COMBINED FEATURE EXTRACTION FOR FAÇADE RECONSTRUCTION

Susanne Becker, Norbert Haala

Institute for Photogrammetry, Universitaet Stuttgart
Geschwister-Scholl-Str. 24D, 70174 Stuttgart
Forename.Lastname@ifp.uni-stuttgart.de

KEY WORDS: Three-dimensional, Point Cloud, Urban, LIDAR, Modelling, Façade Interpretation

ABSTRACT:

Within the paper, the combined application of terrestrial image and LIDAR data for façade reconstruction is discussed. Existing 3D building models as they are available from airborne data collection are additionally integrated into the process. These given models provide a priori information, which efficiently supports both the georeferencing of the terrestrial data and the subsequent geometric refinement. Approximate orientation parameters for the terrestrial LIDAR measurements are provided by suitable low-cost components. Exact georeferencing is then realised by an automatic alignment to the building models, which are given in the required reference coordinate system. The automatic relative orientation of the terrestrial images is implemented by tie point matching. A modified version of this matching process is then used to align these images to the terrestrial LIDAR data, which were already georeferenced in the preceding step. After this fully automatic orientation process, the given 3D model is refined by the extraction of window structures from the LIDAR point clouds and façade images.

1. INTRODUCTION

Terrestrial LIDAR is frequently used for the collection of highly detailed 3D city models. Urban models are already available for a large number of cities from aerial data like stereo images or airborne LIDAR. However, while airborne data collection is especially suitable to provide the outline and roof shape of buildings, terrestrial data collection from ground based views is especially suitable for the refinement of building facades. Thus, terrestrial and aerial data provide complementary information during 3D city model generation. In our approach, this is realised by using given 3D building models from aerial data collection as a priori information during geocoding of the terrestrial data. This automatic alignment for both the terrestrial LIDAR and image data is one of the main focuses of this paper. In the second part of the paper, the combination of LIDAR and image data for façade modelling while using the given 3D models as reference surfaces will be discussed.

Spatially complex areas like urban environments can only be completely covered by terrestrial laser scanning (TLS) if data collection is realised from different viewpoints. Usually, scans from different viewpoints are aligned based on tie and control point information measured at specially designed targets. These targets are manually identified while a refined measurement is performed automatically. In contrast, our approach allows for fully automatic registration and georeferencing by matching the point clouds from terrestrial laser against the corresponding faces of the given 3D building model. This can be implemented by the standard iterative closest point algorithm introduced by (Besl & McKay 1992) since a coarse alignment of the scans is available. For this purpose, the position and orientation of the scanner is determined simultaneously to point measurement by integrated GPS and digital compass.

One of the main applications of 3D city models is the generation of realistic visualisations. This requires a suitable texture mapping for the respective building surfaces in addition to geometric data collection. Thus, in order to simultaneously capture corresponding colour information, a digital camera is directly integrated in some commercial 3D systems. However, this limits the camera viewpoints to the laser scanning stations, which might not be optimal for the collection of high quality image texture. Additionally, laser scanning for the documentation of complex object structures and sites frequently has to be realised from multiple viewpoints. This can result in a relatively time consuming process. For these reasons, the acquisition of object geometry and texture by two independent sensors and processes to allow for an image collection at optimal positions and time for texturing will be advantageous. Even more important, images collected from multiple terrestrial viewpoints can considerably improve the geometric modelling based on the TLS data.

Captured images can be directly linked to the 3D point cloud if the camera is directly integrated to the laser scanner and a proper calibration of the complete system is available. In contrast, for independent viewpoints of camera and laser, the combined evaluation requires a suitable co-registration process for the respective range and image data sets. 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), which extracts points and suitable descriptions for the following matching based on histograms of gradient directions. By these means robust automatic tie point measurement is feasible even for larger baselines.

We use this operator to align both the terrestrial images and the terrestrial LIDAR data. First, a bundle block adjustment based on the matched key points between the digital images is realised. The resulting photogrammetric network is then transferred to object space by additional tie points which link the digital images and the TLS data that were georeferenced in the preceding step. 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, which are usually measured in addition to the run-time during scanning, represent the backscattered energy of the respective laser footprints. Thus, the intensities are exactly coregistrated 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 the key point operator.

While the automatic georeferencing of the different data sets will be discussed in Section 2, their combined use for a refined 3D modelling will be presented in Section 3. The benefit of using both image based measurements and densely sampled point clouds from terrestrial laser scanning is demonstrated for automatic façade refinement by the extraction of window structures.

