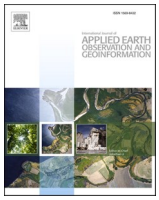
International Journal of Applied Earth Observation and Geoinformation 119 (2023) 103324

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



# Semantic-guided 3D building reconstruction from triangle meshes

Senyuan Wang [a], Xinyi Liu [a,*], Yongjun Zhang [a,*], Jonathan Li [b], Siyuan Zou [a], Jipeng Wu [c], Chuang Tao [d], Quan Liu [d], Guorong Cai [e]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, HB 430079, China
[b] Department of Geography and Environmental Management and Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[d] Shanghai Weizhizhuoxin Information Technology Co. Ltd., Shanghai 200120, China
[e] School of Computer Engineering, Jimei University, Xiamen, FJ 361021, China

ARTICLE INFO

*Keywords:*
Semantic-guided
Building reconstruction
Triangle meshes
Space partition with contour
Structural recovery

ABSTRACT

Planar primitives tend to be incorrectly detected or incomplete in complex scenes where adhesions exist between different objects, resulting in topology errors in the reconstructed models. We propose a semantic-guided building reconstruction method known as semantic-guided reconstruction (SGR), which is capable of achieving the independence and integrity of building models in two key stages. In the first stage, the space partition is represented by a 2.5D convex cell complex and is capable of restoring planar primitives that are easily lost and can further infer the potential structural adaptivity. The second stage incorporates semantic information into a graph-cut formulation that can assist in the independent reconstruction of buildings while eliminating interference from the surrounding environment. Our experimental results confirmed that the SGR method can authentically reconstruct weakly observed surfaces. Furthermore, qualitative and quantitative evaluations show that SGR is suitable for reconstructing surfaces from insufficient data with semantic and geometric ambiguity or semantic errors and can obtain watertight models considering fidelity, integrity and time complexity.

## 1. Introduction

The efficient reconstruction of 3D building models in an automated manner is in constant demand for applications such as smart cities (Yang and Lee, 2019; Allam et al., 2022), environmental analysis (Badach et al., 2020), and entertainment (Matrone et al., 2020; Biljecki et al., 2015; Hackel et al., 2017). At present, as the construction of smart cities transforms from digital twins to the Metaverse, the concept of "3D Real Scene" has been proposed in China (Xu et al., 2021), aiming to build real 3D models in digital space from geographic entities in the real world.

Building reconstruction is a tricky task because of its complex structure and delicate topology. The data are often obtained from laser scanning or photogrammetry, containing noise and occlusions (Han et al., 2021). However, the acquisition costs of point clouds through Airborne Laser Scanning (ALS) are high; points on façades are often missing owing to flight altitude. Urban 3D model reconstruction based on oblique photogrammetry (Zhang et al., 2020; Xiao, 2019; Bas and Ok, 2021; Zhang et al., 2022; Zou et al., 2023) is currently one of the most popular reconstruction methods. Oblique photogrammetry mainly uses multiview aerial images as input and generates a triangular mesh through a series of steps. Compared with point clouds, the triangular mesh has a better shape description, and some noise points have been eliminated in its generation process (Vu et al., 2011). Therefore, we studied building reconstruction based on a triangular mesh. Considering the planar characteristics of man-made buildings, an intuitive approach (Liu et al., 2019; Nan and Wonka, 2017; Bauchet and Lafarge, 2020; Li et al., 2019) is to fit the original data with planes and capture the relationships between them as much as possible, or further infer the potential relationship. Based on further observations of its structure, a building typically has piecewise planar rooftops and vertical walls. The essence of boundary representation modeling (Chen et al., 2021; Zhou and Neumann, 2010) is to accurately extract the building roof and interior structural lines and organize and optimize them according to the correct topology. Advanced roof plane segmentation technology (Zhang et al., 2022) has further improved the accuracy of such methods. However, the modeling method based on the above idea is sensitive to noisy and missing data, which can easily lead to the low geometric accuracy of key points and boundary lines. In addition, geometry and

* Corresponding authors.
*E-mail addresses:* liuxy0319@whu.edu.cn (X. Liu), zhangyj@whu.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.jag.2023.103324
Received 7 January 2023; Received in revised form 4 April 2023; Accepted 18 April 2023
Available online 2 May 2023

semantic information in real data are often biased or even incorrect. In the most serious cases, some adhesions between the buildings and the surrounding environment cannot be separated, resulting in holes or incoherent parts in a single building extracted from a scene. Song et al. (2020) concluded that reconstructing a façade without sufficient observation data remains an issue owing to the continuing inability to acquire oblique data. At present, construction in the context of the Metaverse not only needs to achieve the purpose of visualization, but also needs to identify the readable and applicable single buildings; in this case, relying exclusively on the geometric structure for the reconstruction process is inadequate. Building a model reconstruction that considers semantic information has received increasing attention. Currently, the buildings and surrounding environment are still attached to each other and are not independent.

In this study, we propose a semantic-guided reconstruction (SGR) method for building modeling. The main contributions of the SGR are as follows:

(1) A novel optimization mechanism incorporating semantic information that can extract building entity models from mixed scenes with semantic ambiguity and adhesion between different objects.
(2) A 2.5D space partition that considers the structure contours to restore weakly observed façades, which may be covered or have serious noise.
(3) A reliable point cloud segmentation process based on quasi-a-contrario theory is extended to the plane segmentation of the mesh to accurately and robustly obtain planar primitives from the high noise and uneven mesh that may be present.

The remainder of this paper is organized as follows. Section 2 provides an overview of related studies. Section 3 describes the implementation of the SGR approach. Section 4 presents and discusses experimental results. Section 5 summarizes our conclusions and directions for future work.

## 2. Related works

The building model of an artificial scene is typically regarded as a collection of piecewise planar elements. As initial data always have defects of type or another, it is necessary to infer or "reconstruct" the object from the initial shape detected (Wang et al., 2020). In this paper, we review current methods for extracting geometric primitives and identifying their combinations to obtain a final model.

**Primitive extraction.** The extraction of primitives aims to extract high-quality geometric shapes from raw input data polluted by noise and outliers, including region growth (Rabbani et al., 2006; Nurunnabi et al., 2012) and RANSAC (Schnabel et al., 2007; Awwad et al., 2010). Both methods decompose the input point clouds or meshes into planar primitives.

**Primitive regularization.** Previous studies in this category aimed to further regularize initial primitives by exploring prior knowledge of global regularity (Oesau et al., 2016; Fang et al., 2018). These methods refine the initial primitives from the noise data and reduce the geometry of the output 3D model and the solution space to be explored by assuming a relationship between the basic primitives, such as the relationship between repetition and global alignment (Li et al., 2011), parallel or coplanar relationship between the initial primitives (Monszpart et al., 2015), and Manhattan hypothesis (Coughlan and Yuille, 2000). However, the resulting model may have gaps, rather than being completely aligned, because each primitive is processed separately.

**Primitive assembling.** Two groups of methods can be applied to assemble the initial primitives into a surface mesh model. Connectivity methods (Bouzas et al., 2020) apply an adjacency graph to analyze the relationship between the initial primitives and then extract the output mesh model (Schindler et al., 2011). These methods have advantages in terms of efficiency; however, they can create incorrect links that directly lead to topology errors or incomplete models.

Slicing methods turn the modeling matter into a labeling problem of the "partition plane" or "partition cell" of the space unit, which has been shown to be helpful for generating a topologically closed 3D building model (Chauve et al., 2010; Mura et al., 2016; Nan and Wonka, 2017; Bauchet and Lafarge, 2020). This group of popular methods can be effective and robust in recovering watertight and concise models from challenging data. However, if a single building is seriously hindered by its surrounding environment or experiences semantic ambiguity, the constraint method based on geometric information cannot remove the interference of the surrounding environment or rebuild the building independently.

**Semantic reconstruction.** The significance of semantic information in reconstruction tasks has received increasingly more attention (Blaha et al., 2016). Initial attempts to combine geometry and semantics for reconstruction were based on a depth map (Ladickỳ et al., 2012; Zheng et al., 2020); later, volume representations (Curless and Levoy, 1996; Savinov et al., 2015) and meshes (Cabezas et al., 2015) were utilized. Most of these approaches require two-dimensional images as input data to completely reconstruct the object and smooth the surrounding environment to obtain a visually appealing model (Holzmann et al., 2018). Semantic modeling methods have recently been proposed. To date, most studies have involved labeling the building components and generating a semantic modeling framework (Coudron et al., 2020; Werbrouck et al., 2020).

## 3. 3D reconstruction of a single building with semantic priors

The input of our SGR method requires that the original mesh has semantic information pertaining to all building and non-building categories. Fig. 1 shows the buildings in yellow and non-buildings in green in the semantic mesh. Our SGR method comprises three main phases: (i) planar primitive segmentation, (ii) space partitioning, and (iii) face selection with semantic priors.

### 3.1. Planar primitive segmentation

The commonly used method for initially defining a structure is to detect the initial primitives in the input mesh. In the classical plane segmentation approach (Schnabel et al., 2007), the emphasis is on clustering point sets rather than accurately fitting planes.

