Grammar-based 3D Facade Segmentation and Reconstruction

Guowei Wan[⋆†] Andrei Sharf[‡]

[†]SIAT, China [‡]Ben-Gurion University [⋆]National Univ. of Defense Tech.

Abstract

Recent advances in scanning technologies allow large-scale scanning of urban scenes. Commonly, such acquisition incurs imperfections: large regions are missing, significant variation in sampling density, noise and outliers. Nevertheless, building facades often consist structural patterns and self-similarities of local geometric structures. Their highly-structured nature, makes 3D facades amenable to model-based approaches and in particular to grammatical representations. We present an algorithm for reconstruction of 3D polygonal models from scanned urban facades. We cast the problem of 3D facade segmentation as an optimization problem of a sequence of derivation rules with respect to a given grammar. The key idea is to segment scanned facades using a set of specific grammar rules and a dictionary of basic shapes that regularize the problem space while still offering a flexible model. We utilize this segmentation for computing a consistent polygonal representation from extrusions. Our algorithm is evaluated on a set of complex scanned facades that demonstrate the (plausible) reconstruction.

1. Introduction

In recent years, there has been an increasing interest in the modeling and reconstruction of digital urban scenes. Rapid advances in laser scanning technology and the proliferation of GIS services such as those offered by Microsoft Virtual Earth or Google Earth have been driving a strong trend towards 3D reconstruction of urban architectural models from satellite and aerial photography combined with street-level laser scanned geometry. The state-of-art in automatic reconstruction from such data allows modeling of the geometry of building layout and ground polygon extrusion with roofs, where the building facades are merely approximated with a small number of textured planes [1, 2]. Reconstructing detailed building structures including facades has remained a challenge.

The main difficulty with scans of large-scale urban environments is data quality (or lack thereof). Large distances between the scanner and scanned objects imply reduced precision and higher level of noise. Furthermore, unlike small-scale scanning, such as those in the Michelangelo project [3], where a scanner can be strategically positioned to achieve the necessary coverage, similar controls are rather limited during large-scale urban scanning. As a result, the obtained point clouds typically exhibit significant missing data due to occlusion, as well as uneven point density.

Nevertheless, building facades often exhibit a structured arrangement consisting of repetition patterns and self-similarities (see Figure 1(a)). Regularity and self-symmetry in urban buildings is not a chance occurrence, but is demonstrated universally across countries and cultures. Such structured arrangements arise from manufacturing ease, build-ability, aesthetics, etc. The highly-structured nature of building facades, makes their treatment amenable to model-based approaches and in particular to grammatical representations as they can naturally capture the hierarchical structure of facades.

While in recent years many techniques have been developed to detect repeated parts in models [4, 5, 6, 7, 8, 9, 10], most of these works do not investigate how to best use the strong regularity and structure present in urban facades. Moreover, most of the techniques are applied in image space by analyzing photometric 2D images sampled over an underlying regular domain. Only few attempts have been made towards detection and utilization of self-similarities directly on 3D scanned geometry [9, 11, 12, 13].

The main difficulty with repetition and pattern analysis lies in the significant variations and lack of perfect regularity that may exist in building facades, even for ones corresponding to simple architectural styles. Procedural models and specifically shape grammars offer a flexible yet powerful tool to account for such variability while being compact

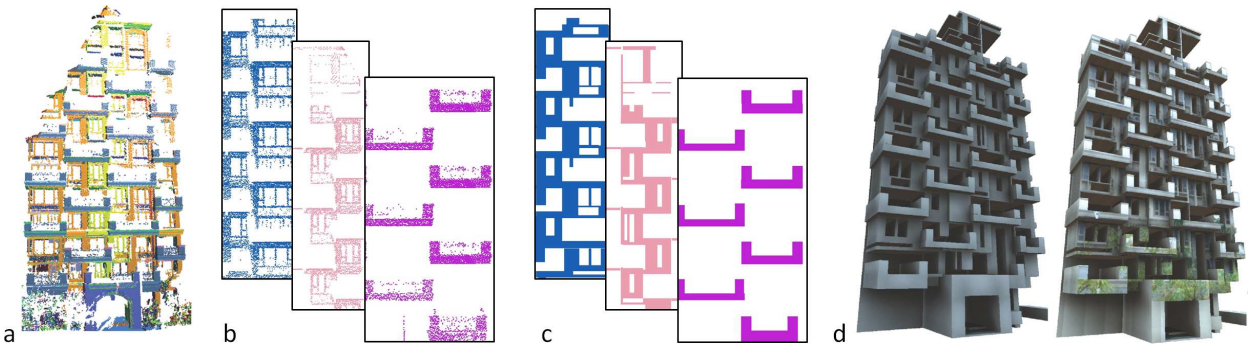

Figure 1: Grammar-based 3D facade reconstruction. Given a 3D LiDAR scan of a building facade (a), we decompose it into depth-layers (only three layers shown) (b) and apply grammar-based segmentation per-layer (c). Detected segments are extruded to reconstruct the 3D polygonal model (d) and textured for visual realism.

and providing a semantic representation of the facade. Recently, several methods have utilized procedural models like shape grammars for computing segmentation [8, 14]. Their idea is to model the space of segmentation solutions using a set of simple procedural rules and a dictionary of basic elements.

