Photogrammetrie • Fernerkundung • Geoinformation 7/2007, S. 485–494, 12 figs.

Detection and Description of Repeated Structures in Rectified Facade Images

Susanne Wenzel, Martin Drauschke & Wolfgang Förstner, Bonn

Keywords: Symmetry, Repeated Structures, Window Arrays, Compact Description, Model Selection

Summary: We present a method for detecting repeated structures, which is applied on facade images for describing the regularity of their windows. Our approach finds and explicitly represents repetitive structures and thus gives initial representation of facades. No explicit notion of a window is used, thus the method also appears to be able to identify other man made structures, e. g. paths with regular tiles.

A method for detection of dominant symmetries is adapted for detection of multiple repeated structures. A compact description of repetitions is derived from translations detected in an image by a heuristic search method and the model selection criterion of the minimum description length (MDL).

Zusammenfassung: *Erkennung und Beschreibung wiederholter Strukturen in entzerrten Fassadenbildern*. In dieser Arbeit wird eine Methode zur Detektion wiederholter Strukturen vorgestellt. Das Verfahren wird auf Fassadenbildern angewendet, um daraus eine Beschreibung der Regelmäßigkeiten der Fenster abzuleiten. Diese Beschreibung definiert eine kompakte Repräsentation der Fassadenstruktur. Es wird kein Vorwissen über Fenster eingeführt, so dass die Methode direkt auf andere von Menschen erzeugte Strukturen übertragbar ist.

Zunächst wird eine Methode zur Detektion dominanter Symmetrien in Bildern vorgestellt und daraus im Anschluss ein Verfahren zur Detektion multipler wiederholter Strukturen abgeleitet. Aus den beobachteten Translationen im Bild wird die kompakte Beschreibung der Fassadenstruktur abgeleitet, über ein Suchverfahren und das Modellauswahlkriterium der minimalen Beschreibungslänge (MDL).

1 Introduction

Symmetric and repeated structures are typical properties of man made objects. Thus, finding such features in a scene may be indicative of the presence of man made objects. Additionally, a compact description of the found regularities can be suitable for the evaluation of lower-level i. e. window detectors, or it can be used as a mid-level feature for model-based learning. Furthermore, the aggregation of repeated structures can improve the matching of objects with such regularities.

Therefore, the intention of this work is, firstly, to check if there are any regularities, and secondly, to infer a compact description of the repeated structure. The description consists of a hierarchy of translations and their appropriate numbers of repetitions.

A typical facade is characterized by perpendicular regularities in horizontal and vertical directions. As we work on rectified images of facades we limit our description on horizontal and vertical directions. The more general approach in (Schaffalitzky & Zisserman 2000) can be used to overcome this limitation. From their approach one

1432-8364/07/2007/0485 $ 2.50

could estimate the two most dominant (not necessarily perpendicular) directions to vanishing points. Concerning facade images, we assume that these two vanishing points always lie in vertical and horizontal direction. Then we are able to determine the affine transformation for image rectification, thus the approach of (Schaffalitzky & Zisserman 2000) would make our method more generic.

A lot of work on the detection of repetitive structures in images has been published within the last years. (Leung & Malik 1996) grouped repeated elements in the context of texture processing, allowing a similar transformation between the items. Another texture-based approach has been proposed by (Hays et al. 2006), where they map repetitive structures of texels within an iterative procedure.

(Tuytelaars et al. 2003) can detect regular repetitions under perspective skew. All mentioned works are limited on the constraint that the repetition by the elements can be described by a single 2-dimensional transformation.

Our work is based upon the approach of (Loy & Eklundh 2006), who proposed a method to detect dominant symmetries in images. The symmetry detection is based on the analysis of feature matches by their location including orientation and scale properties. We adapted this work to find repeated structures. The analysis of feature matches remains and had to be only slightly adjusted on the new problem.

Our method shall be applied for recognizing and outlining of building facades, where we have to face with competing structures of different sizes. Newest techniques for describing facades and their parts are developed by (Ripperda & Brenner 2006) or (Čech & Šára 2007) who use formal grammars. So far, these grammars are too general for our problem, but in future, these approaches might be helpful to combine the description of symmetries and repeated structures.