We did not directly detect primitives owing to the discontinuity and irregularity of the mesh; rather, the QTPS algorithm (Zhu et al., 2021) was extended to the planar segmentation of the mesh data. First, the QTPS algorithm segmented the encrypted vertices of the triangular mesh into planar clusters $P = \{p_v\}$, and then the adjacent faces within a predefined ring of each inner vertex of the planar segment were used as the seed patch. In general, we selected a ring of neighbors of inliers as adjacent faces. Finally, our two-stage progressive region growth strategy was used to cluster patches on the plane; the same a-contrario theory as in the QTPS was used to verify the accuracy of the extended plane to obtain complete and accurate plane segmentation, $P = \{p_{tri}\}$(see Fig. 2 (a)).

### 3.2. Space partitioning with volumetric unit

The slicing method is currently one of the most prevalent reconstruction methods because of its robustness to defective data using a divide-and-conquer strategy. Inspired by the kinetic Shape Reconstruction (KSR) (Bauchet and Lafarge, 2020), we extended and designed a 2.5D partition mode, the essence of which is to generate meaningful partitions efficiently and adaptively. First, we obtained an adaptive polygon partition based on the projection of façade primitives in two-dimensional space and then described the partition of the overall space by considering the non-façade. The final result was a convex polyhedral set.

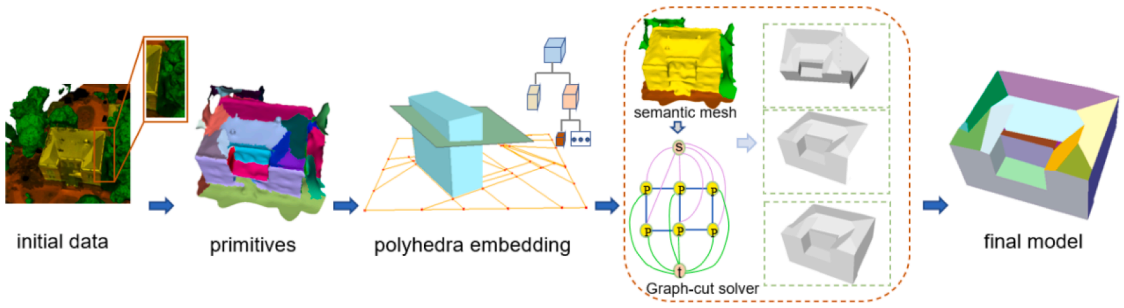



**Fig. 1.** Pipeline of semantic-guided reconstruction.

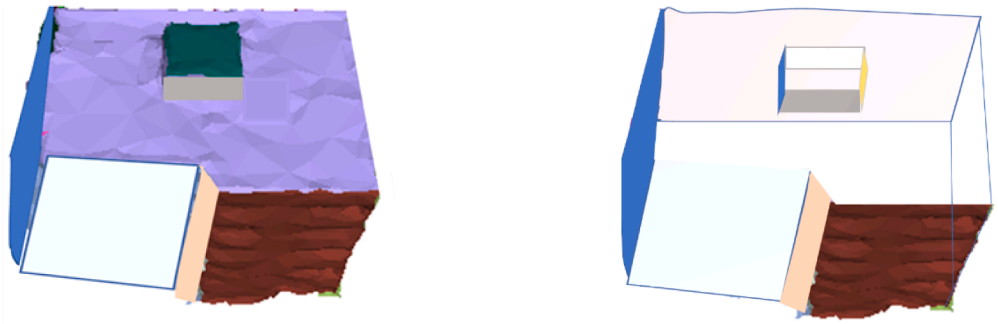



**Fig. 2.** Illustration of the initial planar primitive segmentation.

**Detection of Line-segments:** The Line-Segment Detector (LSD) method (Von et al., 2008) was used to approximate the contour of the two-dimensional projection of the façade primitives to obtain the line segmentation.

**Polygonal partition in dynamic planar graph:** The method of polygon partition guided by line segments is widely used in images (Achanta and Susstrunk, 2017, Duan and Lafarge, 2015) and is also used to vectorize the boundary pixel chain of super voxels. Bauchet and Lafarge (2018) dynamically updated the planar graph by progressively expanding the pre-detected line segments; the restoration of the connection point in the polygon partition was solved in an adaptive manner. This method inspired us to use the detected line segments to guide polygonization in a two-dimensional (2D) space, as follows in Fig. 3.

(i) Initialize the projection of the façade as a graph, $G = (V, E)$, where $V$ is the vertex of the graph, represented by the midpoint of the line segment, and $E$ is the section among the vertices, i.e., the edge.

(ii) A dynamic planar graph $G_t = (V_t, E_t)$ was introduced to generate a polygon partition in 2D. The points in the initial graph (midpoint of the line segment) extend outward at a uniform speed along the two endpoints of the line segment. As shown in Fig. 3 (c), after time t, the ray $l_1$ first intersects with line segment $l_3$ to generate point $p_1$ (a collision) and then continues to extend outward to generate points $p_2$, $p_3$, and $p_4$ successively. Considering the robustness of the defective data and the complexity of the plane partition, the extension stops when the collision number of the line segment reaches two in our experiment. When the line segments are divided and intersected, the graph is updated by inserting new vertices and edges. Therefore, when all line segments in the planar graph no longer extend, the planar partition process is completed. The initial line-segment graph was converted into a new planar graph.

(iii) The polygon partition is stretched to the maximum height of the input mesh to generate a set of polyhedra. Then, the intersection between the non-façade primitive and the polygon in 2D is determined. We note that the non-façade primitives only clip the polyhedra expanded from polygons with 2D intersections. Finally, a 2.5D space partition is generated.

As the dynamic partition is an effective polyhedron embedding, the polyhedral partition has greater shape self-adaptability to generate a watertight and intersection-free model than the tetrahedral partition.

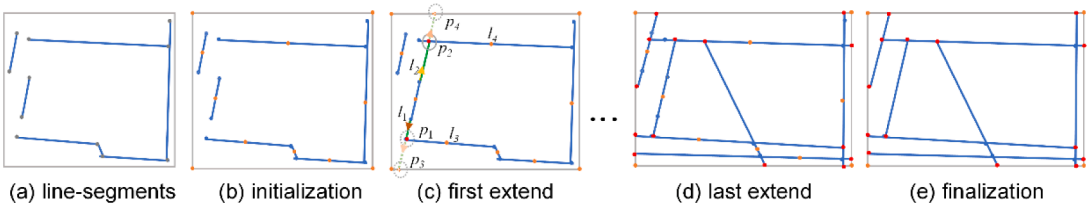



**Fig. 3.** Illustration of the kinetic planar graph: the line segments in (a) are initialized as shown in (b); the extremities (blue dots) of the line segments collide with each other to yield the new nodes (red dots) in (c, d); and the final graph after removing unnecessary nodes is shown in (e). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

### 3.3. Semantic-guided surface extraction

The SGR method was designed to accurately identify the buildings mixed with their surrounding environment according to semantic information and meet the following requirements: (1) the algorithm should be able to discard the parts occluded by the surrounding environment and fill in the holes; and (2) even though there may be parts with semantic errors on the monomer, the complete structure can still be reproduced. The surface extraction of a building is usually transformed into a binary labeling problem of convex polyhedra (Chauve et al. 2010). The cells are labeled as interior or exterior polyhedra using a graph-cut formulation; the output surface becomes the interface between the different labels. This approach designs an energy function related to the semantic information of the original mesh to extract the surface model from the polyhedral embedding. In contrast to assigning an internal and external guess to each polyhedron through ray shooting (Verdie et al. 2015), we adopted a more efficient voting strategy that exploits the semantic information of the initial mesh and the oriented normal of the inliers to assign the label of the polyhedron more robustly.

Let the binary variable $l_i = (in, out)$ encode the inside $(l_i = in)$ and outside $(l_i = out)$ labels of polyhedra $C$. The possible output surface $S = (l_i)_{i \in C}$ is obtained by minimizing the energy function using two terms:

$$\{l_i\} = \text{argmin}E := E_{class\_data} + \lambda E_{class\_area} \tag{1}$$

where $E_{class\_data}$ measures data fidelity and $E_{class\_area}$ evaluates surface complexity, both of which are terms used in $[0, 1]$. $\lambda \in [0, 1]$ is used to balance the two terms.

Data fidelity $E_{class\_data}$ measures the consistency between the inner and outer labels of each polyhedron and the normal direction of the inliers. The aim was to remove the non-building part from the plane of semantic ambiguity. Therefore, the semantic information of the input mesh was introduced to clarify the judgment of consistency. The vertexes of the original mesh are assigned to the plane of the corresponding polyhedron as the inliers of the polyhedron according to the Euclidean distance. The normals can be calculated from the minimum spanning tree and should be as accurate as possible, pointing to the outside of the building surface. Then, the vote for the inlier points of the semantic mesh to express data fidelity is considered, as shown below:

$$E_{class\_data} := \frac{1}{|I|} \sum_{i \in C} \sum_{p \in I_i} d_i^B(p, l_i) \tag{2}$$

where $C$ is the set of polyhedra, $I_i$ is the associated inliers of all the faces of polyhedron $i$, and the value of $|I|$ is twice the number of all the inliers in the original mesh. Specifically, $d_i^B(p, l_i)$ is a voting function related to semantic information that is used to test whether the normal direction of the inner point $p$ is consistent with the label of polyhedron $i$.

