Improved Window Detection in Facade Images

64

Marcel Neuhausen and Markus König

Abstract

A variety of applications require detailed geometric information about a building's hull. In particular, windows are of high interest in this context. However, common 3D models only roughly outline existing buildings' geometrical outward appearance. The field of application is, thus, severely limited by now. In this paper, we propose an approach to the automatic detection of windows in facade images complementing common building models by information derived from detected windows to enable new opportunities and widen the spectrum of applications. Therefore, we apply a soft cascaded classifier to identify windows in patches of facade images. Moreover, a postprocessing is applied to the detections. We initially refine their dimensions and alignment on the facade. From these we infer so far non-detected windows. With an overall detection rate of 95 and 97% precision our proposed detection system yields sufficient results for complementing existing 3D building models by information of the detected windows.

Keywords

Window detection • Cascaded classifier • Urban environments • 3D building reconstruction

64.1 Introduction

In recent years, 3D models of existing buildings have become prominent in several application areas. In entertainment industry these are used to create high immersive virtual sceneries of realistic cities for movies and video games. Online map services integrate 3D urban models into their maps for an improved navigation experience. Such models also proved suitable in various civil engineering tasks. In urban planning, the administration of cities and their infrastructures benefit from georeferenced virtual buildings as well as the management of inner city construction sites and related logistics. Beyond visualization purposes, detailed geometric information about a building's hull can be used to support the automation of planning and assessment tasks or to enable simulations. The spectrum of applications demanding such information of existing buildings is manifold. In this context, especially windows often are of high relevance. As these constitute thermal bridges, windows are a necessary feature in energy efficiency rating. Moreover, windows mainly affect a facade's stiffness so that Earthquake simulations and risk analyses for settlement induced damages demand these to determine a building's stability.

However, 3D building models which are publicly available from land registry offices or web services like OpenStreetMaps often lack relevant details. These commonly only comprise georeferenced geometrical information of buildings in form of coarse block models which are sometimes extended by simplified roof shapes (see Fig. 64.1). Complementing existent building models by windows, thus, is inevitable to facilitate further applications. Hence, we propose an approach to

M. Neuhausen (✉) · M. König
Ruhr-University Bochum, Bochum, 44780, Germany
e-mail: marcel.neuhausen@ruhr-uni-bochum.de

M. König
e-mail: koenig@inf.bi.ruhr-uni-bochum.de

I. Mutis and T. Hartmann (eds.), *Advances in Informatics and Computing in Civil and Construction Engineering*,
https://doi.org/10.1007/978-3-030-00220-6_64

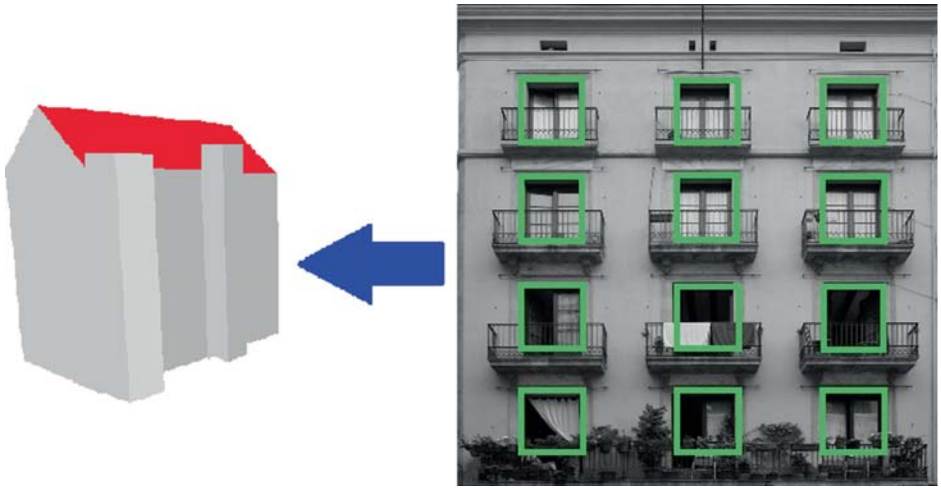

Fig. 64.1 3D block model of a building as commonly provided by public services. Complementing these with window information may facilitate various applications

