FACADE RECONSTRUCTION FROM AERIAL IMAGES BY MULTI-VIEW PLANE SWEEPING

Lukas Zebedin, Andreas Klaus, Barbara Gruber and Konrad Karner

VRVis Research Center
Inffeldgasse 16/2, Graz, AUSTRIA
{zebedin, klaus, gruber, karner}@vrvis.at

KEY WORDS: Building Reconstruction, Aerial Images, Plane Sweeping, Information Fusion, Multi-View Matching

ABSTRACT:

This papers describes an algorithm to estimate the precise position of facade planes in digital surface models (DSM) reconstructed from aerial images using an image-based optimization method which exploits the redundancy of the data set (along and across track overlap). This approach assumes that a facade is a vertical plane and that the heightfield is precise enough to generate hypotheses for the initialization of the optimization algorithm. The initialization is first roughly oriented using the principal line directions of its texture, afterwards a hierarchical algorithm performs a finer optimization to maximize the correlation across different views. The proposed method is applied to real world imagery and its results are shown.

1 INTRODUCTION AND MOTIVATION

Reconstruction of buildings in urban areas from aerial images is a challenging task. Many applications like virtual tourism, urban planning and cultural documentation benefit from a realistic, high-quality city model. There already exist methods to create a dense point cloud of urban scenes using LIDAR scans or dense image matching ((Berthod et al., 1995), (Cord et al., 1998)) which can be used to create a polygonal roof model ((Samadzadegan et al., 2005)), (Vosselman and Dijkman, 2001)), however the estimation of facades poses a separate problem because of the oblique angle at which they are viewed during aerial data acquistion. The optimization employed by the proposed algorithm is image-based.

One critical aspect of building reconstruction is the estimation of the contours of buildings. Many workflows on urban scene reconstruction rely on additional information like a ground-plan ((Brenner, 2000) and (Haala et al., 1998) for example) to delineate the contours of buildings. However, this information is not always available or has to be manually created which is a major drawback if a fully automatic workflow is desirable.

The other possibility is to infer the outlines of buildings by segmenting the DSM into building blocks. This has been done by (Weidner, 1996) and (Vosselman, 1999). The drawback of this approach is obviously the flawed, jaggy nature of the obtained contours. (H. Gross, 2005) tried to alleviate this by fitting rectangles to the outline. Such improvements however can only guess the position of the facades. If the resulting model is afterwards textured, any error in the placement results in skewed and misaligned textures.

This drawback of automatic deduction of outlines can be alleviated by optimizing the position of the outlines as proposed in this paper.

(Coorg and Teller, 1999) presented a similar algorithm which operated on close-range imagery. They, however, relied strongly on horizontal lines in building facades to even initialize their estimates.

The basic idea of plane sweeping was also used in (T. Werner, 2002), but there only a translational plane sweep is considered in terrestial imagery. Also the initialization of the plane sweep is quite different from our approach where vanishing points are being exploited.

(C. Vestri, 2000) discusses a very similar algorithm to the one proposed in this paper, but is based on pointwise reconstruction of a facade. The main difference however is that they use vertical planes which are rotated in 20 degree intervals around the vertical axis to obtain the facade points whereas our algorithm optimizes the rotational and translational component of each facade independently therefore increasing the estimation accuracy. Additionally the pointwise reconstruction performed by them does not exploit the knowledge that the facade is a plane.

This contribution is based on images from the UltraCamD camera from Vexcel Corporation with its multispectral capability. The UltraCamD camera features a multi-head design. It delivers large format panchromatic images composed from nine CCD sensors (11500 pixels across-track and 7500 pixels along-track) and simultaneously recorded four additional channels (red, green, blue and NIR) at a frame size of 3680 by 2400 pixels. The image data used comprise the panchromatic high resolution images as well as the low resolution multispectral images.