The paper is organised as follows: In Section 2, we sketch the algorithm by (Loy & Eklundh 2006) for the detection of dominant symmetries in images and our modification towards the detection of several dominant symmetries in images. Then we describe our further enlargement of this approach for the detection of multiple repeated elements in Section 3. In Section 4, we present our approach for the derivation of the compact description of the structure of found repeated elements. A short conclusion is given in Section 5.

2 Detection of Dominant Symmetries

Loy & Eklundh (2006) proposed a method for finding dominant symmetries in images. We give a brief summary on this method to clarify the underlying idea of our approach described in Section 3. The principal functionality is sketched in Fig. 1.

Firstly, they detect prominent features by the SIFT operator (Lowe 2004). Note, the approach is not restricted to the SIFT operator, any rotationally invariant method that detects distinctive points with good repeatability and that gives a feature descriptor suitable for matching can be used.

Every SIFT feature is described by its location (row, column, scale and orientation) and by the descriptor, which encodes the gradient content in the local image patch, normalised with respect to the feature's orientation.

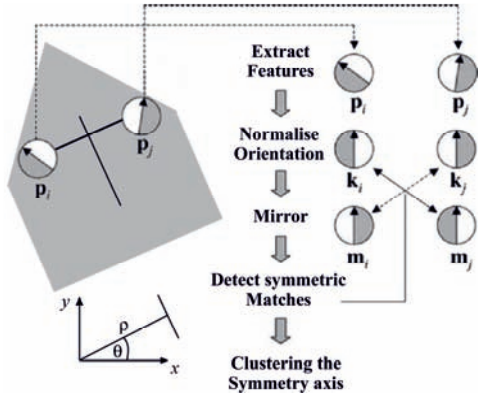

Fig. 1: Principal functionality of detecting dominant symmetries, partly taken from (Loy & Eklundh 2006). The Hessian normal form of the symmetry axis (θ, ρ) is derived from the feature pair p_i and p_j.

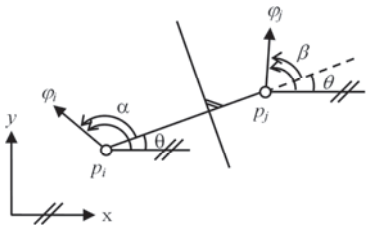

Fig. 2: Illustration of the functionality of the angle-weight, according to (2).

Flipped versions of the features are obtained after resorting the descriptor elements (Wenzel 2007). They subsequently match the sets of original features and their flipped versions to get pairs of potential symmetric features. Every pair is represented by the Hessian normal form of their symmetry axis with their normal and distance from origin. These coordinates are clustered over these parameters to find dominant symmetries among the found features.

The quality of symmetry M of each feature pair p_i und p_j is measured by

$$M_{ij} = \Phi_{ij} S_{ij}, \tag{1}$$

where Φ_{ij} and S_{ij} are two weights defined as

$$\begin{aligned}\Phi_{ij} &= -\cos(\varphi_i + \varphi_j - 2\theta) \\ &= -\cos(\alpha + \beta)\end{aligned} \tag{2}$$

and

$$S_{ij} = \left[\exp\left(\frac{-|s_i - s_j|}{\sigma_s(s_i + s_j)}\right)\right]^2. \tag{3}$$

The angle-symmetry-weight $\Phi_{ij} \in [-1, 1]$ returns a high value for those feature pairs whose orientations are symmetrical with respect to the proposed symmetry axis, cf. Fig. 2. The angles α and β add up to $180°$, if the orientations are exactly symmetrical in respect to the proposed symmetry axis.

The scale-weight $S_{ij} \in (0, 1]$ is used for limiting the differences between both features with respect to their characteristic scales s_i and s_j. Larger differences can be tolerated by increasing the parameter σ_s. Loy & Eklundh (2006) introduced another weight with respect to the distance between both features, but this is only advantageous, if one would like to insert prior knowledge of the observed object. Since we want to look for all kind of symmetries within facade images, we do not use this weight.

