International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B3, 2012
XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia

# GESTALT GROUPING ON FAÇADE TEXTURES FROM IR IMAGE SEQUENCES: COMPARING DIFFERENT PRODUCTION SYSTEMS

E. Michaelsen [1], D. Iwaszczuk[2], L. Hoegner[2], , B. Sirmacek[3], U. Stilla[2]

[1] Fraunhofer-IOSB, Gutleuthausstrasse 1, 76275 Ettlingen, Germany, eckart.michaelsen@iosb.fraunhofer.de
[2] Technische Universitaet Muenchen (TUM), Photogrammetry and Remote Sensing, 80333 Muenchen, Germany
[3] German Aerospace Center (DLR), Remote Sensing Technology Institute, 82234 Wessling, Germany

**Commission III, WG III/4**

**KEY WORDS:** Façade recognition, thermal imagery, production systems

**ABSTRACT:**

The façades of buildings are almost always organized according to Gestalt principles such as *good continuation*, *repetition in similarity*, or *symmetry* etc. Coding such principles in production systems yields a very flexible frame to explore the usefulness of such principles in automatic façade understanding. Capturing images and image sequences of façades in the thermal domain and understanding such data is of importance e.g. for energy saving. In this contribution two different production systems are compared using the same data and interpreter.

## 1. INTRODUCTION

### 1.1 Thermal Textures

Thermal textures on façades of buildings are of growing interest. In thermal infrared (IR) images damages and weak spots in building hull (especially in building insulation) and heat waste can be observed. Thanks to combination with a 3D building model the spatial reference of the IR images is given. In urban areas with narrow streets it is often not possible to capture a whole façade in one frame. Then the picture can be stitched from a video. Figure 1 shows such image which actually results from texturing a 3D building model by projecting suitable image information from a vehicle mounted thermal video camera (Hoegner & Stilla U, 2007).

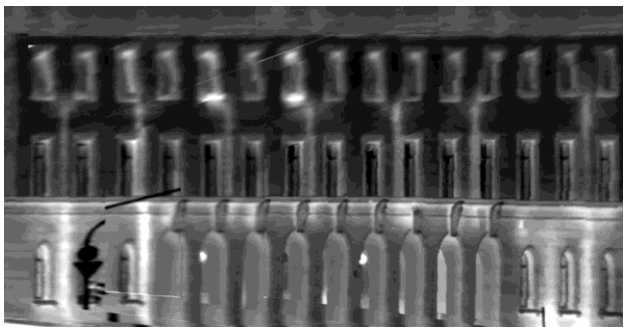

Figure 1. A thermal façade texture

Of course artefacts from stitching cannot be completely avoided. But, obviously in such textures heat leakages can be detected and the heat bridges can be stored together with 3D building data. Our goal is to proceed in automation of this process.
There are two reasons why windows should be detected in IR textures and excluded for the inspection. First, in thermal imagery glass reflects the surrounding, e.g. sky, a neighbouring building and trees, and shows false results for the temperature measurements. Second, windows can influence automatic heat leakage detection (Hoegner & Stilla U, 2009) and lead to false results. Therefore a method for window detection in IR images is needed.

### 1.2 Gestalt Grouping

Façades are man-made objects and display strong gestalt structure, such as ordering according to lattice or symmetry principles. This also holds for the thermal spectral domain. The laws of Gestalt grouping are known for about hundred years, namely "good continuation", "repetition in similarity", "symmetry – mirror or rotational" and proximity (Wertheimer, 1923). Automatic Gestalt grouping can e.g. be performed by tensor voting or accumulator methods. The state-of-the-art has e.g. been presented at the symmetry competition along with the CVPR (Liu & Rauschert, 2011).

