City3D: Large-scale Urban Reconstruction from Airborne Point Clouds

Jin Huang, Jantien Stoter, Ravi Peters, Liangliang Nan
Delft University of Technology
{J.Huang-1, J.E.Stoter, R.Y.Peters, liangliang.nan}@tudelft.nl

Abstract

We present a fully automatic approach for reconstructing compact 3D building models from large-scale airborne point clouds. A major challenge of urban reconstruction from airborne point clouds lies in that the vertical walls are typically missing. Based on the observation that urban buildings typically consist of planar roofs connected with vertical walls to the ground, we propose an approach to infer the vertical walls directly from the data. With the planar segments of both roofs and walls, we hypothesize the faces of the building surface, and the final model is obtained by using an extended hypothesis-and-selection-based polygonal surface reconstruction framework. Specifically, we introduce a new energy term to encourage roof preferences and two additional hard constraints into the optimization step to ensure correct topology and enhance detail recovery. Experiments on various large-scale airborne point clouds have demonstrated that the method is superior to the state-of-the-art methods in terms of reconstruction accuracy and robustness. In addition, we have generated a new dataset with our method consisting of the point clouds and 3D models of 20k real-world buildings. We believe this dataset can stimulate research in urban reconstruction from airborne point clouds and the use of 3D city models in urban applications.

1. Introduction

Digitizing urban scenes is an important research problem in computer vision, computer graphics, and photogrammetry communities. 3D models of urban buildings have become the infrastructure for a variety of real-world applications such as visualization [45], simulation [36, 41, 48], navigation [8], and entertainment [18]. These applications typically require high-accuracy and compact 3D building models of large-scale urban environments.

Existing urban building reconstruction methods strive to bring in a great level of detail and automate the process for large-scale urban environments. Interactive reconstruction techniques are successful in reconstructing accurate 3D building models with great detail [29, 30], but they require either high-quality laser scans as input or considerable amounts of user interaction. These methods can thus hardly be applied to large-scale urban scenes. To facilitate practical applications that require large-scale 3D building models, researchers have attempted to address the reconstruction challenge using various data sources [4, 6, 16, 22, 23, 25, 40, 49]. Existing methods based on aerial images [6, 16, 23] and dense triangle meshes [40] typically require good coverage of the buildings, which imposes challenges in data acquisition [53]. Approaches based on airborne point clouds alleviate data acquisition issues. However, the accuracy and geometric details are usually compromised [4, 22, 25, 49].

Following previous works using widely available airborne point clouds, we strive to recover desired geometric details of real-world buildings while ensuring topological correctness, reconstruction accuracy, and good efficiency.

The challenges for large-scale urban reconstruction from airborne point clouds include:

**Building instance segmentation.** Urban scenes are populated with diverse objects, such as buildings, trees, city furniture, and dynamic objects (e.g., vehicles and pedestrians). The cluttered nature of urban scenes poses a severe challenge to the identification and separation of individual buildings from the massive point clouds. This has drawn considerable attention in recent years [31, 38].

**Incomplete data.** Some important structures (e.g., vertical walls) of buildings are typically not captured in airborne point clouds due to the restricted positioning and moving trajectories of airborne scanners.

**Complex structures.** Real-world buildings demonstrate complex structures with varying styles. However, limited cues about structure can be extracted from the sparse and noisy point clouds, which further introduces ambiguities in obtaining topologically correct surface models.

In this work, we address the above challenges with the following strategies. Firstly, we address the building instance segmentation challenge by separating individual buildings using increasingly-available vectorized building footprint data. Secondly, we exploit prior knowledge about the structures of buildings to infer their vertical planes.
Based on the fact that vertical planes in airborne point clouds are typically walls connecting the piecewise planar roofs to the ground, we propose an algorithm to infer the vertical planes from incomplete point clouds. Our method has the option to extrude outer walls directly from the given building footprint. Finally, we approach surface reconstruction by introducing the inferred vertical planes as constraints into an existing hypothesis-and-selection-based polygonal surface reconstruction framework [28], which favors good fitting to the input point cloud, encourages compactness, and enforces manifoldness of the final model (see Fig. 1 for an example of the reconstruction results). The main contributions of this work include:

- a robust framework for fully automatic reconstruction of large-scale urban buildings from airborne point clouds.
- an extension of an existing hypothesis-and-selection-based surface reconstruction method for buildings, which is achieved by introducing a new energy term to encourage roof preferences and two additional hard constraints to ensure correct topology and enhance detail recovery.
- a novel approach for inferring vertical planes of buildings from airborne point clouds, for which we introduce an optimal-transport method to extract polylines from 2D bounding contours.
- a new dataset consisting of the point clouds and reconstructed surface models of 20k real-world buildings.

2. Related Work

A large volume of methods for urban building reconstruction has been proposed. In this section, we mainly review the techniques relevant to the key components of our method. Since our method relies on footprint data for extracting building instances from the massive point clouds of large scenes, and it can also be used for footprint extraction, we also discuss related techniques in footprint extraction.

**Roof primitive extraction.** The commonly used method for extracting basic primitives (e.g., planes and cylinders) from point clouds is random sample consensus (RANSAC) [15] and its variants [34, 54], which are robust against noise and outliers. Another group of widely used methods is based on region growing [10, 32, 37], which assumes roofs are piece-wise planar and iteratively propagates planar regions by advancing the boundaries. The main difference between existing region growing methods lies in the generation of seed points and the criteria for region expansion. In this paper, we utilize an existing region growing method to extract roof primitives given its simplicity and robustness, which is detailed in Rabbani et al. [32].

**Footprint extraction.** Footprints are 2D outlines of buildings, capturing the geometry of outer walls projected onto the ground plane. Methods for footprint extraction commonly project the points to a 2D grid and analyze their distributions [27]. Chen et al [10] detect rooftop boundaries and cluster them by taking into account topological consistency between the contours. To obtain simplified footprints, polyline simplification methods such as the Douglas-Peucker algorithm [13] are commonly used to reduce the complexity of the extracted contours [23, 44, 47]. To favor structural regularities, Zhou and Neumann [50] compute the principal directions of a building and regularize the roof boundary polylines along with these directions. Following these works, we infer the vertical planes of a building by detecting its contours from a heightmap generated from a 2D projection of the input points. The contour polylines are then regularized by orientation-based clustering followed by an adjustment step.

**Building surface reconstruction.** This type of methods aims at obtaining a simplified surface representation of buildings by exploiting geometric cues, e.g., planar primitives and their boundaries [12, 20, 25, 42, 46, 50]. Zhou and Neumann [51] approached this by simplifying the 2.5D TIN (triangulated irregular network) of buildings, which may result in artifacts in building contours due to its limited capability in capturing complex topology. To address this issue, the authors proposed an extended 2.5D contouring method with improved topology control [52]. To cope with missing walls, Chauve et al. [9] also incorporated additional

---

1Our code and dataset will be published after paper acceptance.
primitives inferred from the point clouds. Another group of building surface reconstruction methods involves predefined building parts, commonly known as model-driven approaches [21, 43]. These methods rely on templates of known roof structures and deform-to-fit the templates to the input points. Therefore, the results are usually limited to the predefined shape templates, regardless of the diverse and complex nature of roof structures or high intra-class variations. Given the fact that buildings demonstrate mainly piecewise planar regions, methods have also been proposed to obtain an arrangement of extracted planar primitives to represent the building geometry [5, 14, 24, 28]. These methods first detect a set of planar primitives from the input point clouds and then hypothesize a set of polyhedral cells or polygonal faces using the supporting planes of the extracted planar primitives. Finally, a compact polygonal mesh is extracted from the hypothesized cells or faces. These methods focus on the assembly of planar primitives, for which obtaining a complete set of planar primitives from airborne point clouds is still a challenge.

In this work, we extend an existing hypothesis-and-selection-based general polygonal surface reconstruction method [28] to reconstruct buildings that consist of piecewise planar roofs connected to the ground by vertical walls. We approach this by introducing a novel energy term and a few hard constraints specially designed for buildings to ensure correct topology and decent details.