In this paper we utilize a grammar-based segmentation for 3D LiDAR scans of urban facades. Since LiDAR scans are typically noisy, we apply a scan filter in a preprocessing step using [12] and perform depth-layer decomposition to convert 3D data into 2.5D and reduce complexity. We compute the grammar-based segmentation for each layer similar to Teboul et al. [14] by fitting an optimal instance of a particular grammar. We utilize the segmented shapes and reconstruct the 3D polygonal model by extruding facade elements across depth-layers. Texture can be applied to polygonal model using existing commercial tools for visual realism (see Figure 1(d)). Our work's main contribution is the utilization of grammar rules as prior for segmenting and reconstructing scanned facades. We define a probability map to determine the relationship between elements of the grammar and the scanned data. The segmentation task can be considered as an optimization problem of a sequence of derivation rules with respect to a given grammar. The outcome of this method is an accurate segmentation of the facade with basic elements that can be extruded to reconstruct the polygonal model.

2. Related Work

Given the large volume of work on urban modeling, we refer the reader to the recent survey by Vanegas et al. [15] for a comprehensive coverage. Here we focus on previous works most closely related to ours, in particular those addressing repetition detection, segmentation reconstruction of urban scenes.

Our method is heavily based on the existence of a structured arrangement of elements and repeated parts in urban facades. Symmetries, repetitions, and regularity have been extensively studied in the context of image analysis, and to a lesser extent for 3D geometry, with application towards procedural modeling, scan completion, and improvement. Schaffalitzky and Zisserman [16] automatically detect repeated image elements in planes, which are typical in urban facades. Korah and Rasmussen [17] address the problem of automatically detecting 2D grid structures such as windows on building facades from images. Pauly et al. [9] introduce a transform domain analysis coupled with a non-linear optimization to detect regular k-parameter structures in 3D shapes. Recently, Shen et al. [18] suggested to adaptively partition 3D urban facades into a set of concatenated and interlaced grids.

Symmetry analysis and repetition detection methods have been extensively researched over the last years. Mueller et al. [8] use mutual information to identify a repetitive pattern on a regular grid. The result is then turned into a grammar formulation. Similarly Teboul et al. [14] suggest a grammar-based segmentation of 2D facades. The grammar yields segmentations of facades into semantic parts (e.g. walls, windows, balconies, etc). They use a random exploration of the grammar space to optimize the sequence of derivation rules. In our work, we use a similar grammar for the segmentation of 3D point sets (decomposed into 2.5D depth layers). Wu et al. [19] present a feature-based method that extracts repetition and symmetry hypotheses from a rectified image. They make simplifying assumptions such as constant repetition height, and no gaps between floors to find the repetitive pattern. In [20] repeated elements

are detected using a region growing starting from the constellations of the matched line features. Such approaches are based on the very strong assumption of global correlation between repetitive elements in the scene. Furthermore, they assume that repetitions can be parameterized by simple and somewhat regular patterns. In contrast, we use a grammar-based model which is more flexible to account for variability in the facade while regularizing segmentation.

Most repetition and segmentation techniques do not investigate means to use the detected structures for data enhancement. Vieira et al. [21] presented a method for scan super-resolution by registration of low resolution models with high-resolution exemplars. Recently, user-assisted global repetition patterns have been used for effectively consolidating sparse building scans by Zheng et al. [12]. We use the consolidation algorithm in a preprocessing step for improving scan quality. Li et al. [13] present a method that couples two input modalities, 2D images and 3D range scans for accurate repetition detection and 3D geometry enhancement.

Works on reconstruction of urban scenes have mostly been based on collections of photos [22, 23, 24, 25, 26, 27] or multi-view video [28], relying on photogrammetric reconstruction and image-based modeling techniques. Debevec et al. [4] propose an interactive image-based modeling method that exploits characteristics of architectural objects coupling an image-based stereo algorithm with manually specified 3D model constraints. More recently, Sinha et al. [25] present an interactive modeling system using unordered sets of photographs, leveraging the piecewise-planarity of architectural models. Xiao et al. [26] efficiently model facades from images by decomposing facades into rectilinear elementary patches. Later they extend the semantic segmentation and analysis approach to more general scenes, to produce visually compelling results by imposing strong priors of building regularity [2]. Special shape symmetries can also be leveraged to model architectural objects from a single image [29]. Recently, a semi automatic reconstruction method [11] has been presented which fits geometric parametric primitives to scanned facades through interactive user guidance. In contrast, our method automatically reconstructs the polygonal model without any manual assistance by utilizing a grammar based segmentation. Nevertheless, our reconstruction is restricted by the instantiation of the grammar rules to generate a feasible segmentation of the scanned facade.

3. Overview

In our work, we investigate means to explore, detect, and use structured geometry for reconstructing urban facades acquired using state-of-the-art LiDAR scanners. The acquired data comes in the form of point clouds, and lack any segmentation or high-level structural information. Such data quality makes it challenging to detect a set of structured elements, their reoccurrences and their reconstruction

Several recent works [8, 14] have demonstrated success in segmenting 2D facade images using grammar-based approaches. Nevertheless, migration to 3D scans is not straight forward. In [9, 12], a limited repetition detection in 3D scans was demonstrated, while making strong assumptions on the global and simple pattern of repetitions.

We introduce a grammar-based segmentation of 3D scans following the concept presented by Teboul et al. [14]. While their method applies to 2D photographs, we present an algorithm for scanned facades that are inherently 3D and typically sparse and noisy. We initially decomposed the data into planar depth layers (i.e. convert 3D data into 2.5D components) and define grammar rules to segment layers into foreground and background parts. Our method performs in a *top-down* manner to avoid the effect of over-segmentation. Except for the preprocessing step which requires a minimal user interaction for bootstrapping point-set enhancement, our pipeline is fully automatic (see Figure 1).

The scanning process mainly yields sampling of the frontal structures of the facades. We recover missing non-frontal parts, by extrusions. We utilize the segmented parts across depth-layers to compute their correspondence and then extrude between them. Hence, the final polygonal model is a valid mesh.

