IMPLICIT SHAPE MODELS, MODEL SELECTION, AND PLANE SWEEPING FOR 3D FACADE INTERPRETATION

Sergej Reznik, Helmut Mayer

Institute of Photogrammetry and Cartography, Bundeswehr University Munich, Germany
Sergiy.Reznik|Helmut.Mayer@unibw.de

KEY WORDS: Facade Interpretation, Implicit Shape Models, Model Selection, Plane Sweeping, Markov Chain Monte Carlo

ABSTRACT:

In this paper we address the automatic 3D interpretation of facades from terrestrial image sequences making two novel contributions: First, we employ Implicit Shape Models (Leibe and Schiele, 2004) coherently for the detection as well as for the delineation of windows, allowing to learn the appearance of windows and their outline from training data. Second, we use model selection to choose the most appropriate model for the configuration of windows in terms of rows or columns. These components are complemented by plane sweeping for the 3D determination of the windows or rows / columns made up from them. Results show the feasibility of the approach.

1 INTRODUCTION

Facade interpretation from terrestrial wide-baseline image sequences has been a focus of research since the seminal paper of (Dick et al., 2004). They interpreted 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 to add and delete new parameters and therefore also objects. The results are impressive though restricted to a limited number of objects as the models are generated manually. A more geometric approach is taken by (Werner and Zisserman, 2002). They make use of the regular structure of buildings, particularly the existence of vanishing points. Specific geometric regularities such as the symmetries of dormer windows are used to obtain a high-quality textured model. Yet, the existence of these regularities is presumed to be known.

Our first main contribution of this paper lies in employing Implicit Shape Models – ISM (Leibe and 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 and Reznik, 2006), we now delineate the outline of whole windows on any kind of facade via ISM.

The second main contribution can be seen as an inversion and at the same time extension of (Alegre and Dallaert, 2004) and (Brenner and Ripperda, 2006). We invert, as we do not split the facade, but rather detect and delineate objects and group the constituents into rows, columns, and finally also grids. 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. One particular contribution is to show how the likelihood term has to be interpreted to come up with meaningful results for our delineation of windows based on ISM. (Dick et al., 2004) have also used model selection, but to switch between different interpretation for windows, namely with and without an arc, etc.

We assume, that a wide-baseline image sequence is given, and employ (Nistér, 2004), which makes the reconstruction much more stable by additionally presuming that an (approximate) calibration is available. 3D Reconstruction leads to camera parameters and 3D points. From the latter we compute the facade planes via Random Sample Consensus – RANSAC (Fischler and 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) allowing to get rid of partial occlusions. We use a sampling distance of 1 cm to normalize the further processing.

We first describe appearance based detection and delineation of windows on the facade plane images based on ISM in Section 2. Section 3 is devoted to model selection for the decision between a representation based on individual windows 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 4. The paper ends with conclusions.

2 DETECTION AND DELINEATION OF WINDOWS BASED ON IMPLICIT SHAPE MODELS

We employ Implicit Shape Models – ISM (Leibe and Schiele, 2004) for the detection of windows, but also 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 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 Figure 1 a)). Only in elliptical areas around the corners of the outline (cf. Figure 1 d)) Förstner points (Förstner and Gülch, 1987) are extracted. The image patches around the Förstner points shown in Figure 1 b) 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 Figure 1 a). For the delineation, the relation of the patches to the corners of the outline is used marked as blue lines in Figures 1 a) and c).

