on top of our HDS 3000 scanner. Thus, the position and orientation of the scanner can be measured to allow for a direct georeferencing of the different scans. This approximate solution was then refined by an automatic registration of the laser scans against available 3D building models, which are used as reference data for the refinement of the georeferencing process. For this purpose, the standard iterative closest point algorithm introduced by (Besl & McKay 1992) is applied.

Fig. 1 depicts our 3D city model and the collected 3D point cloud after the alignment process. The quality and amount of detail of this 3D building data set is typical for 3D models, like they are available area covering for a number of cities. Our exemplary 3D city model, which is maintained by the City Surveying Office of Stuttgart, features roof faces collected semi-automatically by photogrammetric stereo measurement. In contrast, the outlines of the buildings were captured by terrestrial surveying. Thus, the horizontal position accuracy of facade segments, which were generated by extrusion of this ground plan, is relatively high, despite the fact that they are limited to planar polygons.

After the alignment step, the 3D point cloud and the 3D city model are available in a common reference system. Thus, relevant 3D point measurements can be selected for each building facade by a simple buffer operation. These 3D points are then transformed to a local coordinate system as defined by the facade plane. Fig. 2 shows the

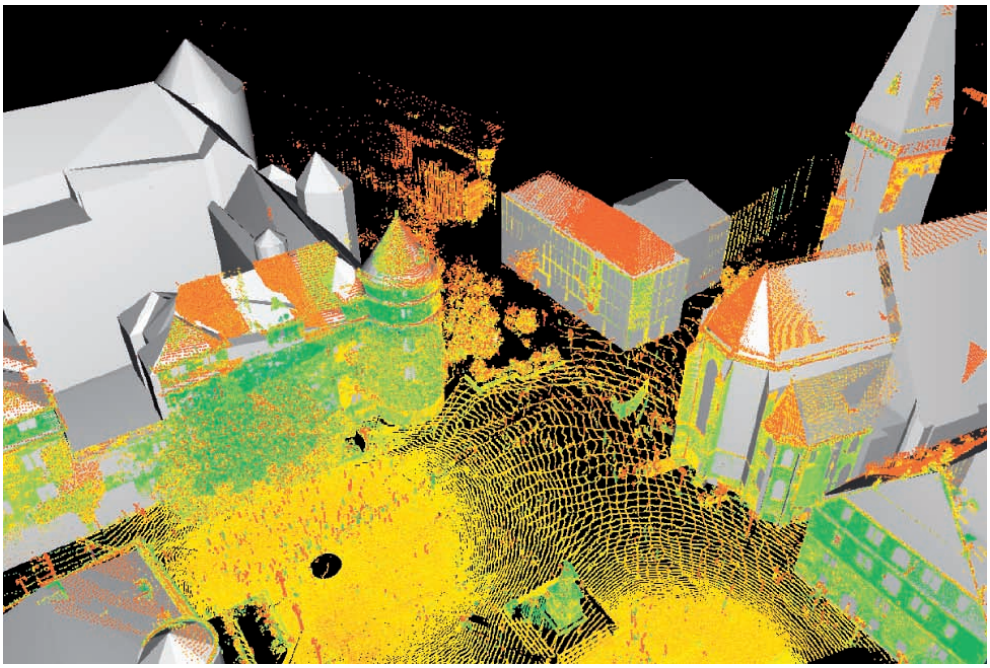

Fig. 1: 3D point cloud from laser scanning aligned with virtual city model.

Fig. 2: 3D point cloud as used for the geometric refinement of the corresponding building facade.

resulting point cloud, which has an approximate spacing of 4 cm.

Since the LIDAR measurements are more accurate than the available 3D building model, the final reference plane is determined from the 3D points by a robust estimation process. After mapping of the 3D points to this reference plane, further processing can be simplified to a 2.5D problem. As an example, while assuming that the refined geometry of the facade can be described sufficiently by a relief, the differences between the measured 3D laser points and the given facade polygon can be interpolated to a regular grid.

2.2 *Alignment of image data*

In order to integrate image data into the facade reconstruction, the images have to be oriented against each other and registered according to the already georeferenced laser point cloud in a following step. Image orientation is usually performed by means of a bundle block adjustment, which accurately estimates the orientation parameters. While tie points are necessary for connecting the images, control point information is needed for the georeferencing. Aiming at a fully automatic reconstruction process, both tie points and control points are to be derived automatically.

In our approach, first, a bundle block adjustment based on the matched key points between the digital images is implemented. The resulting photogrammetric network is then transferred to object space by additional tie points which link the digital images and the georeferenced terrestrial laser scan data. For this purpose, the feature extraction and matching is also realised using the reflectivity images as they are provided from the laser scanner. These reflectivity images are usually measured in addition to the runtime of the respective laser pulses during scanning. Since they represent the backscattered energy of the respective laser footprints, these intensities are exactly co-registered to the 3D point measurements. Despite the differences between these reflectivity images and the images captured by a standard digital camera with respect to spectral band width, resolution and imaging geometry they can be matched against each other automatically by LOWE's key point operator (Lowe 2004).

2.2.1 Image to image registration

Image to image registration based on tie points is a prerequisite step for photogrammetric 3D modelling. In the recent past, much effort has been made to develop approaches that automatically extract such tie points from images of different types (short, long, and wide baseline images) (Remondino & Ressl 2006). While matching procedures based on cross-correlation are well suited for short baseline configurations, images with a more significant baseline are typically matched by means of interest points. However, these techniques would fail in case of wide baseline images acquired from considerably different viewpoints. The reason is large perspective effects that are caused by the large camera displacement. Points and corners cannot be reliably matched. Therefore, interest point operators have to be replaced by region detectors and descriptors. As an example, the Lowe operator (Lowe 2004) has been proved to be a robust algorithm for wide baseline matching (Mikolajczyk & Schmid 2003).

