Photogrammetrie • Fernerkundung • Geoinformation 3/2008, S. 187–196, 14 figs.

Implicit Shape Models, Self-Diagnosis, and Model Selection for 3D Facade Interpretation

SERGEJ REZNIK & HELMUT MAYER, Neubiberg

Keywords: Facade Interpretation, Implicit Shape Models, Self-diagnosis, Model Selection, Plane Sweeping, Markov Chain Monte Carlo

Summary: This paper addresses the automatic 3D interpretation of facades from terrestrial image sequences making three novel contributions: First, we employ Implicit Shape Models (LEIBE & SCHIELE 2004) coherently for the detection as well as for the delineation of windows, learning the appearance of windows and their outline from training data. Second, window hypotheses are validated by means of self-diagnosis based on the assumption of a possibly strong similarity of individual windows. Third, we use model selection to choose the most appropriate model for the configuration of windows in terms of rows or columns. These contributions are complemented by plane sweeping for the 3D determination of the windows or the rows / columns made up from them. Results show the potential of the approach.

Zusammenfassung: *Implicit Shape Models, Selbstdiagnose und Modellauswahl für die 3D Interpretation von Fassaden*. Dieser Artikel zielt mit drei neuen Beiträgen auf die automatische Interpretation von Fassaden aus terrestrischen Bildsequenzen: Erstens werden Implicit Shape Models (LEIBE & SCHIELE 2004) kohärent sowohl für die Detektion als auch für die Bestimmung der Umrisse von Fenstern verwendet. Das Aussehen der Fenster und ihre Umrisse werden aus Trainingsdaten gelernt. Zweitens werden Fensterhypothesen mittels Selbst-Diagnose auf Grundlage der Annahme einer z.T. starken Ähnlichkeit individueller Fenster validiert. Drittens wird Modellauswahl genutzt, um das am besten geeignete Modell für die Konfiguration der Fenster in Form von Zeilen oder Spalten auszuwählen. Diese Beiträge werden durch Plane Sweeping für die 3D Bestimmung der Fenster oder der aus ihnen gebildeten Zeilen oder Spalten ergänzt. Die Ergebnisse zeigen das Potential des Ansatzes.

1 Introduction

The inclusion of structured facades extends the modelling of buildings towards highly detailed visualizations suitable for applications ranging from architectural planning to the production of movies. By interpreting the parts constituting facades in terms of their semantics it becomes possible to interact with them, e. g., making it feasible to open windows or doors.

The interpretation of facades from terrestrial images and wide-baseline image sequences has been a focus of research since the seminal work of DICK et al. (2004). They interpret buildings in line with the trend in computer vision towards statistical generative models. Particularly, they employ Reversible Jump Markov Chain Monte Carlo – RJMCMC (GREEN 1995) allowing for the addition and deletion of new parameters and therefore also objects. The results are convincing though restricted to a limited number of objects as the models are complex and generated manually. A more geometric approach is taken by WERNER & ZISSERMAN (2002). They make use of the regular structure of buildings, especially the existence of orthogonal vanishing points. Geometric regularities such as symmetries of dormer

1432-8364/08/2008/187 $ 2.50

windows are used to obtain a high-quality textured model. Becker & Haala (2007) show that by combining laser and image data with rectangular cell decomposition, realistic 3D interpretations of facades can be generated. Müller et al. (2007) and Van Gool et al. (2007) present impressive results for facade interpretation from single images exploiting common repetitions of windows and balconies by means of architectural shape grammars. They particularly show how depth layering can be performed automatically if substantial perspective effects exist in an image.

Our first contribution of this paper lies in employing Implicit Shape Models – ISM (Leibe & Schiele 2004) coherently for the appearance based detection as well as for the delineation of windows. While we used information of corners to delineate windows only on dark facades and employed black rectangles for bright facades in (Mayer & Reznik 2006), we now delineate the outline of whole windows on any kind of facade via ISM.

Our second contribution has been inspired by (Hinz & Wiedemann 2004). The basic idea is to validate weak hypotheses based on self-diagnosis of the generated hypotheses making use of the fact that windows on a facade look often very similar.