For the retrieval, i.e., for the detection of the windows, Förstner points are extracted with the same parameters as for training, but in the whole image (cf. Figure 4 a)). Patches around the points with a size of 35 pixels are then matched via cross correlation to all patches in the training data. If the cross correlation coefficient is above an empirically found threshold of 0.75, the match
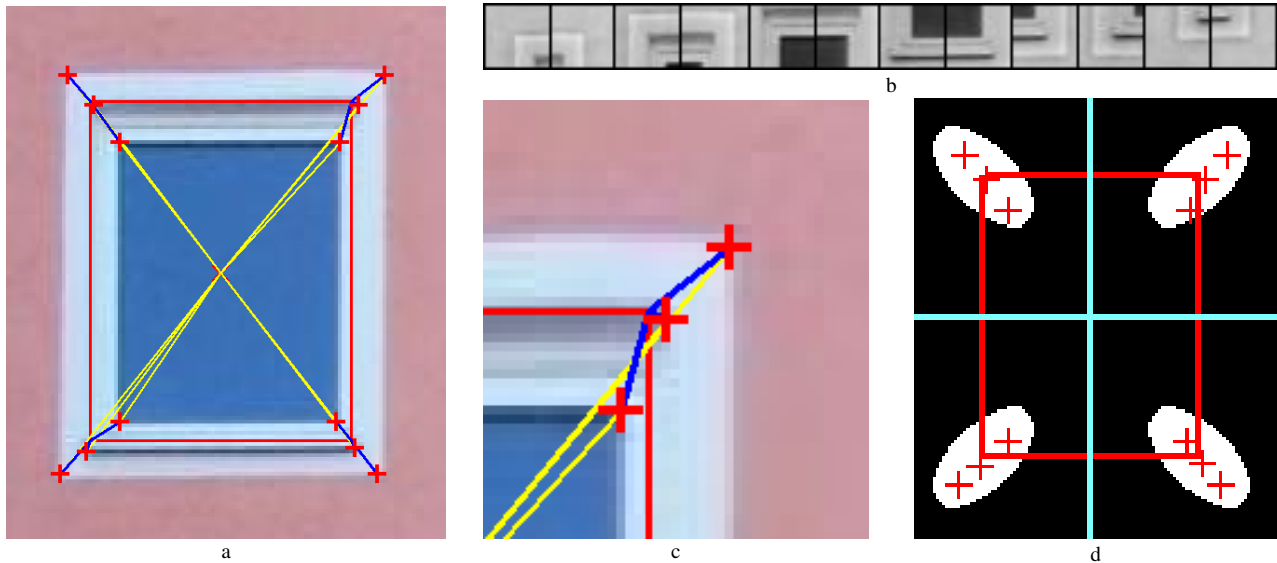
173

Figure 1: Training – a) Window with manually given outline of window (red rectangle), Förstner points at corners of window outline (red crosses) as well as their relation vector 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 outline (white) where Förstner points are extracted.

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 regarded as hypotheses for windows. The patches which led to the maxima are hypotheses for corners of 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 Figure 1 a) and c). E.g., the point marked in red in the upper left corner of the dark window pane in Figure 4 a) has been matched by cross correlation to the training patch marked in red just left above the "b" of Figure 4 b). Figure 4 c) 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. Figures 4 c) and d)). The result is a hypothesis for parts of the window outline.

The hypotheses for window outlines as in Figure 4 d) are accumulated over all points in the given image and all training patches. The result is a distribution for the window outline as in Figure 2 a) which is finally smoothed (cf. Figure 2 b)) and normalized by setting the largest value in the window to one.

Figures 3 and 5 give two results for distributions of window outlines. The distributions for the window outline are input to a Markov Chain Monte Carlo – MCMC (Neal, 1993) Maximum A Posteriori (MAP) estimation procedure. The employed prior punishes too small and too wide or too high windows. The likelihood function is the sum over the distribution along the window outline (e.g., cf. red line in Figure 6 a)).