Our algorithm works in the following key stages:

- Enhance input scan by consolidation using [12].
- Decompose facade into disjoint depth layers.
- Compute grammar-based segmentation per depth layer.
- Compute polygonal model by extruding between corresponding segments across depth-layers.

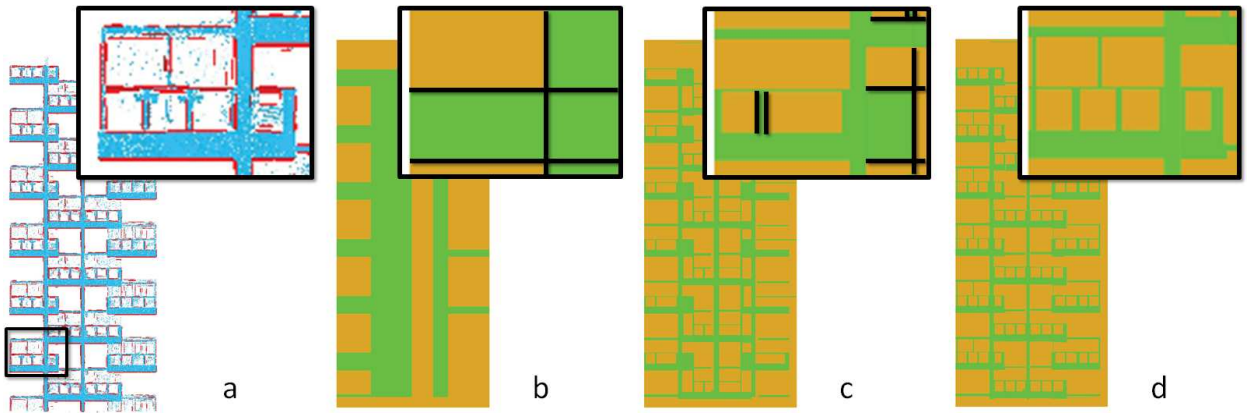

Figure 2: Split rules applied consecutively to segment a layer into foreground (green) and background (orange) segments recursively (top row shows a zoom-in). Using detected edges in the input scan (a), a sequence of split rules marked as black lines (b)- (e), recursively segment the data into subregions.

4. Grammar-based Segmentation

Our input data consists of large-scale scans of building facades acquired using a street-level LiDAR mounted vehicle. The acquisition process generates noisy point sets that may consist of large missing due to occlusions and scanner accuracy. Such data quality makes it challenging to process it as is.

Preprocessing. In a preprocessing step we use [12] for enhancing the point cloud by utilizing global repetitions. The user manually marks candidate structures in the input which are then used for finding global reoccurrences. The multiplicity of the geometry is fused together and projected to a base-geometry. Consolidation performs using a non-local filter that is applied to complete missing parts and remove outliers using information from remote regions.

Depth-layer decomposition. Initially, we segment the point-set into planes, and associate each point with its plane normal. Using RANSAC, we detect and extract planes similar to [30] (see Figure 1(b)). Points are normalized to $(-1,1)$ and the distance threshold for RANSAC we used is 0.005. We reject planes that are below a threshold in RANSAC voting (50 in our experiments).

We prune outlier planes arising out of sparse sampling and noise using the *Manhattan-world* prior [31, 32], which biases the planes to lie along one of three principal directions. We estimate the three major axes by clustering the detected planes. Planes are represented as 3D vectors of their orientation (nx, ny, nz). *KNN*-clustering in 3D space then selects the three dominant orthogonal clusters. Finally, all planes in a principal cluster are aligned with the cluster center parameters.

Layer decomposition is obtained simply by taking all planes in a cluster corresponding to one of three principal directions. In a principal direction, planes differ only by their offset, yielding a decomposition of the scanned data into planar layers with an *offset-depth* in that direction. We essentially convert the unstructured 3D point cloud into a structured set of 2.5D depth layers. The number of depth layers per direction is typically less than five.

Next, we compute two types of 3D edges: (a) edges extracted by identifying intersections of nearby planes, and (b) in-plane edges are extracted by applying the edge detector of [33] to the points in each plane (see Figure 2(a)). We prefer edges aligned to three principal directions thus pruning out spurious edges, e.g., those arising due to scanner beam motion path.

Grammar Definition. A shape grammar G applies on a planar (2D) depth layer (see Figure 2) consists of a dictionary of basic shapes $S = s_1, s_2, ..., s_k$ and a set of rules $R = r_1, r_2, ..., r_n$. Following *Manhattan-world*, we assume structures to align with three principal axes. Therefore, we restrict our dictionary basic shapes to axis aligned rectangles s defined by five parameters $s = (x, y, w, h, t)$, denoting center position (x, y), width and height (w, h) and type (foreground/background) t.

We define only one rule called *split rule* that splits a basic rectangle into nine subparts. We do not handle floors and rooftops in our segmentation and therefore a generic split rule was sufficient for handling all our facades. A split rule r contains four parameters (h_1, h_2, v_1, v_2), which denote positions of the two horizontal and vertical segmenting lines respectively. In order to reduce the grammar's degrees of freedom, we restrict split positions to coincide with horizontal and vertical edges of the facade. Thus, a split operation is guaranteed to generate foreground structures (in case of no edges inside a rectangle, it will not split and stay empty).

Shapes are manipulated through replacing rules, that turn a left-hand side (LHS) into some shapes on the right-hand side (RHS) under a possible condition. We usually write a derivation rule as: $LHS \rightarrow RHS$. The derivation process takes an atomic shape called axiom, and keeps replacing the existing atomic shapes using grammar rules constructing a derivation tree. The final segmentation is defined by the leaves of this tree.