Fig. 3 shows images from a calibrated camera (NIKON D2x Lens NIKKOR 20 mm). For the automatic provision of tie points the SIFT (scale invariant feature transform) operator has been applied to extract and match key points. Wrong matches were removed by a RANSAC based estimation (Fischler & Bolles 1981) of the epipolar geometry using Nister's five point algorithm (Nistér 2004). Finally, the image orientations were determined from 2079 automatically extracted tie points.

2.2.2 Image georeferencing

The provision of control point information, which is necessary for the determination of the orientation parameters, typically involves manual effort if no specially designed

Fig. 3: Image data for photogrammetric modelling.

Fig. 4: Measured laser reflectivities as 3D point cloud (left) and 2D image representation (right).

targets are used. The reason is that object points with known 3D coordinates have to be manually identified in the images by a human operator. The idea to automate this process is linking the images to the georeferenced LIDAR data by a matching process (Böhm & Becker 2007) which is similar to the automatic tie point matching as described in Section 2.2.1.

Common terrestrial laser scanners sample object surfaces in an approximately regular polar raster. Each sample provides 3D coordinates and an intensity value representing the reflectivity of the respective surface point. Based on the topological information inherent in data acquisition, the measured reflectivity data can be depicted in the form of an image. This allows for the application of image processing tools to connect the images captured by the photo camera to the LIDAR data.

Fig. 4 (left) shows the laser point cloud of an already georeferenced scan. The image representation derived from the reflectivity values is given in Fig. 4 (right). Each pixel with a valid laser reflectivity value refers to

Fig. 5: Key point correspondences for the laser reflectivity image (left) and one of the photographs (right).

the 3D coordinates of the related sample point. Thus, obtained point correspondences between the laser image and the photos provide control point information which is necessary for the determination of the photos' orientation parameters.

However, images generated from laser reflectivities considerably differ from images that have been captured by standard digital photo cameras. This is due to differences in the imaging geometry, the sampling distance and the spectral band width of the measured reflectivities. For these reasons, the determination of point correspondences between a laser reflectivity image and a photograph requires an algorithm which is insensitive to changes in illumination and scale and uses region descriptors instead of edge detectors. Fig. 5 depicts the laser reflectivity image (left) and one of the photographs captured by the NIKON camera (right) in real proportions. In order to have similar intensity values in both images, only the green channel of the photograph has been considered for the determination of corresponding points. The resulting key points were extracted and matched by means of the SIFT implementation provided by (Vedaldi 2007). Using default settings 492 key points are detected in the laser reflectivity image and 5519 in the photograph. Of those 31 are matched to corresponding key points represented by the red dots and lines in Fig. 5.

In a next step, wrong matches are to be removed by a RANSAC based computation of a closed form space resection (Zeng & Wang 1992). For this purpose, the SIFT point correspondences are used as control point information. The accuracy of the resulting orientation parameters strongly depends on the configuration of the SIFT point correspondences. If the points are close together, the solution of the space resection becomes unstable and the uncertainty of the SIFT point coordinates (Remondino & Ressl 2006) leads to significant variations in the orientation parameters. However, when given a sufficient number of well-distributed points, the obtained orientation parameters will be accurate enough to serve as approximate values for a final bundle adjustment, allowing for a fully automatic orientation process. For our data set, the average standard errors of the finally estimated orientation parameters are $\sigma_X = 7.6\,\text{cm}$, $\sigma_Y = 5.6\,\text{cm}$, $\sigma_Z = 8.1\,\text{cm}$, $\sigma_{az} =$

0.167°, $\sigma_{el} = 0.164^\circ$, $\sigma_{roll} = 0.066^\circ$. The average precision of the computed object coordinates is $\sigma_X = 3.3\,cm$, $\sigma_Y = 4.7\,cm$, $\sigma_Z = 2.1\,cm$.

3 Facade Reconstruction

Based on georeferenced terrestrial LIDAR and image data, the existing coarse building model can be geometrically refined by the integration of window objects. Usually, tools for the generation of polyhedral building models are either based on a constructive solid geometry (CSG) or a boundary representation (B-Rep) approach. In Section 3.1, the pros and cons of both techniques will be discussed briefly. This will motivate our new approach for facade reconstruction, which is based on cell decomposition as an alternative form of solid modelling. Afterwards, our two-step approach for facade reconstruction will be presented in Section 3.2 and 3.3. While in the first step windows and doors are modelled from the LIDAR data (cf. Section 3.2), the window frames are further refined by photogrammetric analysis in the second step (cf. Section 3.3).