3   MODEL SELECTION: INDIVIDUAL WINDOWS, ROWS, AND COLUMNS

In the preceding Section we have described how to detect and delineate individual windows such as in Figure 7 a). Yet, windows are usually not arranged randomly, but in rows, columns, or grids.
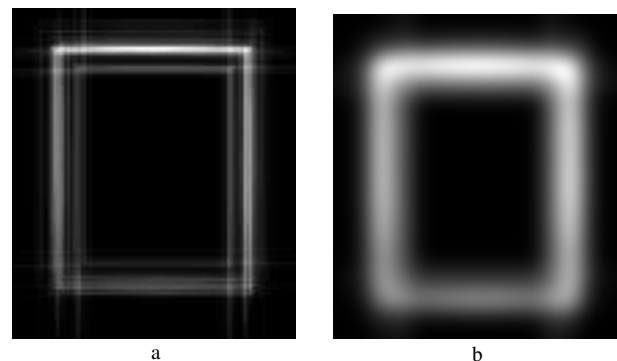

Figure 2: Distribution for window outline – a) accumulation; b) smoothing
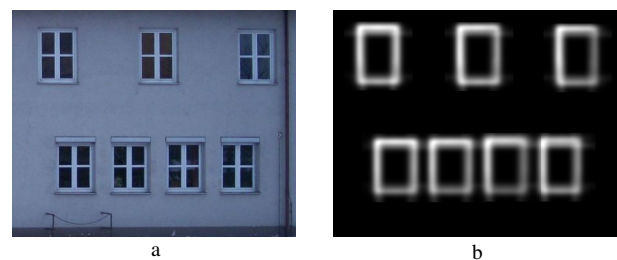

Figure 3: a) Facade and b) distribution for window outlines

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 represent a facade by means of individual windows or by rows or columns of windows. E.g., Figure 7 shows a configuration which can be represented adequately by means of columns, but not in terms of rows. Basically, in terms of an optimum fit described in the form of the likelihood always the individual windows will be preferred as they can optimally adapt to the data. Thus, one needs a way to reward arrangements of objects and one way to do this is to consider that they can be described by smaller number of parameters.
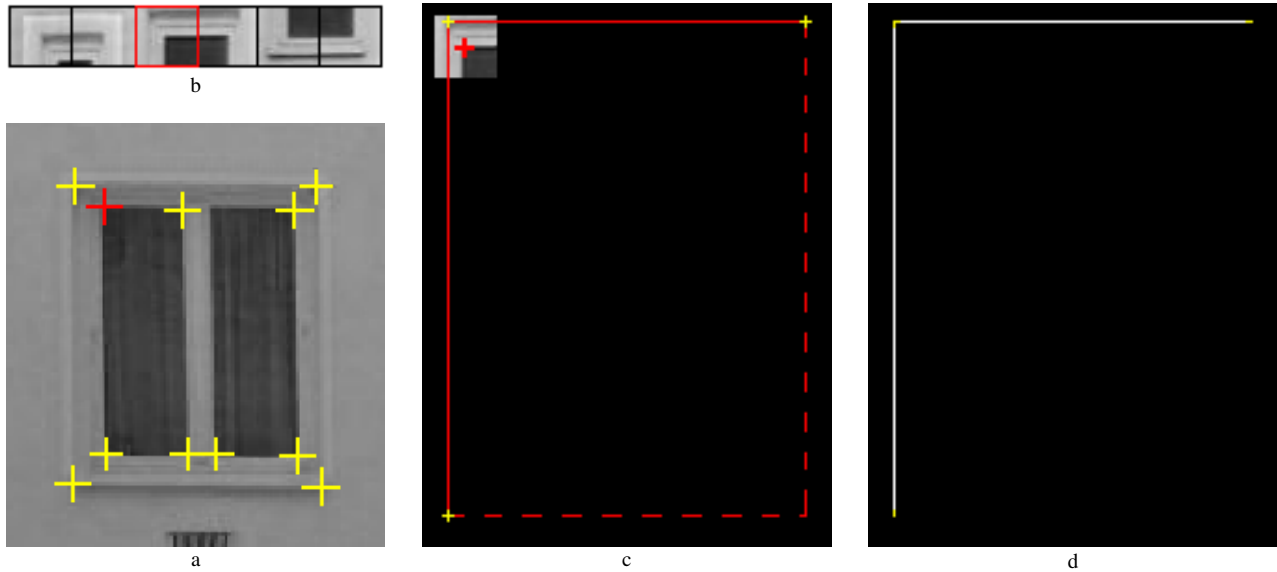
Figure 4: Retrieval – a) Förstner points; b) Training patches with the patch just left above the "b" being matched to the red cross at the upper left corner of the window pane in a); c) Relation of the center of the patch (red cross) to the window outline in the training data (left cross for position – lengths of sides from training data); d) Hypothesis for parts of the window outline
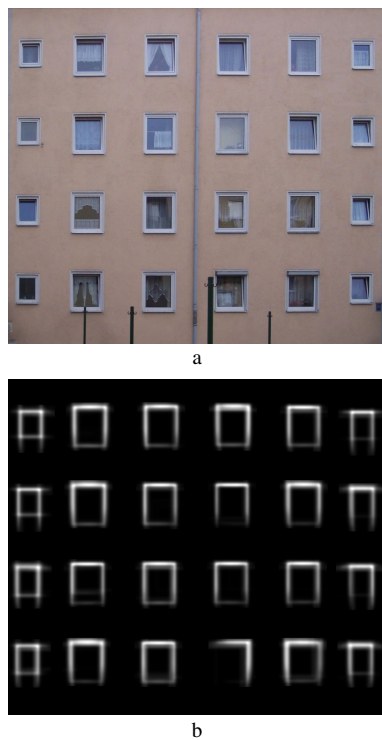