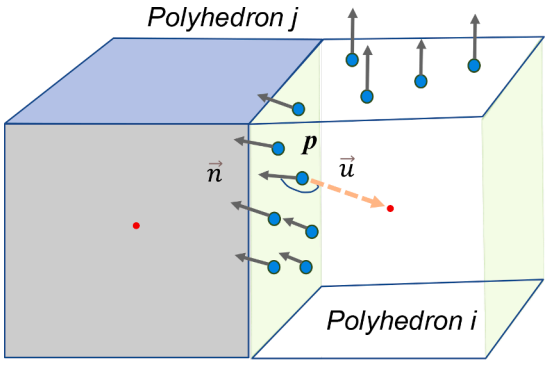



**Fig. 4.** Graphical representation of the consistency of the normal vector. The red point represents the center of the polyhedron; the black arrow represents the normal vector of interior point $p$; and the red arrow represents the vector where $p$ points to the center. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

As shown in Fig. 4, $\vec{n}$ and $\vec{u}$ are the normal vectors of interior point $p$ and the vector from the point $p$ to the center of the corresponding polyhedron, respectively. When the inner product of $\vec{n}$ and $\vec{u}$ is less than zero, this point is generally marked as in, and vice versa. We prefer to assign the labeling polyhedron $i$ as "in" and polyhedron $j$ as "out".

When considering semantic information, we fully exploit the contradiction between its semantic attributes and statistics of its inliers in a convex polyhedron. We may find some phenomena that can easily cause labeling errors; for example, if a polyhedron after a partition contains only one vertex whose normal is toward the outside of the surface, the polyhedron is marked as "out" according to the voting mechanism and is finally selected. However, the results may not be reliable in terms of the number of vertices included or the accuracy of the normal. In addition, a polyhedron cannot be labeled only according to the normal information of the inlier point, and is not sufficient to settle the issue of semantic ambiguity on the plane. Therefore, the related constraints are required. When the statistics for vertices are introduced, the variables $ratio_{f\_P}$, $ratio_{p^B\_all}$, $ratio_{p^B\_P}$, $d_{|in-out|}$, and $n_p$ are defined and graphically illustrated in Fig. 5. Here, $ratio_{f\_P}$ is the ratio of the number of faces labeled as a building class to the total number of faces contained in the polyhedron. If there are three or more inliers of the building class on the face of the polyhedron, it is considered as the face of the building class. Next, $ratio_{p^B\_all}$ is the ratio of the number of building class points in each convex polyhedron to the number of points in the mesh. Then, $ratio_{p^B\_P}$ represents the proportion of building points in each convex polyhedron. To a certain extent, the above three variables reflect whether the convex polyhedron is a part of the building when there is interference from the surrounding environment. We used $d_{|in-out|}$ to count the difference between the number of normals of the points toward the interior of the surface and the exterior of the surface in each polyhedron. Ideally, the normals of all inliers of each polyhedron face the interior or exterior of the surface. If the difference is small, there may be an error in the normals of the inliers of a polyhedron or the difference may indicate that the polyhedron is in a semantically ambiguous part. Finally, $n_p$ is the number of inliers in each convex polyhedron.

Therefore, we propose the concept of the contradiction domain, which mainly refers to the scope that should not belong to the building. The contradictions explored include the following cases:

1. From the perspective of the proportion of the faces covered by the building points, if $ratio_{f\_P}$ is greater than $\theta_{ratio_{f\_P}}$, then the building points cover most of the faces in a convex polyhedron; thus, the polyhedron belongs to the building part. Theoretically, the building points it contains should occupy a certain proportion of the original points unless $ratio_{p^B\_all}$ is relatively small in scale. This condition mainly removes polyhedra whose building points are scattered on several patches, and where the number of points labeled as buildings accounts for a small proportion. This condition may often appear in the boundary part of the semantic ambiguity and may be fragmented.
2. With respect to the building importance of each polyhedron, i.e., the proportion of the building points to all the inliers it contains, if $ratio_{p^B\_all}$ is greater than $\theta_{ratio_{p^B\_all}}$, then this polyhedron contains a certain building point, and all the inliers in the same polyhedron should also have a higher proportion. However, if the value of $ratio_{p^B\_P}$ is very small, the polyhedron likely contains both building points and other types of points.

To balance the influence of the normal consistency and semantics of the inliers on the labels of the polyhedron, we set a range within which the labeling of the polyhedron was clear and the error was small. This range is represented by the following Eq. (3):

$$\frac{d_{|in-out|}}{\max_P\{in,out\}} > ratio_{p^B\_P}(n_p > 3) \tag{3}.$$

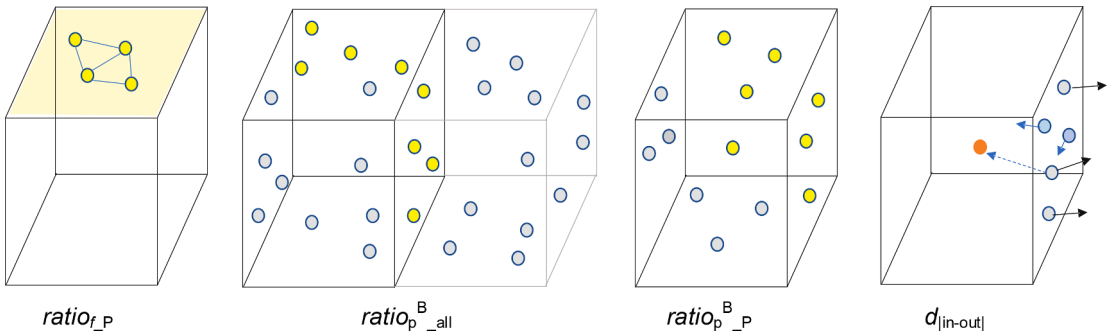



**Fig. 5.** Illustration of the defined variables. The yellow points represent the building points while the gray points represent the inliers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The difference $d_{|in-out|}$ between the inliers labeled with "in" and "out" is more significant than the proportion of building points in the polyhedron. To avoid the voting errors caused by a convex polyhedron containing a vertex, Eq. (3) was used when the number of inliers is not less than three. If one of the following two conditions is satisfied, it can be used as the harmony zone:

1. By satisfying Eq. (3), more than half of the inliers belong to the building class of a polyhedron. For example, $ratio_{p^B\_P} > 0.8$. A polyhedron meeting this condition is mainly covered by building points.
2. By satisfying Eq. (3), $ratio_{f\_P}$, $ratio_{p^B\_all}$, and $ratio_{p^B\_P}$ should all be greater than a certain value in their respective ranges.
3. If the value of $d_{|in-out|}$ is large enough, then the number of inliers labeled with "in" or "out" in the polyhedron will vary greatly. In this case, the effect of semantics on labels should be appropriately reduced. For example, $ratio_{f\_P}$ is greater than a threshold such as 0.04;

The above uncertainty threshold in the definition of contradiction and harmony domains is only a conceptual introduction. See the parameter settings for the subsequent experiments. Based on the above, the voting function is determined as follows:

$$d_i^B(p,l_i) = \begin{cases} d_i(p,l_i), & C_i \in Z_{harmony} \\ 1_{\{l_i=in\}}, & C_i \in Z_{\text{contradiction}} \\ d_i(p,l_i)\cdot ratio_{p^B\_all}, & C_i \in Z_{else} \end{cases} \tag{4}$$

$$d_i(p,l_i) = \begin{cases} 1_{\{\overrightarrow{n}\cdot\overrightarrow{u}>0\}}, \ if\ l_i = in \\ 1_{\{\overrightarrow{n}\cdot\overrightarrow{u}<0\}}, \ if\ l_i = out \end{cases} \tag{5}$$

where $p$ is the interior point of the polyhedron, $l_i$ is the label of the polyhedron, $\overrightarrow{n}$ is the normal of $p$, and $\overrightarrow{u}$ is the vector that $p$ points to the center of the polyhedron.

$d_i(p,l_i)$ is an indicator function when $\overrightarrow{n}\cdot\overrightarrow{u} > 0$, $d_i(p,l_i = in)$ returns a penalty of 1 whereas $d_i(p,l_i = out) = 0$. The binary voting mechanism only requires that the normal points are half the space of the separating surface, which is robust to the imprecise normal directions.

When voting for each interior point, if the point belongs to the harmony region, the voting function is determined only according to the consistency of the normal vector. If the point satisfies the condition of the contradiction field, the label of this polyhedron is forced to be "in" regardless of the normal vector of the interior point. For the remaining ambiguous areas, the normal vector consistency and semantic proportion should be jointly considered.

Second, $E_{class\_area}$ measures the complexity of the output surface, which traditionally is measured by calculating the area of the output surface. Our innovation is that we introduce the semantic information of the original mesh on the output surface. Compared to other methods, we target the real surface area covered by the building class, expressed by Eq. (6):

$$E_{class\_area} = \frac{1}{A}\sum_{\widetilde{ij}} a_{\Delta}^{B}\cdot 1_{\{x_i \neq x_j\}} \tag{6}$$

where the value of $A$ is defined as the sum of the areas of all the divided patches, which is a normalization factor; $i \sim j$ represents two adjacent polyhedra $i$ and $j$, and $a_{\Delta^B}$ denotes the area of the triangular mesh (the semantic label is the building), which is coplanar between polyhedra $i$ and $j$. To maintain a balance between the number of output surfaces and geometric errors, in our experiments, we set the value of $\lambda$ to 0.5. Finally, the energy function was solved using a graph-cut algorithm (Boykov et al., 2004).