3. Methodology

3.1. Overview

The proposed approach takes as input a raw airborne point cloud of a large urban scene and the corresponding building footprints, and it outputs 2-manifold and watertight 3D polygonal models of the buildings in the scene. Fig. 2 shows the pipeline of the proposed method. It first extracts the point clouds of individual buildings by projecting all points onto the ground plane and collecting the points lying inside the footprint polygon of each building. Then, we reconstruct a compact polygonal model from the point cloud of each building.

Our reconstruction of a single building is based on the hypothesis-and-selection-based framework of PolyFit [28], which is for reconstructing general piecewise-planar objects from a set of planar segments extracted from the point cloud. Our method exploits not only the planar segments directly extracted from the point cloud but also the vertical planes inferred from the point cloud. From these two types of planar primitives, we hypothesize the faces of the building. The final model is then obtained by choosing the optimal subset of the faces through optimization.

The differences between our method and PolyFit are: 1) our method is dedicated to reconstructing urban buildings and it makes use of vertical planes as hard constraints, for which we propose a novel algorithm for inferring the vertical planes of buildings that are commonly missing in airborne point clouds. 2) We introduce a new roof preference energy term and two additional hard constraints into the optimization to ensure correct topology and enhance detail recovery. In the following sections, we detail the key steps of our method with an emphasis on the processes that differ from PolyFit [28].

3.2. Inferring vertical planes

With airborne point clouds, important structures like vertical walls of a building are commonly missed due to the
restricted positioning and moving trajectories of the scanner. In contrast, the roof surfaces are usually well captured. This inspired us to infer the missing walls from the available points containing the roof surfaces. We infer the vertical planes representing not only the outer walls but also the vertical walls within the footprint of a building. We achieve this by generating a 2D rasterized height map from its 3D points and looking for the contours that demonstrate considerable variations in the heigh values. To this end, an optimal-transport method is proposed to extract closed polylines from the contours. The polylines are then extruded to obtain the vertical walls. The process for inferring the vertical planes is outlined in Fig. 2 from (d) to (f).

Specifically, after obtaining the point cloud of a building, we project the points onto the ground plane, from which we create a height map. To cope with the non-uniform distribution of the points (e.g., some regions have holes while others may have repeating points), we construct a Triangulated Irregular Network (TIN) model using 2D Delaunay triangulation. The TIN model is a continuous surface and naturally completes the missing regions. Then, a heightmap is generated by rasterizing the TIN model with a specified resolution \( r \). The issue of small holes in the heightmaps (due to uneven distribution of roof points) is further alleviated by image morphological operators while preserving the shape and size of the building [17]. After that, a set of contours are extracted from the heightmap using the Canny detector [7], which serves as the initial estimation of the vertical planes. We propose an optimal-transport method to extract polylines from the initial set of contours.

**Optimal-transport method for polyline extraction.** The initial set of contours are discrete pixels, denoted as \( S \), from which we would like to extract simplified polylines that best describe the 2D geometry of \( S \). Our optimal-transport method for extracting polylines from \( S \) works as follows. First, a 2D Delaunay triangulation \( T_0 \) is constructed from the discrete points in \( S \). Then, the initial triangulation \( T_0 \) is simplified through iterative edge collapse and vertex removal operations. In each iteration, the most suitable vertex to be removed is determined in a way such that the following conditions are met:

- the maximum Hausdorff distance from the simplified mesh \( T_0 \) to \( S \) is less than a distance threshold \( \epsilon_d \).
- the increase of the total transport cost [11] between \( S \) and \( T_0 \) is kept at a minimum.

In each iteration, a vertex satisfying the above conditions is removed from \( T_0 \) by edge collapse, and the overall transportation cost is updated.

As the iterative simplification process continues, the overall transportation cost will increase. The edge collapse operation stops until no vertex can be further removed, or the overall transportation cost has increased above a user-specified tolerance \( \epsilon_c \). After that, we apply an edge filtering step [11] to eliminate small groups of undesirable edges caused by noise and outliers. Finally, the polylines are derived from the remaining vertices and edges of the simplified triangulation using the procedure described in De Goes et al. [11]. Compared to De Goes et al. [11], our method not only minimizes the total transport cost but also provides control over local geometry, ensuring that the distance between every vertex in the final polylines and the initial contours is smaller than the specified distance threshold \( \epsilon_d \).

