The University of
**Nottingham**

UNITED KINGDOM · CHINA · MALAYSIA

Neuhausen, Marcel and Koch, Christian and König, Markus (2016) Image-based window detection: an overview. In: 23rd International Workshop of the European Group for Intelligent Computing in Engineering, 29 June - 1st July 2016, Krakow, Poland.

**Access from the University of Nottingham repository:**
http://eprints.nottingham.ac.uk/34763/1/Neuhausen_etal_EG-ICE_2016.pdf

# Image-based Window Detection – An Overview

Marcel Neuhausen*, Christian Koch**, Markus König*
*Ruhr-University Bochum, **The University of Nottingham
marcel.neuhausen@ruhr-uni-bochum.de

**Abstract.** Automated segmentation of buildings' façade and detection of its elements is of high relevance in various fields of research as it, e. g., reduces the effort of 3 D reconstructing existing buildings and even entire cities or may be used for navigation and localization tasks. In recent years, several approaches were made concerning this issue. These can be mainly classified by their input data which are either images or 3 D point clouds. This paper provides a survey of image-based approaches. Particularly, this paper focuses on window detection and therefore groups related papers into the three major detection strategies. We juxtapose grammar based methods, pattern recognition and machine learning and contrast them referring to their generality of application. As we found out machine learning approaches seem most promising for window detection on generic façades and thus we will pursue these in future work.

## 1. Introduction

In recent years, the Building Information Modelling concept of using 3 D building models for construction and maintenance purposes has increasingly developed towards models of entire cities as civil engineering goes beyond the scope of single buildings. Moreover, models of buildings are required in other application areas such as virtual and augmented reality, navigation or simulation.

By now, 3 D city models often are publicly available by services like Google Earth or can be requested from land-registries. However, since such models are commonly released as coarse block models, they often lack details, such as the shape of roofs, façades, and façade elements, which are essential in some civil engineering tasks. For most applications these models are insufficient since façade elements constitute the major part of a building. In building recognition as often used for navigation and localization, e. g., tourist guidance (Ali et al., 2007), façade elements and windows in particular play a major role. The same applies to other tasks as simulations of the stability of buildings. Especially mechanized tunnelling projects may benefit from more detailed models as they ease risk assessment of settlement-induced damages inflicted to urban structures (Obel et al. 2016). In particular, the so-called wall-openings coefficient, meaning the ratio of openings to wall, is a crucial factor in risk assessment. Openings within a solid wall decrease its stiffness and thus increases the vulnerability to settlements. Windows, in this regard, account for the major part of wall-openings.

Façade reconstruction including window detection is, hence, a crucial issue which is by far not solved (Musialski et al., 2013). In this paper, we present an overview of the major strategies for automated window detection on façades and discuss their advantages and limitations regarding tunneling projects.

## 2. Related Work

In the past decades an extensive body of literature arose concerning building reconstruction and particularly façade reconstruction. The approaches made in this field of research are highly diverse and manifold. However, some effort to categorize existing methods has already been made. Usually approaches are mainly differentiated by the type of input data. Baltsavias (2004),

Brenner (2005), and Haala and Kada (2010) provide surveys focussing on airborne imagery and laser scans. Although rudimentary building reconstruction is possible, this data source is certainly not well suited for façade elements recognition. The papers of Baltsavias (2004) and Brenner (2005), hence, concern the detection of large structures like buildings and streets and the reconstruction of simple buildings' shapes, respectively. On the contrary, Haala and Kada (2010) also include approaches using terrestrial laser scans to gain more detailed geometrical information. They also present approaches developing façade's texture mapping to reconstructed 3 D models of buildings. In the comprehensive survey of approaches on building and façade reconstruction of Musialski et al. (2013), the authors cover both airborne and terrestrial imagery and laser scans. Additionally, they differentiate between automatic and semi-automatic approaches. Although they discuss approaches of pattern detection, matching as well as façade parsing, the paper focuses on reconstruction, while façade element detection is only briefly mentioned.