## 4. Experimental results and discussion

### 4.1. Experimental settings

The proposed SGR approach was implemented in C++ using the CGAL and VCG libraries. The details of the pipeline of the proposed algorithm are as follows.

(1) Line segmentation expands outward with rays until collisions occur to generate a two-dimensional polygon partition. The collision number $K$ plays a trade-off role between model fineness and the robustness to missing data, which is set by the user. In the experiments, the value of $K$ is set to 2.

(2) The thresholds of several of the variables we used to define the contradiction domain and harmony domain are worth discussing specifically. The three variables, $ratio_{f\_P}$, $ratio_{p^B\_all}$ and $ratio_{p^B\_P}$, appear in both the contradiction domain and harmony domain and the threshold settings are different in the two domains.

We explored the significance of each parameter in comparison experiments. Table 1 lists the parameter settings and Fig. 6 shows a comparison of our visualization results for the different parameters. In Table 1, (a)-(d) show the input parameters of the visualization results in the first row of Fig. 6; (e)-(h) correspond to the second row. The reduction in the threshold value of $ratio_{f\_P}$ led to the expansion of the scope of the contradiction domain; thus, there was a structural loss in the

**Table 1**
Settings for the input parameters corresponding to Fig. 6(a)-(h).

| | contradiction domain | | | harmony domain | | | |
|---|---|---|---|---|---|---|---|
| data | $ratio_{f\_P}$ | $ratio^{B}_{p\text{all}}$ | $ratio^{B}_{pP}$ | $ratio_{f\_P}$ | $ratio^{B}_{p\text{all}}$ | $ratio^{B}_{pP}$ | $d_{\text{in-out}}$ |
| a | 0.8 | 0.05 | 0.8 | 0.3 | 0.02 | 0.2 | 50 |
| b | 0.2 | 0.05 | 0.8 | 0.3 | 0.02 | 0.2 | 50 |
| c | 0.4 | 0.15 | 0.8 | 0.3 | 0.02 | 0.2 | 50 |
| d | 0.8 | 0.05 | 0.8 | 0.01 | 0.08 | 0.02 | 50 |
| e | 0.8 | 0.05 | 0.8 | 0.3 | 0.02 | 0.2 | 50 |
| f | 0.8 | 0.05 | 0.8 | 0.02 | 0.008 | 0.08 | 50 |
| g | 0.8 | 0.05 | 0.8 | 0.3 | 0.02 | 0.2 | 10 |
| h | 0.5 | 0.3 | 0.3 | 0.3 | 0.02 | 0.2 | 30 |

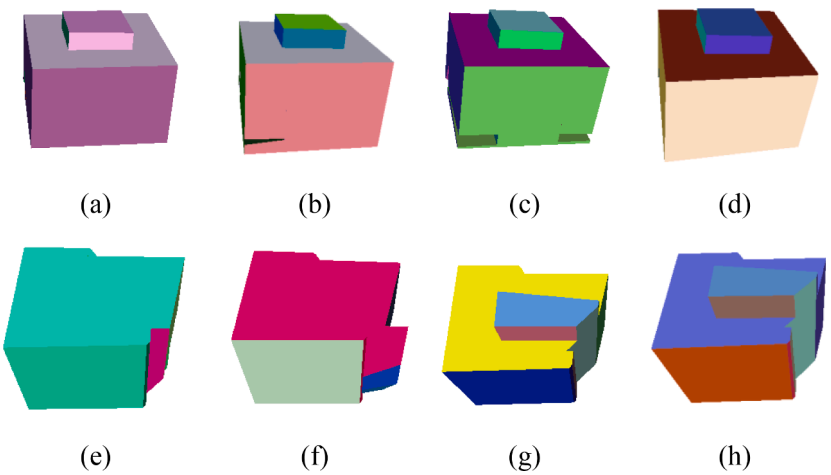



**Fig. 6.** Final model generated from the parameters in Table 1. Different results were generated by adjusting the values of the variables in the defined fields respectively.

result model (see Fig. 6 (b)). According to the statistical information, decreasing the values of the three variables resulted in more polyhedra belonging to the harmony domain. The final model contained additional parts that were not building frames, as shown in Fig. 6 (f). In summary, the increase in variables in the contradiction domain led to the loss of structure in the final model; the reduction in variables in the harmony domain resulted in the retention of more structural details. Therefore, the goal is to find an appropriate threshold that can accurately remove non-building parts adhering to the building frame in the final model reconstruction.

(3) The distance threshold in the plane fitting process is related to the average distance between the original mesh vertices. In this experiment, it was manually set to 0.15 m according to the geometric precision of the mesh.

### 4.2. Experimental data

The experimental data included two mesh datasets containing semantic information. The first dataset depicts a university area in Guangzhou, China, and the second is the SUM dataset (Gao et al. 2021). First, we processed the multi-view aerial oblique image using the ContextCapture (B.SYSTEMS, 2016) commercial software to obtain a triangular mesh model. Then, based on the depth neural network U-Net (Ronneberger et al., 2015), we integrated the geometric features of the mesh, spectral features, and façade value features and introduced an attention mechanism to classify the meshes to obtain a triangular mesh with semantic information. Both datasets contained six semantic categories: terrain, vegetation, buildings, water, vehicles, and ships. Fig. 7 illustrates the two datasets.

### 4.3. Baselines

We contrasted the results of our SGR method with the results of the following three state-of-the-art algorithms: structure-aware method (Bouzas et al., 2020), PolyFit (Nan and Wonka, 2017), and kinetic shape reconstruction (KSR) (Bauchet and Lafarge, 2020). The main feature of the structure-aware method is the use of structural awareness to restore the original primitive and its adjacency. PolyFit obtains a watertight structure by extending and intersecting the initial primitives. KSR uses a dynamic data structure to adaptively divide the space, thereby providing high elasticity to occluded and missing data.

### 4.4. Results and discussion

We evaluated the reconstruction results from different degrees of interference of non-building semantic labels on single buildings and the complexity of the buildings themselves.

**Reconstruction results**. The input data in Fig. 8(a)-(c) are from the Guangzhou survey area. Based on the initial data in Fig. 8(a), the single building structure is complete, but its independent extraction and reconstruction are damaged because of adhesion with the ground. The data in Fig. 8(b) have semantic classification errors; the single building has missing data and curling due to occlusion. The structure of the original data in Fig. 8(c) is more complex than before; semantic classification errors exist over a large area of the façade. The results in Fig. 8 (a)-(c) show that our SGR algorithm completely reconstructed the building from the original mesh with missing and incorrect semantic classification, even generating the details of a building with complicated structures, as shown in Fig. 8(c). The data in Fig. 8(d) and (g) are from

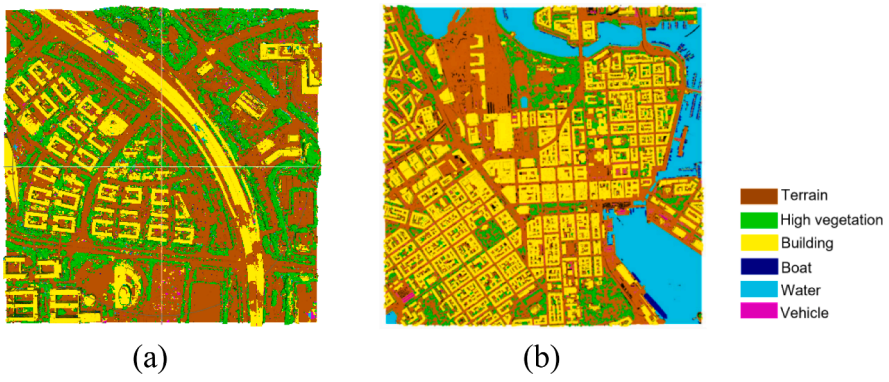



**Fig. 7.** Experimental mesh data. Scene data (a) is from the self-test data and other data (b) is from the SUM public data set.

