Structure-driven Facade Parsing With Irregular Patterns

Jinglu Wang* Chun Liu† Tianwei Shen* Long Quan*
The Hong Kong University of Science and Technology* NavInfo Co., Ltd†
{jwangae,tshenaa,quan}@cse.ust.hk* liuchun@navinfo.com†

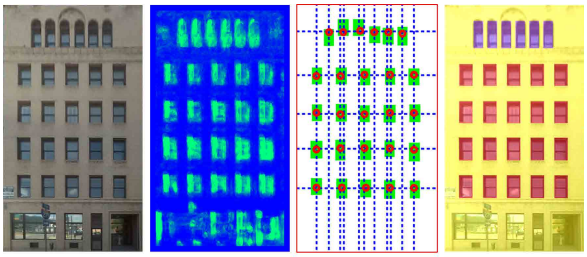
Abstract

We propose a novel method for recognizing irregular patterns in facades. An irregular pattern is an incomplete 2D grid, representing the placements of repetitive structural architectural objects (e.g., windows), which is capable of being generalized to a variety of facade structures. To effectively recognize such a pattern, we jointly model objects and object structures in a unified Marked Point Process framework, where the architectural objects are abstracted as sparsely populated geometric entities and the pairwise spatially interactions are modeled as elliptical repulsion fields. To optimize the proposed model, we introduce a structure-driven Monte Carlo Markov Chain (MCMC) sampler, by which the irregular pattern hypotheses are iteratively constructed in a bottom-up manner and verified in a top-down manner. The solution space is explored more efficiently for fast convergence. Extensive experiments have shown the efficiency and accuracy of our method of parsing a large category of facades.

1. Introduction

3D urban reconstruction [12, 11, 21] has attracted much attention since the launch of many popular on-line 3D geographic services. With the rapid growth of aerial radiometry, many astonishing 3D maps are available to the public, but fail to capture enough details at the ground level. Therefore, facade analysis, which can drive 3D reconstructions at the ground level in rich details, is in great need.

Generally, semantic facade analysis is a process to decompose facades into semantic architectural objects in geometric forms. Although numerous methods have been proposed, it is still challenging due to three major limitations. (1) There is a lack of generality in facade structure representation. In the current literature, facades are mostly represented by a single pattern in strict alignments [16, 10, 17] or a specific architectural style [14], which can not adapt to various real facades. (2) The object-level facade modeling is missing so that the analysis is not stable under heavy pixel-wise noise and occlusion. (3) It is time consuming when involving grammar splitting methods which have huge solution space.



(a) (b) (c) (d)

Figure 1. Workflow: (a) input image; (b) object (window) probability map; (c) 2D incomplete grid facade structure representation. (d) regulation result using multi-lattice structure.

To address the aforementioned limitations, we propose a novel method to recognize irregular object patterns, which are generally modeled in an incomplete and non-aligned grid, in facades. The overview of our method is illustrated in Fig. 1. By any supervised learning techniques, we first produce a object probability map (Fig. 1b). Then we model facade objects and object structures jointly in a Marked Point Process [1, 7] framework. Since facade objects are sparsely populated, we define the pairwise spatially interactions between objects as elliptical repulsion fields to model object spacing flexibly. By such definition, our joint object and structure model contains arbitrary numbers of objects, arbitrary numbers of repetitions (Fig. 1c), and customized interaction functions. To optimize such a complex model, we design a structure-driven MCMC sampler which encodes the global grid layout in a top-down manner to explore the solution space efficiently. In addition, any further regulation can be performed based on our detected objects, such as the low rank constraint [6] or grammars [10]. An example of regulated objects clustered in multiple groups are shown in Fig. 1d.

In summary, our contributions are: 1) a general facade structure representation that allows facade objects to sit on incomplete grids and to be of different sizes, 2) a Marked Point Process that models sparsely populated objects, and 3) an optimization of a structure-driven MCMC sampler in a bottom-up/top-down inference manner.

1.1. Related Work

With different modeling assumptions, researchers took two major categories of approaches on facade parsing: bottom-up and top-down approaches.