### 1.3 Related Work

In the last decade façades classification has drawn particular attention. Pu & Vosselman (2009) fused laser data and close-range images to reconstruct building façade details. They extracted windows and doors in both close-range optical and laser images by using Hough accumulation of lines. Detected windows and doors helped them to register close-range optical and laser images. That shows the importance of the facade classification study in three-dimensional city modeling. Burochin et al. (2009) proposed a segmentation method to detect repetitive structures like windows in close-range optical images. For segmentation they defined a model by considering shape and reflectance of a window. Then they applied matching process to find correspondence between model and image. In (Ali et al., 2007), a summary of the researches on window detection has been given. They also proposed a window detection system based on cascade classifiers. In a following study, Ali et al. proposed a system to detect windows in laser scanner data. They use depth variations to detect windows (Ali et al., 2008). Lee & Nevatia (2004) proposed a robust system to detect windows in optical images. They extracted window

boundaries searching for structures that satisfy regularity and symmetry rules. In addition to that, they extract three-dimensional models of windows by searching for image features. Teboul et al. (2010) used shape grammars towards fixed tree representations which are able to capture a wide variety of building topologies for detailed facade segmentation. They obtained very high performance even for buildings which are partially occluded or which appear under different illumination conditions. Ripperda (2008) reversible jump Markov chain Monte Carlo (rjMCMC) for the estimation of optimal parameters for the windows and uses a formal grammar to describe their behaviour. Mayer & Reznik (2006) propose combination of Markov Chain Monte Carlo with information from Implicit Shape Models and with Plane Sweeping as well. Tanks to this they achieve a 3D interpretation of building facades determining windows and their 3D extent. In Mayer & Reznik (2008) the method is extended with self-diagnosis and model selection to choose the most appropriate model for the configuration of windows in terms of rows or columns.

Most of these algorithms are computationally expensive and not suitable for real time applications. Sirmacek (2011a) proposed a segmentation and graph theory based facade classification method with emphasis on real-time requirements. However, this method requires very uniform and also correctly ortho-rectified color images as input.

In Europe there has been a joint research effort on façade classification called eTRIMS (Foerstner et al., 2009). A special role plays the syntactic formulation of Gestalt laws. Inside eTRIMS such approach has been formulated in (Tylecek &Sara, 2011) using stochastic grammars for the description, and random sampling as search method.

A similar formulation uses production systems as declarative knowledge representation and special interpreters for search. This has the advantage of clear modularity and explicit declarative inclusion or exclusion of particular constraints or recursive principles. Thus comparison of their benefits or cons is facilitated. E.g. Matsuyama & Hwang (1990) have proposed the SIGMA system for automatic understanding of aerial images of man-made objects. This system featured declarative knowledge coding using production rules and a special interpretation scheme quite similar to the one used here. Unfortunately this work has not been continued. Another such approach was called BPI (Stilla & Michaelsen, 1997) and this is being continued as the GESTALT system (Michaelsen et al., 2010).

In Sirmacek (2011b) the usage of L-shaped feature primitives is proposed for window and door detection from thermal facade images. Iwaszczuk et al. (2011) suggest using local dynamic threshold and masked correlation for corner detection and orders detected window candidates into row and columns. In this paper we would like to merge the idea of detecting primitives (corners and L-shapes) with gestalt rules to find windows from thermal images robustly.

This contribution is organized as follows: Section 2 presents production systems in general and two special systems are presented coding the likely organization of windows on facades. Section 3 comparatively studies the behaviour of these systems on example data obtained in the city of Munich. The work closes with a discussion on the results and an outlook for future work in Section 4.

## 2. PRODUCTION SYSTEMS

Structural knowledge e.g. about the part-of hierarchies of man-made objects, about geometric properties of their mutual arrangements, and about their appearance can be coded in a declarative way using systems of production rules.

### 2.1 Extraction of Primitives

Prerequisite to all syntactic work on images is segmentation for primitive objects. Here a corner detector based on a masked correlation which consists of "on" and "off" fields and of "don't care" areas is applied. Masked correlation was originally applied by (Stilla, 1993) to recognise stamped characters. The advantage of this method is, that can cope with blurred edges. We adapt the idea of masked correlation to search for the changes in the intensity between façade and window. We place a "don't care" area between "on" and "off" fields, which helps to avoid blurring on the edges. The correlation coefficient $c$ is calculated using

$$c = a \cdot \frac{1}{\sqrt{\frac{m}{m_-}\left(\frac{\sigma_+}{\bar{g}_+ - \bar{g}_-}\right)^2 + \frac{m}{m_+}\left(\frac{\sigma_-}{\bar{g}_+ - \bar{g}_-}\right)^2 + 1}}$$