Split operations subdivide a rectangle into subparts that coincide with horizontal and vertical edges in the data. In our grammar setting, a splitting (LHS) rule can be derived into $0-9$ subdivisions. Of course that if more than nine subdivisions are required, we apply our splitting rule recursively to subregions (see zoom-ins in Figure 2).

Different derivation trees lead to different geometric instances and segmentations. In the derivation process we allow shapes to emerge whereas a lot of them are very unlikely to represent real buildings. Finding the best segmentation relies on an energy function optimization of grammar parameters as defined next.

Segment Energy. Grammar-based factorization implies that a facade geometry is equivalent to a specific sequence of M rules or policy π. As a consequence, segmenting the facade is viewed as choosing the M rules which generate a facade that best fits the scanned data $\mathcal{P}$. The energy quantifies the appropriateness of a given policy π with respect to the data.

The leaf nodes of the facade tree provide a segmentation of the facade into regions R_s, where $s = \{F, B\}$ is a label denoting foreground/background type. Given a derivation tree π yielding a segmentation, the energy is based on a data fitting probability term. For each region R_c, we compute the probability $prob(x \in c)$ of each pixel x belonging to the class c as defined by the policy π. Note that since we operate on 2D planar depth-layers independently, we have an underlying regular grid of pixels $x \in \Re^2$.

We define $prob(x \in c)$ as:

$$prob(x = F) = \begin{cases} 0.99 & d(x,p) < \varepsilon \\ 1 - \frac{d(x,p)-\varepsilon}{\delta-\varepsilon} & \varepsilon < d(x,p) < \delta \\ 0.01 & \text{else} \end{cases}$$

where $prob(x = B) = 1 - prob(x = F)$, and $d(x,p)$ is the shortest Euclidean distance from pixel x to a scan point $p \in \mathcal{P}$.

The probability of the whole region R_c belonging to class c can be computed as the joint probability of all its underlying pixels. Assuming that pixels have independent labels, the joint probability given a policy π is:

$$prob(R_c \in c|\pi) = \prod_{x \in R_c} prob(x \in c)$$

and using Bayes' theorem we write the likelihood function $\zeta()$ as:

$$\zeta(\pi|R_c \in c) = \prod_{x \in R_c} prob(x \in c)$$

we convert it into log-likelihood :

$$\log \zeta(\pi|R_c \in c) = \sum_{x \in R_c} \log prob(x \in c)$$

then, the energy of a whole policy π is computed as the sum of energies of all regions R_i in the segmentation provided by π:

$$E(\pi) = \sum_i \log \zeta(\pi|R_i \in c_i) = \sum_i \sum_{x \in R_i} \log prob(x \in c_i)$$

Thus, for a scanned facade layer, we can compare energies of two different policies π_1 and π_2 as they are both defined as a sum of positive numbers over the same data.

Energy Maximization using Random Walk. Once we have defined the energy function $E(\pi)$ above, our procedural segmentation problem can be casted as a maximum likelihood estimator of π:

$$\pi^* = \arg \max_{\pi \in G(S,R)} E(\pi)$$

The maximization problem is highly combinatorial since the number of parameters may vary (e.g. two policies may have different tree topologies and depths). We apply a random walk algorithm for computing the minimum energy. Starting from an initial seed policy, we randomly consider additional policies in the neighborhood of the seed. If we find a new policy with a higher energy score, we choose it and continue exploring the neighborhood of the new policy. In order to sample a new rule r_1 in the neighborhood of a rule r_0, we simply seek for edges in the data in the neighborhood of r_0 with same orientation as r_0.

The initial seed is chosen randomly by applying arbitrary splits (see Figure 2(b)). At the beginning of the process, we allow the optimization algorithm to search for rules far from the initial seed. As the optimization procedure progresses, we constrain the search for the optimal solution to be closer to the current seed. Nevertheless, we still explore the space far from the current seed from time to time in order to avoid local minima. The random walk in this context means that we randomly generate additional candidate subdivisions in a local region and pick the best one. Note that even if we choose a subdivision which assigns wrong labels (background/foreground), additional subdivisions of that subregion will correct it. Nevertheless, in this case the derivation tree depth may be more than the optimal solution.

Typically, in approximately 50 iterations the algorithm converges towards a minimum. We stop before if further minimization iterations can not reduce energy term and we reached the maximum derivation depth (of 7).

Extrusion of hidden parts. Since our grammar dictionary contains of rectangular axis-aligned shapes, the result is a segmentation of depth-layers into rectangular foreground and background elements. Nevertheless we do not perform any symmetry analysis on the scanned data and our grammar segmentation does not account for repetitive geometry. Still, we consolidate the 3d points by repetitive elements in preprocessing step and reoccurring structures will have the same point distribution and therefore the same geometry too.

We can easily create polygonal fragments from the foreground segments. Nevertheless, side- and horizontal-planes (e.g. balcony sides and floor respectively) are completely or partially missing due to occlusion from the scanner beam. We therefore use the detected fragments to compute extrusions between them. We compute edges of the segmented fragments and for each pair of fragments, find co-planar edges that lie on side or horizontal planes. We require that edges completely overlap in the common plane in order to avoid non-rectangular extrusions. We extrude polygons between corresponding edges and obtain a complete polygonal model that reconstructs the facade.

The final polygonal mesh is a collection of polygons we compute from layers and extrusion. Since the model describes the facade of a building and its elements, the reconstructed mesh is a polygon soup with coincident polygons, t-junctions, which is not necessarily closed (see Figure 1).