The Hessian normal form of the symmetry axes (θ, ρ) of all found potential symmetric feature pairs are accumulated with respect to their weightings in a two dimensional array. The result is a two dimensional histogram of the sum of symmetry measures over the parameters θ and ρ of the symmetry axes. Dominant symmetries of an image ap-

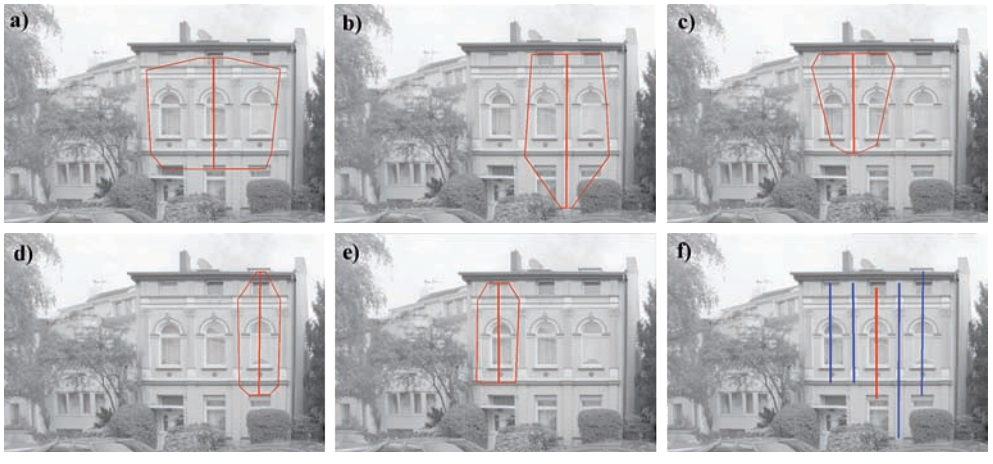

Fig. 3: Results for symmetry detection: Exactly five dominant symmetry axes were found.
a)–e) Single results for detected symmetries with convex hulls of involved features.
f) Combination of all found symmetries.

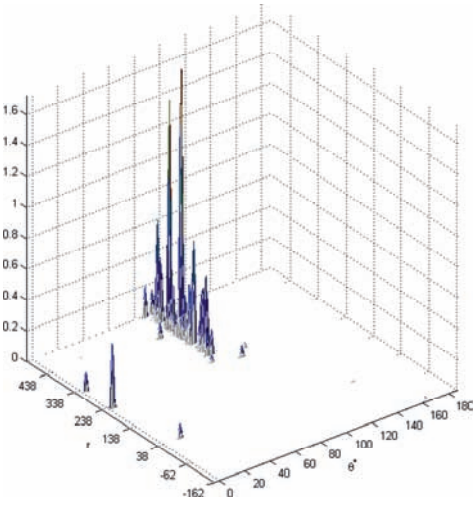

Fig. 4: 2D-histogram over the polar coordinates of the symmetry axes for the example from Fig. 3.

pear as relative maxima of this histogram. In contrast to (Loy & Eklundh 2006) where the goal was to find only the major symmetry, we search for all significant symmetries by investigating all peaks of the histogram which are supported by at least t feature pairs. In our experiments we choose $t = 4$.

Fig. 4 shows the histogram in respect to the facade of Fig. 3. This facade is described solely by symmetries with vertical symmetry axes. Therefore the histogram has its global maximum at ($\theta = 90^\circ$, $\rho = 391pix$) and additional local maxima along the 90° grid line.

Fig. 3f) shows the five detected symmetries in one image. In Fig. 3a)–e) we show each of the detected symmetry axis together with the convex hull of its supporting feature points. For this example, we detected 1617 features which form 151 potential symmetrical feature pairs. The major symmetry axis, cf. Fig. 3a), is supported by 34 feature pairs. The other four symmetry axes shown in Fig. 3b)–e) are supported by 21, 13, 13 and 9 feature pairs, respectively.

All of the detected symmetries lie in the building facade, other objects of the image do not disturb the symmetry detection. Furthermore, the convex hulls of the involved features in all symmetries lead directly to the image region, which is characterized by the symmetrical structures.