The third contribution can be seen as an inversion and at the same time extension of (Alegre & Dallaert 2004, Brenner & Ripperda 2006, and Ripperda & Brenner 2007). We invert, as we do not split the facade, but rather detect and delineate objects and group the constituents into rows and columns. We extend the above work as we employ model selection based on Akaike's Information Criterion (AIC) to compare different groupings. Basically, individual windows always lead to the best likelihood as they can adapt to the individual shapes of windows. Only by taking into account the lower number of parameters for rows, columns, etc., they will prevail. A particular contribution is to show how the likelihood term has to be interpreted in terms of the (minimum) size to be sampled to obtain meaningful results. (Dick et al. 2004) have also used model selection, but to switch between different interpretations for windows, namely with and without arc, etc. In this paper also first results for facades with different distances between windows for different parts of the facade are presented.

We assume, that a wide-baseline image sequence is given, and employ given (approximate) calibration information via the five-point algorithm (Nistér 2004), which makes the reconstruction much more stable. 3D Reconstruction leads to camera parameters and 3D points. From the latter we compute the facade planes via Random Sample Consensus – RANSAC (Fischler & Bolles 1981). We orient the planes using the vertical vanishing points in the images, again employing RANSAC. All images looking at a particular facade are projected on its plane and combined using a consensus-based approach (Mayer 2007) getting rid of partial occlusions. We use a manually defined sampling distance of 1 cm to normalize the further processing. Thus, for the remainder of the article all facade plane images are assumed to be vertically oriented and normalized to a resolution of 1 cm.

We first describe the appearance based detection and delineation of windows on the facade plane images based on ISM in Section 2. Section 3 is devoted to self-diagnosis for the validation of window hypotheses, while Section 4 deals with model selection for the decision between representations based on individual or rows or columns of windows. Plane sweeping leading to the determination of the depth, i. e., the 3D shape of windows, is described in Section 5. The paper ends with conclusions.

2 Detection and Delineation of Windows Based on Implicit Shape Models

We employ Implicit Shape Models – ISM (Leibe & Schiele 2004) for the detection of windows and for the delineation of their outline. For training we cut out image patches containing windows, in our case 120 windows of modern type. We note that none of the windows shown in our results is part

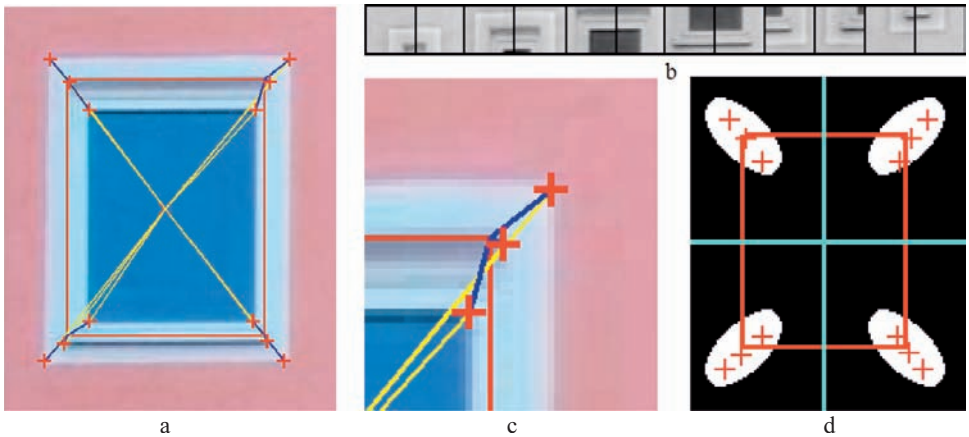

Fig. 1: Training – a – Image patch with manually given outline of window (red rectangle), Förstner points at corners of window outline (red crosses) as well as their vectors to the center of the window (yellow lines) and to their corresponding corner of the outline (blue short lines); b – Image patches around Förstner points; c – Detail of a) focusing on the relation of Förstner points to the corner of the window; d – Elliptical areas around window corners (white) where Förstner points are extracted.

of the training set and that we use the patches as well as their horizontally mirrored versions, making the algorithm more invariant to the viewing direction. The rectangular outlines of the windows are manually delineated (cf. red rectangle in Fig. 1a). Only in elliptical areas around the corners of the outline with radii 20 and 10 pixels / cm for the major and the minor axis (cf. Fig. 1d) Förstner points (FÖRSTNER & GÜLCH 1987) are extracted. The image patches around the Förstner points shown in Fig. 1b are the basis for the appearance based detection of windows together with their arrangement relative to the center of the window computed from the manually delineated outline marked as yellow lines in Fig. 1a.