Bottom-up approaches exploit low-level features to detect facade elements. The facade structures are then inferred by clustering these elements. R. Tyleček et al. [13] modeled windows based on edges and considered the facade structure as rigid or weakly aligned grids. Another strategy is to parse the facades based on the discovery of symmetries in facade images. Zhao et al. [24] used the transform space voting to extract the lattice structures. Similar works can be found in [22, 23, 18] for discovering near-regular structures. In general, these bottom-up approaches rely heavily on local features and are sensitive to occlusion and noise.

Top-down approaches encode facade decomposition knowledge as priors and use them to regularize facade interpretation. A common practice is to use parametric grammar rules. Thus the facade parsing task is to discover both the parametrized grammar rules and the parameters to yield a pre-defined facade layout. Recently, Teboul et al. [16] developed a system based on 2D split grammar and obtained excellent results on Haussmannian buildings. The extension of this work to asymmetric splits and 3D data were presented in [14, 15, 9]. al. [3] retrieve a parsing that respects common architectural constraints. Besides, Zhu et al. [19] proposed a top-down/bottom-up inference algorithm with a generative attribute grammar to describe man-made scenes.

2. An Object-level Model for Facade Parsing

Because the facade objects, i.e., windows, are sparsely and structurally populated, we jointly model the objects and structures in the Marked Point Process [7]. The objects are modeled as a rectangle configuration $\mathbf{X} = [x_1, \cdots, x_i, \cdots, x_n]$. The position for each rectangle x_i is sampled over the probability map, and the width w_i and height h_i are uniformly sampled from a predefined range. The energy function used for penalizing improper configurations is an association of two terms.

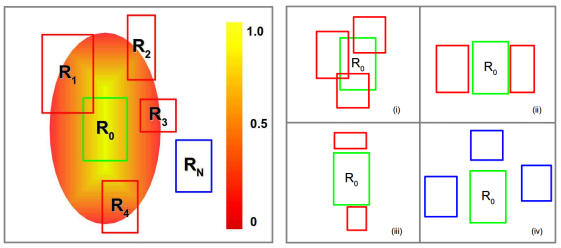
$$U(X) = U_{prior}(X) + \alpha U_{data}(X), \quad \alpha \in (0,1) \quad (1)$$

where U_{prior} denotes the prior term encodes local interactions which correspond to the spacing between windows and the global layout which concerns with window alignment and repetition patterns, U_{data} the data term measuring the coherence of the sampled rectangles with respect to the probability map, α the weighting coefficient.

By minimizing the energy function, the optimal set of rectangles is obtained, e.g., green rectangles in Fig. 1c.

2.1. Prior Term

We construct the prior term $U_{prior}(X) = U_{local}(X) + U_{global}(X)$ for a configuration of rectangles X based on a local term $U_{local}(X)$ and a global term $U_{global}(X)$.



(a) (b)

Figure 2. Rectangle Interaction. (a) The neighborhood of R_0 is marked with an ellipse and the color indicates the penalty according to the configuration. R_N does not interact with R_0. R_1 overlaps with R_0. Other rectangles are too close to R_0. (b) Four types of rectangle configuration: (i) overlapping; (ii) horizontally too close; (iii) vertically too close; (iv) favored configuration.

Local Interaction. The interacting rectangle x_j is chosen to be the closest neighbor of current rectangle x_i. Particularly, we propose a repulsion term U_R between pairwise neighboring rectangles to avoid dense arrangement (Fig. 2). Thus the local interaction term U_{local} is the sum of repulsion terms U_R among all pairwise interactions in the current configuration X, i. e., $U_{local} = \sum_{i,j} U_R(x_i, x_j)$.

For each rectangle x_i, we establish an elliptical repulsion region (Fig. 2a) using a horizontal repulsion distance δ_x and a vertical one δ_y. If the neighboring rectangle x_j invades into this region ($t_x < \delta_x$ and $t_y < \delta_y$, where t_x and t_y are the horizontal distance and the vertical distance of the closest point to the ellipse center), it will be penalized with a cost $U_R(x_i, x_j)$ based on the distance between the invading rectangle and the ellipse center $U_R(x_i, x_j) = 1.0 - \frac{t_x^2}{\delta_x^2} - \frac{t_y^2}{\delta_y^2}$, Otherwise, the cost is 0.

Global Layout. The global layout is not directly tractable from data. Instead, we apply the global layout constraint during the structure-driven optimization in a bottom-up/top-down inference manner [5]. In each optimization iteration, we construct a global layout hypothesis from the already detected object and verify the structure by object detection on this structure. In this way, the global layout constraint is implicitly encoded.