3.1 *Cell decomposition versus CSG and B-Rep*

Within B-Rep algorithms the surface boundaries of a solid are generated from measured points, lines or planes, leading directly to a boundary representation. For instance, if the reconstruction of a building facade is based on 3D point clouds from terrestrial laser scanning, state-of-the-art meshing techniques can be used to convert the point clouds to polygonal meshes. Depending on the density of the underlying point cloud, these approaches are suitable for the reconstruction of free-form objects such as complex protrusions, ornaments or the stonework. While such representations allow for good visualizations of the facade's relief (Haala et al. 2006), an interpretation of the reconstructed model concerning different architectural structures is not supported. This is due to the fact that a dense polygon mesh does not distinguish between meaningful entities, for example facade, window, balcony, etc. Furthermore, in consequence of limitations in point sampling distance and accuracy of LIDAR sensors, special characteristics of man-made objects such as right angles, parallel lines or coplanar faces are not maintained during the modelling process. In order to make the reconstruction result more robust against error-prone data, a number of building reconstruction approaches first extract planar regions from the point cloud which are then intersected to generate polyhedral building structures. However, the combination of these segments to generate topological correct boundary representations is difficult to implement (Rottensteiner 2001). This task is additionally aggravated if geometric constraints such as meeting surfaces, parallelism and rectangularity have to be guaranteed for respective segments.

Such regularization conditions can be met easier when object representations based on CSG are used (Brenner 2004). CSG based techniques combine simple primitives by means of regularized Boolean set operators, allowing for a powerful and intuitive modelling (Mäntylä 1988). CSG representations are also always valid since the simple primitives are topologically correct and this correctness is preserved during their combination by the Boolean operations. Additionally, the implicit geometric constraints of these primitives like parallel or normal faces of a box type object make the parameter estimation quite robust. This is especially important for reconstructions based on error prone measurements. However, while CSG is widely used in computer aided design, most visualization and simulation applications require the additional derivation of a boundary representation. While this so-called boundary evaluation is not difficult conceptually, its correct and efficient implementation can be complicated. Error-prone measurements, problems of numerical precision and unstable calculation of intersections can considerably hinder the robust generation of a valid object topology.

The application of cell decomposition can help to facilitate these problems. Cell de-

composition is a special type of decomposition models, which subdivides the 3D space into relatively simple solids. Similar to CSG, these spatial-partitioning representations describe complex solids by a combination of simple, basic objects in a bottom up fashion. In contrast to CSG, decomposition models are limited to adjoining primitives, which must not intersect. The basic primitives are thus 'glued' together, which can be interpreted as a restricted form of a spatial union operation. The simplest type of spatial-partitioning representations is exhaustive enumeration. There the object space is subdivided by non overlapping cubes of uniform size and orientation. In this sense, cell decomposition is similar to a spatial occupancy enumeration, where the object space is subdivided by non overlapping cubes of uniform size and orientation. Nevertheless, cell decompositions are based on a variety of basic cells, which may be any objects that are topologically equivalent to a sphere i.e. do not contain holes. This allows for a simplified combination of the respective elements, while the disadvantages of exhaustive enumeration like large memory consumption and the restricted accuracy of the object representation can be avoided.

In solid modelling, cell decomposition is mainly used as auxiliary representation for specific computations (Mäntylä 1988). However, it can also be efficiently applied to the automatic reconstruction of topologically correct building models at different levels of detail. This has already been proved for generalisation approaches (Kada 2007). But, as it will be demonstrated in the following sections, cell decomposition is also suitable to convert an existing coarse building model to a more detailed representation. Using appropriate partition planes for the cell generation, geometric constraints such as parallelism or rectangularity can be integrated very easily. Furthermore, constraints concerning neighboured cells can be treated in a simple manner, since the cell decomposition process implicates topology information for the generated cells.

3.2 *Facade refinement by terrestrial LIDAR*

The idea of the first part of our reconstruction algorithm is to segment a 3D object with a flat front face into 3D cells. Each 3D cell represents either a homogeneous part of the facade or a window area. Therefore, they have to be differentiated based on the availability of measured LIDAR points. After this classification step, window cells are eliminated while the remaining facade cells are glued together to generate the refined 3D building model. The difficulty is finding planar delimiters from the LIDAR points that generate a good working set of cells. Since our focus is on the reconstruction of the windows, the planar delimiters have to be derived from the 3D points that were measured at the window borders. These points are identified by a segmentation process.

3.2.1 Cell generation

As it is visible for the facade in Fig. 2, usually fewer 3D points are measured on the facade at window areas. This is due to specular reflections of the LIDAR pulses on the glass or points that refer to the inner part of the building and were therefore cut off in the pre-processing stage. If only the points are considered that lie on or in front of the facade, the windows will describe areas with no point measurements. Thus, our point cloud segmentation algorithm detects window edges by these no data areas. In principle, such holes can also result from occlusions. However, this is avoided by using point clouds from different viewpoints. In that case, occluding objects only reduce the number of LIDAR points since a number of measurements are still available from the other viewpoints.

During the segmentation process, four different types of window borders are distinguished: horizontal structures at the top and the bottom of the window, and two vertical structures that define the left and the right side. For instance, the edge points of a left window border are detected if no neighbour