$$a = \operatorname{sgn}\left(\rho_+ - \rho_-\right) \cdot \operatorname{sgn}\left(\bar{g}_+ - \bar{g}_-\right) \qquad (1)$$

where $\rho_+$ – value of "*on*" mask, $\rho_-$ – value of "*off*" mask, $\bar{g}_+$ – mean value of intensity values in the image covered by "*on*" mask, $\bar{g}_-$ – mean value of intensity values in the image covered by "*off*" mask, $m_+$ – number of "*on*" pixels in the mask, $m_-$ – number of "*off*" pixels in the mask, $m$ – number of "*on*" and "*off*" pixels in the mask, $\sigma_+$ – standard deviation of intensity values covered by "*on*" mask, $\sigma_-$ – standard deviation of intensity values covered by "*off*" mask.

Four corner types are assumed: upper left, upper right, lower right and lower left. Each type is correlated with the whole image and pixels which result in a correlation coefficient higher than our detection threshold are selected. The selected pixels are coded with the orientation attribute of primitive instances ("upper left", "upper right", "lower right" and "lower left") and with its correlation coefficient $c$. In Fig. 2 exemplary corner detection is presented. For the red, green, blue and yellow pixels the correlation coefficient was higher than the detection threshold. Colours encode the orientation attribute. For a typical façade image some 20,000 such pixels remain from texture of e.g. in this case 1024 x 524.

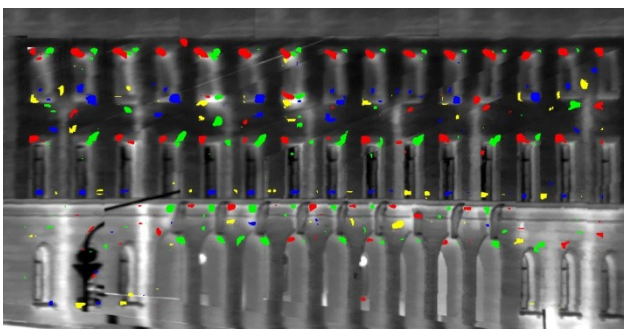

Figure 2. Corners extracted in a thermal façade textures: red - "upper left", green - "upper right", blue - "lower right" and yellow - "lower left"

It is obvious that for each corner perceived by a human observer several such pixels cluster together. Following Marr's principle of avoidance of early decisions these objects would be entered into the production system as primitives – using an according clustering production as first knowledge source. However, this

overloads the computational resources necessary for the following reasoning currently available. So we decided to perform a non-maximum suppression as is usually performed in computer vision. Only 365 L-primitive instances remain which are displayed on black ground in Fig. 3.

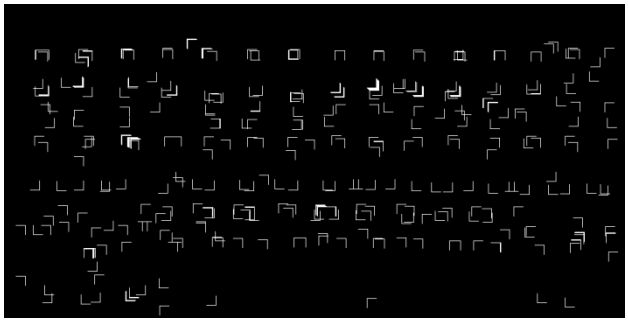

Figure 3. Primitives extracted after non-maximum suppression

From this figure the reader may estimate what is lost during the primitive extraction phase. Recall that the following symbolic process only sees these data, not the image itself.

### 2.2 Two Different Production Systems

The paper reports experiments with different types of production rules representing lattice grouping and symmetry grouping. Three variants are compared – see Tab. 1:

1) *"canonical"* the natural common sense part-off hierarchy is: A façade consists of a vertical column of (two or three) horizontal rows of (e.g. a dozen) windows; these windows are of same size and shape; each window consists of an upper U-structure and a lower U-structure and each of these consists of two L-primitives in according symmetric convex configuration. A careful look at the set of primitives given e.g. in Fig. 3 shows that only a minority of the windows perceived by humans in Fig. 2 allow reduction to instances Rectangle according to this system. Most often something is missing or badly displaced. Accordingly, the grouping of non-trivial Row and Lattice instances will also fail. There is little sense in trying automatic interpretation with this system.