2. DATA PREPARATION AND ORIENTATION

Within our investigations, a standard digital camera NIKON 2Dx was used for image collection, while the Leica HDS 3000 scanner was used for LIDAR measurements. This scanner is based on a pulsed laser operating at a wavelength of 532 nm. It is able to acquire a scene with a field of view of up to 360° horizontal und 270° vertical in a single scan. The typical stand-off distance is 50 to 100 meters, but measurements of more than 200 meters are possible. The accuracy of a single point measurement is specified with 6 mm. Within our tests the facades of the historic buildings around the Schillerplatz were recorded. In order to be able to reconstruct the scene in detail, the resolution on the facades was chosen to about ten centimetres, which is typical for this type of application. To prevent holes in the point cloud i.e. due to occlusions by the monument placed in the centre of the square, the dataset is composed of three separate 360°-scans from varying stations.

For direct georeferencing of the terrestrial scans, a low-cost GPS and a digital compass were additionally mounted on top of the HDS 3000 laser scanner. Digital compasses such as the applied TCMVR-50 can in principle provide the azimuth at a standard deviation below 1°. However, these systems are vulnerable to distortion. Especially in build-up areas the Earth's magnetic field can be influenced by cars or electrical installations. These disturbances usually reduce the accuracy of digital compasses to approximately 6° (Hoff and Azuma, 2000). The used low cost GPS receiver mounted on top of the digital compass is based on the SIRF II chip. Since it was operated in differential mode, the EGNOS (European Geostationary Navigation Overlay Service) correction signal could be used. By these means the accuracy of GPS positioning can be improved from 5-25m to approximately 2m. The vertical component of the low-cost GPS measurement was discarded and substituted by height values from a Digital Terrain Model.

For our test area the geometry of the respective buildings is already available from a 3D city model, which is maintained by the City Surveying Office of Stuttgart. In the following this building geometry is used both for georeferencing and refined modelling. The quality and amount of detail of this data set is typical for such 3D models, like they are available area covering for a number of cities. For the applied city model the roof faces were 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 façade segments, which were generated by extrusion of this ground plan, is relatively high, despite the fact that they are limited to planar polygons.

2.1 Georeferencing of LIDAR data

A global orientation of the laser scanner head in WGS 84 is measured by the low-cost GPS in combination with the digital compass. This approximate solution is further refined using the iterative closest point (ICP) algorithm introduced by (Besl & McKay 1992). The result of the direct georeferencing is used as an initial value for the iterative registration of the laser scans. Once the registration of the TLS data has converged, it is kept fixed. Then the complete dataset is registered with the city model using the same algorithm. Since the initial approximation of the direct georeferencing is within the convergence radius of the ICP algorithm, this approach allows for an automated geo-referencing of TLS data (Schuhmacher & Böhm 2005).

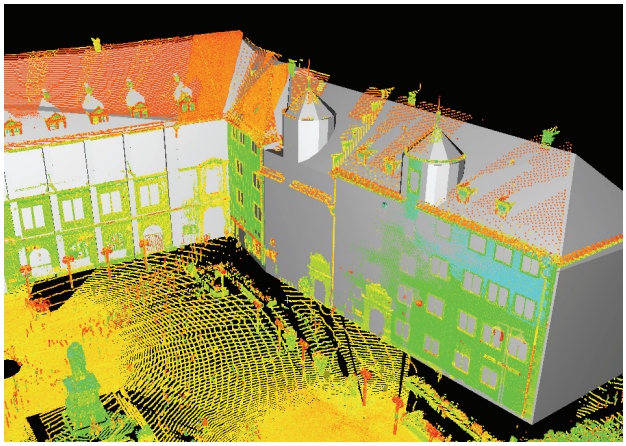

Figure 1: 3D point cloud from laser scanning aligned with a virtual city model.

As it is demonstrated in Figure 1, after this step the 3D point cloud is available in the reference system as provided by the 3D city model.

2.2 Alignment of image data

The integration of image data into the façade reconstruction requires image orientation in a first step. The images have to be aligned with each other and registered according to the already georeferenced laser point cloud. This is usually performed by means of a bundle block adjustment providing accurate estimates of 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.

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).

Figure 2. Image data for photogrammetric modelling.

Figure 2 shows images from a calibrated camera (NIKON D2x Lens NIKKOR 20mm). 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 (Nister 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 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.