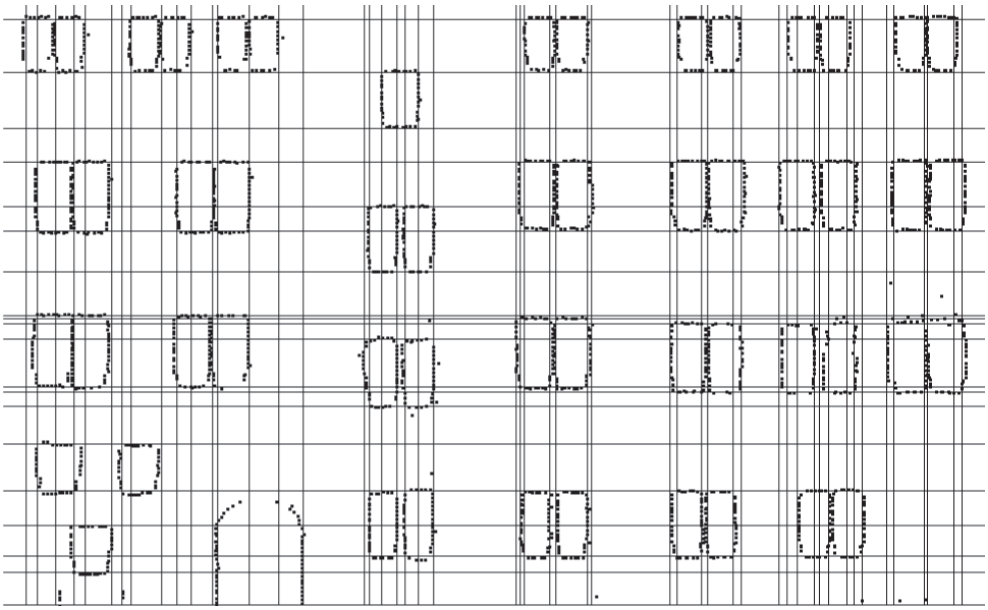

Fig. 6: Detected edge points and window lines.

measurements to their right side can be found in a pre-defined search radius at the facade plane. The search radius should be set to a value a little higher than the scan point distance on the facade. In a next step, horizontal and vertical lines are estimated from non-isolated edge points. Fig. 6 shows the extracted edge points at the window borders as well as the derived horizontal and vertical lines. Based on these window lines, planar delimiters can be generated for a subsequent spatial partitioning.

Each boundary line defines a partition plane, which is perpendicular to the building facade. For the determination of the window depth, an additional partition plane is estimated from the LIDAR points measured at the window crossbars. These points are detected by searching a plane parallel to the facade, which is shifted in its normal direction. The set of all partition planes provides the structural information for the cell decomposition process. Therefore, it is used to intersect the existing building model producing a set of small 3D cells.

3.2.2 Classification of 3D cells

In a next step, the generated 3D cells have to be classified into building and non-building fragments. For this purpose, a 'point-availability-map' is generated. It is a binary image with low resolution where each pixel defines a grid element on the facade. The optimal size of the grid elements is a value a little higher than the point sampling distance on the facade. Grid elements on the facade where LIDAR points are available produce black pixels (facade pixels), while white pixels (non-facade pixels) refer to no data areas. Of course, the already extracted edge points and the resulting structures in Fig. 6 are more accurate than the rasterized point-availability-map. However, this limited accuracy is acceptable since the binary image is only used to classify the 3D cells, which are already created from the detected horizontal and vertical window lines. The classification is implemented by computing the ratio of facade to non-facade pixels for each generated 3D cell. Cells including more than 70 % facade pixels are defined as facade solids, whereas 3D cells

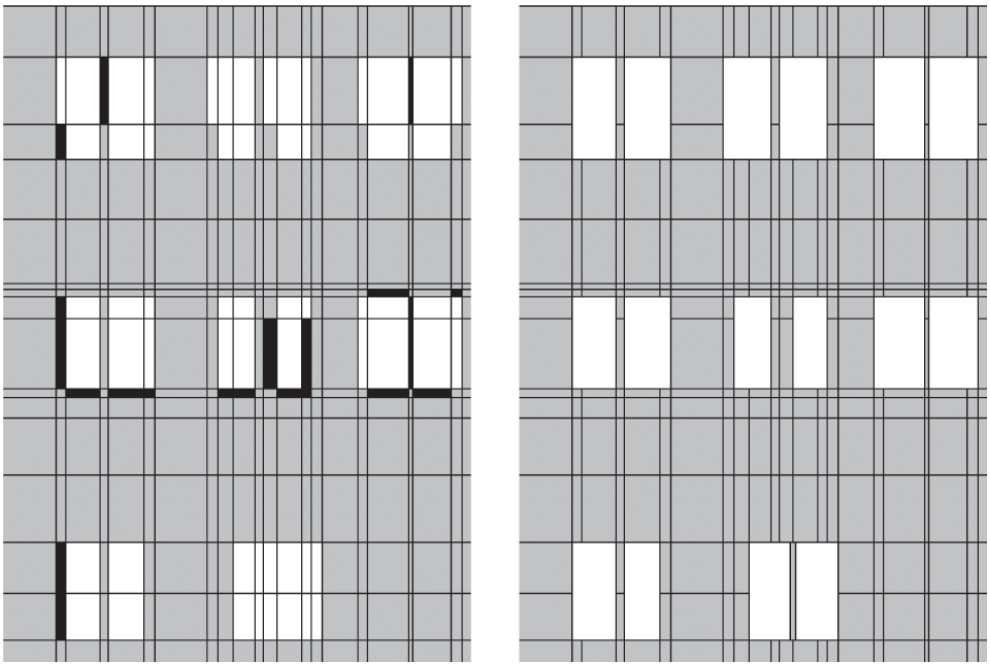

Fig. 7: Classification of 3D cells before (left) and after enhancement (right).