2) Experience shows that often one corner is missing, while other corners appear multiply in displaced versions. There is a standard approach to cope with such situation: The symmetry axes of one vertical U-structure and one horizontal U-structure are intersected. Additionally, one side of these structures must be quite close to one of the other, and of course again convexity is demanded. Thus even incomplete windows can be instantiated, and attributed with height and width. But they will be instantiated multiply, and for this reason a clustering production is included, that fuses several such adjacent Intersect instances into one Rectangle object. The rest of the system – namely grouping into rows and lattices is the same. All systems used here first group in horizontal direction and then in vertical. Experiments with this system are reported below.

3) The third variant attempts to group the window corners into rows first. This has the advantage that some of the corners have higher probability of appearing than others (according to their orientation). Quite long such rows can be grouped, and thus the generator vector (shift from one window to the next) can be estimated with good precision. Then from two such rows a row of U-structures can be built simultaneously with all parts in one, and afterwards a row of windows with common width and height for all windows which are part of it. So this follows a different part-of hierarchy than the one used above. This follows the idea that two nearby rows of structures having the same generator with high accuracy probably result from the same repetitive pattern. We can be more liberal with biased displacements such as shear and un-biased displacements will be averaged out by the previous grouping. Experiments with this system are also reported below.

| Left-hand | Right-hand | constraint |
|---|---|---|
| U-structure | L-primitive, L-primitive | symm. & convex |
| Rectangle | U-structure, U-structure | symm. & convex |
| Row | Rectangle, Rectangle | horizontal proximity |
| Row | Row, Rectangle | good continuation |
| Lattice | Row, Row | vertical proximity |
| Lattice | Lattice, Row | good continuation |
| System *"canonical"* | | |

| | | |
|---|---|---|
| U-structure | L-primitive, L-primitive | symm. & convex |
| Intersect | U-structure, U-structure | prox. & orthogonal |
| Rectangle | Intersect, …, Intersect | proximity |
| Row | Rectangle, Rectangle | horizontal proximity |
| Row | Row, Rectangle | good continuation |
| Lattice | Row, Row | vertical proximity |
| Lattice | Lattice, Row | good continuation |
| System *"windows first"* | | |

| | | |
|---|---|---|
| L-Row | L-primitive, L-primitive | horizontal proximity |
| L-Row | L-Row, L-primitive | good continuation |
| U-Row | L-Row, L-Row | parts(sym. & conv.) & similar generator |
| Row | U-Row, U-Row | orthogonal & similar generator |
| Lattice | Row, Row | vertical proximity |
| Lattice | Lattice, Row | good continuation |
| System *"L-rows first* | | |

Table 1. Production systems

### 2.3 Automatic Interpretation

**Search:** The grouping uses the interpretation system proposed by Michaelsen et al. (2011) which is a successor of the BPI system (Stilla & Michaelsen, 1997). Two types of productions are feasible: Normal form productions and cluster productions. Only one cluster production rule is used here (third of the *"windows first")*, all others are normal forms. Each production tests a geometrical constraint on the right hand side objects and in case of success infers and assesses a new left hand side object. Primitives must be assessed by the extraction process. The assessments are important because the search of the interpreter is mainly assessment driven. Optional top-down acceleration of the search is possible and recommended. The search can be terminated either by exhausting all possibilities, or after a time limit is reached, or when the first target object is found.

**Decision:** As result of a search a set of non-primitive instances has been accumulated. A decision procedure must be defined selecting from these a single or a small sub-set that can serve as result e.g. for the next step of the analysis. First one or few object classes are picked; here these are Row and Lattice objects. From these first the best object is selected; here the Lattice instance containing most windows, and among these the one that is best assessed by the search process, and if there is no lattice than the best Row instance. All instances similar to this one are suppressed by local inhibition, and then the next best is picked, and so forth. Such rank ordering of accumulated interpretation results follows von Hansen et al. (2006). In the

following Section 3 the best five instances in this rank order are displayed.