2.2. Data Term

The data term combines a set of quality functions, rectangle rate R, contrast rate C, homogeneity rate H and area rate A, in order to ensure that good rectangle candidates are always well positioned (locations) and scaled (width and height) with respect to the probability map (Fig. 3). For each rectangle x_i, the data term is defined as:

$$U_d(x_i) = \max[R(x_i), C(x_i)] * H(x_i) * A(x_i) \quad (2)$$

The total data term is the sum of all individual data terms, namely, $U_{data}(X) = \sum_i U_d(x_i), x_i \in \mathbf{X}$.

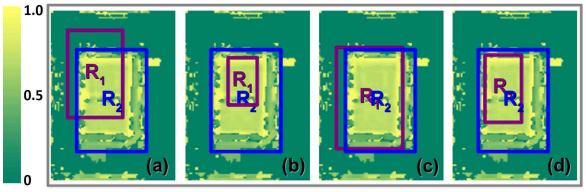


Figure 3. Data term considerations over an object probability map. Two rectangle proposals R_2 (favored) and R_1 (undesired) are shown in four configurations corresponding to different considerations: (a) R_2 should have a large integral over the probability map (rectangle rate); (b) R_2 should be more differentiable from neighborhood by contrast comparing to R_1 (contrast rate); (c) R_2 should have a higher homogeneity to be well positioned over the probability map (homogeneity rate); (d) R_2 should have a large support (area rate).

Rectangle Rate. Rectangle rate R is used to quantify the confidence of a rectangular object candidate over the probability map. We first define the rectangle ratio $r \in [0, 1]$ as the intensity of per unit area. Following previous work [6], the quality function maps r to the rectangle rate $R \in [-1, 1]$, taking the form

$$R(x_i) = \begin{cases} 1 - \theta \times (r_i/T)^2 & \text{if } r_i < \mathrm{T}, \\ -(r_i/T)^2 & \text{if } r_i \geq \mathrm{T} \end{cases} \quad (3)$$

where T is a threshold, θ a weighting coefficient.

Contrast Rate. A good rectangle candidate should have strong contrast to its neighbors over the probability map. To measure the contrast ratio, we take Bhattacharya distance in considering the two Gaussian mixture distributions of the probabilities over the rectangle region A inside the rectangle and the surrounding region B outside the rectangle in a small neighborhood. The threshold d_0 is used to calibrate the contrast ratio [20] (Equation 4, μ and σ are mean and standard deviation respectively) onto contrast rate $C \in (-1, 1]$ (Equation 5).

$$d_B(A, B) = \frac{(\mu_A - \mu_B)^2}{2 \times \sqrt{\sigma_A^2 + \sigma_B^2}} - log(\frac{\sigma_A \sigma_B}{\sigma_A^2 + \sigma_B^2}) \quad (4)$$

$$C(x_i) = \begin{cases} 1 - \frac{d}{d_0} & \text{if } d < d_0, \\ exp(\frac{-(d-d_0)}{d_0} - 1) & \text{otherwise} \end{cases} \quad (5)$$

Homogeneity Rate. The homogeneity rate H is a measure to ensure that the probabilities inside the rectangle are consistent. It is defined by counting the number of intensities with a small deviation to the mean intensity over the probability map, $H(x_i) = 1_{x_i \in N}, N = \{p | v(p) \in B_\sigma(\mu)\}$, where p is a pixel inside the rectangle x_i over the probability map, $v(p)$ the intensity value, μ the mean probability inside the rectangle, and σ the threshold parameter.

Area Rate. We define the area rate A to be the ratio between area size of the current rectangle to the maximum allowed rectangle size. It is used to encourage rectangle as large as possible. For a rectangle x_i with its width $w \in [W_{min}, W_{max}]$ and its height $h \in [H_{min}, H_{max}]$, the area rate is $A(x_i) = w * h/(W_{max} * H_{max})$.

3. Structure-driven Optimization

The optimization of the energy function proposed in Equation 1 is not simple because of large variations of sizes, arbitrary number of objects, arbitrary numbers of alignments and arbitrary number of repetitive patterns. Hence the searching space is extremely large. Moreover, this energy function is not guaranteed to be convex with customized object interactions. Thus, we consider a MCMC [4] sampler to optimize the model. Particularly, we adopt the multiple-birth-and-death (MBD) algorithm [20] and propose a structure-driven MCMC sampler to optimize the energy defined in Equation 1.