For the retrieval, i. e., for window detection, Förstner points are extracted with the same parameters as for training, but in the whole image (cf. Fig. 2a). Patches around the points with a size of 35 pixels are matched via cross correlation of the images trans-

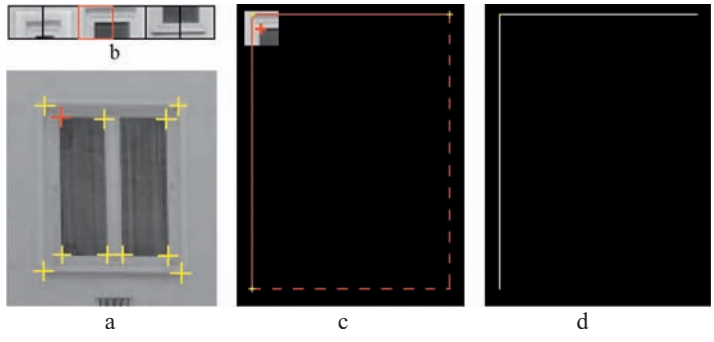

Fig. 2: Retrieval – a – Förstner points; b – Training patches with the patch just left above the ''b'' being matched to the red cross at the upper left corner of the window pane in a; c – Relation of the center of the patch (red cross) to the window outline in the training data (left cross for position – lengths of sides from training data); d – Hypothesis for parts of the window outline.

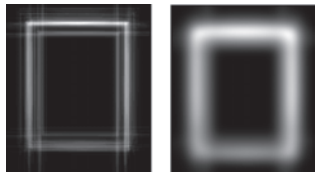

Fig. 3: Distribution for window outline – after accumulation (left) and after smoothing (right).

formed to grayscale to all patches in the training data. If the cross correlation coefficient is above an empirically found threshold of 0.75, the match is accepted and the vector relating the training patch to its center is used to generate a hypothesis for the center of the window in an initially empty accumulation image. The hypotheses are integrated via a Gaussian of the average size of the windows used for training and local maxima of the resulting function are hypotheses for windows. The patches which led to the maxima are employed to delineate the corners of the window outlines.

To precisely delineate the windows, we employ the relation between the centers of the training patches and the given outline of the windows marked as blue lines in Fig. 1a and c. E.g., the point marked in red in the upper left corner of the dark window pane in Fig. 2a has been matched by cross correlation to the training patch marked in red just left above the "b" of Fig. 2b. Fig. 2c shows how the center of the patch marked by a thick red cross is related to the corner of the outline of the window marked by a small yellow cross. From the corner of the outline the two neighboring sides of the rectangle from the training data are drawn (cf. Fig. 2c and d). The result is a hypothesis for parts of the window outline.

The hypotheses for window outlines, as, e. g., Fig. 2d, are accumulated over all points and all training patches that led to the maximum for the window. The result is a distribution for the window outline as in Fig. 3 left which is finally smoothed (cf. Fig. 3 right) and normalized by setting the largest value in the window to 1.

The parameters of the rectangles representing the windows are estimated from the distributions for the window outlines interpreted as likelihood and priors for the window shapes by Markov Chain Monte Carlo – MCMC (Neal 1993) Maximum A Posteriori (MAP) estimation. The employed priors punish too small and too wide or too high windows. The likelihood is the sum over the distribution along the window outline (cf. red lines in Fig. 4b).

3 Self-Diagnosis for the Validation of Window Hypotheses

The proposed algorithm works only well for high quality images and simple facades. If this is not the case, it might detect and delineate false hypotheses, e. g., doors or other rectangular objects. The algorithm also does not deal well with partially occluded windows.