3 Detection of Repeated Structures

We adapted the basic idea of clustering feature pairs within a single image to detect repeated structures. Obviously, the flipping of the feature descriptors can be omitted. Instead, we match the detected features with each other, such that we find pairs of very similar features, similar with respect to orientation and scale. Additionally, the weight according to orientation is adapted to our purpose. Thus, the angle-symmetry-weight Φ_{ij} is simplified to the angle-weight $\Phi^*_{ij} \in [-1, 1]$

$$\Phi^*_{ij} = \cos(\varphi_i - \varphi_j) \tag{4}$$

and it supports mostly those feature pairs with similar orientation. Thus, the quality of repetition M* is measured by

$$M^*_{ij} = \Phi^*_{ij} S_{ij}. \tag{5}$$

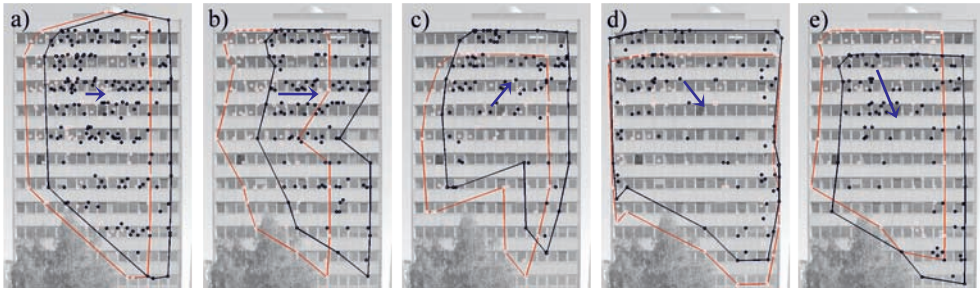

Fig. 5: Five most dominant translations for this image. The features involved and their boundaries[1] are shown together with the translation vector between the red and black groups.

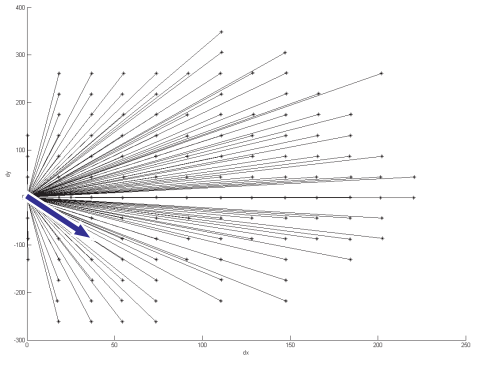

Fig. 6: Observed translations of 122 detected repeated groups for the example from Fig. 5 (i. e. the blue arrow represents the translation found in Fig. 5e).

Again clustering over directions and amount of translations yields the dominant translations in the image. Dominant translations in the image correspond to the maxima of the histogram of the repetition measure. Furthermore, we focus on those translations which are supported by at least t feature pairs. Again we choose $t = 4$. Fig. 5 shows the five most dominant repeated structures for this example. In each case, the red features are matched to the black features by the same translation, indicated by the blue arrows. Furthermore, the boundaries[1] of both feature groups are represented for better visualisation. For this example altogether 122 repeated groups were detected[2].

For better demonstration of these results Fig. 6 shows all detected translations as plot of translation vectors. This representation shows clearly the regularity in the detected translations.

Fig. 7 presents another example, where the facade is not as simple as in Fig. 5. At a first glance, there is a simple grid of windows, but a closer look reveals three differ-

[1] Instead of using the convex hull of the points we use an algorithm, that minimises the polygon area by allowing only line lengths of (*longest line* / *2*) of the convex hull polygon.

[2] We selected the matching criterion of the Lowe-matcher with *distRatio* = *0.9* very sensitively concerning variances (shade, curtains etc.) of the facade elements. Thus, relatively large distances between the descriptors of the features lead to a positive match. For more details about the parameters for the matching of two SIFT feature descriptors, especially about *distRatio* (Lowe 2004).

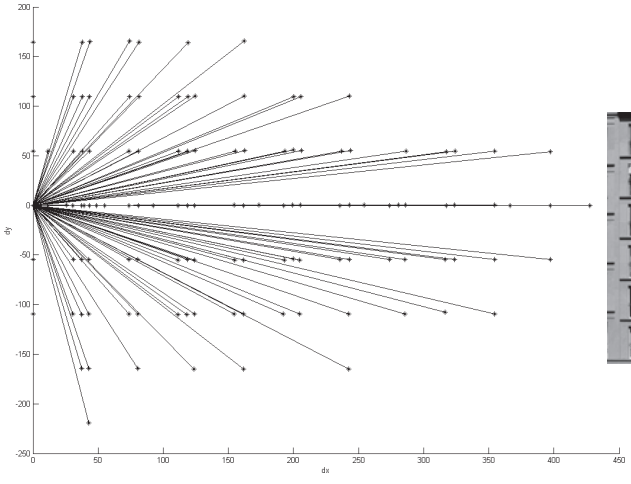

Fig. 7: Another example for the detection of repeated structures, with more complex pattern of detected translation. The windows do not form a regular grid. There are three different distances between the window columns, so that there are different superposed grids that are concatenated to the whole structure. There were 2349 keypoints found with 7939 matches and 116 repeated constellations among them.