**Regularity enhancement.** Due to noise and uneven point density in the point cloud, the polylines generated by the optimal-transport algorithm are unavoidably inaccurate and irregular (see Fig. 3 (a)), which often leads to artifacts in the final reconstruction. We alleviate these artifacts by enforcing structure regularities that commonly dominate urban buildings. We consider the structure regularities defined by Li and Wu [26], namely horizontal, vertical, parallelism, collinearity, orthogonality, and Z-symmetry. We propose a clustering-based method to identify the groups of line segments that potentially satisfy these regularities. Our method achieves structure regularization by two steps: clustering and adjustment.

**Clustering.** We cluster the line segments of the polylines generated by the optimal-transport algorithm based on their orientation and pairwise Euclidean distance [35]. The pairwise Euclidean distance is measured by the minimum distance between a line segment and the supporting line of the other line segment.

**Adjustment.** For each cluster that contains multiple line segments, we compute its average direction. Then each line segment in the cluster is adjusted to align with the average direction. In case the building footprint is provided, the structure regularity can be further improved by aligning the segments with the edges in the footprint. After average adjustment, the near-collinear, near-orthogonal line segments are adjusted to be perfectly collinear and orthogonal, respectively (we use an angle threshold of 20°).

After regularity enhancement, the vertical planes of the

![Figure 3](image-url)
building can be obtained by vertical extrusion of the regularized polylines. The effect of the regularity enhancement is demonstrated in Fig. 3, from which we can see that it significantly improves structure regularity and reduces the complexity of the building outlines.

### 3.3. Reconstruction

Our surface reconstruction involves two types of planar primitives, i.e., vertical planes inferred in the previous step (see Sec. 3.2) and roof planes directly extracted from the point cloud. Unlike PolyFit [28] that hypothesizes faces by computing pairwise intersections using all planar primitives, we compute pairwise intersections using only the roof planes, and then the resulted faces are cropped with the outer vertical planes (see Fig. 2 (g)). This process ensures that the roof boundaries of the reconstructed building can be precisely connected with the inferred vertical walls. Besides, since the object to be reconstructed is a real-world building, we introduce a roof preference energy term and a set of new hard constraints specially designed for buildings into the original formulation. Specifically, our objective for obtaining the model faces \( F^* \) can be written as

\[
F^* = \arg \min_X \lambda_d E_d + \lambda_c E_c + \lambda_r E_r, \tag{1}
\]

where \( X = \{ \chi_i | \chi_i \in \{0, 1\} \} \) denotes the binary variables for the faces (1 for selected and 0 otherwise). \( E_d \) is the data fitting term that encourages selecting faces supported by more points, and \( E_c \) is the model complexity term that favors simple planar structures. For more details about the data fitting term and the model complexity term, please refer to Nan and Wonka [28]. In the following part, we mainly elaborate on the new energy term and hard constraints.

**New energy term: roof preference.** We have observed in rare cases that a building in aerial point clouds may demonstrate more than one layer of roofs, e.g., semi-transparent or overhung roofs. In such a case, we assume a higher roof face is always preferable to the ones underneath. We formulate this preference as an additional energy term called roof preference, which is defined as

\[
E_r = \frac{1}{|F|} \sum_{i=1}^{|F|} \chi_i \cdot \frac{z_{\text{max}} - z_i}{z_{\text{max}} - z_{\text{min}}} \tag{2}
\]

where \( z_i \) denotes the \( Z \) coordinate of the centroid of a hypothesized face \( f_i \). \( z_{\text{max}} \) and \( z_{\text{min}} \) are respectively the highest and lowest \( Z \) coordinates of the building points. \(|F|\) denotes the total number of hypothesized faces.

**New hard constraints.** We impose two hard constraints to enhance the topological correctness of the final reconstruction.