## 3. EXPERIMENTS

Images of the temperature on facades in Munich have been obtained such as displayed in Fig. 1. All the systems are evaluated on the primitives displayed in Fig.2. There is little chance in obtaining any usable result with the system "canonical" system presented in Tab. 1. E.g. most of the windows are incomplete, and some primitives are badly displaced. Results of running system 2 and 3 of Table 1 are given below. The computational effort has been fixed by stopping the search after the same fixed time. The five best ranking results are displayed in Figs. 4 and 5.

### 3.1 Results with the "windows first" system

From the 365 L-primitive instances 291 U-structure, 331 Intersect, 139 Rectangle, 31053 Row, and 0 Lattice objects have been inferred.
Figure 4 displays a result obtained after searching the data using the production system that looks for windows by clustering intersections of nearby orthogonal U-structures. With our preliminary parameter setting this can find almost all parts of the salient upper window row (one window in the middle and one on the right margin missing). Row gestalts are indicated by yellow rectangles connected by a blue line. The best gestalt contains nine windows; second best in rank is a four window row on the right side, where the generator is found with good precision also, but the window sizes are a little too small. Third and fourth in rank are coincident with parts of the best, each containing only six windows.
The middle row of windows appears badly disturbed. The fifth in rank Row gestalt sees four wide windows there. It guesses a generator which is in 4/3 harmony with the correct one. This has got to be regarded as failure. Without knowledge on the particular form or size of the windows this cannot be avoided. Looking on the primitives in Figure 3 only even a human observer would be tempted to see such illusory gestalt. Below, on the first floor nothing is found. Since neither the generator nor the window size matches no Lattice object can be instantiated.

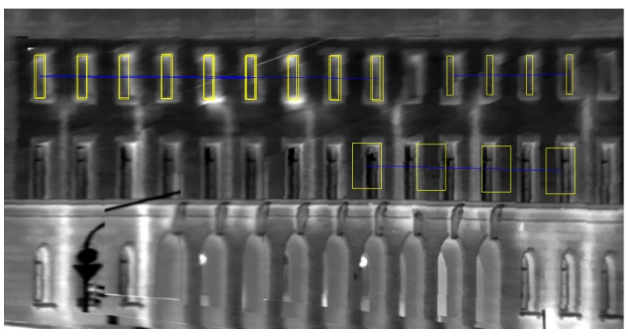

Figure 4. Result with "windows first" productions

### 3.2 Results with the "L-row first" system

From the 365 L-primitive instances 23381 L-Row, 9422 U-Row, 18864 Row, and 176 Lattice objects have been inferred with the same computational effort as in Sect. 3.1 – i.e. 300 seconds in eight parallel threads and resorting the queue after 64 hypotheses, using pure bottom-up data-driven control.

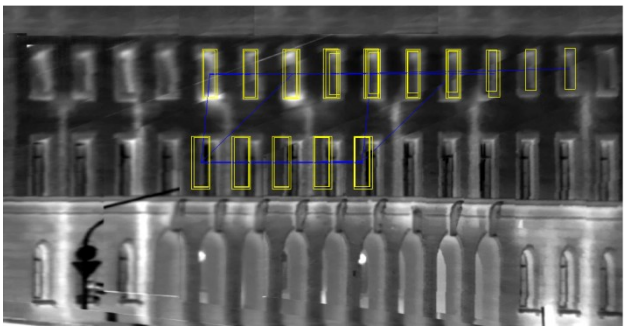

Figure 5. Result with "rows first" productions

Figure 5 displays again the best five resulting instances, two lattices and three rows. Both lattices contain 2x5 windows and the vertical spacing is estimated roughly correct. The upper row alone is less complete than in Section 3.1 (one row of eight members and one of seven with overlap). But there are also results with correct generator and window size on the middle row (one row of five members). Still, the phase of the middle rows is estimated wrongly; they are all displaced left and a little upward. Below, on the first floor again nothing is found.