window detection in facade images sufficiently supplementing such models to facilitate the automation of civil engineering tasks like energy efficiency ratings. We show that our detection system based on a soft cascaded classifier yields reliable results despite the windows' poorness of features. For this, we train the classifier on a set of windows taken from facade images of buildings in different countries and calibrate it towards a low false positive rate. For detection, we apply a sliding window approach which scans a facade image and passes patches to the classifier. To reduce the windows' variety in appearance emerging from perspective distortion due to different camera angles, we operate on rectified images for classifier training as well as for detection. We subsequently improve our results via a postprocessing based on the set of obtained detections. Therefore, we initially refine the geometry of detected windows by adjusting their dimension towards actual edges in the facade image. Ensuing, we cluster them regarding size and appearance. We align the windows within the clusters in a grid-like manner to infer potential positions of so far non-detected windows from this refined set. At these positions, we apply a second classifier which is calibrated less conservatively to also identify windows possessing less clear evidence.

We evaluate our detection system on rectified facade images of buildings from different countries. We and that our detector itself already yields adequate results to provide an initial assessment of a building. Nevertheless, the geometrical accuracy as well as the provided detection rate are too low to complement given models with the detected windows to a meaningful extent. Though, by refining the detections and inferring further windows, we can increase both detection rate and accuracy so that complementing such models by the resulting windows becomes reasonable.

64.2 Background

Building reconstruction in general became a wide field of research in the last decades and a large body of literature arose around it. Nevertheless, window detection as a subtopic has, by now, sparsely been addressed and still remains a challenging task.

Many applications such as simulations or analyses focus on several buildings in a city district. Thus, input data for detection has to be available for large areas. Aerial imagery and laser scans which are often used in building reconstruction satisfy this need but are impractical for window detection due to the oblique top view and spatial resolution. Becker and Haala [4] proposed an approach relying on terrestrial LiDAR point clouds. Exploiting that emitted light of the LiDAR sensor passes through glas, they identify windows by no-measurement areas in the facade's plane. Due to reflections of the laser on window panes, Ali et al. [1] improved this method by an adaptive distance threshold of adjacent measurements. Although, LiDAR based approaches o er reasonable detection methods for windows, point clouds are not publicly available for large areas yet and ought to be acquired at first. Albeit StreetMapper [3] may facilitate data gathering of point clouds for urban areas, it remains to be costly and involves high effort.

For this reason, ground view images qualify best for our purpose as they are already available from web services for most regions or, elsewise, can be gathered at low costs and effort. Grammar based approaches split facade images taken from ground view perspective into increasingly smaller semantic units to detect windows. Formal grammars are applied to subdivide a facade according to symmetries and repetitions [10, 11]. Teboul et al. [12], alternatively, developed a shape grammar including semantic relationships between facade components into the subdivision process. These approaches yield sufficient results on highly regular facades as they depend on symmetries and uniformity. Another approach superimposes histograms of horizontal and vertical edges [7] so that resulting peaks indicate the windows' locations. Haugeard et al. [6] advanced this by superimposing histograms separately for each potential row of windows. However, Meixner et al. [8] found that such approaches perform well if facades are plain and regular but will fail as soon as facades' complexity increases due to irregular window patterns or extensions like balconies and awnings.

Yang et al. [14] proposed an approach to identify windows by their inherent characteristics regardless of the facades' appearance. For this, they apply a randomized decision forest and evaluate the suitability of various image features. Ali et al. [2] proposed a similar approach applying a Viola-Jones detector [13] to facade images. However, their reported detection rates are ineligible for 3D building model enrichment. In related work [9] the cascaded classifier of the Viola-Jones detector is substituted by a soft cascaded approach [5]. Although this improves the detection rate, results are still not applicable for our purpose.