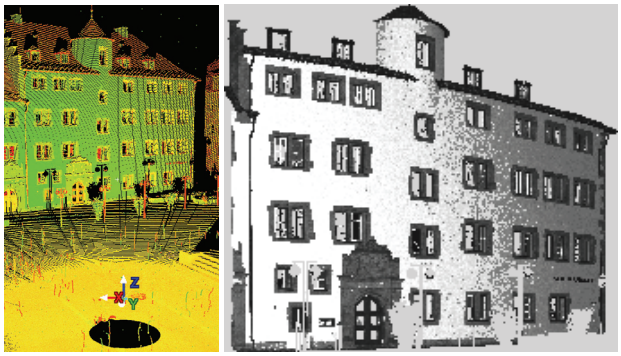

Figure 3. Measured laser reflectivities as 3D point cloud (left) and 2D image representation (right).

Figure 3 (left) shows the laser point cloud of an already georeferenced scan. The position of the laser scanner is marked by the coordinate axes of the local scanner system. The image representation derived from the reflectivity values is given in Figure 3 (right). Each pixel with a valid laser reflectivity value refers to 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 photo cameras. On the one hand, the laser intensities represent the reflectivity of the measured surface only in a narrow wavelength range (for example 532 nm for the HDS 3000). Furthermore, the viewing direction and the direction of illumination are identical in case of laser scanning. By contrast, photo cameras usually work with ambient light sources which may cause shadow areas on the object and therefore lead to grey value edges in the photograph. On the other hand, the laser image is not based on central projection but on polar geometry. Thus, like it is visible in the right image of Figure 3, straight 3D lines appear curved in the reflectivity image. Another aspect is the sampling distance, which is often much higher for a laser scan compared to the spatial resolution of a photo captured by a camera. 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.

Figure 4 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 Figure 4. Due to the decreasing reflectivity values in the right part of the laser image, correct matches could be found only on the left part of the building façade.

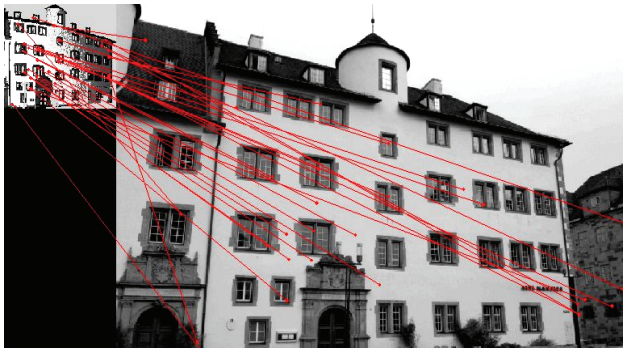

Figure 4. Key point correspondences for the laser reflectivity image (left) and one of the photographs (right).

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. However, the accuracy of orientation parameters obtained from a minimal set of points strongly depends on the point configuration. 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. Therefore, it is difficult to find the correct solution within the RANSAC process. In order to improve the accuracy of the key point positions and the derived orientation parameters, the initial set of point correspondences is augmented: For each pair of key points, new point correspondences are generated by randomly shifting the key point in the photograph by a few pixels. Out of these additional point correspondences only the one is kept which contributes to

the best solution for the exterior orientation. Beyond that, for a further stabilisation of the RANSAC process, a priori information on the ground height is integrated. Assuming a smooth terrain in front of the building, only those solutions are considered, where the positions of the camera and the laser scanner differ less than 1m in height. In this way, about 22% of the key point matches are confirmed as valid correspondences.

The resulting approximate orientation parameters for the photographs are then refined in a final bundle adjustment. For this purpose the Australis software package was used. The average standard errors of the estimated orientation parameters are σ_X = 7.6cm, σ_Y = 5.6cm, σ_Z = 8.1cm, σ_{az} = 0.167°, σ_{el} = 0.164°, σ_{roll} = 0.066°. The average precision of the computed object coordinates is σ_X = 3.3cm, σ_Y = 4.7cm, σ_Z = 2.1cm.

3. FAÇADE RECONSTRUCTION

After this georeferencing process, the collected terrestrial data sets are aligned to the existing building model, which is provided from the existing 3D city model. Thus, both the LIDAR point clouds and the images can be used to enhance this coarse model. In the following, this is demonstrated exemplarily for the geometric refinement of the building façade by a two-step approach. First, the LIDAR point clouds are used to decompose the given building model into 3D cells to additionally represent façade structures like windows and doors. This cell decomposition, which can be used very effectively to represent building models at multiple scales (Haala et al 2006), is then refined in a second step by photogrammetric analysis of the images. Thus, the amount of detail is further increased for the window frames while profiting from the higher resolution of the image data.