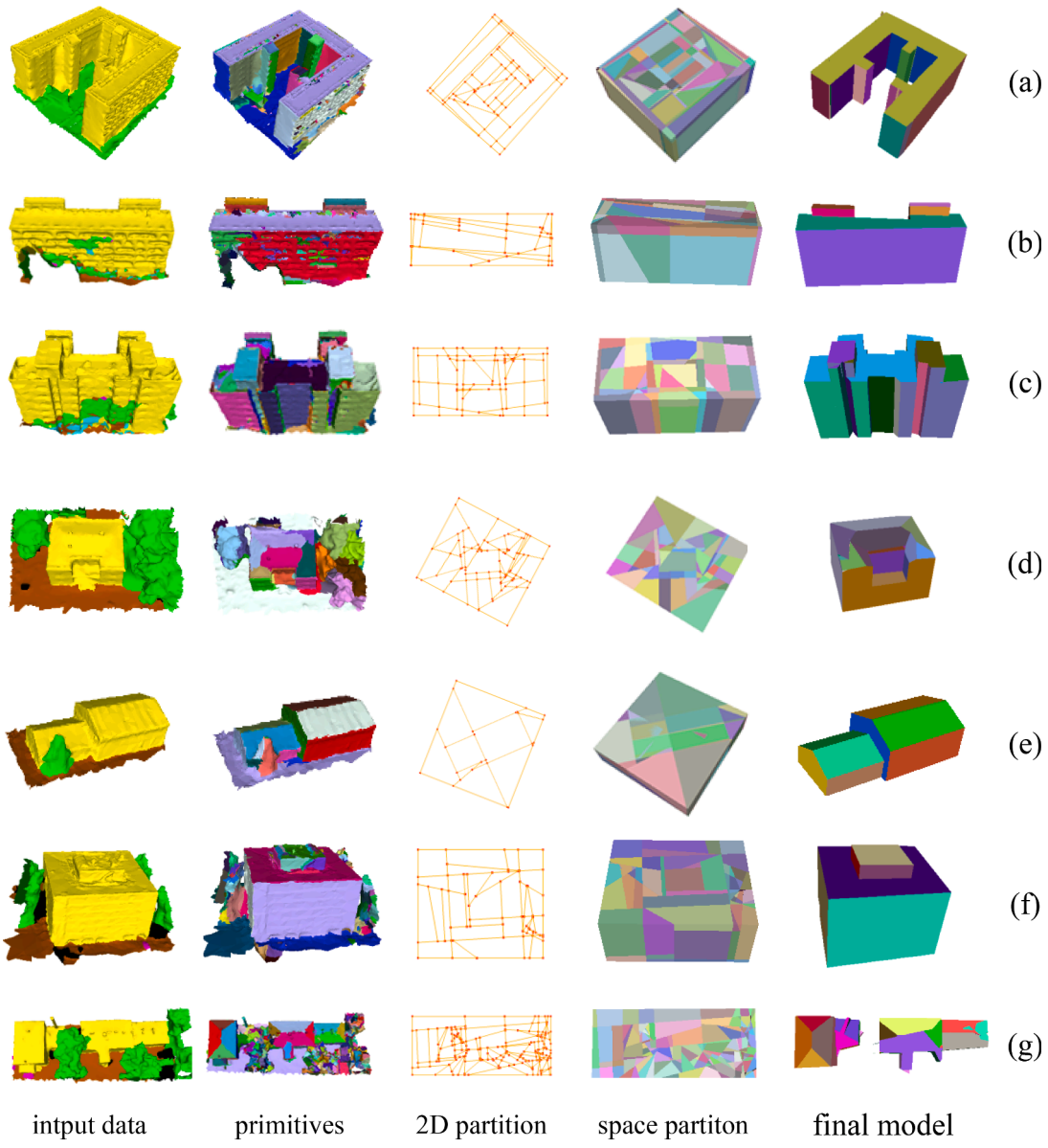



**Fig. 8.** Reconstruction of a single building from mesh data with semantic information. The original data is disturbed by the surrounding environment.

the public dataset SUM and are mainly affected by the surrounding non-buildings. However, some of the data were significantly obscured, as shown in Fig. 8(d), such as simple roof structures (Fig. 8(d)) and gable roofs (Fig. 8(e)). The experimental results showed that the SGR can reconstruct single buildings independently of vegetation occlusion. In particular, although the entire wall of the original data in Fig. 8(d) was covered by vegetation, SGR accurately identified and completely restored the missing wall. In Fig. 8(g), our method accurately identified the sheltered façade, restored its structure, and reconstructed two buildings connected by trees.

Table 2 presents a quantitative evaluation of the SGR method. From the number of input and output triangular meshes, SGR can achieve a lightweight model. The number of polyhedra generated by space partitioning partially depends on the structural complexity of the original data, as shown in Fig. 8(a) and Fig. 8(g). In terms of running time, the proposed SGR basically processed single buildings in approximately one minute, but required longer run times for complex structures. We note that the space partition process is a time-consuming step in the entire process. The overall runtime includes the entire process from the input to the output.

**Table 2**
Statistics on the example are shown in Fig. 8.

| Mesh (Fig. 8) | (a) | (b) | (c) | (d) | (e) | (f) | (g) |
|---|---|---|---|---|---|---|---|
| #input trimesh | 108 k | 51 k | 48 k | 59 k | 11 k | 45 k | 94 k |
| #planar shapes | 315 | 135 | 283 | 132 | 22 | 70 | 294 |
| #number of polyhedra | 114 | 103 | 126 | 227 | 67 | 130 | 781 |
| #output facets | 204 | 159 | 395 | 64 | 70 | 144 | 488 |
| Graph cut(sec) | 0.03 | 0.02 | 0.03 | 2.04 | 0.01 | 0.04 | 0.01 |
| 2.5D partitioning (sec) | 4.42 | 3.85 | 4.8 | 12.78 | 4.68 | 10.32 | 68.42 |
| face selection (sec) | 0.03 | 0.02 | 0.07 | 1.37 | 0.02 | 0.01 | 0.14 |
| Total time (sec) | 34.75 | 21.06 | 27.72 | 67.67 | 10.41 | 32.46 | 578.62 |

**Comparisons.** The main purpose of this group of experiments was to compare the effects of various methods on the reconstruction of single buildings under occlusion and to observe the effects of SGR's introduction of semantic information into the reconstruction process.

Buildings were sheltered by trees to varying degrees, as shown in Fig. 9 (a) and (b). As shown in Fig. 9(b), in the case of severe occlusion, the structure-aware method and PolyFit method were unable to produce results. KSR aims to reconstruct the geometry and structure so that the surrounding trees are reconstructed together and the goal of separating the buildings from the surroundings is not achieved. On the other hand, SGR accurately identified the interface between buildings and trees in a heavily obscured environment and rebuilt single buildings without interference from trees. However, because SGR depends on the accuracy of the initial segmentation, the upper block cannot be reproduced. In Fig. 9(c), two buildings were connected by trees, which is a relatively complicated scene, but SGR still located the façade between the building and trees and restored the occluded part, generating a visually accurate and attractive building model. Overall, the SGR method generated independent and complete building models from mixed data for buildings and the surrounding environment.

Table 3 lists the statistics comparing the efficiency of these methods and the accuracy of the model for the quantitative results. Efficiency was expressed by the overall run time of the algorithm. Model accuracy was measured by calculating the root mean square error (RMSE) distance of the model to the original mesh (excluding non-buildings):

$$RMSE = \sqrt{\frac{\sum_{\forall v \in m_{ori}^{B}} \min_{P_{\Delta tri} \in M} \left\| v - P_{\Delta tri} \right\|^{2}}{\left| \{ v | \forall v \in m_{ori}^{B} \} \right|}} \tag{7}$$

where $v$ is a vertex labeled building in the original mesh and $P_{\Delta tri}$ is the planar primitive of the final model $M$ composed of a triangular mesh. Here, the distance from the vertex to the model is considered the minimum Euclidean distance to $P_{\Delta tri}$, while $\left| \{ v | \forall v \in m_{ori^B} \} \right|$ is the amount of building vertices in the original mesh.

In Table 3, while the KSR method offers a significant advantage in terms of runtime, the accuracy of the model is relatively low when the buildings are seriously obscured. The PolyFit method requires a long runtime and produces poor quality results. The structure-aware method required less runtime, but was only effective when the data quality was very good. Both the structure-aware and polyFit methods failed to generate results for the data in Fig. 9(c). While our SGR method had a longer runtime, it was better than PolyFit. The accuracy of the model was relatively stable, regardless of whether the data were less or severely occluded.

**Effect of semantic information.** The experiment was divided into two groups (semantic and non-semantic information) with the input as the original complete data and non-building data, respectively.

**Table 3**
Quantitative evaluation of the different reconstruction methods.

| Method | Fig. 9(a) | | Fig. 9(b) | | Fig. 9(c) | |
|---|---|---|---|---|---|---|
| | time (s) | RMSE (cm) | time(s) | RMSE (cm) | time (s) | RMSE (cm) |
| Structure-aware | 23.2 | 12.33 | 14.3 | 9.15 | – | – |
| PolyFit | 49.2 | 11.50 | 2604 | 5.08 | – | – |
| KSR | 3.77 | 8.04 | 27.98 | 10.88 | 237.99 | 19.76 |
| SGR | **45.69** | **10.66** | **195.08** | **7.17** | 578.62 | **5.84** |

In Fig. 10, the visualization results for the building attached to the surrounding environment are almost the same as those with non-building parts removed (as shown in Fig. 10, arrow 1). This indicates that semantic information played a partial role in filtering the surrounding environment. Arrow 2 shows the influence of the original data, with or without semantic information, on the reconstruction results. There were redundant fragments in the reconstruction results without semantic information. In the grouping experiment without semantic information, if the non-building part was removed, there were incorrect parts in the experimental results, as indicated by arrow 3. These results show that an ideal result cannot be achieved by simply removing the non-building parts for reconstruction through the pretreatment step. In the grouping experiment without semantic information, there were obvious incorrect details in the visualization results of both the original data and the data with the non-building parts removed (arrow 4). Overall, the results of the grouping experiments without semantic information are obviously inferior to those with semantic information. In general, the semantic information not only filtered the reconstruction of buildings, but also improved the structural correctness.

To further illustrate the role of semantic information in guiding the reconstruction of buildings, the experimental setup used data containing non-building parts of different degrees as input, as shown in Fig. 11(b) and Fig. 11(c). Fig. 12 intuitively presents the difference between Semantic information participation and non participation in the modeling process.