- **Single roof.** This constraint ensures that the reconstructed 3D model of a real-world building has a single layer of roofs, which can be written as,

\[
\sum_{\chi \in V(f_i)} \chi_i = 1, (1 \leq i \leq |F|)
\]

where \( V(f_i) \) denote the set of hypothesized faces that have overlap with face \( f_i \in F \) in the vertical direction.

- **Face prior.** This constraint enforces that for all the derived faces from the same planar segment, the one with the highest confidence value is always selected as a prior. Here, the confidence of a face is measured by the number of its supporting points. This constraint can be simply written as

\[
\chi_i = 1,
\]

where \( \chi_i \) is the variable whose value denotes the status of the most confident face \( f_i \) of a planar segment. This constraint resolves ambiguities if two hypothesized faces are near coplanar and close to each other, which preserves finer geometric details. The effect of this constraint is demonstrated in Fig. 4.

The final surface model of the building can be obtained by solving the optimization problem given in Eq. (1), subject to the single roof and face prior hard constraints.

### 4. Results and Evaluation

Our method is implemented in C++ using CGAL [1]. All experiments were conducted on a desktop PC with a 3.5 GHz AMD Ryzen Threadripper 1920X and 64 GB RAM.

**Test datasets.** We have tested our method on three datasets of large-scale urban point clouds including more than 20k buildings.
AHN3 [2]. A country-wide airborne point cloud dataset covering the entire Netherlands, with an average point density of \(8 \text{ points/m}^2\). The corresponding footprints\(^2\) of the buildings are obtained from the Register of Buildings and Addresses (BAG) [3], with an accuracy of 30 cm.

DALES [39]. A large-scale aerial point cloud dataset consisting of forty scenes spanning an area of 10 km\(^2\), with instance labels of 6k buildings. The average point density is 50 points/m\(^2\). No footprint data is available in this dataset.

Vaihingen [33]. An airborne point cloud dataset published by ISPRS, which has been widely used in semantic segmentation and reconstruction of urban scenes. We use in our experiments its training set that contains footprint information and covers an area of 399 m \(\times\) 421 m with 753k points. The average point density is 4 points/m\(^2\).

### 4.1. Reconstruction results

**Visual results.** We have used our method to reconstruct more than 20k buildings from the aforementioned three datasets. For the AHN3 [2] and Vaihingen [33] datasets, the provided footprints were used for both building instance segmentation and extrusion of the outer walls. Our inferred vertical planes were used to complete the missing inner walls. For the DALES [39] dataset, we used the provided instance labels to extract building instances, and we used our inferred vertical walls for the reconstruction.

Fig. 1 shows the 3D reconstruction of all buildings in a large scene from the AHN3 dataset [2]. The reconstructed building models are simplified polygonal meshes with an average face count of 34. To better reveal the quality of our reconstructed building models, we demonstrate in Fig. 5 a set of individual buildings reconstructed from the three test datasets. From these visual results, we can see that although the buildings have diverse structures of different styles, and the input point clouds have varying densities and different levels of noise, outliers, and missing data, our method succeeded in obtaining visually plausible reconstruction results. These experiments also indicate that our approach is successful in inferring the vertical planes of buildings from airborne point clouds and it is effective to include these planes in the 3D reconstruction of urban buildings.

**Quantitative results.** We have also evaluated the reconstruction results quantitatively. Since ground-truth reconstruction is not available for all buildings in the three datasets, we chose to use the commonly used accuracy measure, Root Mean Square Error (RMSE), to quantify the quality of each reconstructed model. In the context of surface reconstruction, RMSE is defined as the square root of the average of squared Euclidean distances from the points to the reconstructed model. In Tab. 1, we report the statistics of our quantitative results on the buildings shown in Fig. 5. We can see that our method has obtained good reconstruction accuracy, i.e., the RMSE for all buildings is between 0.04 m to 0.26 m, which is quite promising for 3D reconstruction of real-world buildings from noisy and sparse airborne point clouds. Observed from the number of faces column of Tab. 1, our results are simplified polygonal models and are more compact than those obtained from commonly used approaches such as the Poisson surface reconstruction method [19] (that produces dense triangles). Tab. 1 also shows that the running times for most buildings are less than 30 seconds. The reconstruction of the large