3.1 Façade Refinement by Terrestrial LIDAR

As a first step of the LIDAR based refinement of the building façade, suitable 3D point measurements are selected by a simple buffer operation. While assuming that the façade can be described sufficiently by a relief, the vertical distances between the measured 3D laser points and the given façade polygon can be used to generate a 2.5D representation, or can even be interpolated to a regular grid. Thus, further processing like the following segmentation is simplified considerably by such a mapping of the 3D points against the given façade plane.

3.1.1 Point cloud segmentation

The LIDAR measurement will be used to decompose the coarse 3D building with a flat front face into suitable façade cells. For this purpose, planar delimiters are derived by segmenting LIDAR points measured at the window borders.

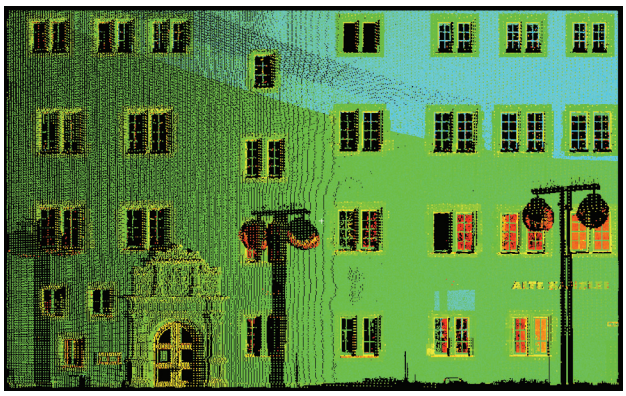

Figure 5. 3D point cloud as used for the geometric refinement of the corresponding building façade.

Figure 5 shows a point cloud, which was selected for a building façade based on the alignment to the virtual city model as depicted in Figure 1. This 2.5D representation clearly demonstrates that usually fewer points are measured at window areas than the façade of a building. 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 façade, the windows will describe areas with no point measurements. Taking advantage thereof, our point cloud segmentation algorithm detects window edges, which are defined by these no data areas. In principle, such holes can also result from occlusions. This is avoided by using point clouds from different viewpoints, though. In that case, occluding objects only reduce the number of LIDAR points since a number of measurements are still available from the other viewpoints.

Our segmentation process differentiates four types of window borders: horizontal structures at the top and the bottom of the window, and two vertical structures that define the left and the right side. As an example, the edge points of a left window border are detected if no neighbour measurements to their right side can be found in a pre-defined search radius at the façade plane. We used a search radius a little higher than the scan point distance on the façade, otherwise, no edge points would be found at all.

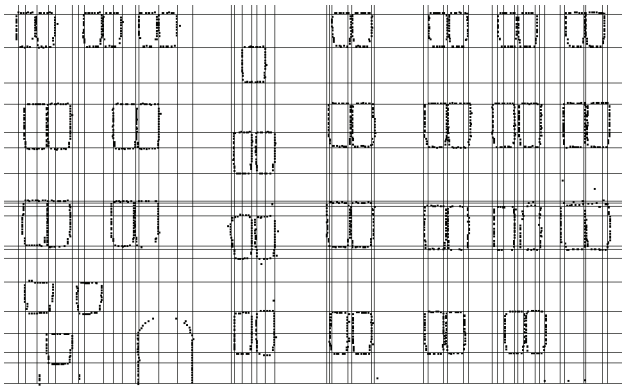

Figure 6. Detected horizontal and vertical window lines.

Based on the extracted edge points, which are depicted in Figure 6, the window borders can be determined in the following step.

3.1.2 Spatial-Partitioning of the building façade

Within this step, horizontal and vertical lines are estimated from non-isolated edge points. As it is also visible in Figure 6, these boundary lines are then used to decompose the building façade in suitable cells. Each of these cells represents either a homogeneous part of the façade or a window area. After a classification based on the availability of measured LIDAR points, window cells can be eliminated from the façade and the refined 3D building model is generated.

The separation of cells into building and window fragments is based on a 'point-availability-map'. This low resolution binary image provides pixels which either represent façade regions, where LIDAR points are available, or areas with no 3D point measurements. This image is then used to compute the ratio of façade to non-façade pixels for each facade cell as required for the following classification. A refined classification is implemented based on neighbourhood relationships and constraints concerning the simplicity of the resulting window objects. Uncertain cells are for example classified depending on

their neighbours in order to align and adapt proximate windows in horizontal and vertical direction. Within this step, convex window objects can additionally be guaranteed.