Owing to the absence of the façade, the structure was automatically inferred during the reconstruction process to generate the results shown in Fig. 11(a). If the semantic information is correct, the non-building

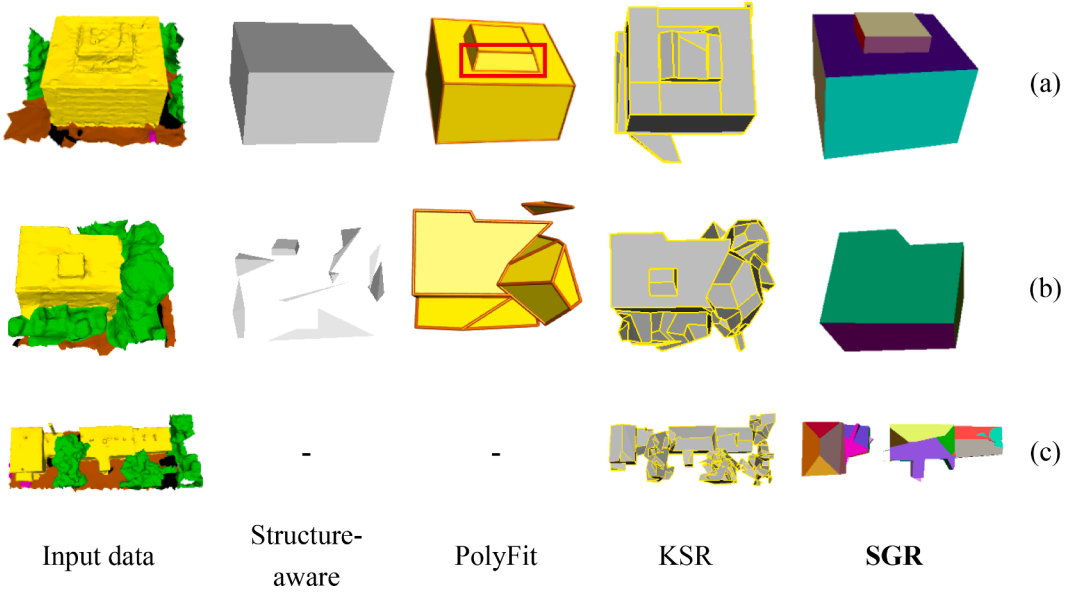



**Fig. 9.** Comparison of the reconstruction methods.

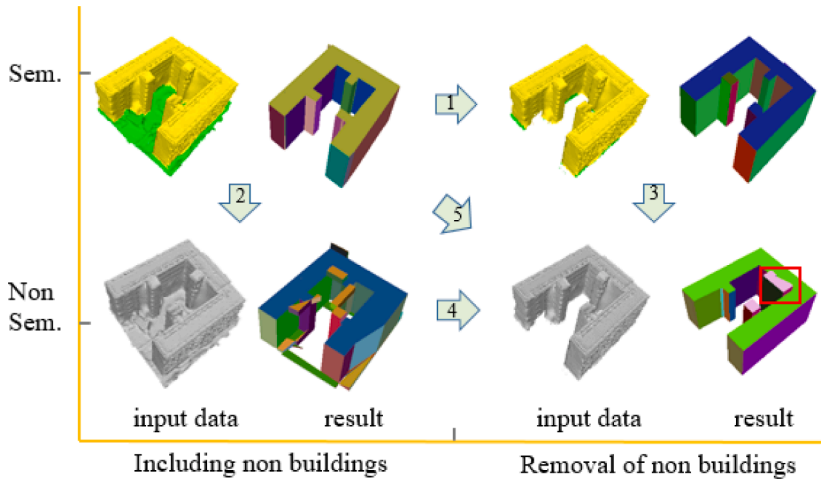



**Fig. 10.** Impact of semantic information on the model.

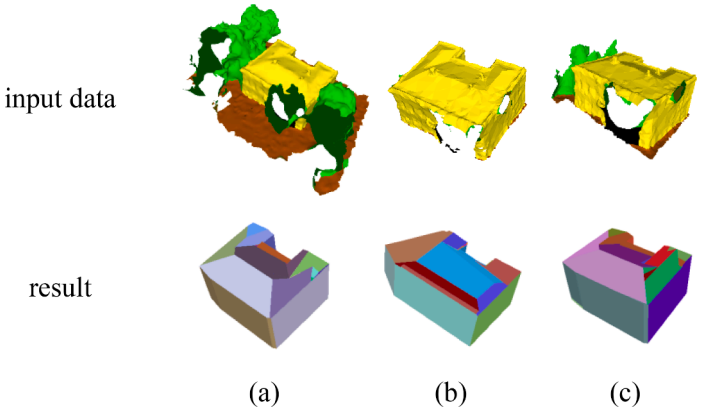



**Fig. 11.** Influence of non-building interference on reconstruction.

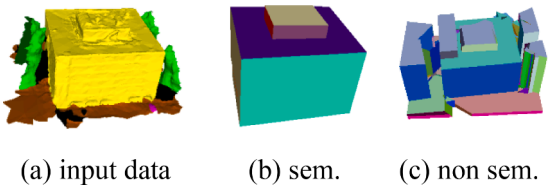



**Fig. 12.** The Comparative experiment of the proposed method with and without semantics. (a) input data; (b) use semantic information for reconstruction; (c) the reconstruction process does not use semantic information.

part is directly removed (as shown in the input data in Fig. 11(b)); a severely occluded façade is not generated, resulting in structural errors on the left side of the building. The input data in Fig. 11(c) does not completely remove the non-building parts, as shown in Fig. 11(b); some non-building parts are retained on the severely occluded façade. From the experimental results, although there was non-building interference, the façade on the left side of the building was correctly generated. In conclusion, when buildings are seriously disturbed by the surrounding environment, semantic information plays a more significant role in guiding the reconstruction process and is of great significance in removing non-building parts and completing missing parts in building monomer reconstruction.

**Effect of primitive segmentation.** We explored the influence of the initial primitive segmentation on the reconstruction process and results. Fig. 13 shows the visualization results.

In the initial primitive segmentation stage, the parameters were changed to generate different segmentation results. According to the statistical results, the number of roofs detected was also different, which was the key factor causing the different results in the final model. From the results, the more primitives that were detected, the more facets that were included in the final generated model with increased runtime. Furthermore, in practical applications, the number of initial primitives can be appropriately controlled according to the need for detailed retention, which can improve the efficiency.

**Impact of parameters.** Our SGR method has two main parameters: 1) the threshold of the roof area and 2) the value of $\lambda$ in the energy function.

The roof played an important role in the detailed retention and run time of the resulting model. To capture the main roof structure and effectively obtain the framework of buildings, we set the area threshold of the roofs.

Four thresholds were set, as shown in Fig. 14. As the threshold decreased, more roof details were retained. In addition, the runtime increased as the threshold increased. The area threshold of the roof should be within a certain range to ensure certain structural details, as well as efficiency.

We also explored the impact of $\lambda$ on the reconstruction, including the number of output facets and geometric errors. Gradually increasing the value of $\lambda$ and the number of output patches caused a significant decreasing trend (blue curve in Fig. 15). The value of $\lambda$ was 0.4–0.8 to balance the simplicity and accuracy of the model. The geometric error was calculated as the RMSE of the Euclidean distance from the vertex of the original mesh to the final primitive model (as described in Eq. (7)).

**Potential for detail retention.** To confirm the potential of our SGR

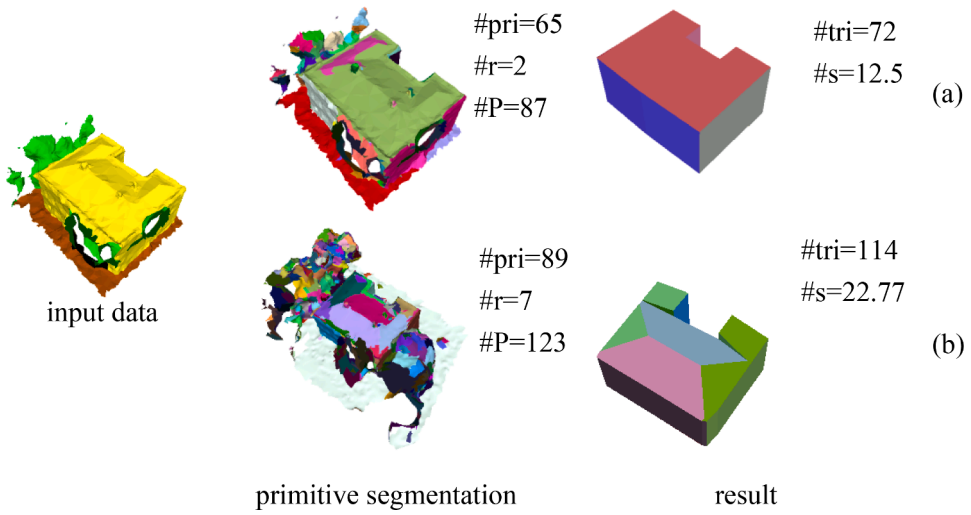



**Fig. 13.** Effect of the initial primitive on the reconstructed model.

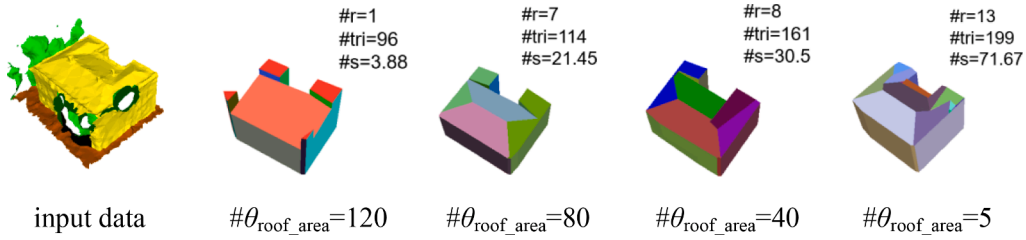



**Fig. 14.** Influence of the area threshold of the roof. If the threshold value is too high, the building structure will be simplified; otherwise, more details will be retained.