## 3. Approaches

In the following we concentrate on approaches based on terrestrial imagery, as we focus on window detection in this paper. Indeed laser scan generated point clouds simplify the detection of façade elements and especially windows (van Gool et al., 2007). However, the acquisition is very time-consuming and expensive as for risk assessment in tunnelling projects models of entire cities or at least of areas along the tunnelling alignment are required. Façade images, contrariwise, are mostly publicly available from Google Street View and other services or may be easily gathered manually or even semi-automatic by drone flights as Freimuth and König (2015) showed.

Most approaches using ground view imagery can roughly be categorized by three main strategies, although some combine multiple strategies, which interdicts a clear disjunction in general. However, for comparison reasons the following partition by strategies facilitates a contrasting juxtaposition. We, therefore, classify multiple strategy approaches by their prominent method.

The remainder is structured as follow: section 3.1 gives an overview of approaches applying grammars to façade images. Section 3.2 summarizes pattern recognition approaches and section 3.3. relates to window detection using machine learning techniques. Accordingly, in section 4 we discuss the limitations of these strategies with respect to their generality of application. Finally, section 5 concludes our findings and makes proposals for further research.

### 3.1 Grammar Based Approaches

The most prominent methods in grammar based façade reconstruction are shape and splitting grammars. These grammars make use of assumptions to symmetry such as grid-like arrangement of façade elements or their usually rectangular shapes. In the context of several comprehensive publications Ripperda and Brenner (2009) proposed an approach to apply a formal grammar to façades for automatic reconstruction. Exploiting the assumptions mentioned above, Ripperda (2008) designed a context-free grammar splitting a façade into its symmetric and repeating areas, respectively. To optimize the grammar's segmentation of a façade, the authors proposed a *reversible-jump Markov chain Monte Carlo* (rjMCMC) method sampling over numerous runs (Ripperda and Brenner, 2006). Thus, façade elements are automatically derived. To improve their approach, Ripperda and Brenner (2007) extended their approach by a data driven rule extension. Müller et al. (2007) proposed an approach which splits the façade into floors and tiles by means of symmetry information. Ensuing, it subdivides the tiles further

into smaller rectangular regions and clusters them into groups. The extracted architectural elements can then be matched with 3 D objects of a database. For façade reconstruction the windows' depth information has to be explicitly specified via interaction with users. Compared to the former, Teboul et al. (2010) defined a more detailed shape grammar that not only splits a façade into rectangles but models sematic relationships of façade parts and their elements, e. g., the relationship of attic to roof or window. The authors understand façade segmentation by use of a grammar as finding a specific sequence of rules. Thus, they use an energy minimization scheme determining the best fitting manifestation on particular images. In a last step, they classify each segmented part of the image by a randomized decision forest which labels each pixel according to its probability of class affiliation. In a subsequent approach, Teboul et al. (2013) are content with a simple shape grammar that only allows iteratively splitting rectangles in smaller ones in order to focus on improving the parsing of facades. The parsing is then done by a reinforcement learning algorithm. As the one of Ripperda and Brenner (2006), many approaches fall back on sampling algorithms like MCMC for the optimization of parse trees. Riemenschneider et al. (2012) pointed out that sampling such a large solution space holding a complex structure is not a sufficient method. For this reason they propose the *Cocke-Younger-Kasami algorithm* (CYK) for efficiently parsing context-free grammars. In their approach the authors define a grammar allowing for irregular lattices splitting the façade into rectangular regions of different sizes. Similar to Teboul et al. (2010) they assign the most probable label of class affiliation to each pixel using a merit function that describes the likelihood of each pixel belonging to a class.

To stochastically sample parse trees or finding the most probable parse tree of an ambiguous grammar for a given input, probabilistic grammars provide the method of choice. In contrast to deterministic grammars these extend former by stochastic components. Each production rule in a grammar's set of rules additionally possesses a likelihood of appearance that, as consequence, induces a probability distribution on the generated output. Assuming façades to be structured highly regular and symmetric like it can be found at large office buildings, Alegre and Dellaert (2004) developed a stochastic context-free grammar which only contains horizontal and vertical splits of the façade. This is used to derive a Bayesian stochastic model which depicts a hierarchical partitioning of disjoint blocks of façade images. For approximating the posteriors of partitions the authors apply MCMC sampling. Tyleček and Šára (2012) propose a novel approach by relinquishing to specify a grammar with an explicit set of production rules. Instead of this they develop a stochastic model in a tree-like structure acting as grammar. For optimizing the grammar's manifestation they use a general rjMCMC framework.