3.1.3 Model Refinement

Finally, the façade geometry is modelled by eliminating the classified window cells from the existing coarse building model. For this purpose, a plane parallel to the facade at window depth is determined from LIDAR points measured at the window crossbars.

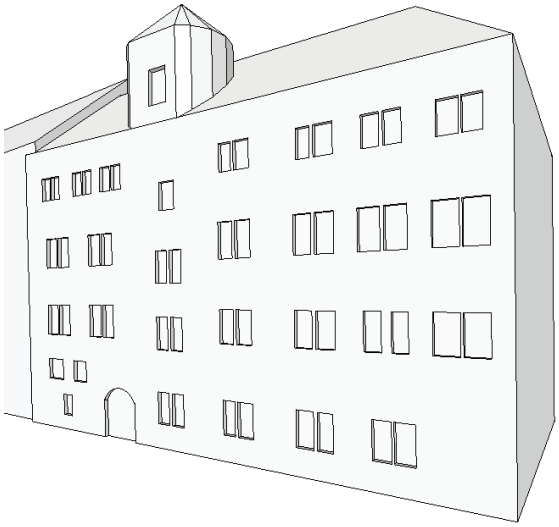

Figure 7. Refined facade of the given building model.

As depicted in Figure 7, the classified facade cells are then carved out from the building model at this window depth. While the windows are represented by polyhedral cells, also curved primitives can be integrated in the reconstruction process as demonstrated by the round-headed door of the building. Furthermore, our approach is not limited to the modelling of indentations like windows or doors. Details can also be added as protrusions to the façade.

3.2 Image based Facade Refinement

For our data set, the point sampling distance of terrestrial laser scanning was limited to approximately 10cm. Thus, smaller structures can not be detected. However, the amount of detail can be increased by integrating image data in the reconstruction process. This is exemplarily shown for the reconstruction of window crossbars.

3.2.1 Derivation of 3D edges

By matching corresponding primitives, the georeferenced image data is used to derive the required 3D information. In order to reconstruct linearly shaped façade detail such as crossbars, 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 can be matched. However, frequently occurring façade structures, such as windows and crossbars, hinder the search for corresponding edges. Therefore, the boundaries of the already reconstructed windows are projected into both images. Only the 2D edges within these regions are further processed. Thus, possible mismatches are reduced, even though, they cannot be avoided completely. Figure 8 depicts the selected 2D edges for an exemplary window in both images.

Remaining false correspondences result in 3D edges outside the reconstructed window. Therefore, these wrong edges can be easily identified and removed. In addition, only horizontal and vertical 3D edges are considered for the further reconstruction process. The reconstructed wrong (green) and correct (red) 3D edges are shown in local façade coordinates in Figure 9. The position of the window that has been derived from the LIDAR data is illustrated in black.

Figure 8. Selected 2D edges for a window in both images.

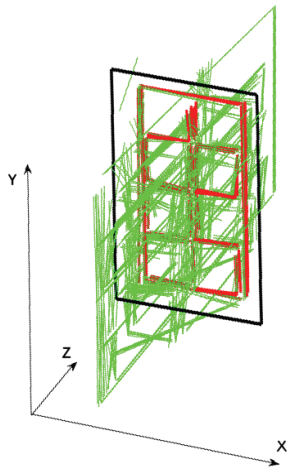

Figure 9. Wrong (green) and correct (red) 3D window edges.

3.2.2 Reconstruction of additional façade 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. For this reason, 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 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 Figure 10.

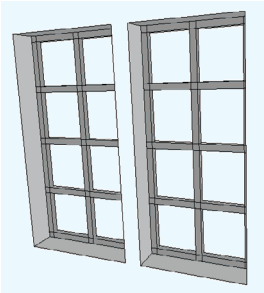

Figure 10. Reconstructed crossbars for two windows.

Most crossbars can be reconstructed reliably. However, problems may arise for windows that are captured under oblique views. This is due to perspective distortions or occlusions making it difficult to detect 2D edges at crossbars (Figure 11). Consequently, only a reduced number of 3D edges can be generated thereof in those areas.