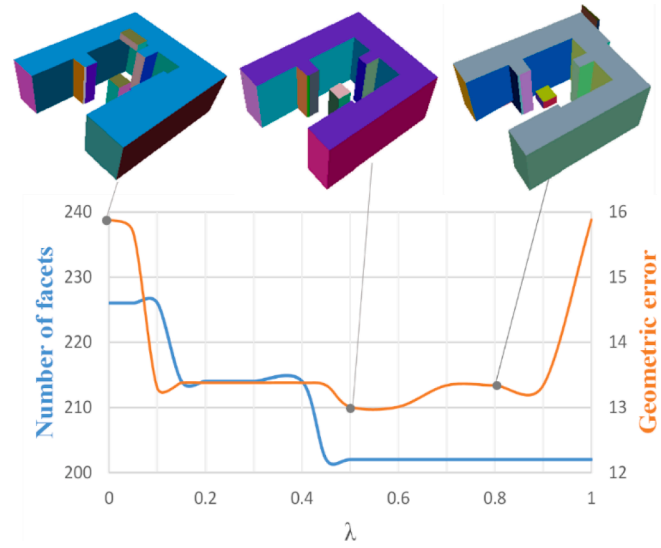



**Fig. 15.** Influence of parameter $\lambda$. As the value of $\lambda$ increases, the output model becomes simpler and the number of output faces (blue curves) decreases. The values that are too small or too large will make the geometric error larger. The unit of geometric error is cm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

method in retaining details, we selected data for the comparison experiment. In Fig. 16, the red boxes in the original data indicate the building details. The structure-aware method aimed to extract the main structure to the maximum extent. The visualization results of our method generated details that were consistent with the original data. We

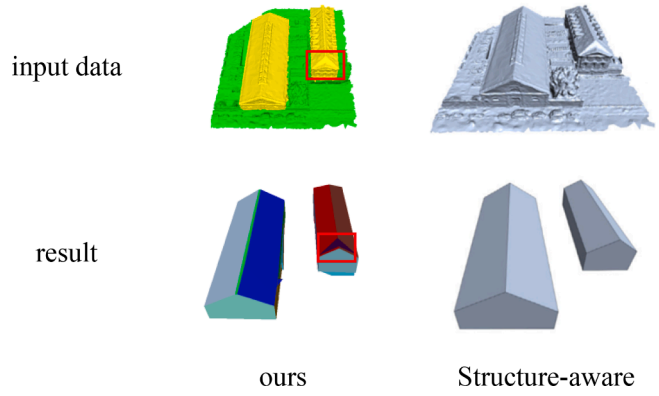



**Fig. 16.** Potential to generate detail.

suggest that more details will be retained based on their robustness and accuracy.

**Limitations.** As the original data inevitably contain noise and occlusions or are missing, primitive objects in the real object may not be recovered in the original input data, making it undetectable. Structural deficiencies may also exist in the final model. Moreover, the noise and uneven density of the original mesh may affect the accurate calculation of the vertex normals, particularly on the occluded plane, which is important for the extraction of the final surface. If the error is significant, it is necessary to manually correct the normal direction.

## 5. Conclusions

We proposed a semantic-guided building reconstruction (SGR)

method for building objects using fuzzy boundary semantics. SGR makes full use of the semantic information of the original building data and statistical information of the semantic information on the boundary. In addition, SGR adaptively expands the façade in 2D space to explore possible intersection relations and creates a space partition for reconstruction. It can capture and recover occluded or missing façade information by combining the semantic information and the façade projection on the boundary. The SGR can also ensure the efficiency and scalability of its algorithm by adaptively generating space partitions. Our experiments showed that SGR could reconstruct single buildings independently and cost-effectively from various objects and scenes and multiple data sources, with different degrees of semantic interference to buildings and varying complexities of building structures. In our experiments, when multiple buildings were occluded by vegetation, our SGR method reduced the interference of the surrounding environment and rebuilt individual buildings separately.

In the future work, we will use the proposed method to make a city-level building model dataset, and further explore the the use of deep learning in the automatic reconstruction task to better compensate for imperfect local structures and thereby improve the accuracy of the model.

*CRediT authorship contribution statement*

**Senyuan Wang:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Xinyi Liu:** Methodology, Writing – review & editing. **Yongjun Zhang:** Writing – review & editing. **Jonathan Li:** Writing – review & editing. **Siyuan Zou:** Writing – review & editing. **Jipeng Wu:** Writing – review & editing. **Chuang Tao:** Funding acquisition, Resources, Supervision. **Quan Liu:** Funding acquisition, Resources, Supervision. **Guorong Cai:** Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 42192581, 42201474, 41971424), the Science and Technology Major Project of Hubei Province (Grant No. 2021AAA010), Zhizhuo Research Fund on Spatial-Temporal Artificial Intelligence (Grant No. ZZJJ202206), and GuangDong Basic and Applied Basic Research Foundation(Grant No. 2022A1515110128).

## References

Achanta, R., Susstrunk, S., 2017. Superpixels and polygons using simple non-iterative clustering. In: Proc. CVPR. 4651–4660.

Allam, Z., Sharifi, A., Bibri, S.E., Jones, D.S., Krogstie, J., 2022. The metaverse as a virtual form of smart cities: opportunities and challenges for environmental, economic, and social sustainability in urban futures. Smart Cities. 5 (3), 771–801. https://doi.org/10.3390/smartcities5030040.

Awwad, T.M., Zhu, Q., Du, Z., Zhang, Y., 2010. An improved segmentation approach for planar surfaces from unstructured 3D point clouds. The Photogram. Rec. 25 (129), 5–23. https://doi.org/10.1111/j.1477-9730.2009.00564.x.

Badach, J., Voordeckers, D., Nyka, L., Van Acker, M., 2020. A framework for air quality management zones-useful GIS-based tool for urban planning: Case studies in Antwerp and Gdańsk. Build. Envir. 174, 106743. https://doi.org/10.1016/j.buildenv.2020.106743.

Bas, S., Ok, A.O., 2021. A new productive framework for point-based matching of oblique aircraft and UAV-based images. The Photogram. Rec. 36 (175), 252–284. https://doi.org/10.1111/phor.12374.

Bauchet, J.-P., Lafarge, F., 2018. Kippi: Kinetic polygonal partitioning of images. In: Proc. CVPR. 3146–3154. https://doi.org/10.1109/CVPR.2018.00332.

Bauchet, J.-P., Lafarge, F., 2020. Kinetic shape reconstruction. ACM Trans. Graph. 39 (5), 1–14. https://doi.org/10.1145/3376918.

Biljecki, F., Stoter, J., Ledoux, H., Zlatanova, S., Çöltekin, A., 2015. Applications of 3D city models: State of the art review. ISPRS Int. J. Geo-Inf. 4 (4), 2842–2889. https://doi.org/10.3390/ijgi4042842.

Blaha, M., Vogel, C., Richard, A., Wegner, J.D., Pock, T., Schindler, K., 2016. Large-scale semantic 3d reconstruction: an adaptive multi-resolution model for multi-class volumetric labeling. In: Proc. CVPR. 3176–3184. https://doi.org/10.1109/CVPR.2016.346.

Bouzas, V., Ledoux, H., Nan, L., 2020. Structure-aware building mesh polygonization. ISPRS J. Photogramm. Remote Sens. 167, 432–442. https://doi.org/10.1016/j.isprsjprs.2020.07.010.

Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Trans. Patt. Anal. Mach. Intell. 26 (9), 1124–1137.

Cabezas, R., Straub, J., Fisher, J. W., 2015. Semantically-aware aerial reconstruction from multi-modal data. In: Proc. ICCV. pp. 2156–2164. https://doi.org/10.1109/ICCV.2015.249.

Chauve, A.-L., Labatut, P., Pons, J.-P., 2010. Robust piecewise-planar 3d reconstruction and completion from large-scale unstructured point data. In: Proc. CVPR. 1261–1268. https://doi.org/10.1109/CVPR.2010.5539824.

Chen, Z., Li, D., Fan, W., Guan, H., Wang, C., Li, J., 2021. Self-attention in reconstruction bias U-Net for semantic segmentation of building rooftops in optical remote sensing images. Remote Sens. 13 (13), 2524. https://doi.org/10.3390/rs13132524.

Coudron, I., Puttemans, S., Goedemé, T., Vandewalle, P., 2020. Semantic Extraction of Permanent Structures for the Reconstruction of Building Interiors from Point Clouds. Sensors. 20 (23), 6916. https://doi.org/10.3390/s20236916.

Coughlan, J., Yuille, A.L., 2000. The Manhattan world assumption: Regularities in scene statistics which enable Bayesian inference. In: Proc. NeurIPS. 809–815.

Curless, B., Levoy, M., 1996. A volumetric method for building complex models from range images. In: Proc. SIGGRAPH. pp. 303–312.