The data set used in this paper to compute the depicted results was acquired in Summer 2005 over the inner city of Graz, Austria. It consists of 155 images flown in 5 strips. The along-track-overlap of this data set is 80%, the across-track overlap is approximately 60%. The ground sampling distance is around 8cm.

2 FACADE OPTIMIZATION

The algorithm for obtaining optimized facades can be decomposed into three distinct steps: first some hypotheses have to be found. Those estimated facades are then refined in such a way, that they are parallel to the true facade. In the last step the fine-grained optimization using multi-view correlation is performed.

2.1 Input Data

The optimization algorithm is image-based, therefore a precise orientation of the imagery is of utmost importance. The average back projection error is of utmost importance to enable convergence of the optimization. Theoretically two views of a plane are enough to calculate the correlation score, however in case of occlusions and in order to increase stability more views can be used. Therefore the data acquisition is also critical to the success of the optimization because only views are usable where the facade lies near the border of the image. The reason for this is the fact that aerial images have a very limited visibility of vertical planes as in the center of each image the perspective projection is comparable to a orthographic projection which hides all vertical planes . This assumption requires that flight altitude, velocity, focal length and along/across-track overlap are carefully chosen to provide also data redundancy for facades.

Another prerequisite is the DSM which is used to initialize the hypothesis for facades. For the experiments conducted for this paper, a plane sweeping approach was chosen which is improved and densified by applying an iterative and hierarchical multi-view matching algorithm based on homographies. A more detailed description of this algorithm implemented on graphics hardware can be found in (Zach et al., 2003).

The building block layer is based on a land use classification and describes the position of buildings within the scene. The land use classification used for this data set is a supervised classification that includes a training phase and that runs automatically afterwards. The classification results comprise classes like buildings, streets or other solid objects with low height, water, grass, tree or wood, as well as soil or bare earth. The classification is based on support vector machines and is described in detail in (Gruber-Geymayer et al., 2005).

2.2 Initialization

The initial estimates of the position of facades is obtained by applying a Canny edge detector to the heightfield. Those edgels are afterwards chained together to form lines. One important parameter of this line extraction is the minimum length of each line, as longer lines tend to be more stable in the optimization performed in a later phase.

The line extraction is aided by the land use classification which assigns a label to each pixel in the heightfield. These labels are used to restrict line extraction to regions near buildings.

The result of this procedure is illustrated in Figure 1. Note that only lines near the building are extracted whereas there are no lines near the tree in the inner courtyard of the building.

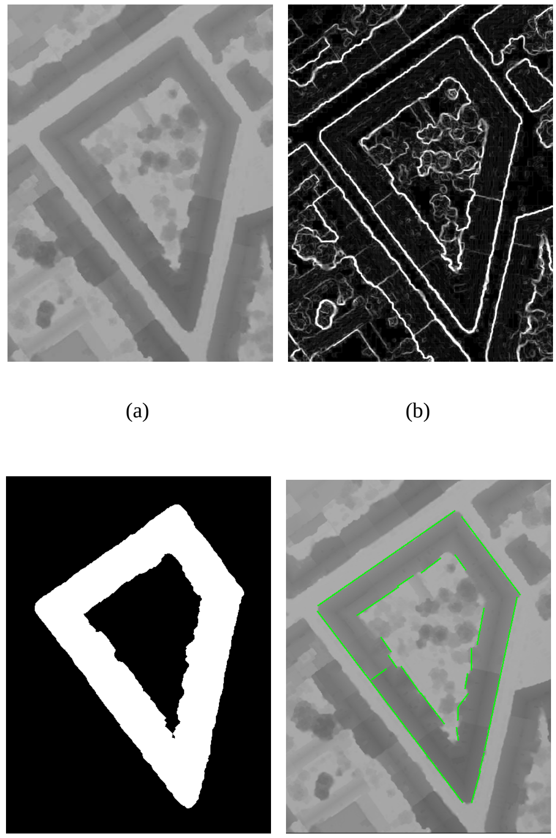
These lines in 2D are then extended to 3D planes by estimating the minimum and maximum height from the surrounding area in the heightfield. A small margin is subtracted from the top and bottom of the plane to account for possible occlusions near the roof (protrusion of the eave line) and the ground.