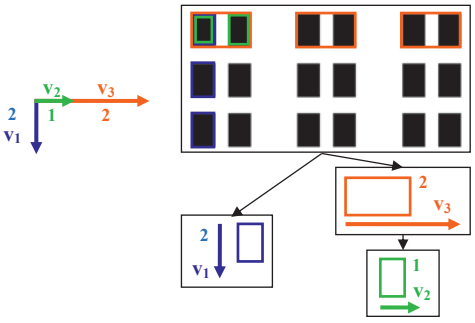

Fig. 8: A typical facade is characterized by perpendicular regularities through rows and columns. In this example, there is a hierarchy of repetition for the horizontal direction and a single twofold repetition for the vertical direction. Thus in horizontal direction the compact description of the structure consists of a hierarchy ($K = 2$) of basis elements with the amount of the translation and the number of repetitions.

ent distances between the window columns. The outer column pairs are quite close to each other. They form a frame for the middle region, where you can see a catenation of two grids of different size. Hence, the observations of translations, shown on the left of Fig. 7, do not look as regular as Fig. 6.

We look for a compact description of these repetitions which exactly depicts the regularity and underlying pattern, respectively.

4 Derivation of the Compact Description

Because we work on rectified images, the main directions of the translations run parallel to the image borders. Therefore, we can reduce the search for a suitable basis to separate searches in the horizontal and in the vertical directions. Then, a typical facade is characterised by perpendicular regularities through rows and columns. There may be different types of repeated elements where bigger elements are compositions of smaller elements. Thus, the repeated elements can be represented in a hierarchical order per direction with depth K, which forms a hierarchical basis. This is illustrated in Fig. 8. Note, that we do not restrict the repeated elements to have a certain shape.

Due to the reduction on horizontal and vertical directions, we project all translations on the dx and dy axes and treat these new translations as our observations $d_i\,(i = 1:n)$. Thus, for the x- and the y-direction, respectively, they can be described as a linear combination of basis translations v_k and the appropriate coefficients α_k, the number of repetition, through

$$d_i = \sum_{k=1}^{K} (\alpha_k \cdot v_k) + \varepsilon_i\,. \tag{6}$$

The depth K of the hierarchical basis corresponds to the number of elements of the linear combination. A priori the value of K is unknown, but we assumed typical urban facades in its complexity do not exceed the value $K_{max} = 4$. Neither the integer-valued coefficients α_k nor the real-valued basis translations v_k are known. Furthermore, each observation d_i is afflicted with a residual ε_i.

We look for a hierarchical basis, consisting of K basis elements, which explain the observed translations in the best possible way, including the minimisation of the residues and the complexity K of the solution.

Since we could not find a direct solution for this problem, we decided for a heuristic procedure. As one can see from Fig. 6 and Fig. 7, the distances between two neighboured observations seem to be good candidates for the minimal translations. Therefore, we determine the differences between all observed translations. We compute a histogram via these second differences of the positions. The highest peaks of this histogram are potential candidates for the basis translations that we look for. This is illustrated in Fig. 9 for the observations in x-direction for the example from Fig. 7. The first plot shows the observations of translations projected on the x-axis. Again this plot clarifies, that the underlying pattern is not always as clear as in Fig. 6. But the second plot of Fig. 9, the histogram over the differences of observed translations, shows the effect of considering the differences. The