64.3 Methodology

In this paper, we propose a window detection system providing reasonable results to enable accurate 3D building model enrichment. Our system comprises three main components as illustrated in Fig. 64.2. First, we preprocess the facade images to reduce variability in the input images. We, then, apply our detector scanning an image and passing image patches to the soft cascaded classifier for identification. Overlapping detections are eventually merged to a single window hypothesis. Finally, a postprocessing is applied which initially refines current detections. Potential further window positions on the facade are, then, inferred based on the alignment of the refined detections. Ensuing, we investigate the inferred regions more closely to also detect less recognizable windows. In the following we describe the three components in more detail.

Although the soft cascaded classifier in general is robust with respect to high variability of positive examples [5], in combination with the windows' poorness of image features, the detection is highly complicated. An elimination of unnecessary variability beforehand is, hence, reasonable. Different camera angles immensely contribute to the variation due to the perspective distortion. Thus, a normalization of the windows' appearance improves the detection results [9]. For this, we semi-automatically rectify the facade images before detection such that windows are ensured to be of rectangular shape.

For detection, we apply a sliding window approach as proposed by Viola and Jones [13]. A sub-window slides across the entire image in multiple scales and step sizes. After each run through the image the sub-window is scaled by a factor of 1.25. Depending on this, the shifting step size starting with $\Delta = 1.0$ scales by $\lfloor s\Delta + 0.5 \rfloor$, where s denotes the current sub-window's scale. As windows are of various shapes, different kinds of sub-windows are required. We found that using two rectangular sub-windows—one upright and one lying—in addition to a squared sub-window allow to sufficiently cover most window shapes occurring on common urban facades. At each sub-window's position the underlying image patch is passed to the classifier resolving the presence of a window in each patch. For this, we apply a soft cascaded classifier which is composed of serially connected weak classifiers. Since a window mainly consist of a transparent pane and a frame, an identification by the frames' edges and changes in brightness between frame and pane seems reasonable. Thus, we consider to use thresholded Haar-like features as weak classifiers in our setup. Positively classified regions in the image are collected and treated as window hypotheses. Since the classifier is insensitive to small translations and changes in scale, there may be

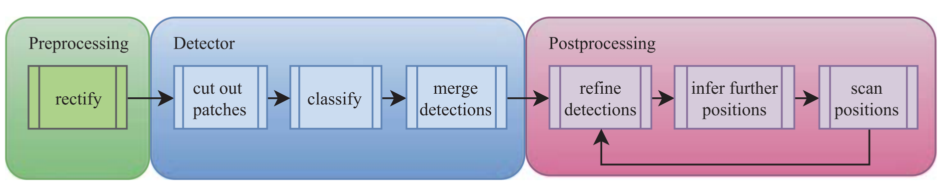

Fig. 64.2 Concept of the entire detection system proposed in this paper. Facade images are initially rectified. Then, the detector is applied. Finally, we refine the detections and search for further windows

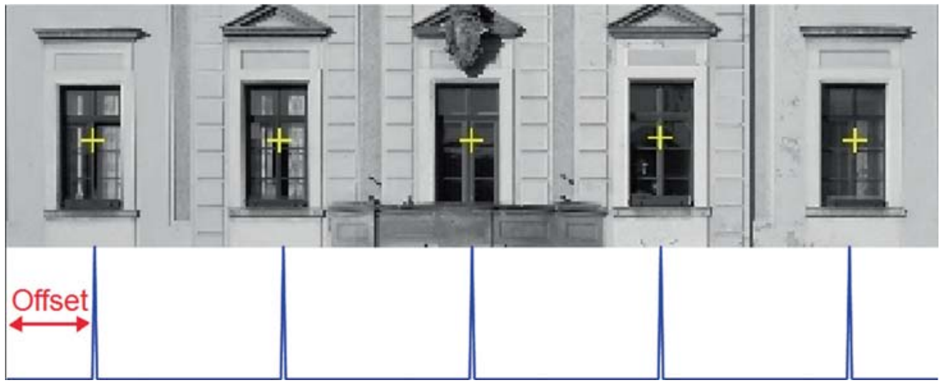

Fig. 64.3 Dirac comb (blue) shifted by an offset (red) to match the windows' centers (yellow crosses)

multiple overlapping detections around the actual windows' locations. Similar occurs if actual windows are aligned in close proximity. To distinguish between different windows but at the same time aim for only one detection per window, we subsequently merge detections which at least overlap to 60%. Position and size of such overlapping detections are averaged resulting in a single window hypothesis.