3.1. Structure-driven MCMC sampler

It has been shown that the birth rate in the MBD algorithm can be non-homogeneous and does not have impact on the convergence to the global minimum but only the convergence speed inn the work [20]. Besides the normal MBD move, we propose three additional moves based on MBD to speed up the algorithm. The convergence is still guaranteed because only the birth mechanism is changed. The convergence analysis is shown is the appendix.

By randomly switching among the three moves, the structure hypotheses are dynamically updated for more efficient solution space searching. We consider that such structure-driven birth mechanism is equivalent to a dynamic birth map.

Multiple-birth-and-death (MBD). In this move, we sample object positions from the probability map since it directly corresponds to the data confidence. The object sizes are uniformly sampled from a pre-defined range.

Lattice-driven MBD. We propose lattice structures for each object candidate in the current iteration in terms of four neighbors including top, bottom, left and right (Fig. 4 top row). The horizontal and vertical displacements are uniformly selected from horizontal and vertical generator lists which are obtained from the object distances. The sizes are the same as the seed object. This move ensures that missing objects can be recovered in a small neighborhood.

Grid-driven MBD. We also propose multiple objects in grid structures (Fig. 4 bottom row). We extract all X and Y coordinates from current object list and construct a grid structure by all coordinate enumerations. Then we add objects centering on the grid with uniformly sampled sizes. This move ensures that missing objects can be recovered in the grid structures.

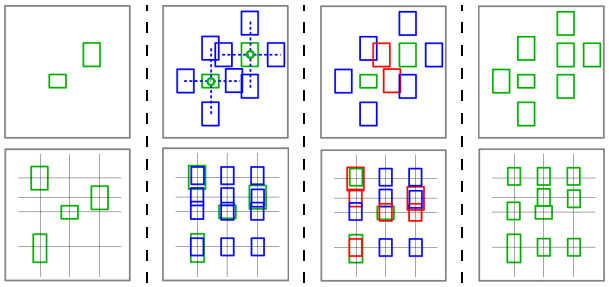


Figure 4. Lattice-driven MBD (top row) and grid-driven MBD (bottom row): (column a) rectangles (green) obtained from previous iteration; (b) proposing rectangles (blue) in lattice (top) / grid (bottom) structures; (c) object rectangles (red) removing by death; (d) rectangles (green) obtained in the current iteration.

Regulation. After the MCMC algorithm converges, we get the irregular patterns (Fig. 1c). To regulate the final results and discover window structures. We first group the detected windows as multiple lattices according to their sizes and alignments, shown as different colors in Fig. 1d. Then, we use the low rank constraint [6] and the grammar that completes a full 3-layer facade analysis [10] for the regulation.

4. Experiment

Demonstrative Results. We test our proposed method on two image sets, the publicly available "ECP" facade dataset [18, 24, 16] and the "Boulevard" dataset which is captured by ourselves including several building styles and with heavy pixel-wise occlusion (102 facade images from Boulevard Saint Michel in Paris). For these images, the probability maps are produced by supervised image classification with random forest [2]. Some demonstrative results are shown in Fig. 5.

We also use 3D laser scans and multi-view images to evaluate our method. For multi-view images, we use Qusi-dense structure from motion [8] to reconstruct 3D points. To apply our method on 3D data, we first generate 2D depth maps by triangulating the scanned/reconstructed 3D points. Then, the depth maps serve as probability maps. The window detection results are shown in Fig. 6 and Fig. 7.

Qualitative Comparison. The grammar based parsing [14, 16] has been introduced to facade analysis. In comparison, we can deal with structures represented by both regular and irregular split grammar methods, shown in Fig. 8.

Quantitative Comparison. The comparison is focused on window detection. Without any architectural structure assumption, our recognition rate is 80%, comparable to all state-of-art methods in "ECP" dataset and superior over them in "Boulevard" dataset. Our computation time is around 2 minutes, close to 2∼7 minutes from [18]. We have also applied the low rank constraint [6] and a simple grammar to as architectural level process in order to complete a full 3-layer facade analysis [10]. The detailed comparison is listed in Table 1.