## 3.2 Pattern Recognition

As windows are mostly assumed to be rectangles arranged in a grid-like manner parallel to the façade's contour, some approaches detect windows by patterns of horizontal and vertical edges in façade images. Lee and Nevatia (2004) proposed an approach to window detection on rectified façade images which was later resumed by Meixner et al. (2011) for interpreting façade elements. In the image, they project horizontal and vertical edges, respectively, and interfered resulting histograms with each other. At emerging peak's location window candidates are supposed and finally verified based on the beforehand extracted edges. Since this method only detects rectangular shaped windows, Lee and Nevatia (2004) extended their approach to arch windows by allowing the detected window's top edge to become curved and again verifying it with the image's edges. To apply window detection to more irregular facades, Recky and Leberl (2010) picked up Lee and Nevatia's (2004) concept. They deduce a subdivision of the facade into levels from vertical edge projection of the façade image and subsequently applied a horizontal edge projection separately on each level to further divide the levels into blocks. By

applying a *k*-Means clustering by color to these blocks, Recky and Leberl (2010) identify if a block belongs either to the façade or to a window. Final window detections are obtained by joining adjacent window blocks. Korah and Rasmussen (2008) discussed a method that focuses more on the grid-like alignment of windows on a façade. Therefore they hypothesize rectangles from all pairs of parallel edges in the image resulting in a large quantity of hypotheses. As windows are supposed to be aligned in a grid, a best possible lattice of a subset of the extracted rectangles is constructed by means of a MCMC optimization procedure.

Instead of explicitly assuming the shape of windows to be rectangular, the *implicit shape models* (ISM) introduced by Leibe et al. (2004) combine the recognition and segmentation of objects in one process. These models consist of a codebook prototypically defining local features and a probability distribution determining their spatial positions. Reznik and Mayer (2008) applied ISM for window detection. Interest points in combination with their geometric arrangement around the window's center constitute the ISM which is then learned from a manually labelled training set. They generate window hypotheses by cross correlating image patches around interest points with the training data. Retrieved good hypotheses are used in a self-diagnosis manner to validate weak hypotheses. For refinement of the windows' alignment, model selection is done based on the assumption that windows are arranged in rows and columns.

Usually façade elements occur recurrently on facades. Thus, elements' alignment on facades can be understood as repeating patterns. An approach to window detection as a subtask of façade reconstruction, relying on repetitions on the façade is proposed by van Gool et al. (2007). They tackle the problem by a bifid strategy approach depending on the quality of the input images. As for their method of fully automatic reconstruction, sufficient perspective effects in the images are required elsewise they switch to the approach of Müller et al. (2007) (see section 3.1). If sufficient perspective is available, interest points occurring repeatedly over the façade are computed and grouped to infer the camera's calibration. Window detection is then done by energy minimization using graph-cut optimization. In a concluding refinement step van Gool et al. (2007) use shape priors for vertical and horizontal alignment.

### 3.3 Supervised Machine Learning

In supervised machine learning in the context of computer vision object recognition is achieved by training a classifier using image features of a labelled data set. By training, classifiers learn a function to separate object classes. In this field *support vector machines* (SVM), building a kind of algorithms able to cope with high dimensional inputs, are widely deemed to yield good classification results (Yang et al., 2012). Inspired by Lee and Nevatia (2004) and Recky and Lerberl (2010), Haugeard et al. (2009) utilize the fact that windows mostly are aligned to floors and, thus, they split facades into floors by thresholding vertical edge histograms. On each floor they project horizontal edges to a histogram to infer window candidates. Edges of these candidates are extracted from the image and represented in a graph. Windows are then classified by a SVM using a kernel operating in inexact graph matching such that windows emerge as a sub-graph of the graph of all contours extracted from the façade image (Musialski et al., 2013).