The postprocessing serves a dual purpose. On the one hand, we refine the dimensions of the current window hypotheses and, on the other hand, we search for further windows in the facade image which have not been detected so far. Due to the detection method, the detector's results match the actual windows edges only approximately. To provide precise detections, we refine the hypotheses' dimensions towards actual edges in the image. As proposed by Lee and Nevatia [7] we perform a one dimensional search for supporting image edges for each side of a window hypothesis. For this, we generate a candidate edge from each hypothesis' side. While we translate these edges pixelwise in orthogonal direction, we search for best evidence. Supporting evidence consists of line segments which confirm the generated line's position. These segments are projected onto the candidate edge and the particular coverages are summed up to an evidence score. By this refinement based on a single window scope, the detections become more accurate. Non-window edges on the facade may, though, interfere a more exact adjustment. Considering the appearance of windows which are of the same kind and their alignment on the facade, allows to further refine the hypotheses' dimensions and position. For this purpose, hypotheses are clustered regarding their texture and size as proposed by Lee and Nevatia [7]. After clustering, the sizes of the hypotheses are adjusted by averaging their dimensions within each cluster. Following, we align each clusters' elements in rows and columns by repositioning elements which are arranged similarly in horizontal and vertical direction, respectively.

By the previous refinement similar detections are identified which are grouped into rows and columns. This enables to find missing windows of the same size in the spaces between current detections of each row or column. The windows' alignment on a facade can be described by a Dirac comb shifted by an offset as shown in Fig. 64.3. Irregularities in the alignment as well as false detection may hinder determining optimal function parameters. To properly estimate o set and period of the Dirac comb, we use a RANSAC algorithm. The positions obtained by the resulting functions' peaks reveal potential further windows. On common urban facades windows may occur which break the periodical pattern. Thus, we only consider potential windows if their regions do not overlap with previously detected windows whereas potential windows overlapping each other do not require a special treatment.

Since we cannot assume a facade to be regular, the presence of a window has yet to be proven based on evidence in the image despite that it has not been detected before. There are two major reasons for the detector missing a window. First, the sliding window might not have hit the window precisely enough through its step size. Second, image features responses might be too low so that they did not reach the weak classifiers' thresholds. From the previous step, we received precise information about positions and sizes of potential windows so that we can scan these regions more specifically. To overcome the second reason, the image patches are classified by a less strict soft cascade which tolerates lower feature responses. As a consequence, it is also more prone to yield false positive detection so that it should only be applied on certain small image regions.

The inferred positions determined by the RANSAC algorithm depend on the current set of detections and subdivision into rows and columns. Windows found while postprocessing may, thus, give new hints for further potential window positions by constituting new rows in the windows' alignment. As a consequence, we iteratively repeat the proposed postprocessing until no new window hypothesis is found anymore.

64.4 Experiments

In the following experiments we investigate the quality of our proposed detection system. At first, to verify the improvement of using a soft cascaded classifier on rectified facade images over the Viola-Jones detector approach proposed by Ali et al. [2], we contrast their reported results with the performance of our detector without postprocessing. In a second experiment we evaluate the benefit of our proposed postprocessing. For this, we compare the results of our detector obtained in the first experiment to the detection results of our entire system on the same facade images.

As comparability between the detection results has to be ensured, the evaluation metrics proposed in [2] are used for both experiments. Since precision with regard to the windows' size and position plays a major role for our application purposes, more specifically, we use the single window evaluation method. Accordingly, a detection is only marked as true positive if it is found inside a ground truth labeled region or at maximum overlaps it by 5 pixels in each direction. The detection, in addition, has to cover the bounding box of the ground truth label at least to a certain fraction. Again, owed to the required precision we allow a minimum coverage of 75%. A detection is marked as false positive if it covers less than 5% of a ground truth label.