Duan, L., Lafarge, F., 2015. Image partitioning into convex polygons. In: Proc. CVPR. 3119–3127.

Fang, H., Lafarge, F., Desbrun, M., 2018. Planar shape detection at structural scales. In: Proc. CVPR. 2965–2973. https://doi.org/10.1109/CVPR.2018.00313.

Gao, W., Nan, L., Boom, B., Ledoux, H., 2021. SUM: A benchmark dataset of Semantic Urban Meshes. ISPRS J. Photogramm. Remote Sens. 179, 108–120. https://doi.org/10.1016/j.isprsjprs.2021.07.008.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K., Pollefeys, M., 2017. Semantic3d. net: A new large-scale point cloud classification benchmark. arXiv preprint arXiv:1704.03847.

Han, J., Zhu, L., Gao, X., Hu, Z., Zhou, L., Liu, H., Shen, S., 2021. Urban scene lod vectorized modeling from photogrammetry meshes. IEEE Trans, Image Proces. 30, 7458–7471. https://doi.org/10.1109/TIP.2021.3106811.

Holzmann, T., Maurer, M., Fraundorfer, F., Bischof, H., 2018. Semantically aware urban 3d reconstruction with plane-based regularization. In: Proc. ECCV. pp. 468–483. https://doi.org/10.1007/978-3-030-01264-9_29.

Ladickỳ, L., Sturgess, P., Russell, C., Sengupta, S., Bastanlar, Y., Clocksin, W., Torr, P.H., 2012. Joint optimization for object class segmentation and dense stereo reconstruction. Int. J. Comp. Vis. 100 (2), 122–133. https://doi.org/10.1007/s11263-011-0489-0.

Li, M., Rottensteiner, F., Heipke, C., 2019. Modelling of buildings from aerial LiDAR point clouds using TINs and label maps. ISPRS J. Photogramm. Remote Sens. 154, 127–138. https://doi.org/10.1016/j.isprsjprs.2019.06.003.

Li, Y., Wu, X., Chrysathou, Y., Sharf, A., Cohen-Or, D., Mitra, N.J., 2011. Globfit: Consistently fitting primitives by discovering global relations. ACM Trans. Grap. 30 (4), 1–12. https://doi.org/10.1145/2010324.1964947.

Liu, X., Zhang, Y., Ling, X., Wan, Y., Liu, L., Li, Q., 2019. TopoLAP: topology recovery for building reconstruction by deducing the relationships between linear and planar primitives. Remote Sens. 11 (11), 1372. https://doi.org/10.3390/rs11111372.

Matrone, F., Lingua, A., Pierdicca, R., Malinverni, E., Paolanti, M., Grilli, E., Remondino, F., Murtiyoso, A., Landes, T., 2020. A benchmark for large-scale heritage point cloud semantic segmentation. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLIII-B2-2020, 1419–1426. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-1419-2020.

Monszpart, A., Mellado, N., Brostow, G.J., Mitra, N.J., 2015. RAPter: rebuilding man-made scenes with regular arrangements of planes. ACM Trans. Graph. 34 (4), 103–1101. https://doi.org/10.1145/2766995.

Mura, C., Mattausch, O., Pajarola, R., 2016. Piecewise-planar reconstruction of multi-room interiors with arbitrary wall arrangements. Comp. Graph. Forum. 35 (7), 179–188. https://doi.org/10.1111/cgf.13015.

Nan, L., Wonka, P., 2017. Polyfit: Polygonal surface reconstruction from point clouds. In: Proc. CVPR. pp. 2372–2380. https://doi.org/10.1109/ICCV.2017.258.

Nurunnabi, A., Belton, D., West, G., 2012. Robust segmentation in laser scanning 3d point cloud data. In: Proc. DICTA. https://doi.org/10.1109/DICTA.2012.6411672.

Oesau, S., Lafarge, F., Alliez, P., 2016. Planar shape detection and regularization in tandem. Compu. Graph. Forum. 35 (1), 203–215. https://doi.org/10.1111/cgf.12720.

Rabbani, T., Van Den Heuvel, F., Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. 36 (5), 248–253.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: Proc. MICCAI. pp. 234–241.
Savinov, N., Ladicky, L., Hane, C., Pollefeys, M., 2015. Discrete optimization of ray potentials for semantic 3d reconstruction. In: Proc. CVPR. pp. 5511–5518.
Schindler, F., Wörstner, W., Frahm, J.-M., 2011. Classification and reconstruction of surfaces from point clouds of man-made objects. In: Proc. CVPRW. pp. 257–263.
Schnabel, R., Wahl, R., Klein, R., 2007. Efficient ransac for point-cloud shape detection. Compu. Graph. Forum. 26 (2), 214–226. https://doi.org/10.1111/j.1467-8659.2007.01016.x.
Song, J., Xia, S., Wang, J., Chen, D., 2020. Curved buildings reconstruction from airborne LiDAR data by matching and deforming geometric primitives. IEEE Trans. Geosci. Remote Sens. 59 (2), 1660–1674. https://doi.org/10.1109/TGRS.2020.2995732.
B. SYSTEMS, 2016. ContextCapture software. https://www.bentley.com/zh/products/product-line/reality-modeling-software/contextcapture.
Verdie, Y., Lafarge, F., Alliez, P., 2015. LOD generation for urban scenes. ACM Trans. Graph. 34 (3), 1–14. https://doi.org/10.1145/2732527.
Von Gioi, R.G., Jakubowicz, J., Morel, J.-M., Randall, G., 2008. LSD: A fast line segment detector with a false detection control. IEEE Trans. Patt. Anal. Mach. Intell. 32 (4), 722–732. https://doi.org/10.1109/TPAMI.2008.300.
Vu, H.H., Labatut, P., Pons, J.P., Keriven, R., 2011. High accuracy and visibility-consistent dense multiview stereo. IEEE Trans. Patt. Anal. Mach. Intell. 34 (5), 889–901. https://doi.org/10.1109/TPAMI.2011.172.
Wang, S., Cai, G., Cheng, M., Junior, J.M., Huang, S., Wang, Z., Su, S., Li, J., 2020. Robust 3D reconstruction of building surfaces from point clouds based on structural and closed constraints. ISPRS J. Photogramm. Remote Sens. 170, 29–44. https://doi.org/10.1016/j.isprsjprs.2020.09.004.
Werbrouck, J., Pauwels, P., Bonduel, M., Beetz, J., Bekers, W., 2020. Scan-to-graph: Semantic enrichment of existing building geometry. Autom. Constr. 119, 103286. https://doi.org/10.1016/j.autcon.2020.103286.
Xiao, X., 2019. Oblique photogrammetry based scene 3D reconstruction with structure sensing functions. Acta Geod. Cartogr. Sin. 48 (6), 802.
Xu, Y., Zhang, J., Zhao, H., Chen, C., Mao, W., 2021. Research on quality framework of real scene 3D model based on oblique photogrammetry. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. https://doi.org/10.5194/ISPRS-ARCHIVES-XLIII-B3-2021-791-2021.
Yang, B., Lee, J., 2019. Improving accuracy of automated 3-D building models for smart cities. Int. J. Digit. Earth. 12 (2), 209–227. https://doi.org/10.1080/17538947.2017.1395089.
Zhang, C., Fan, H., 2022. An improved multi-task pointwise network for segmentation of building roofs in airborne laser scanning point clouds. The Photogram. Rec. 37 (179), 260–284. https://doi.org/10.1111/phor.12420.
Zhang, X., Zhao, P., Hu, Q., Ai, M., Hu, D., Li, J., 2020. A UAV-based panoramic oblique photogrammetry (POP) approach using spherical projection. ISPRS J. Photogramm. Remote Sens. 159, 198–219. https://doi.org/10.1016/j.isprsjprs.2019.11.016.
Zhang, Y., Zou, S., Liu, X., Huang, X., Wan, Y., Yao, Y., 2022. LiDAR-guided stereo matching with a spatial consistency constraint. ISPRS J. Photogramm. Remote Sens. 183, 164–177. https://doi.org/10.1016/j.isprsjprs.2021.11.003.
Zheng, T., Zhang, G., Han, L., Xu, L., Fang, L., 2020. Building fusion: semantic-aware structural building-scale 3d reconstruction. IEEE Trans. Patt. Anal. Mach. Intell. pp. 2328–2345. https://doi.org/10.1109/TPAMI.2020.3042881.
Zhou, Q.-Y., Neumann, U., 2010. 2.5d dual contouring: A robust approach to creating building models from aerial lidar point clouds. In: Proc. ECCV. pp. 115–128.
Zhu, X., Liu, X., Zhang, Y., Wan, Y., Duan, Y., 2021. Robust 3-D plane segmentation from airborne point clouds based on quasi-a-contrario theory. IEEE J-STARS. 14, 7133–7147. https://doi.org/10.1109/JSTARS.2021.3093576.
Zou, S., Liu, X., Huang, X., Zhang, Y., Wang, S., Wu, S., Zheng, Z., Liu, B., 2023. Edge-preserving Stereo Matching using LiDAR Points and Image Line Features. IEEE Geosci. Remote. Sens. Lett. 20, 1–5. https://doi.org/10.1109/LGRS.2023.3239030.

## Further reading

Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. IEEE Trans. Patt. Anal. Mach. Intell. 23 (11), 1222–1239. https://doi.org/10.1109/34.969114.