## 4. DISCUSSION AND CONCLUSION

Obviously, such data are not easy to be parsed. But we can state the following: An object with non-trivial part-of structure – such as a façade – may be decomposed in different ways using different kinds of non-terminal objects in between. Here we have given three different decompositions of façade objects coded as production systems. It turns out that while the systems seem equivalent in the generative right-to-left direction – e.g. for use for façade rendering - they do not yield the same behaviour in the reducing direction, i.e. for recognition. In fact those systems that code natural, common sense decomposition such as *"a façade consists of a stack of rows; each row consists of windows of equal size; each window consist of an upper and a lower U-structure matching; and each such U-structure is made up of two matching L-primitives"* won't work for recognition at all.
If one is determined to use a production system in reducing direction performing recognition by parsing real data – in particular data that contain a large portion of additional clutter primitives and also lots of omissions, such as from thermal mosaics – the decomposition into non-primitives must be chosen with care and different possibilities should be tested including non-intuitive decompositions such as grouping the primitives into rows first and composing the windows afterwards simultaneously on all windows of a row.
The presented results strongly depend on the reliability of the corner detector. A huge number of false and missing detections of the corners can lead to errors in whole algorithm. Poor detection rate of the presented method (Figure 4 and Figure 5) is related to the texture reconstruction techniques, which sticks many images of a video to one image. In this process small distortions cannot be avoided. Considering Figure 3 also a human would have difficulties to recognize windows. Improving the matching between frames would reduce distortions and would deliver better results.
The results also depend on the constraints given in the rightmost column of Table 1. These contain threshold parameters to be chosen by an expert familiar with the issue.

The systems compared here use of course the same constraints – where possible. But some predicates do appear only in one or the other variant, and an unskilled setting in one variant might lead to an unfair comparison. This can be fixed when all such parameters are optimally chosen based on sufficient and representative data labelled by experts.
Minor further dependence of the results may be seen in different parameters used in the interpretation search – e.g. the number of parallel threads, overall time-limit, or top-down settings. The latter is switched off here, in order to improve comparability. And the computational effort was chosen large enough so that little influence can be assumed. Experience shows that also the setting of the parameters of the decision step is of little influence to the result.

### 4.1 Outlook

More experiments are needed, in particular also regarding row (and lattice) grouping according to the constant double ratio principle of pinhole projection. This could be performed on the original images, avoiding any re-sampling. Hopefully the displacement problems are not so bad in that case.
Missing detections might be treated by extrapolation. That is by prolonging the best gestalts and thus generating hypotheses about the position and sizes of the windows with high precision. Then an appearance model can be averaged from the gray values found at the known positions and matched with the values found at hypothesis locations.
The a vertical constraint demanding that window columns should also be vertically grouped may be added, fostering acceptable results on difficult data, such as here in the middle row. And the interpreter is too slow. There should be ways of improving it by hash techniques etc.

## References

Ali, H., Seifert, C., Jindal, N., Paletta, L., and Paar, G. (2007) Window detection in facades, *14th International Conference on Image Analysis and Processing (ICIAP)*, vol. 1, pp. 837–842.

Ali, H., Ahmed, B., and Paar, G., (2008) Robust window detection from 3d laser scanner data. *Congress on Image and Signal processing (CISP'08)*, vol. 2, pp. 115–118.

Burochin, J., Tournaire, O., and Nicolas, P. (2009) An unsupervised hierarchical segmentation of a facade building image in elementary 2d models. *ISPRS Workshop on Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation (CMRT'09)*.

Foerstner, W., Neumann, B., Sara, R., Petrou, M., Hotz, L. (2009): eTRIMS - E-Training for Interpreting Images of Man-Made Scenes. http://www.ipb.uni-bonn.de/projects/

von Hansen, W., Michaelsen, E., Thoennessen, U. (2006): Cluster analysis and priority sorting in huge point clouds for building reconstruction. In: Tang, Y.Y. (ed.): *ICPR 2006,* vol. 1., pp. 23-26.