5. Implementation and Results

We tested our algorithm on a large number of data-sets experimenting with different kind of building facades, from tall to medium height buildings. Figure 1 shows the results of our method on a complex facade with fine details. Our grammar based segmentation yields detailed facades even in the presence of large noise and missing parts (top left floors are completely missing in scan (Figure 1 (a)). Note that the extruded model reconstruct a complete facade which can be further textured for visual realism purposes.

We solve the energy function maximization in a reduced solution space as we restrict split rules to coincide with detected edges. Therefore, performance-wise the optimization procedure is very fast. The time of segmentation relies essentially on the number of layers and the complexity of the structure. We ran our algorithm on a machine of 2.83 GHz Intel Core Q9550 with 4GB RAM and our performance stayed in the range of only few minutes. We summarize these timings in Table 1.

Figure 3 show results of applying our method on various scanned buildings containing large noise and missing parts. In Figure 3(a), our method detected and reconstructed the fine rail details in the balconies. In 3(b), we show zoomins of two depth layers that contain significant outliers and noise. The grammar-based segmentation is robust

Dataset	#pts.	grammar segmentation	model reconstruction
Fig. 1	104K	219s	22s
Fig. 3(a)	815K	746s	47s
Fig. 3(b)	3393K	470s	21s
Fig. 3(c)	1919K	676s	25s
Fig. 3(d)	1688K	517s	22s
Fig. 6	109K	243s	19s

Table 1: Performance statistics on various data.

since it regularizes the reconstruction process; it avoids overfitting to noisy data by restricting derivation depth and utilizing priors from the grammar dictionary. Figure 3(c) shows the application of our method to a very tall building. Large outliers are caused here by the large distances from the scanner and reflections from windows. The optimal segmentation of the layers has successfully removed this noise. In Figure 3(d) we demonstrate the segmentation and reconstruction of a building consisting of a very detailed and non-regular floor level. Typically, facades' floor levels consist of detailed and complex geometry that is occluded by sidewalk obstacles and trees. Our method was able to reconstruct to some extent the floor level data.

In Figure 4 we compare our method with the state-of-the-art method of Li et al. [13] by showing side-by-side results and zoom-ins. Figure 4(a) compares the reconstruction of a detailed complex facade consisting of balconies and windows at several depths. Note that our reconstructed fine geometric details are completely missing in their method. Figure 4(b) demonstrates the regularization and flexibility effects of our method when applied to a highly non-regular facade. Our split grammar is more flexible than global repetition methods to handle non-regular patterns, as shown in the comparison to Li et al. Due to accurate segmentation, our extruded model contains no holes yielding a clean yet detailed reconstruction. Finally, Figure 4(c), shows that even in perfectly regular facades, our grammar method can be better than global methods such as Li et al.

Finally, we demonstrate in Figure 5, the ability of our algorithm to handle complex facades consisting of multiple depth layers and various shape primitives (cylinders in this case). Colors in Figures 5(a) and 5(c) depict several facades and their corresponding reconstruction. In Figure 6, we applied our method on a highly non-regular facade (as can be seen by the complex depth layers in 6(b)). The flexibility of the grammar-based segmentation allows to correctly segment and reconstruct the model.

Conclusions and Future Work. We presented a method for reconstructing building facades from LiDAR scans using a grammar-based segmentation algorithm. The idea is to model the space of segmentation solutions using a set of simple procedural rules and a dictionary of basic elements. Shape grammars offer a flexible yet powerful tool to account for facades variability in comparison with global repetition methods. The key idea is that by utilizing grammar rules and dictionary shapes we reduce and regularize the problem space as demonstrated in our results and comparisons.

We formulate the segmentation problem as a probabilistic fitting problem of foreground and background rectangular shapes generated from grammar derivation rules. We find the best segmentation through a simple maximum log likelihood formulation of the grammar. We then utilize the accurate segmentation to extrude corresponding parts and reconstruct the polygonal model.

Nevertheless, our grammar definition is rather simple as it does not incorporate any high-level priors such as symmetries. We would like to pursuit this direction in our future work as it seems that it could further enhance the method. Additionally, we would like to improve the method's running time by looking for alternative optimization methods rather than random walks, possibly by modeling the segmentation search space in some functional way.

References

[1] L. Zebedin, J. Bauer, K. Karner, H. Bischof, Fusion of feature- and area-based information for urban buildings modeling from aerial imagery, in: Proceedings of the ECCV, 2008, pp. 873–886. 1

[2] J. Xiao, T. Fang, P. Zhao, L. Maxime, L. Quan, Image-based street-side city modeling, ACM Trans Graphics 28 (5) (2009) 1–12. 1, 3

[3] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, D. Fulk, The digital michelangelo project: 3D scanning of large statues, in: ACM SIGGRAPH, 2000, pp. 131–144. 1