The classifier used in our approach for these experiments is trained and calibrated on the dataset described in Sect. 64.4.1. Although our approach consists of two differently strict acting classifiers, it succes to train only one soft cascaded classifier. Differences in the classifiers' behavior can be achieved afterwards by distinct calibrations. For the detector component, we calibrate the classifier to a false positive rate of 2%. This ensures a reasonable detection rate while false positives per image are sufficiently few to not impair the postprocessing immoderately. For the postprocessing component, a calibration to 10% false positives is sufficient, allowing to identify windows with lower feature responses without falsely classifying most facade patches.

64.4.1 Dataset

We use the Ecole Central Paris Facades Database [12] which is one of the most prominent datasets. It provides a variety of facade images from different countries also allowing to investigate the general applicability of a detection approach. The dataset consists of 478 rectified facade images taken in 10 cities in Europe and the US. We cut out 4000 randomly chosen windows from 410 facade images and distribute them equally to the training and calibration of the classifier. Negative samples are generated from image patches of the same facades not containing windows. In the remaining 68 images ground truth was labeled manually. We excluded shop windows from labeling as their appearance is generally too different so that covering these would require another explicitly trained classifier.

Fig. 64.4 Window hypotheses obtained by our detector. Hypotheses are of different sizes and are slightly larger than the actual windows

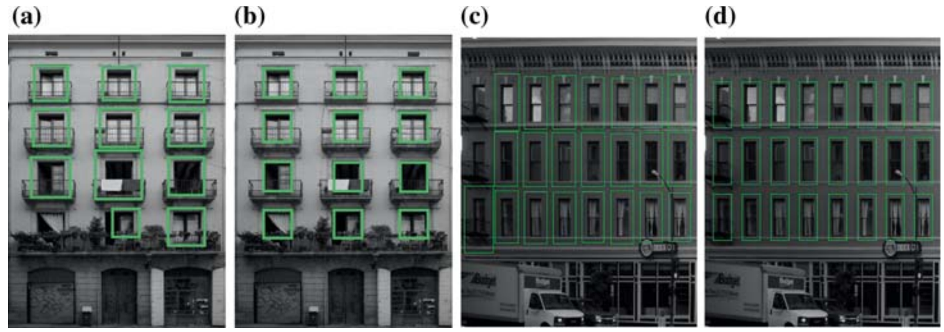

Fig. 64.5 Illustration of postprocessing results in (**b**) and (**d**) in comparison to detector results in (**a**) and (**c**)

64.4.2 Results

In the first experiment, we apply our detector without postprocessing to 68 rectified facade images of the Ecole Central Paris Facades Database. As can be seen in Fig. 64.4 detected regions are of different sizes and slightly larger than the actual windows dimensions. On this dataset our detector yields a detection rate of 84:9% and a false positive rate of 2:1%. This is a significant improvement over the Viola-Jones detector as stated by Ali et al. [2] reporting a maximum detection rate of 57% while 7% of the detections are false positives by means of the same evaluation metrics.

In the second experiment, the detector's results of the first experiment are fed into the postprocessing proposed in this paper. Figure 64.5 illustrates the improvements made by the refinement. By this, detected regions are well adapted and match the actual windows' dimensions much more precisely. Moreover, with a detection rate of 95:2% and 2:5% false positives the detection of further windows by inferring potential positions from current detections leads to a high increase of identified windows while insignificantly increasing false detections.

64.4.3 Discussion

The findings of Neuhausen et al. [9] already indicate an improvement of the soft cascaded classifier over the Viola-Jones detector on window detection. With the achieved detection rate of 84:9% we gain an increase of more than 27% with regard to the best detection results reported by Ali et al. while having a slightly lower false positive rate. Combining the findings with the results of our experiment allows to conclude that the soft cascaded classifier outperforms the conventional Viola-Jones detector on this task. Beyond this, the decrease in variation of the windows' appearance by rectifying the facade images beforehand leads to a further improvement of the detection results.

The second experiment clearly demonstrates the postprocessing's capability of improving the detector's results. We showed that its use increases the true positive rate of our detector by more than 10%. Addition finally, the resulting false positive rate could be kept similarly low. The increase is yet dependent on false positives obtained by the detector. Detected false positives may multiply exponentially if the initial rate is too high. Thus, keeping the false positive rate of the detector as low as possible is a prerequisite. As the experiment shows, a rate of about 2% fulfills this requirement.