*Bag-of-visual-words* (BoW) approaches offer another way to handle multiple different image features at once by integrating them into sparse vectors of feature occurrences. These can then be fed into a classifier. Csurka and Perronnin (2008) described a method making use of BoW for a pixel-wise semantic segmentation of façade images into façade elements. The approach relies on the detection of image patches out of which local image features are extracted and combined to a BoW representation. Classified by sparse logistic regression, the results are assigned back to the level of pixels. Fröhlich et al. (2010) improved this method. Instead of extracting image patches they compute local color features obtained by dense sampling.

Classification is then done by feeding the gained feature vectors into a randomized decision forest.

The previously described classification approaches aim on high dimensional vectors as input. Feature responses have to be explicitly arranged in vector shape. Boosting algorithms contrariwise allow for deriving a strong classifier directly from image features. Originally designed for face detection, Viola and Jones (2001) illustrate a boosted cascade of simple features as a method for object detection which uses image features as classifiers. Features are treated as weak learners and fused by boosting in a way that they form a strong classifier. Ali et al. (2007) applied the approach of Viola and Jones (2001) to detect windows on facades. Divers thresholded Haar-like features (Oren et al., 1997) on different positions within the image act as a pool of weak learners out of which an adapted AdaBoost algorithm (Freund and Schapire, 1997) selects a subset and arranges them in cascading stages to form a strong classifier. For detection a search window is slid over the image and resulting patches are presented to the classifier. To enable detection of different façade elements Drauschke and Förstner (2008) describe an approach similar to the one of Ali et al. (2007) which also builds a strong classifier by AdaBoost. Though, the novel idea is serving different kinds of image features at once to the AdaBoost algorithm. Since the detection and classification of façade elements using supervised learning algorithms are highly dependent on which features are used, Yang et al. (2012) proposed an empirical study on feature evaluation in this context. For region-wise labelling the façade's elements they used randomized decision forests based on different features as classifiers. Martinović et al. (2012) present a similar approach like Yang et al. (2012) but in order to substantiate classification results they augment the decision tree's output with further processing results. In their three layered approach they make use of an oversegmentation of a building's façade image. In a first step a recursive neural network merges oversegmented regions into objects. These hypotheses are then combined with the classification results of a decision tree of integral channel features (see Dollar et al., 2009) learned by discrete AdaBoost. Finally they refine detections by exploiting weak architectural principles, e. g., vertical and horizontal alignment or co-occurrence of objects.

It may be necessary to detect more than one single object class. In terms of façade element recognition this could mean distinguishing between windows, doors, and maybe even more categories. Fidler et al. (2006) established an approach to multiple class categorization based on hierarchical models reducing the computational complexity which such a task usually implicates. Mačák and Drbohlav (2011) adapted the generic approach given by Fidler et al. (2006) for the use of window detection. They applied a difference of Gaussians to extract edge features forming the lowermost layer of a hierarchical window model. From there on they learn following layers such that the respectively superior layer is a composition of instances in local neighborhoods of the underlying layer. The resulting model is then used to detect windows in façade images.

## 4. Discussion

To apply window detection approaches to buildings of different architectural style or even to cities in various countries the methods should well generalize as facades' and windows' appearance vary in both geographical region and year of construction. In the following we discuss the assumptions and limitations of the methods presented above.

**Grammar Based Approaches.** For all grammar based approaches an optimization of the parse tree is inevitable for each presented façade which is a time-consuming procedure if applied on a high quantity of façade images as it is the case in our research. Optimization may either be

done by sampling over all possible trees or by relying on special parsing algorithms like CYK. Especially sampling methods limit the field of application as they are, on the one hand, only applicable to grammars consisting of small rule sets and, on the other hand, potentially get stuck in local optima (Riemenschneider et al., 2012). This mostly leads to simple shape grammars consisting of less rules which only split façade images into rectangles as windows and other façade elements often are assumed to be of rectangular shape. On the other hand, defining more complex grammars may be highly dependent on expert knowledge about building construction. This seems advantageous as there are explicit universal constructions rules. The aesthetic sensibility, however, differs in country which complicates deducing general rules. Moreover, the more a grammar's set of rules is refined the stronger it gets biased from specific façade types and thus is no longer generally applicable as architectural style varies not only among country but also among geographical regions and even among year of the buildings' construction (Haala and Kada, 2010). However, in case of simple shape and splitting grammar as well as in case of complex grammars these approaches only achieve a rough segmentation of façade images. This has to be processed in further steps to both gain proper object contours and determine the objects' class using either image processing or machine learning techniques. Since mostly also a pre-processing of the images is required to obtain symmetry or edge information the grammar and the sampling algorithm, respectively, can be applied to, it is questionable if a detour over grammars yields to be a promising solution.