The solution we have devised to cope with the above shortcomings is to use good hypotheses in order to validate weaker hypotheses. The basic assumption is that at least some windows on a facade are of the

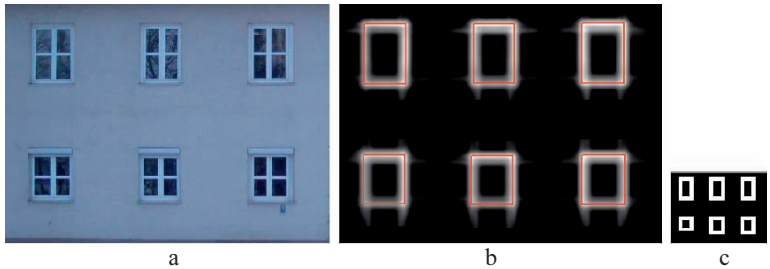

Fig. 4: a – Facade; b – Determination of the likelihood in the distribution for the window outlines (red); c – Minimal size for model selection according to sampling theorem (cf. Section 4).

same type. The validation of hypotheses works as follows: Image patches containing good hypotheses for windows are cross-correlated with the image function around weak hypotheses, determining the optimal location as the maximum. For describing the quality of the hypotheses we use a grade system empirically evaluating all cases with 1, 2, 3, and 4 recognized corners (cf. Fig. 5) based on the number of corners as well as their relation to the outline. The higher the grade of a hypothesis, the better it is evaluated. Based on the grade system we analyze and validate all hypotheses. Results are given in Figs. 6 and 7.

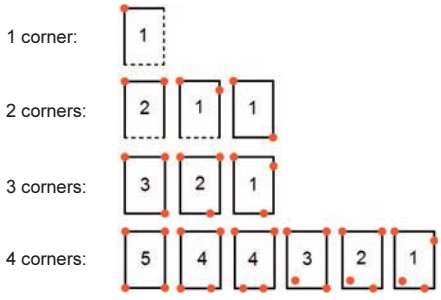

Fig. 5: The grade system for hypotheses – higher grade means better evaluation. Please note that for symmetric configurations only one instance is given.

4 Model Selection: Individual Windows, Rows, and Columns

In the preceding sections we have described how to detect, delineate, and validate individual windows such as in Fig. 8a. Yet, windows are usually not arranged randomly, but in rows, columns, or grids. Rows and columns, in this paper defined to have the same horizontal or vertical distance between windows of the same size, can be built by analyzing the horizontal or vertical arrangement. Yet, it is often not clear if one should

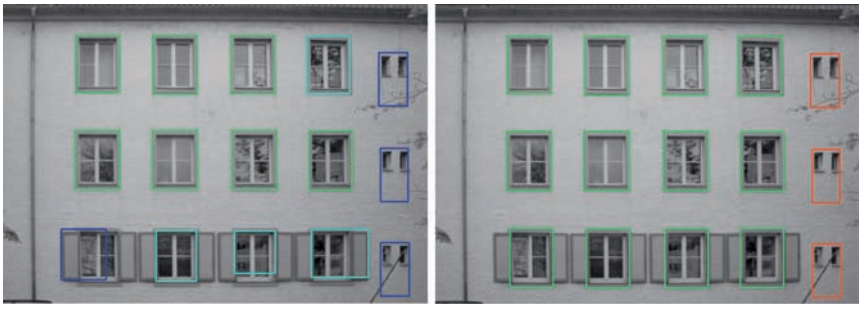

Fig. 6: Validation of hypotheses – left: Before verification. Hypotheses with grade 5 as green, with grade 4 as yellow, with grade 3 as cyan, and with grade 2 as blue rectangles; right: After verification: green – accepted, red – rejected hypotheses.

Fig. 7: Validation of hypotheses – left: Before verification; right: After verification (for colors cf. Fig. 6).

represent a facade by means of individual windows or by rows or columns of windows. E. g., Fig. 8 shows a configuration which can be represented adequately by means of columns, but not rows. Basically, in terms of an optimum fit described in the form of the likelihood always individual windows will be preferred as they can optimally adapt to the data. Thus, one needs a way to reward regular arrangements of objects and one way to do this is to take into account that they can be described by smaller numbers of parameters.