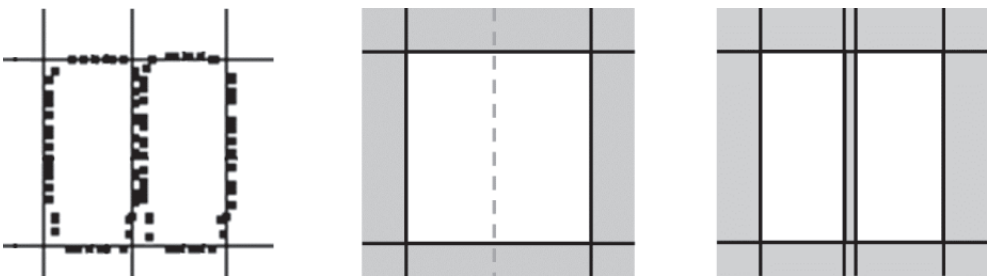

Fig. 8: Integration of an additional facade cell.

with less than 10 % facade pixels are assumed to be window solids. In the left of Fig. 7 these segments are depicted as grey (facade) and white (window) cells.

While most of the 3D cells can be classified reliably, the result is uncertain especially at window borders or in areas with little point coverage. Such cells with a relative coverage between 10 % and 70 % are represented by the black segments in the left of Fig. 7. For the final classification of these cells, neighbourhood relationships as well as constraints concerning the simplicity of the resulting window objects are used. As an example, elements between two window cells are assumed to belong to the facade, so two small windows are reconstructed instead of one large window. This is justified by the fact that facade points have actually been measured in this area. Additionally, the alignment as well as the size of proximate windows is ensured. For this purpose, uncertain cells are classified depending on their neighbours in horizontal and vertical direction. Within this process, it is also guaranteed that the merge of window cells will re-

sult in convex window objects. Fig. 7 (right) illustrates the enhanced classification result.

As it is depicted in Fig. 8, additional facade cells can be integrated easily if necessary. Fig. 8a shows the LIDAR measurements for two closely neighboured windows. Since in this situation only one vertical line was detected, a single window is reconstructed (cf. Fig. 8b). To overcome this problem, the window object is separated into two smaller cells by an additional facade cell. This configuration is kept if facade points are available at this position (cf. Fig. 8c).

3.2.3 Facade modelling

Within the following modelling process, the window cells are cut out from the existing coarse building model. The result of the building facade reconstruction is given in Fig. 9. The front of the pyramidal wall dormer is not considered as being a part of the facade. Therefore, the reconstruction approach is applied on the roof extension, separately.

While the windows are represented by polyhedral cells, also curved primitives can be integrated in the reconstruction process. This is demonstrated exemplarily by the round-headed door of the building. Furthermore, our approach is not limited to the

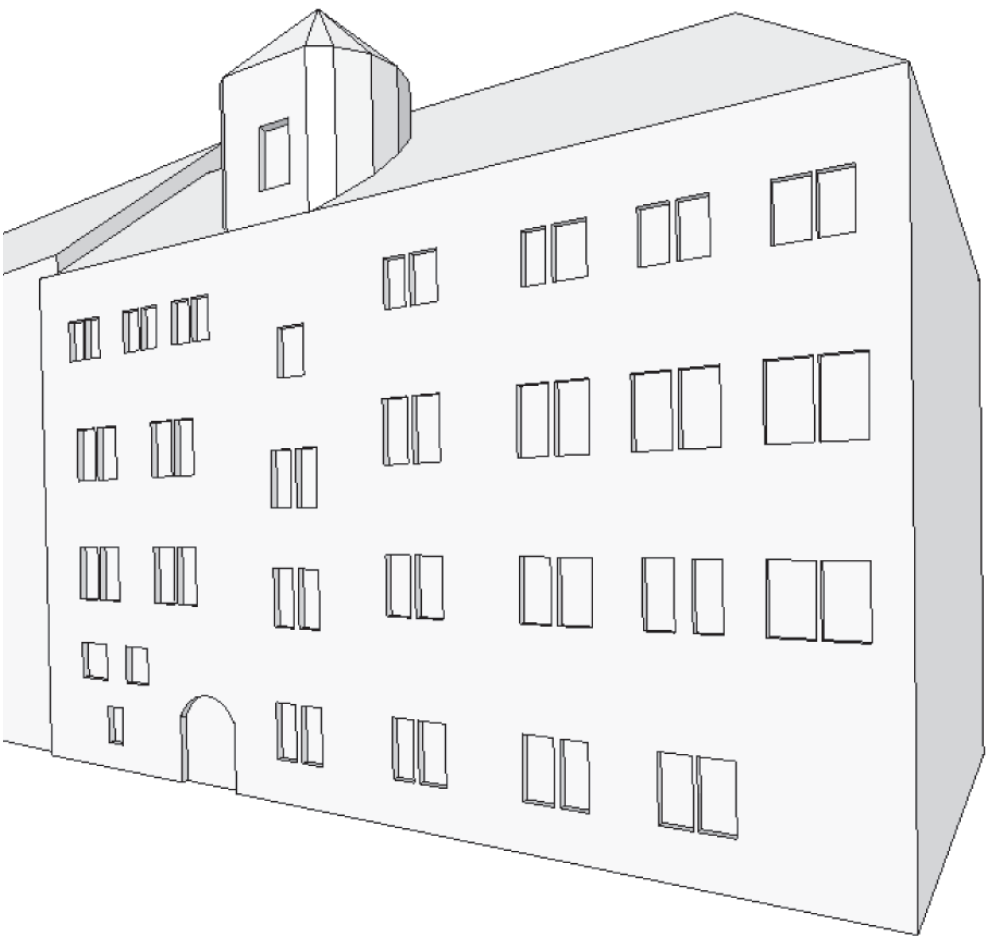

Fig. 9: Refined facade of the reconstructed building.