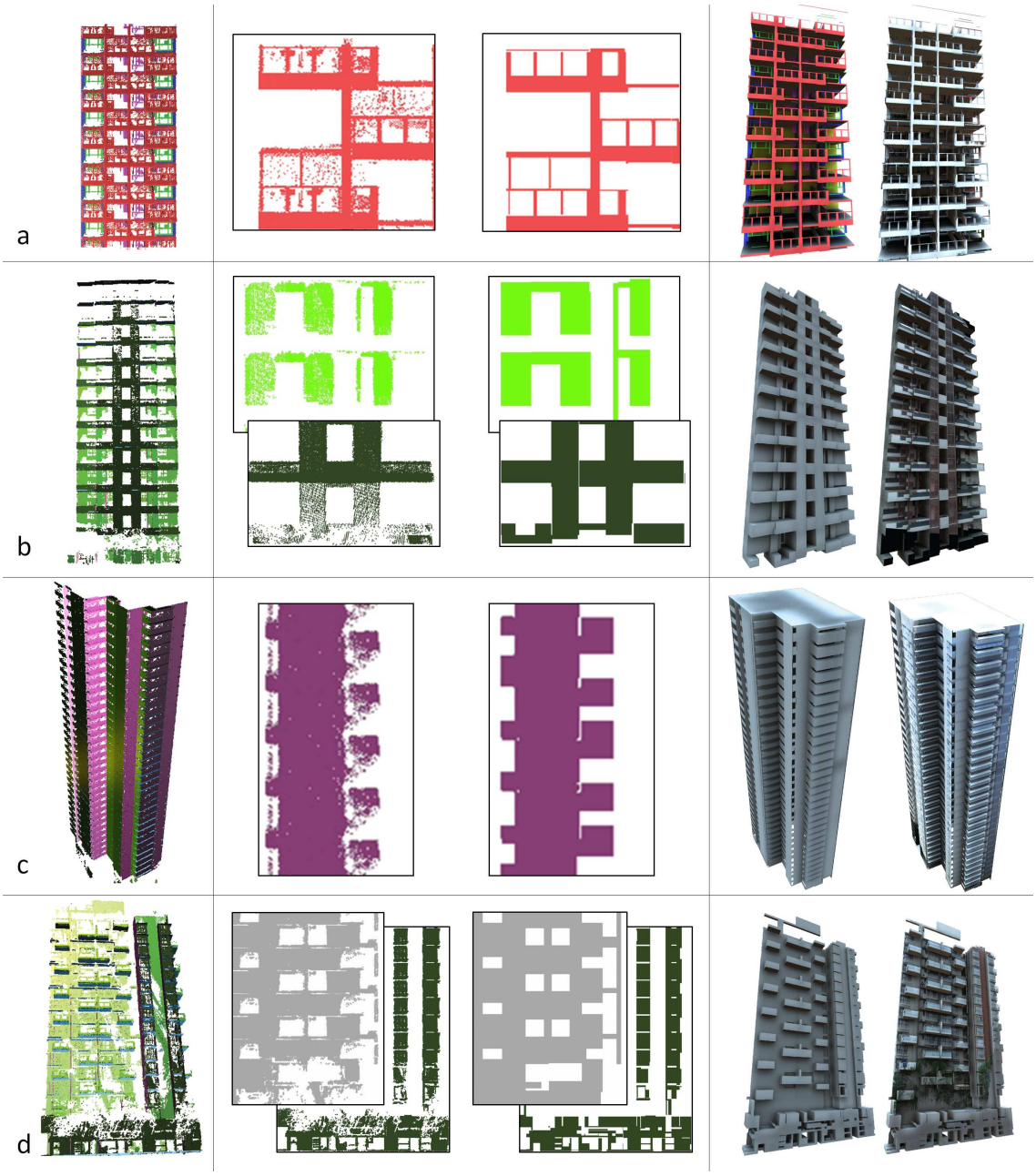

Figure 3: Results of our grammar-based reconstruction on four facades (rows (a)- (d)). In leftmost column is the original point cloud, middle are zoom-ins of depth layers and their grammar based segmentation, rightmost column is the reconstructed model and with applied texture.

[4] P. E. Debevec, C. J. Taylor, J. Malik, Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach, ACM SIGGRAPH 30 (1996) 11–20. 1, 3

[5] J. H. Hays, M. Leordeanu, A. A. Efros, Y. Liu, Discovering texture regularity as a higher-order correspondence problem, in: Proceedings of the ECCV, 2006, pp. 522–535. 1

[6] N. J. Mitra, L. Guibas, M. Pauly, Partial and approximate symmetry detection for 3D geometry, in: ACM SIGGRAPH, Vol. 25, 2006, pp. 560–568. 1

[7] T. Korah, C. Rasmussen, 2d lattice extraction from structured environments, in: ICIP, 2007, pp. 61–64. 1

[8] P. Müller, G. Zeng, P. Wonka, L. J. V. Gool, Image-based procedural modeling of facades, ACM Trans Graphics 26 (3) (2007) 85. 1, 2, 3