2.3 Line Direction Optimization

The first optimization applied to the facade planes tries to align the orientation of real facades and their hypothesis. As a result the plane will be almost parallel to the real facade. The algorithm relies on the fact that facades mainly contain structures which are horizontally or vertically aligned with the facade itself (windows, balconies, signs and alike).

For each facade plane the algorithm first makes a ranking of all available cameras and assigns each one a score. This score is calculated with the following equation:

$$score = normal \cdot (origin - anchor)$$

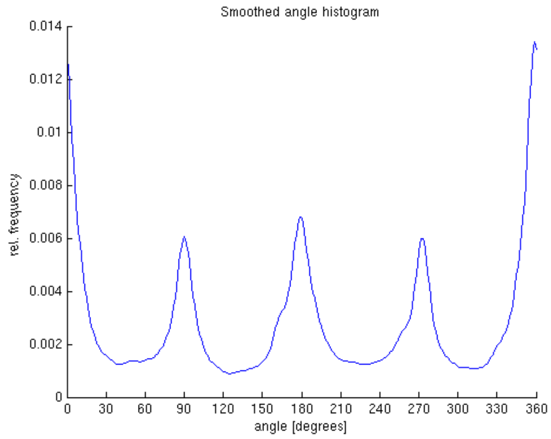


(a) (b) (c) (d)

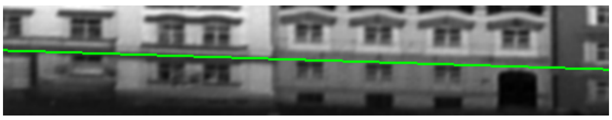
Figure 1: This figure illustrates the line extraction process in the heightfield. (a) shows the original heightfield, (b) depicts the gradient image (Sobel), (c) is the building-layer of the classification for the test area and (d) overlays the extracted lines (green) with the heightfield.

where $normal$ is the normal vector of the facade plane, $origin$ is the position of the camera and $anchor$ is the center of the facade plane.

Once the optimal camera has been determined, the corresponding image is perspectively correctly resampled. A Gaussian filter is then applied to remove small artifacts. For each pixel in the smoothed image the x and y derivative is calculated and stored in a $(\phi, magnitude)$ vector, where ϕ gives the angle of the derivative vector and $magnitude$ its Euclidean length. Subsequently all pairs with a small $magnitude$ are removed. The remaining members of the vector are used to construct an orientation histogram. Each peak in that histogram corresponds to one strong line direction in the texture. This peak estimation is more stable if the histogram is smoothed beforehand. Because of our assumption that a facade contains horizontally and vertically aligned structures, we conclude that the peak closest to zero should in fact be exactly at zero to make the facade plane parallel to the real facade. Figure 2 shows an orientation histogram and the corresponding warped texture. The green line is the estimated principal horizontal line. There are four peaks clearly visible, each accounts for the principal directions (up, down, left, right) of the gradients. To have a parallel facade those four peaks should be at exactly 0, 90, 180 and 270 degrees respectively. The angle histogram enables us to calculate an orientation change which compensates this deviation of the peaks. Figure 3 illustrates this intersection procedure. The detected line direction is used to create a plane which contains the camera center and a line on the facade with this direction. This plane is intersected with a horizontal plane to give the new orientation of the facade estimation.



(a)



(b)

Figure 2: (a) A smoothed orientation histogram with its four distinct peaks in horizontal and vertical direction. (b) shows a part of the corresponding texture with the principal horizontal line direction marked with green.