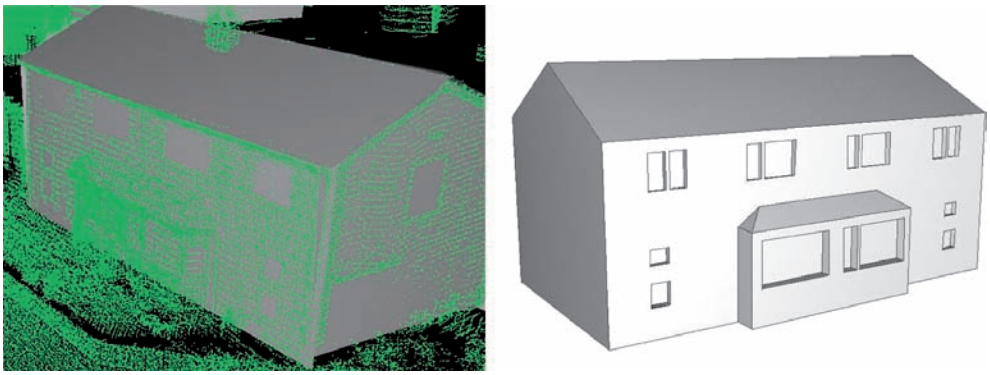

Fig. 10: Point cloud with coarse building model (left) and refined building model after recursive facade reconstruction (right).

modelling of indentations like windows or doors.

As it is visible in Fig. 10, details can also be added as protrusions to the facade. Fig. 10 (left) shows the available 3D point cloud (courtesy of School of Civil Engineering and Geosciences, Newcastle University, UK) as well as the automatically generated coarse building model. Based on this information, LIDAR points that are measured at protrusions can be detected easily since they are not part of the facade plane but lying in front of it. If these points are classified as non-facade points, protrusion areas can be identified in the same way as window regions, just by searching no data areas within the set of points that belong to the facade. The availability of LIDAR points in front of the facade helps to classify the derived 3D cells as protrusion cells. Their extent in the facade's normal direction can be reconstructed by fitting planes to the measured protrusion points. In order to detect further structures within the protrusion, the whole reconstruction algorithm can be applied recursively. The final result of this process is depicted in Fig. 10 (right). It features indentations as well as a protrusion with recursively modelled windows.

3.3 *Facade refinement by photos*

The level of detail for 3D objects that are derived from terrestrial laser scanning is limited depending on the point sampling distance. Small structures are either difficult to detect or even not represented in the data. By integrating image data in the reconstruction process the amount of detail can be increased.

3.3.1 Derivation of 3D edges

Having oriented the image data, 3D information can be derived from corresponding image features in order to reconstruct details of the facade such as crossbars. For this purpose, edge points are extracted from the images by a Sobel filter. These edge point candidates are thinned and split into straight segments. Afterwards, the resulting 2D edges of both images have to be matched. However, frequently occurring facade structures, such as windows and crossbars, hinder the search for corresponding edges. Therefore, the boundaries of the windows that have already been reconstructed from the LIDAR points are projected into both images. Only the 2D edges lying inside these image regions are considered for the following matching process. Fig. 11 depicts the selected 2D edges for an exemplary window in both images. Thus, possible mismatches are reduced, even though, they cannot be avoided entirely. Remaining false correspondences lead to 3D edges lying outside the reconstructed window. Therefore, these wrong edges can be easily identified and re-

Fig. 11: Selected 2D edges for a window in both images.

moved. In addition, only horizontal and vertical 3D edges are considered for the further reconstruction.

3.3.2 Reconstruction of additional facade structures

Photogrammetric modelling allows the extraction of well-defined image features like edges and points with high accuracy. By contrast, points from terrestrial laser scanning are measured in a pre-defined sampling pattern, unaware of the scene to capture. That means that the laser scanner does not explicitly capture edge lines, but rather measures points at constant intervals. Furthermore, laser measurements at edges and corners may provide erroneous and unpredictable results because of the laser beam split that is caused at the object border. For these reasons, the positional accuracy of window borders that are reconstructed from LIDAR points is limited compared to the photogrammetrically derived 3D edges at crossbars. As a consequence, the 3D reconstructions from laser points and images may be slightly shifted. Therefore, the reconstruction of the crossbars is done as follows:

For each window, hypotheses about the configuration of the crossbars are generated and tested against the 3D edges derived from the images. Possible shapes are dynamically generated as templates by recursively dividing the window area in two or three parts. Recursion stops when the produced glass panes are too small for a realistic generation of windows. The minimum width and height of the glass panes are restricted by the same threshold value. After each recursion step, the fitting of the template with the 3D edges is evaluated. The partition is

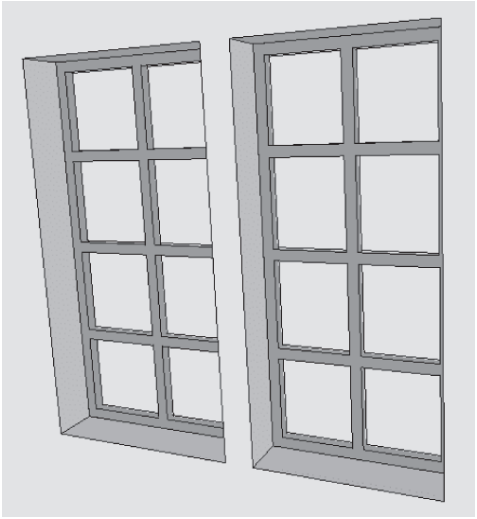

Fig. 12: Reconstructed crossbars for two windows.