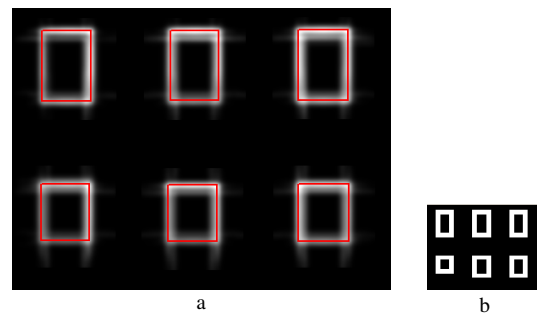Figure 5: a) Facade and b) distribution for window outlines

Figure 6: Determination of the likelihood in the distribution for the window outline – a) Given image with outline in red; b) minimal size

such as (Geman et al., 2002) prefers MDL. Particularly, we use

$$AIC = k - 2n \ln(L)$$

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 five for a row or column (four parameters for window shape plus – horizontal or vertical – spacing). The likelihood is determined in the normalized distribution image described in Section 2 above by means of MCMC. Figure 6 a) shows how the distribution is sampled at one position with the outline given in red. The idea is that every boundary point gives one observation of the likelihood which are multiplied leading to the multiplication factor for the log-likelihood.

A couple of experiments led to experience that it is not sufficient to just sample the given distribution for windows. We concluded that one also has to reduce the determination of the likelihood to a minimal setup. Thus, we derived from the sampling theorem 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. Figure 6 b)) 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 Figure 8. For all three

The above problem is thus regarded as a problem of model selection. Numerous ways 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 adaptation 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 and Suter, 2006). For its simplicity and as we found it to work well for our application, we employ AIC, though recent work on composition
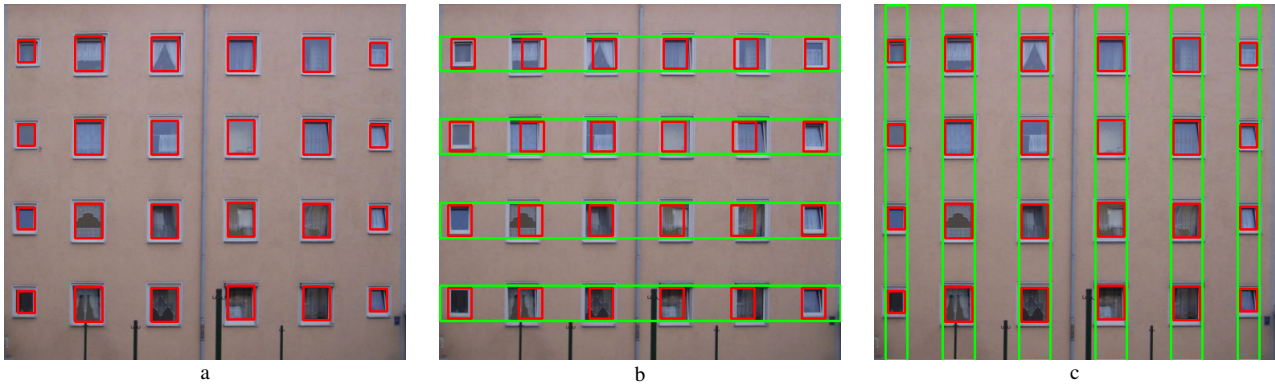
Figure 7: Model Selection – Representation of facade by a) individual windows; b) by rows; c) by columns of windows, the latter consisting of windows with the same size and a constant horizontal or vertical spacing.