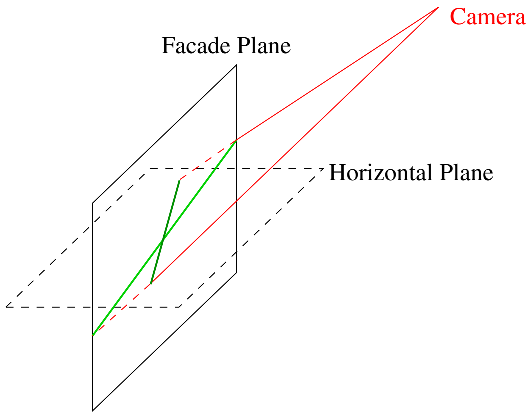


Figure 3: The lines from camera center to the endpoints of the detected line are intersected with the horizontal plane. The new plane defined by this horizontal line is parallel to the real facade.

2.4 Correlation Optimization

In the third and last step the facade plane is further refined to increase the correlation of warped textures from different views. At the beginning the facade plane can not be used to correlate the views at the full resolution level because even an offset of a few pixels may cause a very bad correlation value. Therefore a hierarchical approach is used to overcome this problem. Each warped texture is turned into an image pyramid and starting with the coarsest level the correlation optimization is performed until the highest resolution level is reached. The algorithm is detailed in Algorithm 1. Figure 4 illustrates the process of generating new hypotheses starting with an initial facade plane. The illustration is a top view because it is assumed that facades are always vertical. Figure 6 shows how the optimization on different resolution levels converges to the final position.

The correlation score is calculated using the normalized cross correlation with an adaptive window size depending on the resolution level - on the highest level a smaller window is used as on lower resolution levels. Because of the different resolution the correlation window always covers approximately the same region. Also a correlation truncation (lower boundary) at 0.8 is used to improve the stability of the correlation as explained in (Scharstein and Szeliski, 2002).

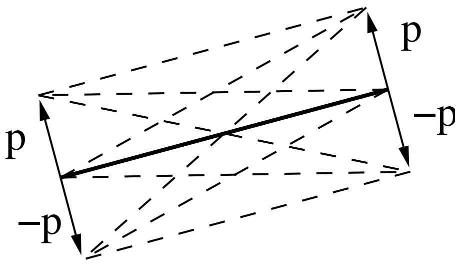


Figure 4: For a given facade plane a translation vector p is calculated which shifts each end of the facade plane and generates therefore eight new hypotheses. New hypotheses are marked with dashed lines.