\[\text{Table 1. Statistics on the reconstructed buildings shown in Fig. 5.}\]

| Dataset | Model | #Points | #Faces | RMSE (m) | Time (sec) |
|---------|-------|---------|--------|----------|------------|
| AHN3    | (1)   | 732     | 23     | 0.07     | 3          |
|         | (2)   | 532     | 42     | 0.12     | 4          |
|         | (3)   | 1165    | 31     | 0.04     | 3          |
|         | (4)   | 20365   | 127    | 0.15     | 62         |
|         | (5)   | 1371    | 48     | 0.04     | 5          |
|         | (6)   | 1611    | 45     | 0.06     | 4          |
|         | (7)   | 3636    | 68     | 0.21     | 18         |
|         | (8)   | 2545    | 52     | 0.04     | 8          |
|         | (9)   | 15022   | 63     | 0.11     | 28         |
|         | (10)  | 23654   | 262    | 0.26     | 115        |
|         | (11)  | 13269   | 102    | 0.11     | 34         |
|         | (12)  | 155360  | 1520   | 0.09     | 2520       |
|         | (13)  | 24027   | 176    | 0.24     | 141        |
|         | (14)  | 28522   | 227    | 0.15     | 78         |
| DALES   | (15)  | 8662    | 39     | 0.04     | 11         |
|         | (16)  | 11830   | 73     | 0.1      | 8          |
|         | (17)  | 10673   | 47     | 0.07     | 7          |
|         | (18)  | 7594    | 33     | 0.07     | 14         |
|         | (19)  | 13060   | 278    | 0.05     | 145        |
|         | (20)  | 11114   | 55     | 0.06     | 24         |
|         | (21)  | 8589    | 51     | 0.06     | 15         |
|         | (22)  | 18909   | 282    | 0.08     | 86         |
| Vaihingen | (23) | 7701    | 51     | 0.24     | 25         |
|         | (24)  | 6845    | 99     | 0.12     | 8          |
|         | (25)  | 1007    | 24     | 0.11     | 2          |
|         | (26)  | 11591   | 206    | 0.17     | 10         |
|         | (27)  | 4026    | 42     | 0.26     | 6          |
|         | (28)  | 5059    | 61     | 0.22     | 9          |

\(^2\)In BAG [3], the outlines of buildings are from rooftops, which is slightly different from footprints. We still use ‘footprint’ in this paper.
complex building shown in Fig. 5(12) took 42 minutes. This is due to the high complexity in the building, resulting in a large optimization problem [28].

**New dataset.** Our method has been applied to city-scale building reconstruction. The results are released as a new dataset consisting of 20k buildings (including the reconstructed 3D models and the corresponding airborne point clouds). We believe this dataset can stimulate research in urban reconstruction from airborne point clouds and the use of 3D city models in urban applications.

### 4.2. Parameters

Our method involves a few parameters that are empirically set to fixed values for all experiments, i.e., the distance threshold $\epsilon_d = 0.25$ and the tolerance for overall transportation cost $\epsilon_c = 2.0$. The resolution $r$ for the rasterization of the TIN model to generate height maps is dataset dependent due to the difference in point density. It is set to $0.20 \, m$ from...
Figure 6. Comparison with 2.5D Dual Contouring (2.5DC) [51] and PolyFit [28] on a single building from the AHN3 dataset [2].

| Method           | #Faces | RMSE (m) | Time (sec) |
|------------------|--------|----------|------------|
| 2.5D DC [51]     | 12781  | 0.213    | 13         |
| PolyFit [28]     | 1848   | 0.242    | 160        |
| Ours             | 2453   | 0.128    | 380        |

Table 2. Statistics on the comparison of 2.5D Dual Contouring [51], PolyFit [28], and our method on the reconstruction of a city block consisting of 160 buildings from the AHN3 dataset [2]. Total face numbers, running times, and average errors are reported.