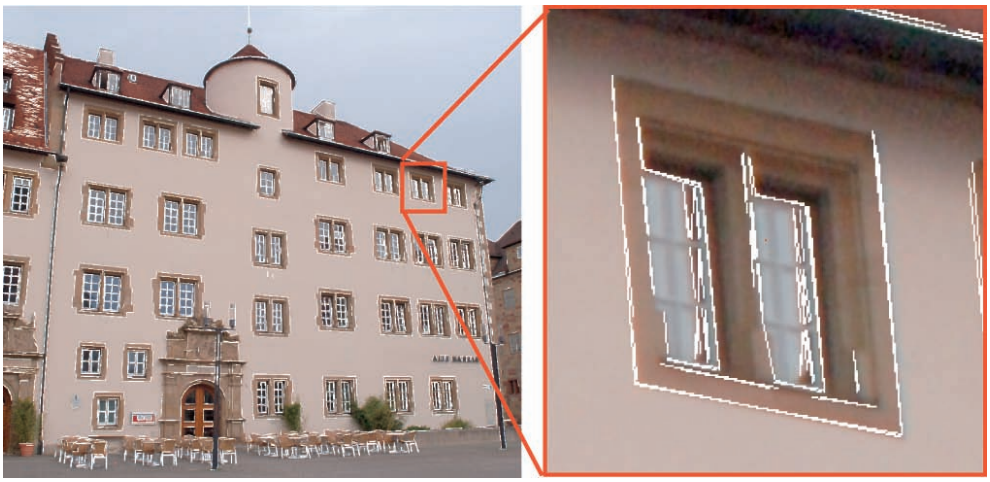

Fig. 13: Detected 2D edges for a window captured under an oblique view.

accepted if 3D edges are available within a buffer area around the dividing line. In a final step, the crossbars and the window frame are modelled. For this purpose, new 3D cells with a pre-defined thickness are generated at the accepted horizontal and vertical division lines as well as at the window borders. The result is exemplarily shown for two windows in Fig. 12.

While most of the crossbars can be reconstructed reliably, problems may arise for windows that are captured under oblique views. Perspective distortions or occlusions make it difficult to detect 2D edges at crossbars (cf. Fig. 13). Consequently, fewer 3D edges can be generated thereof in those areas.

To overcome this problem, neighbourhood relationships are taken into account within the final modelling step. The crossbar configuration is assumed to be equal for all windows of similar size which are located in the same row or column. Based on this assumption, similar windows can be simultaneously processed. Thus, the crossbar reconstruction leads to robust results even for windows that are partially distorted or feature strong perspective distortions in the respective image areas.

Fig. 14 shows the final result of the building facade reconstruction from terrestrial LIDAR and photogrammetric modelling. This example demonstrates the successful detection of crossbars for windows of medium size. However, the dynamic generation of templates even allows for the modelling of large window areas as they often occur at facades of big office buildings.

4 Conclusions

Within the paper, an approach for the refinement of 3D building models was presented. For this purpose, both 3D point clouds from terrestrial laser scanning and image data are applied. Terrestrial LIDAR can be reliably used to reconstruct larger building parts, for example window areas and protrusions. Smaller structures such as window crossbars, which are frequently not available from LIDAR due to the limited point sampling distance, are extracted from images. For the alignment of the LIDAR data and the images acquired from independent camera stations, a marker-free, fully automatic method was proposed.

Our model refinement approach primarily aims at the reconstruction of indentations such as windows and doors. However, as we could show, the algorithm also allows for the modelling of protrusions such as balconies or oriels. The simple integration of