Algorithm 1 Correlation Optimization

Require: At least two views for a facade

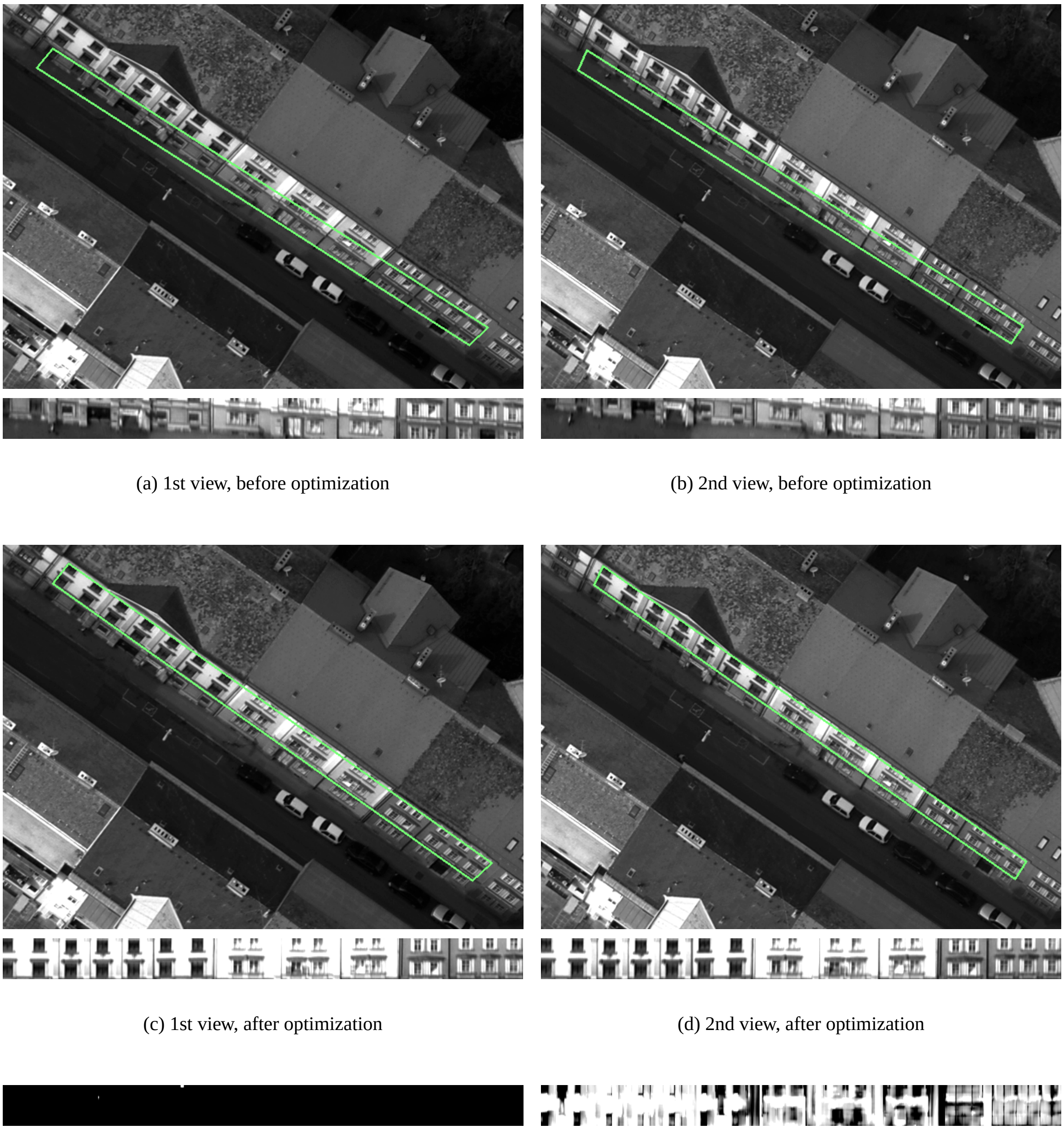
repeat
  calculate a translation vector p normal to the facade plane such that the length of the projection at the current resolution level is approximately one pixel.
  create new hypotheses by moving each end of the facade plane independently back and forth along the translation vector.
  if no higher correlation can be obtained by any hypothesis, switch to a higher resolution level.
until highest resolution level is reached

The quality of the optimization can be judged by the correlation factor. Values of above approximately 0.8 indicate that the estimate snapped to the real facade, whereas lower values may either be due to the fact, that there are occlusions (trees are very disturbing especially in inner courtyards) in the images or that the facade can not be satisfyingly be approximated with one plane because of balconies or depth jumps in the real facade. Figure 5 illustrates an optimization of one facade. Looking at the warped patches one can observe the improvement in positioning the facade.

3 RESULTS AND DISCUSSION

Figure 7 illustrates the result of the optimization on one corner of the building. One can see that the initialization of the facade is in fact the eave line of the roof, whereas the optimization results in the correct position which is slightly translated inwards.

A rendering of the complete building block is depicted in Figure 8. It consists of 21 facades planes and 46 roof planes. The 3D model creation is subject of current research and therefore does not exploit all of the information available. As mentioned in the paragraph above the gap between facade and eave line can be