4.3. Comparisons

We have compared our method with two successful open-source methods, i.e., 2.5D Dual Contouring (dedicated for urban buildings) [51] and PolyFit (for general piecewise-planar objects) [28], on a city block containing 160 buildings from the AHN3 dataset [2]. The input point cloud is sparse and contains only 80,447 points (i.e., on average 503 points per building), and the walls are severely occluded. Fig. 6 shows the visual comparison of one of the buildings. PolyFit assumes a complete set of input planar primitives, which is not the case for airborne point clouds because the vertical walls are often missing. For PolyFit to be effective, we added our inferred vertical planes to its initial set of planar primitives. From the result, we can observe that both PolyFit and our method can generate compact building models, and the number of faces in the result is an order of magnitude less than that of the 2.5D Dual Contouring method. It is worth noting that even with the additional planes, PolyFit still failed to reconstruct some walls and performed poorly in recovering geometric details. In contrast, our method produces the most plausible 3D models. By inferring missing vertical planes, our method can recover inner walls, which further split the roof planes and bring in more geometric details into the final reconstruction. Tab. 2 reports the statistics of the comparison, from which we can see that the reconstructed building models from our method have the highest accuracy. In terms of run time, our method is slower than the other two but it is still acceptable in practical applications (on average 2.4 sec per building).

4.4. Limitations

Our method can infer the missing vertical planes of buildings, from which the outer vertical planes serve as outer walls in the reconstruction. Since the vertical planes are inferred from the 3D points of rooftops, the walls in the final models may not perfectly align with the ground-truth footprints (see the figure on the right). Thus, we recommend the use of high-quality footprint data whenever it is available. Besides, our method extends the hypothesis-and-selection-based surface reconstruction framework of PolyFit [28] by introducing new energy terms and hard constraints. It naturally inherits the limitation of PolyFit, i.e., it may encounter computation bottlenecks for buildings with complex structures (e.g., buildings with more than 100 planar regions). An example has already been shown in Fig. 5 (12).

5. Conclusion

We have presented a fully automatic approach for large-scale 3D reconstruction of urban buildings from airborne point clouds. We propose to infer the vertical planes of buildings that are commonly missing from airborne point clouds. The inferred vertical planes play two different roles during the reconstruction. The outer vertical planes directly become part of the exterior walls of the building, and the inner vertical planes enrich building details by splitting the roof planes at proper locations and forming the necessary inner walls in final models. Our method can also incorporate given building footprints for reconstruction. In case footprints are used, they are extruded to serve the exterior walls of the models, and the inferred inner planes enrich building details. Extensive experiments on different datasets have demonstrated that inferring vertical planes is an effective strategy for building reconstruction from airborne point clouds, and the proposed roof preference energy term and the novel hard constraints ensure topologically correct and accurate reconstruction.
References

[1] CGAL library. *CGAL User and Reference Manual*. CGAL Editorial Board, 5.0.2 edition, 2020. URL https://doc.cgal.org/5.0.2/Manual/packages.html. 5

[2] AHN3. Actueel Hoogtebestand Nederland (AHN). https://www.pdok.nl/nl/ahn3-downloads, 2018. 2, 6, 7, 8

[3] BAG. Basisregistratie Adressen en Gebouwen (BAG). https://bag.basisregistraties.overheid.nl/datamodel, 2019. 6

[4] J.-P. Bauchet and F. Lafarge. City Reconstruction from Airborne Lidar: A Computational Geometry Approach. In *3D GeoInfo 2019 - 14th Conference 3D GeoInfo*, Singapore, Singapore, Sept. 2019. URL https://hal.inria.fr/hal-02295768. 1

[5] J.-P. Bauchet and F. Lafarge. Kinetic shape reconstruction. *ACM Transactions on Graphics (TOG)*, 39(5):1–14, 2020. 3

[6] M. Buyukdemircioglu, S. Kocaman, and U. Isikdag. Semi-automatic 3d city model generation from large-format aerial images. *ISPRS International Journal of Geo-Information*, 7(9):339, 2018. 1