**Pattern recognition.** Approaches focusing on pattern recognition usually rely on a couple of assumptions. Assumptions on the distribution of pixels' intensities, meaning that windows are darker than the surrounding façade, often yield erroneous results (Musialski et al., 2013). However, other approaches make assumptions referring to both the windows' shape and their alignment on the façade. More precisely, like in the approach of van Gool et al. (2007) it is often assumed that windows occur on facades as multiple elements of the same type, i.e. same size and shape, and repeat in fixed displacements in both horizontal and vertical direction. This practice yields remarkable good results as long as facades are, on the one hand, highly regular structured and, on the other hand, broad enough such that façade elements form the presumed patterns. This may be valid for structures like large office buildings but it does not hold for the majority of structures in typical European inner city areas and adjacent regions. In apartment buildings the windows of staircases often deviate in size from residual windows and also hardly break the pattern spanned by them. Detection of such asymmetric patterns cannot be well handled as van Gool et al. (2007) and Meixner et al. (2011) pointed out. Furthermore, most buildings are rather small resulting in less repetitions which also complicates detection. Another drawback Reznik and Mayer (2008) highlighted is that the images' resolution and potential partially occlusions of windows by trees, traffic signs and similar highly impacts the detection's quality. This comes into account right then if publicly available images are used, such as ones of Google Street View, as they are, firstly, usually taken from the street sight such that occlusions occur unavoidably and, secondly, undistorted whereby block artefacts may arise and lower the resolution.

**Supervised Machine Learning.** In contrast to the aforementioned strategies, supervised machine learning approaches rely neither on assumptions of the windows' alignment on the façade as they detect an object by its own characteristic instead of the interplay with other objects, nor do they need explicit prior knowledge about construction details. This is advantageous as both result in a loss of universal applicability. Nevertheless, the detection rate of machine learning algorithms is, amongst others, constrained by the type of image features whose responses are fed in. Features extracted from the input images, thus, should be significant and relevant. This constitutes a challenge in itself, as windows are usually rather poor of outstanding features, except for their frames. However, there are already approaches like Yang

et al. (2012) to cope with this issue. Certainly, for training a supervised machine learning algorithm a labelled set of training images is required which definitely is time-consuming in gathering and setting up correctly. Since these algorithms, however, are able to generalize well, this procedure has only to be made once to detect windows in facades of different construction years or architectural styles.

## 5. Conclusion

We highlighted the meaning of window detection to improve existing 3D city models for the use in civil engineering tasks such as tunnelling. By comparing several approaches of different strategies, we discussed application limitations and concluded that there is, by now, no adequate approach for universal use. Though we showed that machine learning techniques may yield promising results in future since they avoid restricting assumptions and offer an excellent ability of generalization.

We, thus, will focus on machine learning techniques to detect windows in our future work. Particularly, we believe that the approach of Ali et al. (2007) may be developable. An extension of this approach by a pre-processing of the image easing the feature extraction and unifying the windows' characteristics as well as a post-processing to reject misclassification may increase the detection rate. Moreover, we will adapt the approach by substitution of Voila and Jones' (2001) detector by a Soft Cascade detector (Bourdev and Brandt, 2005) as this may also increase the detection rate. For training and testing issues we recommend the eTRIMS data set (Korč and Förstner, 2009) as it provides extensive samples and is already used in several approaches which supports the comparability.

By now, for windows detection mostly edges features are used, in other respects windows seem to be poor of features. Additionally to Yang et al.'s (2012) work, we consider a further evaluation of image features to be desirable since in most facades there are other objects providing even stronger responses on edge features (Recky and Leberl, 2010). Thus, finding sets of features which significantly describe the windows' characteristics is a major task. Features found in this context can also contribute to improve the results of the above mentioned approach.