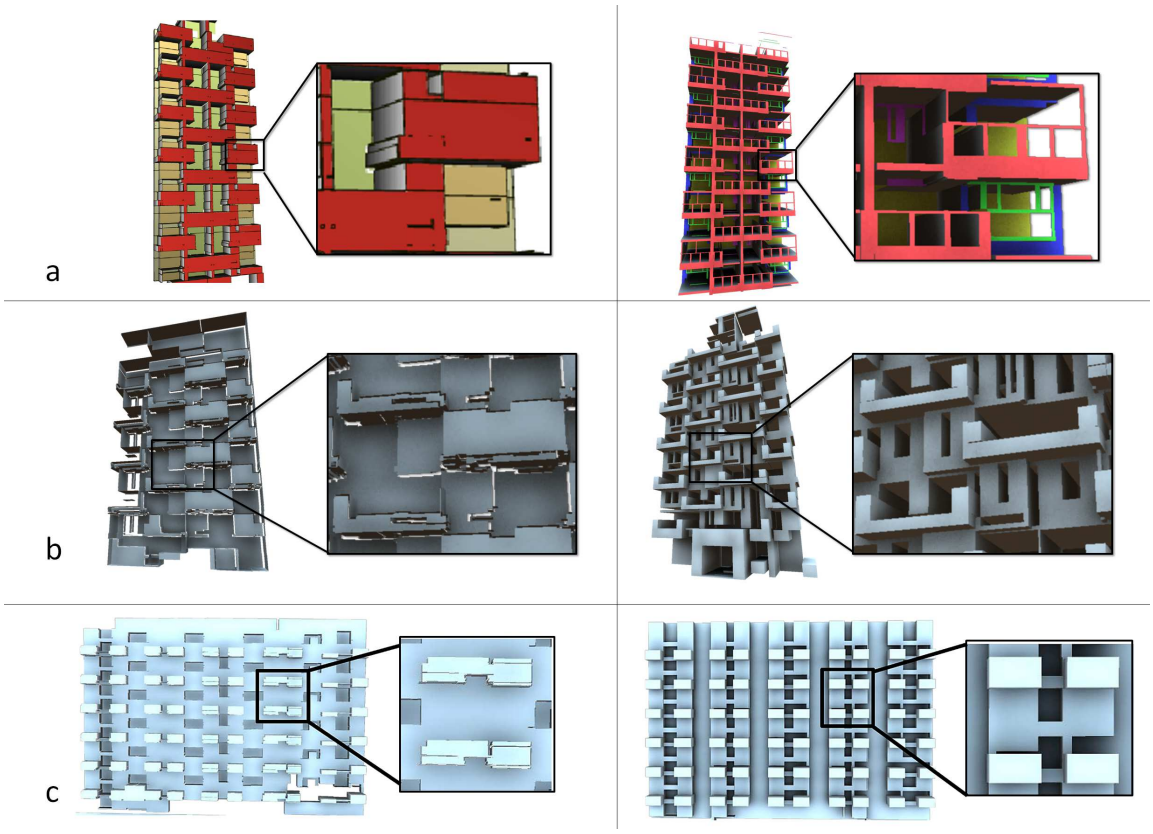

Figure 4: Comparison of Li et al. [2011] (left column) and our method (right column) on three different facades data sets (rows (a)- (c)). Zoom-ins are used to magnify fine geometric detail reconstruction.

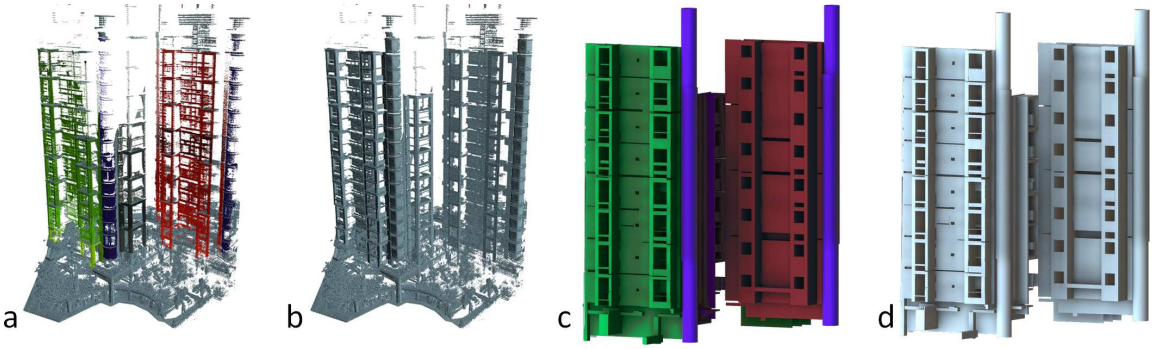

Figure 5: Result of applying our method to a facade containing complex depth layers, rich geometric detail and cylinder-like surfaces. We color facades in the original input (a) and perform a scan filtering and consolidation (b). Our grammar-based reconstruction is colored with corresponding colors to initial facades (c), and the resulting full model (d).

[9] M. Pauly, N. J. Mitra, J. Wallner, H. Pottmann, L. Guibas, Discovering structural regularity in 3D geometry, ACM Trans Graphics 27 (3) (2008) 1–11. 1, 2, 3

[10] P. Musialski, P. Wonka, M. Recheis, S. Maierhofer, W. Purgathofer, Symmetry-based facade repair, in: Vision, Modeling, and Visualization Workshop 2009 (VMV09), 2009, pp. 3–10. 1

[11] L. Nan, A. Sharf, H. Zhang, D. Cohen-Or, B. Chen, Smartboxes for interactive urban reconstruction, ACM SIGGRAPH 29 (2010) 1–10. 1, 3

[12] Q. Zheng, A. Sharf, G. Wan, Y. Li, N. J. Mitra, D. Cohen-Or, B. Chen, Non-local scan consolidation for 3d urban scenes, ACM SIGGRAPH