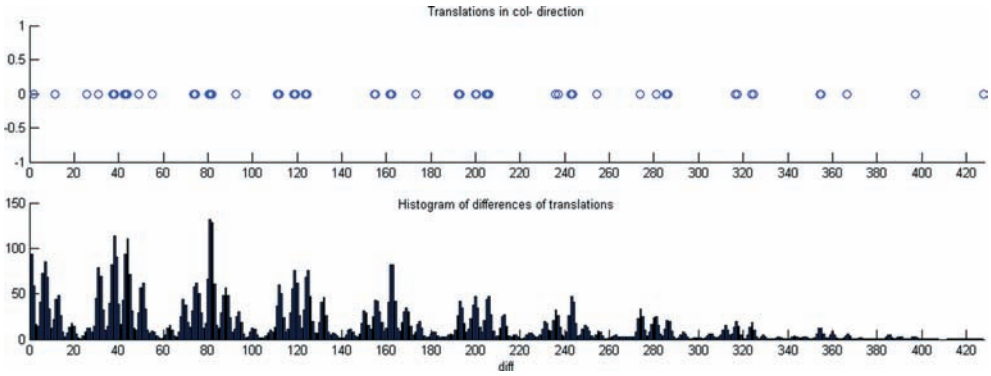

Fig. 9: Visualisation of the process of finding the candidates for basis elements. Here we only show the translations in *x*-direction for the example from Fig. 7. Top: the observations of translations are shown, projected on the *x*-axis. Bottom: histogram of the differences over all observed translations. The highest peaks of this histogram give the candidates for the unknown basis elements.

peaks of this histogram give us potential candidates for basis translation that we will further evaluate.

From the c candidates taken from the histogram of differences of translations, we form all

$$C = \sum_{k=1}^{K_{\max}} \binom{c}{k} \tag{7}$$

combinations of possible bases $\mathbf{v}$. Then, we determine the appropriate coefficients α_k for each of these potential solutions ${}^j\mathbf{v}\ (j = 1:C)$ and for each observation d_i. The residual vector ${}^j\boldsymbol{\varepsilon}$ is obtained for the results of every solution ${}^j\mathbf{v}$. The best solution minimises the residuals with the smallest model complexity.

If a certain data set can be described by a compact model, then only the model parameters and possible deviations of the data from this model need to be encoded. This consideration leads to the minimum description length (MDL) criterion, proposed in (Rissanen 1989):

$$MDL = -\log \prod_{i=1}^{n} P(d_i|\pi) + \frac{K}{2}\log(n)\,. \tag{8}$$

We look for that model (π, K), that describes the observed data $\mathbf{d}$ with the smallest complexity K and the largest data probability

$$\prod_{i=1}^{n} P(d_i|\pi)\,, \tag{9}$$

where π are the parameters of the model. On the assumption of normally distributed residuals the criterion can be represented as

$$MDL = \frac{1}{2}\Omega + \frac{K}{2}\log(n)\,. \tag{10}$$

The consideration of outliers is based on Huber, cf. (Förstner 1989), with the cost-function

$$\Omega = \textstyle\sum_{i=1}^{n} \kappa(\varepsilon_i) \tag{11}$$

and the optimisation function

$$\kappa(\varepsilon) = \begin{cases} T^2 & \text{if } (\varepsilon/\sigma)^2 \geq T^2 \\ (\varepsilon/\sigma)^2 & \text{if } (\varepsilon/\sigma)^2 < T^2 \end{cases}. \tag{12}$$

According to the critical value T traditionally chosen on the basis of the significance level of hypothesis test, we select the threshold value for outliers as $T = 3\sigma$ with $\sigma = 1.5$.

The residuals ${}^j\boldsymbol{\varepsilon}$ are used to determine the MDL value for every possible solution ${}^j\mathbf{v}$. That model $\mathbf{v}$ that gives the smallest MDL value is chosen to be the model that best described the observed translations.

The boundary of the feature pairs, which support the selected model $\mathbf{v}$, defines the region that can be described by these basis elements. Thus, we get a compact description of the repetitive structure in the form of basis elements and the associated regions in the image.