given 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 in Figure 8.

4 3D RECONSTRUCTION VIA PLANE SWEEPING AND RESULTS

The result 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 also determine the 3D extent of the windows on the facade planes. To do so, we follow (Baillard and Zisserman, 1999) and (Werner and Zisserman, 2002) and employ plane sweeping, in this case in the direction of the normal of the individual facade plane. The determination of the depth for individual windows is based on the sum of the least squares differences between the projections of the plane into the individual images 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.

Figure 9 shows four images of a sequence with seven images and Figure 10 the result for three manually coarsely marked facades. In Figures 11 and 12 further results are given showing in both cases two again coarsely marked facades respectively. While for the first three facades rows of windows were chosen by model selection, it decided for the two facades of the second example, that they are better described by means of columns, and for the two facades of the third example it selected individual windows, as the columns with different window sizes do not fit to our models. Please note that our rows and columns 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 Figures 10, 11, and 12.

5 CONCLUSIONS

We have presented two 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 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 a 3D interpretation 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. 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.

On a more global level 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. Another question is a statistically sound link between discriminative and generative modeling such as in (Tu et al., 2005).

REFERENCES

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

Alegre, F. and Dallaert, F., 2004. A Probalistic Approach to the Semantic Interpretation of Building Facades. In: International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, pp. 1–12.

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

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

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

a – **C: 265**, R: 404, W: 359   b – C: 115, **R: 65**, W: 94   c - **C: 76**, R: 83, W: 87
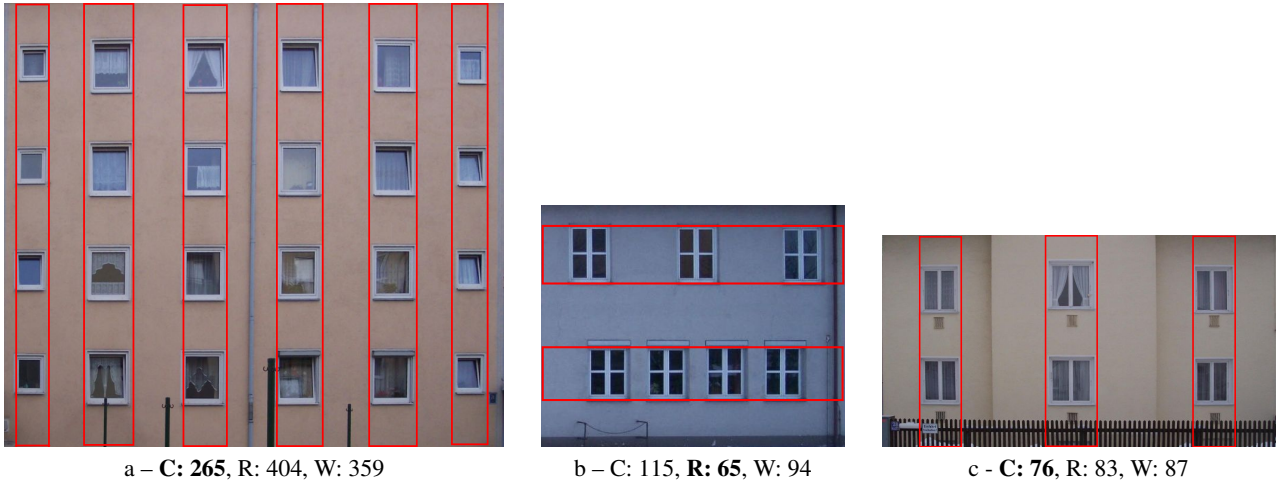
Figure 8: Results for Model Selection using AIC – C: Columns, R: Rows, W: Individual Windows – selected model in bold