In recent years, *deep neural networks* (DNN) attract more and more attention in computer vision for object classification tasks. Szegedy et al. (2013) proposed a method to not only classify objects but also locate them within an image. As such approaches often outperform classical computer vision algorithms, it would be desirable further pursue former and apply them to window detection.

## Acknowledgement

## References

Alegre, F., Dellaert, F. (2004). A Probabilistic Approach to the Semantic Interpretation of Building Facades. In CIPA International Workshop on Vision Techniques Applied to the Rehabilitation of City Centers.

Ali, H., Seifert, C., Jindal, N. and Paletta, L. (2007). Window Detection in Facades. In Proceedings of the IEEE Computer Society Conference on Image Analysis and Processing, pp. 837-842.

Baltsavias, E. P. (2004). Object extraction and revision by image analysis using existing geodata and knowledge: current status and steps towards operational systems. In ISPRS Journal of Photogrammetry and Remote Sensing 58, pp. 129–151.

Bourdev, L., Brandt, J. (2005). Robust Object Detection Via Soft Cascade. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 236-243

Brenner, C. (2005). Building reconstruction from images and laser scanning. In International Journal of Applied Earth Observation and Geoinformation 6 (3–4), 187–198.

Brenner, C., Ripperda, N. (2006). Extraction of Facades using rjMCMC and Constraint Equations. In Proceedings of the ISPRS Symposium on Photogrammetric Computer Vision, pp. 155-160.

Csurka, G., Perronnin, F. (2008). A Simple High Performance Approach to Semantic Segmentation. In Proceedings of the British Machine Vision Conference, pp. 22.1-22.10.

Dollar, P., Tu, Z., Perona, P. and Belongie, S. (2009). Integral channel features. In Proceedings of the British Machine Vision Conference, pp. 91.1-91.11.

Drauschke, M., Förstner, W. (2008). Selecting Appropriate Features for Detecting Buildings and Building Parts. In Proceedings of the Congress of the International Society for Photogrammetry and Remote Sensing, pp. 447-452.

Fidler, S., Berginc, G. and Leonardis, A. (2006). Hierarchical statistical learning of generic parts of object structure. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pages 182–189.

Freimuth, H., König, M. (2015). Generation of Waypoints for UAV-Assisted Progress Monitoring and Acceptance of Construction Work. In Proceedings of the International Conference on Construction Applications of Virtual Reality. Available from: http://www.convr2015.com, accessed February 2016.

Freund, Y. and Schapire, R.E. (1997). A desicion-theoretic generalization of on-line learning and an application to boosting. In Journal of Computer and System Science 55, pp. 119-139.

Fröhlich, B., Rodner, E. and Denzler, J. (2010). A Fast Approach for Pixelwise Labeling of Facade Images. In Proceedings of the IEEE Computer Society International Conference on Pattern Recognition, pp. 3029-3032.

van Gool, L., Zeng, G. van den Borre, F. and Müller, P. (2007). Towards Mass-Produced Building Models. In Proceedings of Photogrammetric Image Analysis Conference, pp. 209-220.

Habib, A. F., Zhai, R. and Kim, C. (2010). Generation of Complex Polyhedral Building Models by Integrating Stereo-Aerial Imagery and Lidar Data. In Photogrammetric Engineering & Remote Sensing 76 (5), pp. 609-623.

Haala N., Kada, M. (2010). An update on automatic 3D building reconstruction. In ISPRS Journal of Photogrammetry and Remote Sensing 65, pp. 570-580.

Haugeard, J., Philipp-Foliguet, S., Precioso, F. and Lebrun, J. (2009). Extraction of Windows in facade using Kernel on Graph of Contours. In Proceedings of the 16th Scandinavian Conference on Image Analysis, pp. 646–656.

Korah, T., Rasmussen, C. (2008). Analysis of Building Textures for Reconstructing Partially Occluded Facades. In Proceedings of the European Conference on Computer Vision, pp. 359-372.