Fig. 10 shows the results of the derivation of compact description for the example from

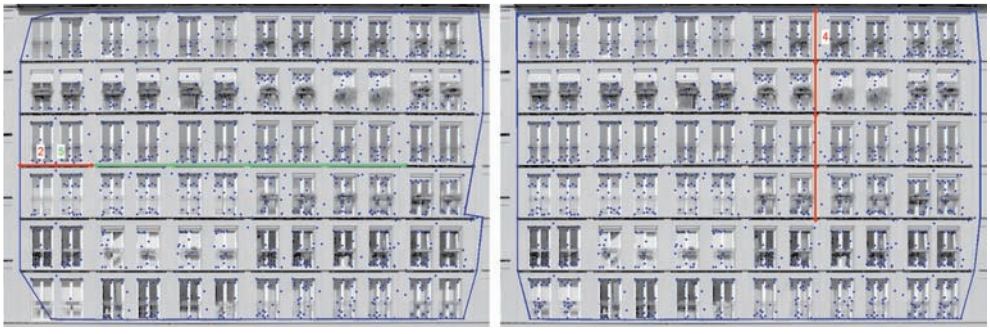

Fig. 10: Results for the compact description of the facade structure for the example from Fig. 7 and Fig. 9, respectively. We show regions with corresponding basis elements for both horizontal and vertical direction.

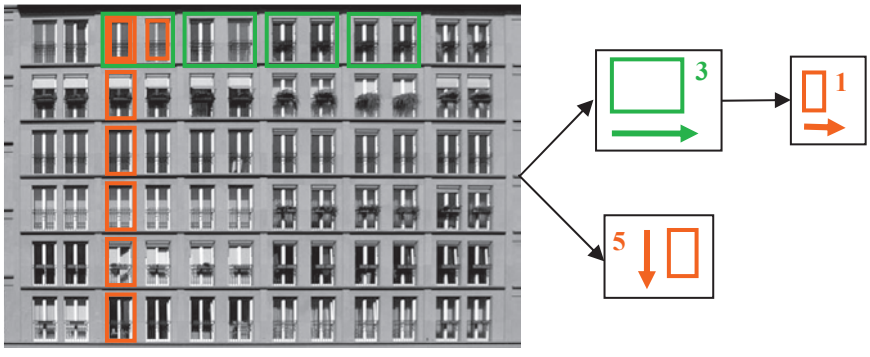

Fig. 11: The interpretation of the result from Fig. 10 by manual inspection. Obviously, the result describes the structure of the middle region of the facade. In vertical direction, this example has a fivefold repetition of a single window. Thus, there are six floors with distance given by the vertical red arrow. In horizontal direction, there is a threefold repetition of a double-window which again can be described by a single repetition of a single window.

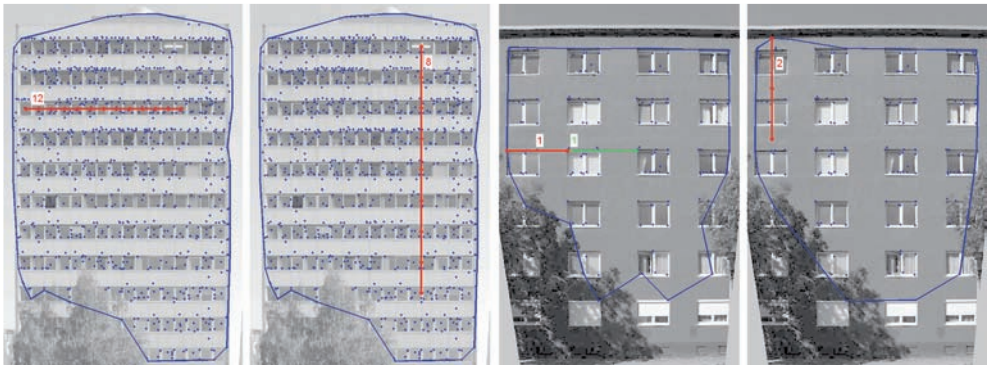

Fig. 12: Two further results of derivation of compact description of structure. On the left we continue the example from Fig. 5 and Fig. 6, respectively. A basis which consists only of one element has been determined for both axis directions. On the right another example of a hierarchical basis in the horizontal direction is presented.

Fig. 7 and Fig. 9, respectively. For the vertical direction a basis has been determined, which consists only of one element. And for the horizontal direction a hierarchical basis with two elements was found, according to the underlying structure of the windows of the middle region. The number of repetitions of basis vectors is given by the maximum number of coefficients for the linear combinations of basis vectors of all observations.