The above problem is thus regarded as a problem of model selection. Numerous means have been devised to balance the complexity of a model, e. g., described by the number of parameters or their accuracy, on one hand and the fit to the data, i. e., the likelihood, on the other hand. Two well known are Minimum Description Length – MDL (Rissanen 1978) and AIC – Akaike's information criterion (Akaike 1973). A very good analysis of the relations of these two means as well as their characteristics, their strengths, and weaknesses can be found in (Schindler & Suter 2006). For its simplicity and as we found it to work well for our application, we employ AIC, though recent work on composition such as (Geman et al. 2002) prefers MDL. Particularly, we use

$$\mathrm{AIC} = \mathrm{k} - 2\mathrm{n}\ln(\mathrm{L}) \qquad (1)$$

with k the number of the parameters of the model, n the number of observations, and L the likelihood of the outline. The number of parameters is four (width, height and center coordinates) for every individual window and six for a row or column (four parameters for window shape plus – horizontal or vertical – spacing and number of elements). The basic idea is to determine the posterior based on the normalized distribution image by means of MCMC as described in Section 2 above. Fig. 4b shows how the distribution is sampled at one position with the outline given in red. Every boundary point gives one observation of the likelihood which is multiplied leading to the multiplication factor for the log-likelihood.

Yet, a couple of experiments made clear that one cannot just sample the given distribution for windows. We found that one has to reduce the determination of the likelihood to a minimal setup. From the sampling theorem we derived that for a window consisting of parallel lines the minimum size is a length of just above three pixels. We accordingly resample the distribution image to this minimum size (cf. Fig. 4c) for the computation of the likelihood for AIC. (Note: For the delineation the original resolution is used to obtain a higher accuracy.)

Results for this procedure are given in Fig. 9. For all three facades consisting of windows with the same size and a constant horizontal or vertical spacing as well as many other facades we tested our procedure on we selected the correct model. If there is an obvious structure on the facade, it is reflected in significantly different AIC values as shown in Fig. 9.

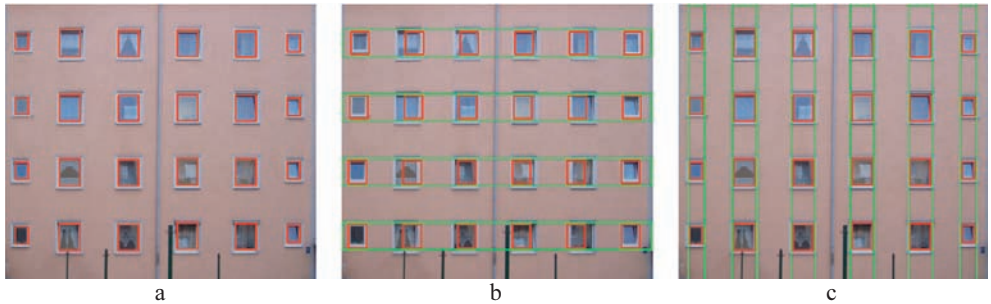

Fig. 8: Model Selection – Representation of facade by a – individual windows; b – rows; c – columns of windows, the latter two consisting of windows with the same size and a constant horizontal or vertical spacing.

a – **C: 265**, R: 404, W: 359 b – C: 115, **R: 65**, W: 94 c – **C: 76**, R: 83, W: 87

Fig. 9: Results for model selection using AIC values – C: Columns, R: Rows, W: Individual Windows. Selected model in bold.

Fig. 10: Results for model selection extending randomly chosen neighbored pairs of windows.

Up to this point we have restricted ourselves to completely evenly spaced rows or columns of windows. To deal with configurations such as in Fig. 10, where the horizontal distances between the windows partly vary, we sample the rows or columns by extending randomly chosen neighbored pairs by other neighboring windows via MCMC until no further window is found anymore which can be linked. Then the next pair is selected, etc. Several start configurations are chosen again randomly and finally the configuration yielding the smallest AIC value is selected. For Fig. 10 it consists of 54 instead of 108 parameters.