(a) 1st view, before optimization

(b) 2nd view, before optimization

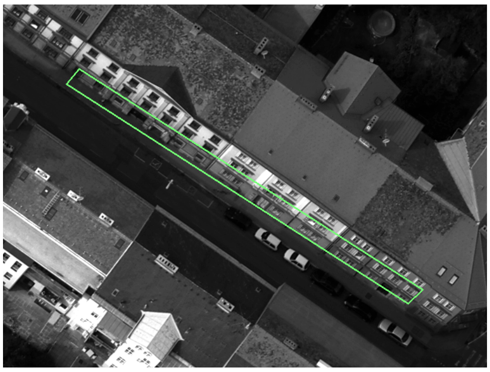
(c) 1st view, after optimization

(d) 2nd view, after optimization

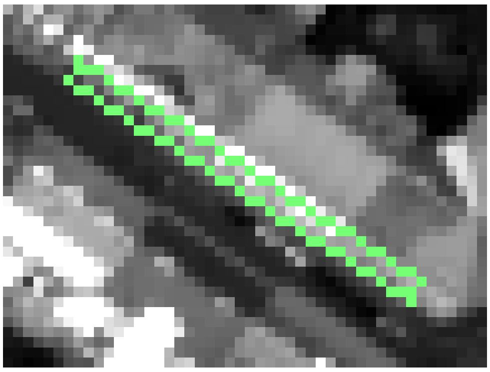
(e) correlation before optimization

(f) correlation after optimization

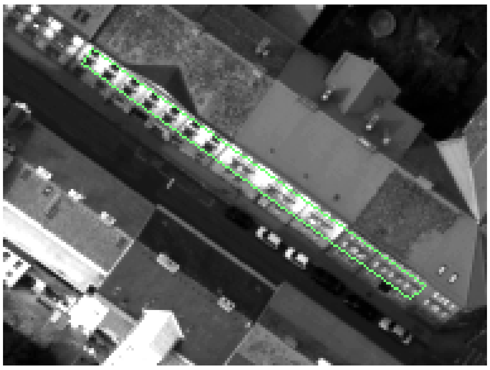
Figure 5: Facade estimation before and after optimization. Two out of three views are shown (left and right). The top two rows represent the initial estimate, the regions marked with the green quadrangle are rectified and shown in the next row. It is clearly visible that the initial estimate deviates from the real facade. After the optimization (third and fourth row) the correct placement can be observed in the rectified images which are nearly identical. This is confirmed by the correlation images (bottom row): the left correlation image shows the correlation for the initial estimate, the right image is calculated after the optimization. The final correlation score is about 0.87.



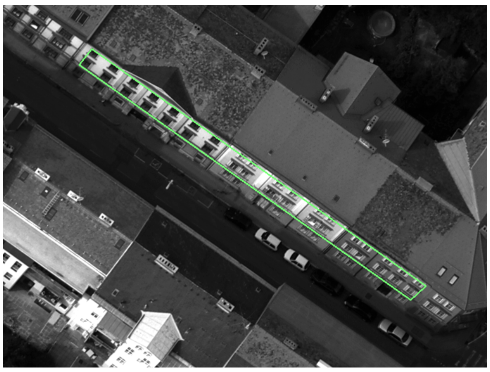
(a)



(b)



(c)



(d)

Figure 6: Four steps in the correlation optimization process: the green lines delineate the estimation after (a) initialization, (b) optimization on the lowest level, (c) medium resolution level and (d) highest resolution level.

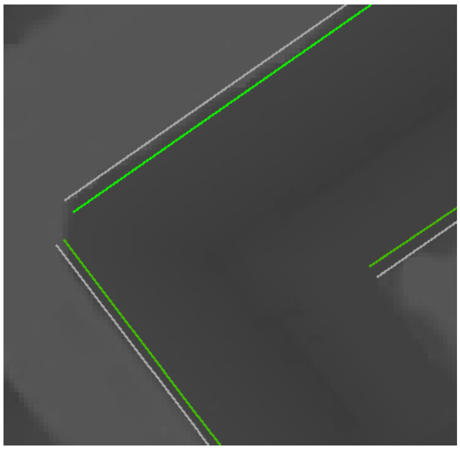


Figure 7: A zoom onto a corner of the building: the gray line denotes the initialization, whereas the green line indicates the position with the optimized correlation. The difference of these positions accounts for the offset between eave line and real facade.

reconstructed (either by comparing the initial estimate and optimized facade or by looking at the correlation image because the correlation will drop where the facade is occluded by the roof) and included in the 3D model. The depicted model lacks this improvement and therefore the roof gets projected onto the facade at the top where in fact the eave line should extend.