[7] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986. 4

[8] C. Cappelle, M. E. El Najjar, F. Charpillet, and D. Pomorski. Virtual 3d city model for navigation in urban areas. *Journal of Intelligent & Robotic Systems*, 66(3):377–399, 2012. 1

[9] A.-L. Chauve, P. Labatut, and J.-P. Pons. Robust piecewise-planar 3d reconstruction and completion from large-scale unstructured point data. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1261–1268. IEEE, 2010. 2

[10] D. Chen, R. Wang, and J. Peethambaran. Topologically aware building rooftop reconstruction from airborne laser scanning point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):7032–7052, 2017. 2

[11] F. De Goes, D. Cohen-Steiner, P. Alliez, and M. Desbrun. An optimal transport approach to robust reconstruction and simplification of 2d shapes. In *Computer Graphics Forum*, volume 30, pages 1593–1602. Wiley Online Library, 2011. 4

[12] P. Dorninger and N. Pfeifer. A comprehensive automated 3d approach for building extraction, reconstruction, and regularisation from airborne laser scanning point clouds. *Sensors*, 8(11):7323–7343, 2008. 2

[13] D. H. Douglas and T. K. Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: the international journal for geographic information and geovisualization*, 10(2):112–122, 1973. 2

[14] H. Fang and F. Lafarge. Connect-and-slice: an hybrid approach for reconstructing 3d objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13490–13498, 2020. 3

[15] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 2

[16] N. Haala, M. Rothermel, and S. Cavegn. Extracting 3d urban models from oblique aerial images. In *2015 Joint Urban Remote Sensing Event (JURSE)*, pages 1–4. IEEE, 2015. 1

[17] H. Huang, C. Brenner, and M. Sester. A generative statistical approach to automatic 3d building roof reconstruction from laser scanning data. *ISPRS Journal of photogrammetry and remote sensing*, 79:29–43, 2013. 4

[18] A. Kargas, G. Loumos, and D. Varoutas. Using different ways of 3d reconstruction of historical cities for gaming purposes: The case study of nafplio. *Heritage*, (23):1799–1811, 2019. 1

[19] M. Kazhdan, M. Bolitho, and H. Hoppe. Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, volume 7, 2006. 6

[20] F. Lafarge and C. Mallet. Creating large-scale city models from 3d-point clouds: a robust approach with hybrid representation. *International journal of computer vision*, 99(1):69–85, 2012. 2

[21] F. Lafarge, X. Descombes, J. Zerubia, and M. Pierrot-Deseilligny. Structural approach for building reconstruction from a single dsm. *IEEE Transactions on pattern analysis and machine intelligence*, 32(1):135–147, 2008. 3

[22] H. Ledoux, F. Biljecki, B. Dukai, K. Kumar, R. Peters, J. Stoter, and T. Commandeur. 3dfier: automatic reconstruction of 3d city models. *Journal of Open Source Software*, 6:2866, 2021. 1

[23] M. Li, L. Nan, N. Smith, and P. Wonka. Reconstructing building mass models from uav images. *Computers & Graphics*, 54:84–93, 2016. 1, 2

[24] M. Li, P. Wonka, and L. Nan. Manhattan-world urban reconstruction from point clouds. In *ECCV*, 2016. 3

[25] M. Li, F. Rottensteiner, and C. Heipke. Modelling of buildings from aerial lidar point clouds using tins and label maps. *ISPRS Journal of Photogrammetry and Remote Sensing*, 154:127–138, 2019. 1, 2

[26] Y. Li and B. Wu. Relation-constrained 3d reconstruction of buildings in metropolitan areas from photogrammetric point clouds. *Remote Sensing*, 13(1):129, 2021. 4

[27] X. Meng, L. Wang, and N. Currit. Morphology-based building detection from airborne lidar data. *Photogrammetric Engineering & Remote Sensing*, 75(4):437–442, 2009. 2
[54] M. Zuliani, C. S. Kenney, and B. Manjunath. The multi-ransac algorithm and its application to detect planar homographies. In *IEEE International Conference on Image Processing 2005*, volume 3, pages III–153. IEEE, 2005.