Korč, F., Förstner, W. (2009). eTRIMS Image Database for interpreting images of man-made scenes. Technical report, Department of Photogrammetry, University of Bonn.

Lee, S.C., Nevatia, R. (2004). Extraction and Integration of Window in a 3D Building Model from Ground View images. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 106-113.

Leibe, B., Leonardis, A. and Schiele, B. (2004). Combined Object Categorization and Segmentation with an Implicit Shape Model. In Proceedings of the Workshop on Statistical Learning in Computer Vision, European Computer Vision Conference, pp. 17-32.

Mačák, J., Drbohlav, O. (2011). Hierarchical shape model for windows detection. In Proceedings of the ÖAGM/AAPR Workshop. Available from: http://oagm2011.joanneum.at, accessed February 2016.

Martinović, A., Mathias, M., Weissenberg, J. and van Gool, L. (2012). A Three-Layered Approach to Facade Parsing. In Proceedings of the European Conference on Computer Vision, pp. 416-429.

Meixner, P., Leberl, F. and Brédif, M. (2011). Interpretation of 2D and 3D Building Details on Facades and Roofs. In Proceedings of the ISPRS Conference on Photogrammetric Image Analysis, pp. 137-142.

Müller, P., Zeng, G., Wonka, P. and van Gool, L. (2007). Image-based procedural modeling of facades. In ACM Transactions on Graphics 26 (3), pp. 85.1-85.9

Musialski, P., Wonka, P., Aliaga, D.G., Wimmer, M., van Gool, L. and Purgathofer, W. (2013). A Survey of Urban Reconstruction. In Computer Graphics Forum 32 (6), pp. 146-177

Obel, M., Mark, P., Neuhausen, M., König, M. and Schindler, S. (2016). Settlement risk assessment by means of categorized surface infrastructure. In Proceedings of the International Conference on Structural Analysis of Historical Constructions. [Accepted]

Oren, M., Papageorgiou, C., Sinha, P., Osuna, E. and Poggio, T. (1997). Pedestrian detection using wavelet templates. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 193-199.

Recky, M., Leberl, F. (2010). Windows Detection Using K-means in CIE-Lab Color Space. In Proceedings of the IEEE Computer Society International Conference on Pattern Recognition, pp. 356-360.

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

Riemenschneider, H., Krispel, U., Thaller, W., Donoser, M., Havemann, S., Fellner, D. and Bischof, H. (2012). Irregular lattices for complex shape grammar facade parsing. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1640-1647.

Ripperda, N. (2008). Grammar Based Facade Reconstruction using rjMCMC. In Photogrammetrie Fernerkundung Geoinformation 2, pp. 83-92.

Ripperda, N., Brenner, C. (2007). Data driven rule proposal for grammar based facade reconstruction. In Proceedings of the ISPRS Conference on Photogrammetric Image Analysis, pp 173-178.

Ripperda, N., Brenner, C. (2009). Application of a Formal Grammar to Facade Reconstruction in Semiautomatic and Automatic Environments. In Proceedings of the AGILE Conference on Geographic Information Science. Available from: http://www.agile-online.org accessed February 2016.

Szegedy, C., Toshev, A. and Erhan, D. (2013). Deep Neural Networks for Object Detection. In Advances in Neural Information Processing Systems 26, pp. 2553-2561.

Teboul, O., Kokkinos, I., Simon, L., Koutsourakis, P. and Paragios, N. (2013). Parsing Facades with Shape Grammars and Reinforcement Learning. In IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7), pp. 1744-1756.

Teboul, O., Simon, L., Koutsourakis, P. and Paragios, N. (2010). Segmentation of Building Facades Using Procedural Shape Priors. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 3105-3112.

Tyleček, R., Šára, R. (2012). Stochastic Recognition of Regular Structures in Facade Images. In IPSJ Transactions on Computer Vision and Applications 4, pp. 63-70.

Viola, P., Jones, M. (2001). Rapid Object Detection Using a Boosted Cascade of Simple Features. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition,             pp. 511-518.

Yang, M, Förstner, W. and Chai, D. (2012). Feature Evaluation for Building Façade Images - An Empirical Study. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B3, pp. 513-518.