5 3D Reconstruction via Plane Sweeping and Results

The results from the above procedure are the outlines of windows on the facade images possibly restricted to form horizontal rows or vertical columns. As we use image sequences as basis, we can determine the 3D extent of the windows. To do so, we follow (Baillard & Zisserman 1999 and Werner & Zisserman 2002) and employ plane sweeping, in this case for planes parallel to their facade plane in the direction of the latter's normal. The determination of the depth for individual windows is based on the sum of the least-squares differences between the projections of the individual images onto the

Fig. 11: Images one, three, five, and seven of sequence Ostbahnhof-1.

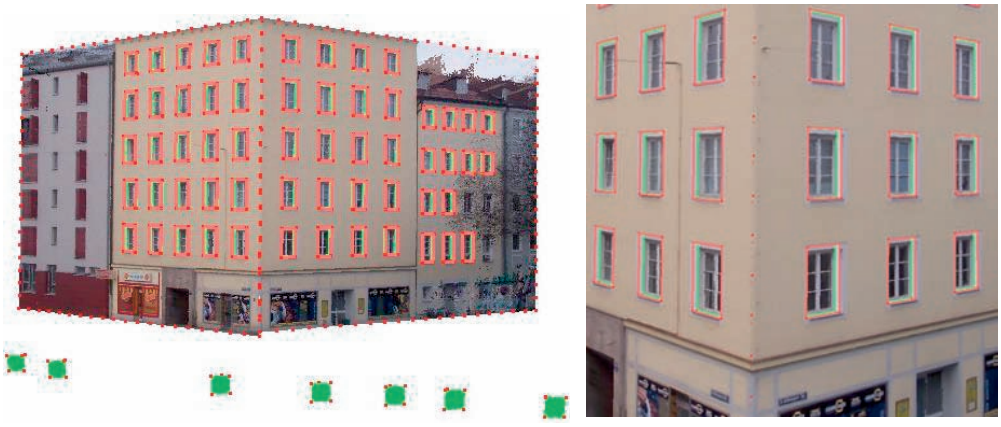

Fig. 12: left: Result for sequence Ostbahnhof-1 (images cf. Fig. 11) – Window outlines for three facades with rows of windows as red rectangles, 3D window positions as green rectangles, and camera positions as green pyramids; right: Detail: part of two walls.

Fig. 13: Result for sequence Bordeaux Square with individual windows constructed from eleven images – explanation cf. Fig. 12.

Fig 14: Result for sequence Ostbahnhof-2 with columns of windows constructed from ten images – explanation cf. Figure 12.

plane to their average image. This is computed for a meaningful range of depth values for windows and the result is the depth value for the minimum of the sum. For rows or columns we sum up the contributions of all images of a row or column at a particular depth.

Fig. 11 shows four images of a sequence with seven images and Fig. 12 the result for three manually coarsely marked facades. Please note that the rows and columns presented in this section consist of windows with the same shape and a constant distance in

either horizontal or vertical direction and we do the selection for the whole facade. The 3D reconstruction was done mostly reliably and accurately and led to the windows behind the facade marked by green rectangles which can be seen in Fig. 12, right. Further results are given in Fig. 13 and 14.

6 Conclusions

We have presented three novel contributions for the interpretation of facades consisting of individual windows, i. e., no glass facades, from terrestrial image sequences, namely the coherent use of Implicit Shape Models for the delineation of windows, self-diagnosis to validate hypotheses for windows, and model selection based on Akaike's information criterion (AIC) for selecting between individual windows and rows and columns constructed from them. Combined with plane sweeping we obtain 3D interpretations of facade planes including the windows.

Concerning future work we think into different directions. First, we need to do model selection for individual rows and columns in a more flexible way by using RJMCMC and use a hierarchical model such as the architectural shape grammars of Müller et al. (2007). Then, we want to create more detailed models of the windows including mullions and transoms, the appearance of both possibly learned in an appearance based hierarchy.

On a more global level we want to integrate other objects such as doors on the ground level but also architectural details around windows possibly including their 3D structure as well as balconies. For the latter plane sweeping might be a solution for some shapes of balconies. We consider Composition Systems (Geman et al. 2002) as an important theoretically sound basis for our hierarchical modeling ranging from the window details to grids made up of windows and other architectural objects. Finally, a statistically sound link between discriminative and generative modeling such as in (Tu et al. 2005) could be advantageous.

Acknowledgment

We thank Deutsche Forschungsgemeinschaft for funding Sergej Reznik under grant MA 1651/10.