4 CONCLUSIONS AND FUTURE WORK

This paper presents a novel approach to improve the location of facade planes using two image-based optimization techniques. The success of such optimizations can easily be judged using the correlation score. The algorithms are outlined and their results are demonstrated using a real world example.

The preliminary results are visually appealing, but further research is required. Especially the exact reconstruction of the offset between eave line and real facade is very promising. The fusion of optimized facade planes, roof planes and offset of the eave lines into a three dimensional model is subject of future research and presents a major step towards fully automated city modelling.

ACKNOWLEDGEMENTS

This work has been done in the VRVis research center, Graz/Austria (http://www.vrvis.at), which is partly funded by the Austrian government research program Kplus. We would also like to thank Vexcel Corporation (http://www.vexcel.com) for supporting this project.

REFERENCES

Berthod, M., Gabet, L., Giraudon, G. and Lotti, J., 1995. High resolution stereo for the detection of buildings. In: A. Grun, O. Kubler and P. Agouris (eds), Automatic Extraction of Man-Made Objects from Aerial and Space Images, Birkhäuser, pp. 135–144.

Brenner, C., 2000. Towards fully automatic generation of city models. In: International Archives of Photogrammetry and Remote Sensing, Commission III, Vol. 33, pp. 85–92.

C. Vestri, F. D., 2000. Improving correlation-based dems by image warping and facade correlation. In: In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), p. 1438 ff.

Figure 8: A 3D rendering of one building with optimized facades.

Coorg, S. and Teller, S., 1999. Extracting textured vertical facades from controlled close-range imagery. In: In Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 625–632.

Cord, M., Paparoditis, N. and Jordan, M., 1998. Dense, reliable, and depth discontinuity preserving dem computation from very high resolution urban stereopairs. In: ISPRS Symposium, Cambridge (England).

Gruber-Geymayer, B. C., Klaus, A. and Karner, K., 2005. Data fusion for classification and object extraction. In: Proceedings of CMRT05, Joint Workshop of ISPRS and DAGM, pp. 125–130.

H. Gross, U. Thoennessen, W. v. H., 2005. 3d-modeling of urban structures. In: Proceedings of the ISPRS Workshop CMRT 2005, pp. 137–142.

Haala, N., Brenner, C. and Statter, C., 1998. An integrated system for urban model generation. In: ISPRS Commission II Symposium, Cambridge, England.

Samadzadegan, F., Azizi, A., Hahn, M. and Lucas, C., 2005. Automatic 3d object recognition and reconstruction based on neuro-fuzzy modelling. In: ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 59, pp. 255–277.

Scharstein, D. and Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: International Journal of Computer Vision, Vol. 47, pp. 7–42.

T. Werner, A. Z., 2002. New technique for automated architectural reconstruction from photographs. In: In Proceedings of the European Conference on Computer Vision (ECCV), pp. 541–555.

Vosselman, G., 1999. Building reconstruction using planar faces in very high density height data. In: Proceedings of the ISPRS Automatic Extraction of GIS Objects from Digital Imagery, pp. 87–92.

Vosselman, G. and Dijkman, S., 2001. 3d building model reconstruction from point clouds and ground plans. In: International Archives of Photogrammetry and Remote Sensing, Vol. 34, pp. 37–43.

Weidner, U., 1996. An approach to building extraction from digital surface models. In: Proceedings of the 18th ISPRS Congress, Commission III, pp. 924–929.

Zach, C., Klaus, A. and Karner, K., 2003. Accurate dense stereo reconstruction using 3d graphics hardware. Eurographics 2003 pp. 227–234.