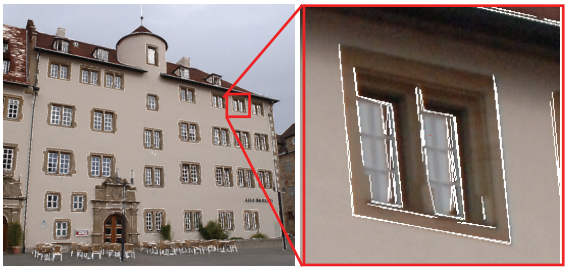

Figure 11. 2D edges for a window under an oblique view.

In order to stabilize the modelling process of crossbars, neighbourhood relationships are taken into account. 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 occluded or feature strong perspective distortions in the respective image areas.

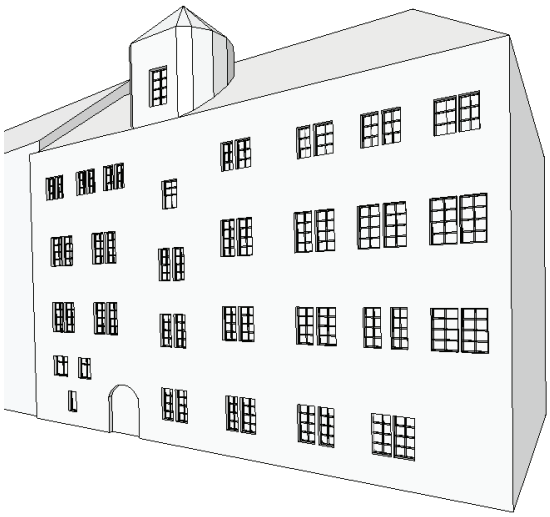

Figure 12. Refined facade with detailed window structures.

The final result of the building façade reconstruction from terrestrial LIDAR and photogrammetric modelling can be seen in Figure 12. 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. CONCLUSION

Within the paper the combined use of terrestrial image and LIDAR data for the extraction of façade geometry was presented. For this purpose a fully automatic geoereferencing of the collected data sets based on SIFT algorithm was realised in a first processing step. As presented, SIFT matching is a promising tool for the marker-free connection of photos and laser data. It is working well in standard scenarios for relative small baselines when the viewing direction of the laser scanner is approximately perpendicular to the dominating object surfaces. In this case, perspective distortions and decreasing reflectivity values in the laser image are negligible. However, problems may arise if the point density of the laser scans is low compared to the spatial resolution of the photograph leading to an instable matching and orientation process.

The refinement of 3D building models is based on a cell decomposition approach. As it was already proved for the automatic generation of topologically correct building models at different levels of detail (Haala et al 2006), this approach allows the simple integration and removal of geometric detail for given building models. Even more important, symmetry relations like coplanarity or alignment can be guaranteed even for larger distances between the respective building parts. Thus, despite of the limited extent of the window primitives, which were extracted from terrestrial LIDAR and images, structural information can be generated for the complete building.

5. REFERENCES

Besl, P. J. & McKay, N. [1992] A method for Registration of 3-D Shapes. IEEE PAMI 14[2], pp. 239-256.

Böhm, J. & Becker, S. [2007]. Automatic marker-free registration of terrestrial laser scans using reflectance features. To appear in Optical 3D, Zurich, Switzerland.

Fischler, M. A. & Bolles, R. C. [1981]. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Comm. Of the ACM, Vol. 24, pp. 381-395.

Haala, N., Becker, S. & Kada, M. [2006]. Cell Decomposition for the Generation of Building Models at Multiple Scales. IAPRS Vol. 36 Part III, Symposium Photogrammetric Computer Vision, pp. 19-24.

Lowe, D. [2004]. Distinctive image features from scale-invariant keypoints. IJCV, Vol. 60(2), pp. 91-110.

Mikolajczyk, K. & Schmid, C. [2003]. A performance evaluation of local descriptors. Proc. Conf. Computer Vision and Pattern Recognition, pp. 257-264.

Nistér, D. [2004]. An efficient solution to the five-point relative pose problem. IEEE PAMI, 26(6), pp. 756-770.

Remondino, F. & Ressl, C. [2006]. Overview and experiences in automated markerless image orientation. IAPRS, Vol. 36, Part 3, pp. 248-254.

Schuhmacher, S. & Böhm, J. [2005]. Georeferencing of terrestrial laserscanner data for applications in architectural modeling. IAPRS Vol. 36, PART 5/W17.

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

Zeng, Z. & Wang, X. [1992]. A general solution of a closed-form space resection. PE&RS, Vol. 58, No. 3, pp. 327-338.