References

Akaike, H., 1973: Information Theory and an Extension of the Maximum Likelihood Principle. – Second International Symposium on Information Theory: 267–281.

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

Baillard, C. & Zisserman, A., 1999: Automatic Reconstruction of Piecewise Planar Models from Multiple Views. – Computer Vision and Pattern Recognition **II:** 559–565.

Becker, S. & Haala, N., 2007: Refinement of Building Facades by Integrated Processing of Lidar and Image Data. – International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **36** (3/W49A): 7–12.

Brenner, C. & Ripperda, N., 2006: Extraction of Facades Using RJMCMC and Constraint Equations. – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **36** (3): 155–160.

Dick, A., Torr, P. & Cipolla, R., 2004: Modelling and Interpretation of Architecture from Several Images. – International Journal of Computer Vision **60** (2): 111–134.

Fischler, M. & Bolles, R., 1981: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. – Communications of the ACM **24** (6): 381–395.

Förstner, W. & Gülch, E., 1987: A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features. – ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland: 281–305.

Geman, S., Potter, D. & Chi, Z., 2002: Composition Systems. – Quarterly of Applied Mathematics **LX:** 707–736.

Green, P., 1995: Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. – Biometrika **82:** 711–732.

Hinz, S. & Wiedemann, C., 2004: Increasing Efficiency of Road Extraction by Self-Diagnosis.

– Photogrammetric Engineering & Remote Sensing **70** (12): 1457–1466.

Leibe, B., Leonardis, A. & Schiele, B., 2004: Combined Object Categorization and Segmentation with an Implicit Shape Model. – ECCV'04 Workshop on Statistical Learning in Computer Vision: 17–32.

Mayer, H., 2007: 3D Reconstruction and Visualization of Urban Scenes from Uncalibrated Wide-Baseline Image Sequences. – Photogrammetrie – Fernerkundung – Geoinformation **3/07:** 167–176.

Mayer, H. & Reznik, S., 2006: MCMC Linked with Implicit Shape Models, and Plane Sweeping for 3D Building Facade Interpretation in Image Sequences. – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences **36** (3): 130–135.

Müller, P., Zeng, G., Wonka, P. & Van Gool, L., 2007: Image Based Procedural Modeling of Facades. – SIGGRAPH / ACM Transactions on Graphics **26** (3): Article 85, 1–9.

Neal, R., 1993: Probabilistic Inference Using Markov Chain Monte Carlo Methods. – Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto.

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.

Reznik, S. & Mayer, H., 2007: Implicit Shape Models, Model Selection, and Plane Sweeping for 3D Facade Interpretation. – International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **36** (3/W49A): 173–178.

Ripperda, N. & Brenner, C., 2007: Data Driven Rule Proposal for Grammar Based Facade Reconstruction. – International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **36** (3/W49A): 1–6.

Rissanen, J., 1978: Modeling by Shortest Data Description. – Automatica **14:** 465–471.

Schindler, K. & Suter, D., 2006: Two-View Multibody Structure-and-Motion with Outliers Through Model Selection. – IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (6): 983–995.

Tu, Z., Chen, X., Yuille, A. & Zhu, S.-C., 2005: Image Parsing: Unifying Segmentation Detection and Recognition. – International Journal of Computer Vision **63** (2): 113–140.

Van Gool, L., Zeng, G., Van den Borre, F. & Müller, P., 2007: Towards Mass-Produced Building Models. – International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **36** (3/W49A): 209–220.

Werner, T. & Zisserman, A., 2002: New Techniques for Automated Architectural Reconstruction from Photographs. – Seventh European Conference on Computer Vision **II:** 541–555.

Address of the Authors:

Dipl.-Phys. Sergej Reznik, Prof. Dr.-Ing. Helmut Mayer, Universität der Bundeswehr München, Institut für Photogrammetrie und Kartographie, D-85577 Neubiberg, Tel.: +49-89-6004-3429 (Mayer), Fax: +49-89-6004-4090, e-mail: Sergiy.Reznik@unibw.de, Helmut.Mayer@unibw.de.

Manuskript eingereicht: Dezember 2007
Angenommen: März 2008