Note, that the features involved (indicated by the blue points) are concentrated on the region of the inner window columns. These are described by only two different distances. The boundary is only disturbed by outliers on the outer left and right. The interpretation of this result by human inspection is given in Fig. 11. Obviously, there is an uncertainty in the number of repetitions compared to the result of Fig. 10, but the amounts of the translations are determined correctly.

Fig. 11 shows the results for further facade images. On the left, where we continue the example from Fig. 5, a basis which consists only of one element has been determined for both axis directions. The boundaries of the features that take part in this basis cover the entire facade region (with exception of the region covered by the tree). On the right of Fig. 11, we present another example of a hierarchical basis in the horizontal direction. The four columns of windows do not have the same distance from each other, but the two window columns on the left have the same distance as the two window columns on the right. Thus, we obtain two different translation vectors according to the real structure of the facade. Note the effects of occlusion and clutter. On the left the found region is disturbed by the tree and on the right the structure is cluttered by the shadow.

5 Conclusions

We showed how the approach from (Loy & Eklundh 2006) can be extended to the detection of multiple repeated groups. From translations detected in an image we derived a model for a compact description of repetitive structures in facade images using a heuristic search method and the criterion of the minimum description length. So far, our algorithm only works on images, which show only one regular part of facades. The matching procedure is very sensitive to repeated objects in the regular part of a facade due to the very generous choice of the matching criterion. Especially, similar structures in the neighbourhood of the facades trouble our approach. We need to refine our method, in particular regarding the robustness against disturbances in the picture.

References

Čech, J. & Šára, R., 2007: Language of the structural models for constrained image segmentation. – Technical Report TN-eTRIMS-CMP-03-2007, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, available at http://www.ipb.uni-bonn.de/projects/etrims/publications.html (last accessed 14.11.2007).

Förstner, W. 1989: Image Analysis Techniques for Digital Photogrammetry. – Schriftenreihe des Instituts für Photogrammetrie der Universität Stuttgart **13**: 205–221.

Hays, J., Leordeanu, M., Efros, A. & Liu, Y., 2006: Discovering Texture Regularity as a Higher-Order Correspondence Problem. – Proceedings of the 9th European Conference on Computer Vision **II**: 522–535.

Leung, T. & Malik, J., 1996: Detecting, localizing and grouping repeated scene elements from an image. – Proceedings of the 4th European Conference on Computer Vision **I**: 546–555.

Lowe, D. 2004: Distinctive Image Features from Scale-Invariant Keypoints. – International Journal of Computer Vision **60 (**2): 91–106.

Loy, G. & Eklundh, J.-O, 2006: Detecting Symmetry and Symmetric Constellations of Features. – Proceedings of the 9th European Conference on Computer Vision **II**: 508–521.

Ripperda N. & Brenner, C., 2006: Reconstruction of Façade Structures Using a Formal Grammar and RjMCMC. – Pattern Recognition – 28th DAGM Symposium, Springer, Berlin, 750–759.

Rissanen, J., 1989: Stochastic Complexity in Statistical Inquiry. – World Scientific: Series in Computer Science **15**.

Schaffalitzky, F. & Zisserman, A. 2000: Planar grouping for automatic detection of vanishing lines and points. – Image and Vision Computing **18** (9): 647–658.

Tuytelaars, T., Turina, A. & Van Gool, L., 2003: Noncombinatorial detection of Regular Repetitions under Perspective Skew. – IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (4): 418–432.

Wenzel, S., 2007: Mirroring and matching of the SIFT-feature descriptors for detection of symmetries and repeated structures in images. – Technical Report TR-IGG-P-2007–04, Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn, available at http://www.ipb.uni-bonn.de/~suse/ (last accessed 25. 09. 2007).

Addresses of the Authors:

Dipl.-Ing. Susanne Wenzel, Dipl.-Inform. Martin Drauschke, Prof. Dr.-Ing. Wolfgang Förstner, Department of Photogrammetry, Institute of Geodesy and Geoinformation, University of Bonn, Nussallee 15, 53115 Bonn, Fax: +49-228-73-2712, Tel.: +49-228-73-2908 (Wenzel), -2901 (Drauschke), -2713 (Förstner), e-mail: susanne.wenzel@uni-bonn.de, martin.drauschke@uni-bonn.de, wf@ipb.uni-bonn.de.

Manuskript eingereicht: August 2007
Angenommen: September 2007