Figure 9: Images one, three, five, and seven of sequence Ostbahnhof-1

Figure 10: Result for sequence Ostbahnhof-1 (images cf. Figure 9) – Window outlines for three facades with rows of windows as red rectangles, 3D window positions as green rectangles, camera positions as green pyramids

Fischler, M. and 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), pp. 381–395.

Förstner, W. and Gülch, E., 1987. A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres

of Circular Features. In: ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data, Interlaken, Switzerland, pp. 281–305.

Geman, S., Potter, D. and Chi, Z., 2002. Composition Systems. Quarterly of Applied Mathematics LX, pp. 707–736.
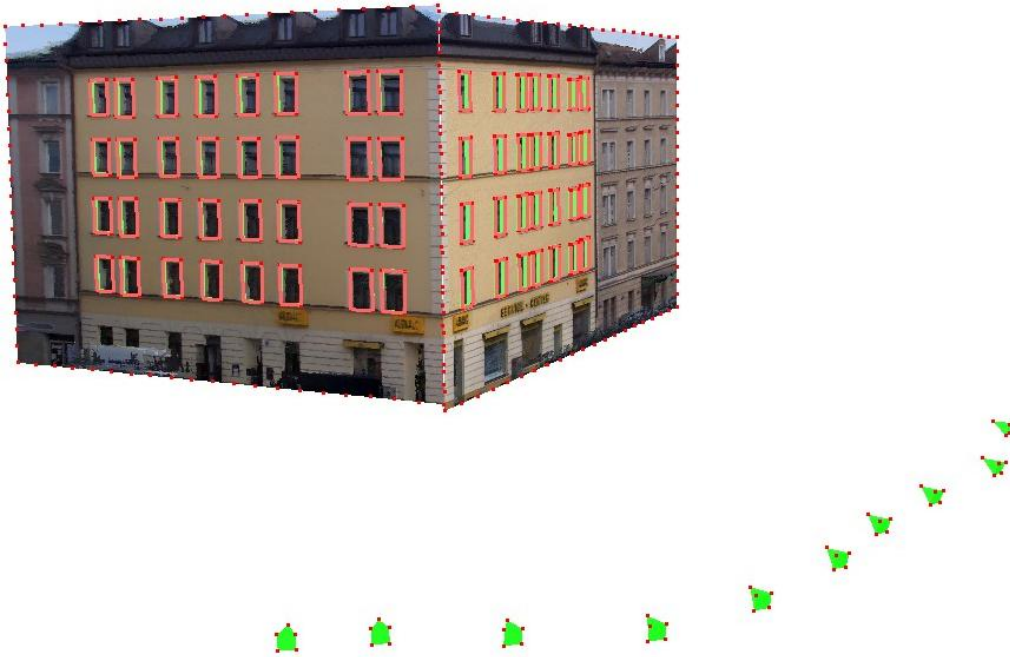
Figure 11: Result for sequence Ostbahnhof-2 with columns of windows constructed from ten images – explanation cf. Figure 10

Figure 12: Result for sequence Bordeaux-3 with individual windows constructed from eleven images – explanation cf. Figure 10

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

Leibe, B. and Schiele, B., 2004. Combined Object Categorization and Segmentation with an Implicit Shape Model. In: ECCV'04 Workshop on Statistical Learning in Computer Vision, pp. 1–15.

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

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

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), pp. 756–770.

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

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

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

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