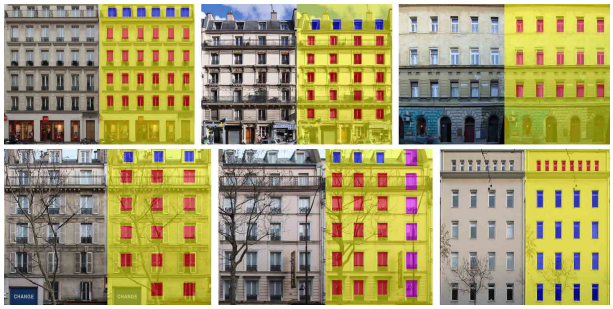


Figure 5. Results on single images from "ECP" (top) and "Boulevard" (bottom) datasets. We can find that our proposed method can handle sever occlusion problems. Different colors denote different lattice structures.

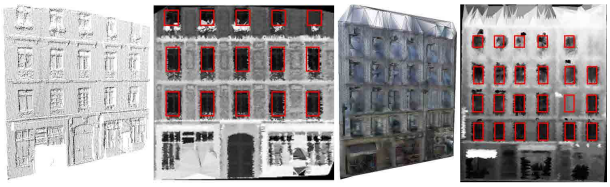


Figure 6. Input laser scans (column 1, 3) and window detection on depth maps (column 2, 4).

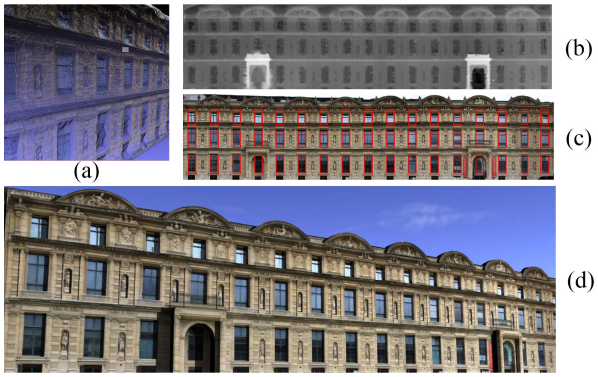


Figure 7. Result on multi-view images: (a) Reconstructed 3D points from multi-view images; (b) depth map; (c) detected windows in fronto-parallel view; (d) image-based building modeling using detected windows.

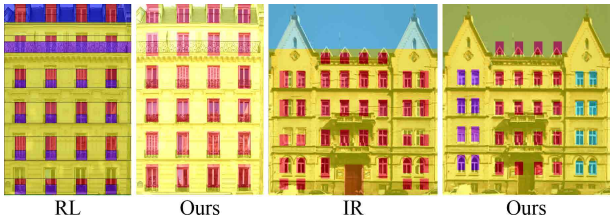


Figure 8. Qualitative comparison with grammar-based methods, "RL" [16] and "IR" [14]. We can find that our method is comparable with grammar-based methods in window detection.

	RL [16]	IL [14]	TL [10]	ours	ours+LR	ours+Gr
ECP	81	68	75	80	87	85
Boulevard	65	N/A	N/A	76	80	78

Table 1. Window detection comparison in pixel average accuracy. "LR" is short for with low rank constraint. "Gr" is short for using grammar constraints.

5. Conclusion

We present a novel method for recognizing irregular patterns which are generalized to a variety of facade structures. We adopt a unified joint object and structure modeling approach based on Marked Point Process framework, and propose a structure-driven optimizer with additional kernel moves to perturb object configurations in rectilinear structures. Extensive experiments and comparisons show that the proposed method can efficiently and accurately extract the rectilinear structures.

Limitations. Our method relies on the probability maps. Thus, if the data is corrupted significantly at object level (completely occlusion or heavy shadows), we will not be able to recover the objects, e.g. the top right window in column four of Fig. 6. Therefore, our method will be slightly less accurate than those obtained with specialized structure priors in specific scenes. However, our work can be considered as an object level processing and can be further integrated to those specific regulation frameworks to improve the detection accuracy.

Acknowledgement

The authors would like to thank all the anonymous reviewers for their constructive comments. This work was supported by RGC-GRF 16208614, ITC-PSKL12EG02.