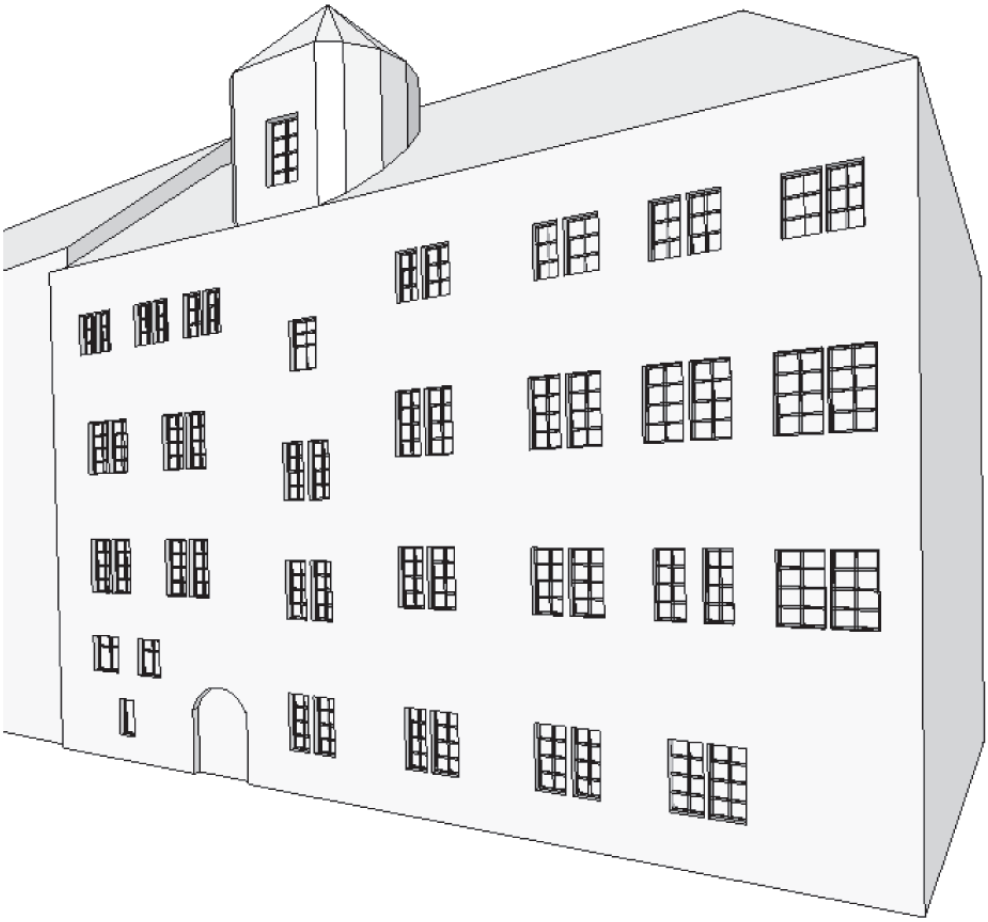

Fig. 14: Refined facade with detailed window structures.

geometric detail is based on the building representation by cell decomposition. Using this modelling technique, topologically correct building models can be automatically created at different levels of detail. Furthermore, as a consequence of the cell generation process, symmetry relations like coplanarity or alignment are guaranteed even for larger distances between the respective building parts. Thus, despite of the limited extent of the window primitives, which were extracted by the analysis of terrestrial LIDAR and images, structural information can be generated for the complete building. In future work, this information will be used to support the facade interpretation at areas where measurements are only available with reduced quality and reliability. For these reasons, in our opinion, this approach has a great potential for processes aiming at the reconstruction and refinement of building models from multiple data sets.

References

Alegre, F. & Dallaert, F., 2004: A Probabilistic Approach to the Semantic Interpretation of Building Facades. – Workshop on Vision Techniques Applied to the Rehabilitation of City Centres: 1–12.

Besl, P. J. & McKay, N., 1992: A method for Registration of 3-D Shapes. – IEEE PAMI **14** (2): 239–256.

Böhm, J. & Becker, S., 2007: Automatic marker-free registration of terrestrial laser scans using

reflectance features. – Optical 3D Measurement Techniques, Zurich, Switzerland: 338–344.

Böhm, J. & Haala, N., 2005: Efficient Integration of Aerial and Terrestrial Laser Data for Virtual City Modeling Using LASERMAPS. – IAPRSIS **36** (3/W19): 192–197.

Brenner, C., 2004: Modelling 3D Objects Using Weak CSG Primitives. – IAPRSIS **35**.

Brenner, C. & Ripperda, N., 2006: Extraction of Facades Using RjMCMC and Constraint Equations. – IAPRSIS **36** (3).

Fischler, M.A. & Bolles, R.C., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. – Communications of the ACM **24**: 381–395.

Früh, C. & Zakhor, A., 2003: Constructing 3D City Models by Merging Ground-Based and Airborne Views. – IEEE Computer Graphics and Applications, Special Issue Nov/Dec.

Haala, N., Kada, M., Becker, S., Böhm, J. & Alshawabkeh, Y., 2006: Graphic Tools for the Generation of Large Scale Urban Scenes. – IAPRSIS **36** (4).

Kada, M., 2007: Scale-Dependent Simplification of 3D Building Models Based on Cell Decomposition and Primitive Instancing. – International Conference on Spatial Information Theory: COSIT '07, Melbourne, Australia.

Kremer, J. & Hunter, G., 2007: Performance of the StreetMapper Mobile LIDAR Mapping System in ''Real World'' Projects. – Photogrammetric Week '07, Wichmann, Heidelberg: 215–225.

Lowe, D., 2004: Distinctive image features from scale-invariant keypoints. – IJCV **60** (2): 91–110.

Mäntylä, M., 1988: An Introduction to Solid Modeling. – Computer Science Press, Maryland, USA.

Mayer, H. & Reznik, S., 2006: MCMC Linked With Implicit Shape Models and Plane Sweeping for 3D Building Facade Interpretation in Image Sequences. – IAPRSIS **36** (3).

Mikolyajczyk, K. & Schmid, C., 2003: A performance evaluation of local descriptors. – Computer Vision and Pattern Recognition: 257–264.

Nistér, D., 2004: An efficient solution to the five-point relative pose problem. – IEEE Transactions on Pattern Analysis and Machine Intelligence **26** (6): 756–770.

Remondino, F. & Ressl, C., 2006: Overview and experiences in automated markerless image orientation. – IAPRSSIS **36** (3): 248–254.

Rottensteiner, F., 2001: Semi-automatic extraction of buildings based on hybrid adjustment using 3D surface models and management of building data in a TIS. – PhD. thesis TU Wien.

Vedaldi, A., 2007: An Implementation of the Scale Invariant Feature Transform. – UCLA CSD Tech. Report N. 070012.

Zeng, Z. & Wang, X., 1992: A general solution of a closed-form space resection. – PE&RS **58** (3): 327–338.

Addresses of the Authors:

Dipl.-Ing. Susanne Becker, PD Dr.-Ing. habil. Norbert Haala, University of Stuttgart, Institute for Photogrammetry, D-70174 Stuttgart, Tel.: +49-711-686-84113, Fax: +49-711-685-83297, e-mail: susanne.becker@ifp.uni-stuttgart.de

Manuskript eingereicht: November 2007
Angenommen: Januar 2008