Concluding, the entire detection system as proposed in this paper achieves a recall of 95:2% with a precision of 97:4%. Furthermore, the detections' bound aries match the actual windows very precisely. These results are sufficient to complement existing building models with precise window information to facilitate a variety of simulation and assessment tasks in civil engineering.

64.5 Conclusion

Building models can be used to simulate a building's behavior so that buildings can be assessed and compared with regard to their characteristics. However, publicly available 3D models usually lack relevant details. Models have, thus, to be complemented by other data.

Referring to an approach detecting windows using a cascaded classifier, in this paper we proposed a detection system improving the stated results. Our system consists of a preprocessing which rectifies facade images before detection to eliminate unnecessary variability in the windows' appearance. Afterwards, a detector scans the image and passes relevant image patches to a soft cascaded classifier. Resulting detections are finally postprocessed. On the one hand, the detections' borders are refined and, on the other hand, further windows are detected. Comparing the results of our detector to these of the Viola-Jones detector highlights that the soft cascade classifier outperforms the conventional cascade on the window detection task. Furthermore, we showed that applying a proper postprocessing to the window hypotheses obtained by our detector tremendously increases detection quality. With a resulting recall of 95:2% and a precision of 97:4% our entire detection system yields suitable results to successfully supplement existing building models by windows.

However, future work has to be done in terms of shop window detection. Moreover, the results of our system could be further improved if false positive hypotheses obtained by our detector were sorted out before postprocessing.

References

1. Ali, H., Ahmed, B., Paar, G.: Robust window detection from 3d laser scanner data. In: Proceedings of the Congress on Image and Signal Processing, pp. 115–118 (2008)
2. Ali, H., Seifert, C., Jindal, N., Paletta, L., Paar, G.: Window detection in facades. In: Proceedings of the Conference on Image Analysis and Processing, pp. 837{842 (2007). https://doi.org/10.1109/iciap.2007.4362880
3. Barber, D., Mills, J., Smith-Voysey, S.: Geometric validation of a ground-based mobile laser scanning system. ISPRS J. Photogramm. Remote Sens. **61**(1), 128–141 (2008)
4. Becker, S., Haala, N.: Renement of building facades by integrated processing of lidar and image data. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **36**, 7–12 (2007)
5. Bourdev, L., Brandt, J.: Robust object detection via soft cascade. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 236–243 (2005)
6. Haugeard, J., Philipp-Foliguet, S., Precioso, F., Lebrun, J.: Extraction of Windows in facade using Kernel on Graph of Contours. In: Proceedings of the 16th Scandinavian Conference on Image Analysis, pp. 646–656. Springer, Berlin Heidelberg, Germany (2009)
7. Lee, S.C., Nevatia, R.: Extraction and integration of window in a 3d building model from ground view images. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 106–113 (2004)
8. Meixner, P., Leberl, F., Bredif, M.: Interpretation of 2d and 3d building details on facades and roofs. In: Proceedings of the Conference on Photogrammetric Image Analysis, pp. 137–142. ISPRS (2011)
9. Neuhausen, M., Martin, A., Obel, M., Mark, P., König, M.: A cascaded classifier approach to window detection in facade images. In: Proceedings of the International Symposium on Automation and Robotics in Construction (2017)
10. Ripperda, N.: Grammar based facade reconstruction using rjmcmc. Photogramm. Fernerkundung Geoinf. **2**, 83–92 (2008)
11. Ripperda, N., Brenner, C.: Application of a formal grammar to facade reconstruction in semiautomatic and automatic environments. In: Proceedings of the AGILE Conference on Geographic Information Science (2009)
12. Teboul, O., Simon, L., Koutsourakis, P., Paragios, N.: Segmentation of building facades using procedural shape priors. In: Proceedings of the Conference on Com-puter Vision and Pattern Recognition, pp. 3105–3112 (2010)
13. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004). https://doi.org/10.1023/b:visi.0000013087.49260.fb
14. Yang, M.Y., Förstner, W., Chai, D.: Feature evaluation for building facade images—an empirical study. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **39**(B3), 513–518 (2012)