Hoegner L, Stilla U (2007) Automated generation of 3d points and building textures from infrared image sequences with ray casting. In: Stilla U, Meyer H, Rottensteiner F, Heipke C, Hinz S (eds) PIA07 - Photogrammetric Image Analysis 2007. International Archives of Photogrammetry, Remote Sensing and Spatial Geoinformation Sciences, Vol 36(3/W49B):65-70

Hoegner L, Stilla U (2009) Thermal leakage detection on building facades using infrared textures generated by mobile mapping. Joint Urban Remote Sensing Event (JURSE 2009). IEEE

Iwaszczuk D., Hoegner L., Stilla U. (2011) Detection of windows in IR building textures using masked correlation. In: Stilla U, Rottensteiner F, Mayer H, Jutzi B, Butenuth M (Eds.) Photogrammetric Image Analysis, ISPRS Conference - Proceedings. Lecture Notes in Computer Science, Vol. 6952, Springer: 133-146

Lee, S. and Nevatia, R. (2004) Extraction and integration if window in a 3d building model from ground view images. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, no. 1, pp. 113–120.

Liu, Y., Rauschert, I. (2011): Symmetry Detection from Real World Images. Competition and workshop along with the CVPR
http://vision.cse.psu.edu/research/symmComp/index.shtml

Matsuyama, T., Hwang, V.S.-S., (1990) *Sigma a Knowledge-based Image Understanding System*. Plenum Press, New York.

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.

Mayer, H. and Reznik, S. (2008) Implicit Shape Models, Self-Diagnosis, and Model Selection for 3D Facade Interpretation. Photogrammetrie - Fernerkundung - Geoinformation 2008 (3), S. 187-196

Michaelsen E., Doktorski, L., Lütjen, K., 2011. An accumulating interpreter for cognitive vision production systems. *Pattern Recognition and Image Analysis*, 21 (3), pp. 410-414.

Michaelsen, E., Stilla, U., Soergel, U., Doktorski, L., 2010. Extraction of building polygons from SAR images. Grouping and decision-level in the GESTALT system. *Pattern recognition letters,* 31 (10), pp. 1071-1076.

Pu, W. and Vosselman, G. (2009) Refining building facade models with images. *ISPRS Workshop on Object Extraction for 3D City Models, Road Databases and Traffic Monitoring - Concepts, Algorithms and Evaluation* (CMRT'09),.

Ripperda, N. (2008) Grammar Based Facade Reconstruction using RjMCMC. PFG Photogrammetrie Fernerkundung Geoinformation. Stuttgart: Schweizerbartsche Verlagsbuchhandlung, vol. 2008(2) pp. 83—92

Sirmacek, B. (2011a) Graph Theory and Mean Shift Segmentation Based Classification of Building Facades. *Joint Urban Remote Sensing Event (JURSE'11)*, Muenchen, Germany.

Sirmacek, B., Hoegner, L., and Stilla, U. (2011b) Detection of windows and doors from thermal images by grouping

geometrical features, *Joint Urban Remote Sensing Event (JURSE'11)*, Muenchen, Germany.

Stilla, U., Michaelsen, E. (1997) Semantic Modeling of Man–Made Objects by Production Nets, in: *Automatic Extraction of Man–Made Objects from Aerial and Space Images (II)*, Ed. by Gruen, A., Baltsavias, E. P., Henricsson, O., Birkhaeuser, Basel, pp. 43–52.

Stilla, U. (1993) *Verfahrensvergleich zur automatischen Erkennung in Metall geschlagener Zeichen*. Karlsruhe: Universität, Fakultät für Elektrotechnik, Dissertation

Teboul, O., Simon, L., Koutsourakis, P., and Paragios, N. (2010) Segmentation of building facades using procedural shape priors. *Computer Vision and Pattern Recognition, CVPR'10*, vol. 1, pp. 3105–3112.

Tylecek, R., Sara, R. (2011): Modeling Symmetries for Stochastic Structural Recognition, ICCV 2011.

Wertheimer M. (1923) Untersuchungen zur Lehre von der Gestalt II. Psychol. Forsch., [Principles of perceptual organization, Trans], vol. 4 . In: Beardslee, D., *Wertheimer, M.* (Eds.), Princeton, NJ, 1958, pp. 115–135.