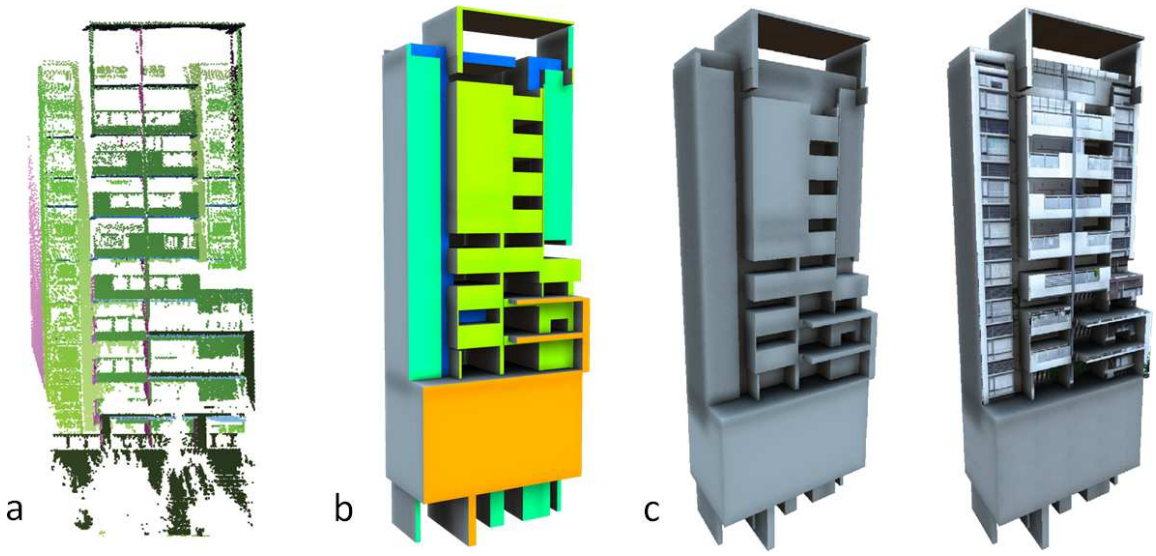

Figure 6: Result of our method applied to a highly non-symmetric facade. The original noisy input scan (a) and the detected complex depth layers colored with unique colors (b). The reconstruction result of our method and the textured version are in (c).

29 (2010) 1–9. 1, 2, 3, 4

[13] Y. Li, Q. Zheng, A. Sharf, D. Cohen-Or, B. Chen, N. J. Mitra, 2d-3d fusion for layer decomposition of urban facades, in: IEEE International Conference on Computer Vision (ICCV), 2011, pp. 1–10. 1, 3, 7

[14] O. Teboul, L. Simon, P. Koutsourakis, N. Paragios, Segmentation of building facades using procedural shape priors., in: CVPR'10, 2010, pp. 3105–3112. 2, 3

[15] C. A. Vanegas, D. G. Aliaga, P. Wonka, P. Mueller, P. Waddell, B. Watson, Modeling the appearance and behavior of urban spaces, in: Computer Graphics Forum, Vol. 29, 2009, pp. 25–42. 2

[16] F. Schaffalitzky, A. Zisserman, Geometric grouping of repeated elements within images, in: Shape, Contour and Grouping in Computer Vision, 1999, pp. 165–181. 2

[17] T. Korah, C. Rasmussen, Analysis of building textures for reconstructing partially occluded facades, in: Proceedings of the ECCV, 2008, pp. 359–372. 2

[18] C.-H. Shen, S.-S. Huang, H. Fu, S.-M. Hu, Adaptive partitioning of urban facades, ACM SIGGRAPH Asia 30 (6) (2011) 1–9. 2

[19] C. Wu, J.-M. Frahm, M. Pollefeys, Detecting large repetitive structures with salient boundaries, in: Proceedings of the ECCV, 2010, pp. 142–155. 2

[20] M. Bokeloh, A. Berner, M. Wand, H.-P. Seidel, A. Schilling, Symmetry detection using line features, Computer Graphics Forum 28 (2) (2009) 697–706. 2

[21] T. Vieira, A. Bordignon, T. Lewiner, L. Velho, Geometry super-resolution by example, in: Sibgrapi 2009 (XXII Brazilian Symposium on Computer Graphics and Image Processing), 2009, pp. 40–47. 3

[22] T. Werner, A. Zisserman, New techniques for automated architecture reconstruction from photographs, in: Proceedings of the ECCV, Vol. 2, 2002, pp. 541–555. 3

[23] A. R. Dick, P. H. S. Torr, R. Cipolla, Modelling and interpretation of architecture from several images, Int. J. Comput. Vision 60 (2) (2004) 111–134. 3

[24] M. Goesele, N. Snavely, B. Curless, H. Hoppe, S. Seitz, Multiview stereo for community photo collections, in: Proc. ICCV, 2007, pp. 1–8. 3

[25] S. N. Sinha, D. Steedly, R. Szeliski, M. Agrawala, M. Pollefeys, Interactive 3D architectural modeling from unordered photo collections, ACM SIGGRAPH 27 (5) (2008) 1–10. 3

[26] J. Xiao, T. Fang, P. Tan, P. Zhao, E. Ofek, L. Quan, Image-based façade modeling, ACM Trans Graphics 27 (5) (2008) 1–10. 3

[27] Y. Furukawa, B. Curless, S. M. Seitz, R. Szeliski, Reconstructing building interiors from images, in: Proc. ICCV, 2009, pp. 80–87. 3

[28] M. Pollefeys, D. Nistér, J. M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S. J. Kim, P. Merrell, C. Salmi, S. Sinha, B. Talton, L. Wang, Q. Yang, H. Stewénius, R. Yang, G. Welch, H. Towles, Detailed real-time urban 3D reconstruction from video, Int. J. Comput. Vision 78 (2-3) (2008) 143–167. 3

[29] N. Jiang, P. Tan, L.-F. Cheong, Symmetric architecture modeling with a single image, ACM SIGGRAPH 28 (2009) 1–8. 3

[30] R. Schnabel, R. Wahl, R. Klein, Efficient ransac for point-cloud shape detection, Computer Graphics Forum 26 (2) (2007) 214–226. 4

[31] J. M. Coughlan, A. L. Yuille, Manhattan world: Compass direction from a single image by bayesian inference, in: Proc. ICCV, 1999, pp. 941–947. 4

[32] Y. Furukawa, B. Curless, S. M. Seitz, R. Szeliski, Manhattan-world stereo, in: Proc. IEEE CVPR, 2009, pp. 1422–1429. 4

[33] R. G. Gioi, J. Jakubowicz, J.-M. Morel, G. Randall, LSD: A fast line segment detector with a false detection control, in: IEEE PAMI, Vol. 32, 2010, pp. 722–732. 4