References

[1] C. Benedek, X. Descombes, and J. Zerubia. Building development monitoring in multitemporal remotely sensed image pairs with stochastic birth-death dynamics. *Pattern Analysis and Machine Intelligence (PAMI)*, 34(1):33–50, January 2012. 1

[2] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. 4

[3] A. Cohen, A. G. Schwing, and M. Pollefeys. Efficient structured parsing of facades using dynamic programming. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3206–3213. IEEE, 2014. 2

[4] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995. 3

[5] F. Han and S. C. Zhu. Bottom-up/top-down image parsing by attribute graph grammar. In *International Conference on Computer Vision (ICCV)*, pages 1778–1785, 2005. 2

[6] T. Han, C. Liu, C. L. Tai, and L. Quan. Quasi-regular facade structure extraction. *Asian Conference on Computer Vision*, 2012. 1, 3, 4

[7] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. Structural approach for building reconstruction from a single dsm. *Pattern Analysis and Machine Intelligence (PAMI)*, 32(1):135–147, January 2010. 1, 2

[8] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *Pattern Analysis and Machine Intelligence (PAMI)*, 27(3):418–433, 2005. 4

[9] A. Martinovic, J. Knopp, H. Riemenschneider, and L. Van Gool. 3d all the way: Semantic segmentation of urban scenes from start to end in 3d. In *Computer Vision and Pattern Recognition (CVPR)*, pages 4456–4465, 2015. 2

[10] A. Martinović, M. Mathias, J. Weissenberg, and L. V. Gool. A three-layered approach to facade parsing. In *European Conference on Computer Vision*, volume 7578, pages 416–429, 2012. 1, 4

[11] P. Müller, P. Wonka, S. Haegler, A. Ulmer, and L. Van Gool. Procedural modeling of buildings. *ACM Transactions on Graphics (TOG)*, 25(3):614–623, July 2006. 1

[12] P. Müller, G. Zeng, P. Wonka, and L. Van Gool. Image-based procedural modeling of facades. *ACM Transactions on Graphics (TOG)*, 26, July 2007. 1

[13] T. Radim and S. Radim. Stochastic recognition of regular structures in facade images. *Transactions on Computer Vision and Applications (CVA)*, 4:63–70, may 2012. 2

[14] H. Riemenschneider, U. Krispel, W. Thaller, M. Donoser, S. Havemann, D. Fellner, and H. Bischof. Irregular lattices for complex shape grammar facade parsing. In *Computer Vision and Pattern Recognition (CVPR)*, 2012. 1, 2, 4

[15] L. Simon, O. Teboul, P. Koutsourakis, L. Van Gool, and N. Paragios. Parameter-free/pareto-driven procedural 3d reconstruction of buildings from ground-level sequences. In *Computer Vision and Pattern Recognition (CVPR)*, 2012. 2

[16] O. Teboul, I. Kokkinos, L. Simon, P. Koutsourakis, and N. Paragios. Shape grammar parsing via reinforcement learning. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2273–2280, 2011. 1, 2, 4

[17] R. Tyleček and R. Šára. Spatial pattern templates for recognition of objects with regular structure. In *Pattern Recognition*, pages 364–374. 2013. 1

[18] R. Tyleček and R. Šára. A weak structure model for regular pattern recognition applied to facade images. *Asian Conference on Computer Vision*, pages 450–463, 2010. 2, 4

[19] T. F. Wu and S. C. Zhu. A numerical study of the bottom-up and top-down inference processes in and-or graphs. *International Journal of Computer Vision*, 93(2):226–252, 2011. 2

[20] D. Xavier, M. Robert, and Z. Elena. Object extraction using a stochastic birth-and-death dynamics in continuum. *Journal of Mathematical Imaging and Vision*, 33(3):347–359, Mar. 2009. 3

[21] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, and L. Quan. Image-based façade modeling. In *ACM Transactions on Graphics (TOG)*, volume 27, page 161. ACM, 2008. 1

[22] C. Yang, T. Han, L. Quan, and C. L. Tai. Parsing façade with rank-one approximation. In *Computer Vision and Pattern Recognition (CVPR)*, 2012. 2

[23] H. Zhang, K. Xu, W. Jiang, J. Lin, D. Cohen-Or, and B. Chen. Layered analysis of irregular facades via symmetry maximization. *ACM Transactions on Graphics (TOG)*, 32(4):104:1–104:10, 2013. 2

[24] P. Zhao and L. Quan. Translation symmetry detection in a fronto-parallel view. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1009–1016, 2011. 2, 4