Semantics-guided spatial data generation for complex large-scale indoor map from 3D colorized point clouds

Zhengwen Wang[1], Xuzhe Wang[2], Juntao Yang[1,*], Te Li[3]
[1]College of Geodesy and Geomatics, Shandong University of Science and Technology, Qingdao 266590, China;
[2]School of Geography and Information Engineering, China University of Geosciences, Wuhan 430074, China;
[3]Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-temporal Big Data Technology, China Railway Design Corporation, Tianjin 300251, China;

Keywords: Point clouds, Floorplan generation, Semantic segmentation, Space subdivision, Indoor scene.

Abstract

With the popularization of the concept of smart cities and the development of indoor positioning and navigation services, as well as the increasing complexity of building structures with the continuous advancement of urbanization, automatic mapping of large-scale indoor spatial data has become the fundamental work for subsequent real-life 3D applications. Therefore, this paper develops a semantics-guided generation method of indoor spatial data for mapping indoor spaces, where the roles of semantics are investigated for the subdivision and reconstruction of indoor spaces. It consists of the following four parts: (1) Semantic segmentation of 3D indoor scene; (2) Storey segmentation using semantics-enhanced height histogram; (3) Semantics-guided room segmentation based on building physical structures; (4) Room-wise boundary optimization using semantics-aware Recursive Search. Both quantitative and qualitative experiments are conducted on two public benchmark datasets: Stanford Large-Scale 3D Indoor Spaces (S3DIS) dataset and Matterport3D dataset. The results demonstrated that our method is capable of reconstructing complicated large-scale indoor scenes with higher robustness and reliability, outperforming existing state-of-the-art algorithms.

1. Introduction

Indoor location services can effectively promote the construction of smart cities (Moreno et al., 2014), enhance people's travel experience (Farahsari et al., 2022), improve the interactive effects of emerging applications such as Augmented Reality (AR)/ Virtual Reality (VR) (Baek et al., 2019), which would be an important research area for future mobile applications. In recent years, indoor map (Otero et al., 2020), as an indispensable supporting data for indoor location services, has become increasingly popular in the field of navigation and location services with the development of digital map technology and mobile Internet technology (Fathalizadeh et al., 2023). Location-based indoor maps are information carriers and tools that abstract and summarize the spatial distribution and semantic characteristics of indoor spaces/objects through graphic symbols, panoramic images, charts, and other forms (Kang et al., 2020). Currently, indoor location services are far from universal primarily because of the lack of indoor maps, which would also be a fundamental issue within the international academic and industrial communities (Jeamwatthanachai et al., 2017; Gao et al., 2018).

The spatial data for indoor map describes the geometric attributes of indoor spaces and the topological information expressing spatial adjacency and connectivity relationships among indoor spaces (Kang and Li, 2017; Zhou et al., 2020), which is the geographic framework and carrier platform of indoor maps. With the rapid development of cities, there are much larger and more complex buildings, where the indoor structure is becoming increasingly complex (Park et al., 2020). It poses great challenges to the acquisition of indoor spatial data. In the past, the manual or interactive methods were mainly used to generate high-quality indoor spatial data. However, it is laborious and time-consuming, which cannot meet the application requirements of mapping the large-scale indoor scenes. Therefore, the automatic acquisition of indoor spatial data is a key issue that urgently needs to be addressed (Yang et al., 2021).

The floorplans, Building Information modelling (BIM) and City Geography Markup Language (CityGML) are usually used to express building objects in the construction fields. Therefore, lots of researches have been devoted to use the floorplans, BIM or CityGML as a valuable information source for generating indoor spatial data. A BIM is a digital representation of the advanced geometric and semantic information of building elements through the entire lifecycle of a building. However, it is lack of the spatial topological information for the indoor navigation task. To better understand the functional relationship, Isikdag et al., (2013) made full use of the semantic and geometric information defined in a BIM Industry Foundation Classes (IFC) to generate a new BIM Oriented Indoor Data model with ISO 19107 compliant representations for supporting indoor navigation. Tashakkori et al., (2015) presented a spatial model using 3D architectural and semantic information from BIM (IFC) for indoor emergency decisions. Also for indoor emergency situations, Boguslawski et al., (2016) produced a full 3D topological network of a building environment in order to find a safy route. For indoor spaces, it is common for the rooms with irregular shapes, which might result in the generation of detours in the topological network. Fu et al., (2020) optimized the straight medial axis transformation method to produce the indoor navigation network derived from the IFC model. As a matter of fact, the emergency-related response ia a process from indoor to outdoor, not only indoor. To connect the indoor and outdoor network, Teo and Cho (2016) presented the entrance-to-street strategy based on the constraint of perpendicular and shorter path conditions to outdoor vertices. With regards to different forms of locomotion (e.g., walking, flying, etc.), the definition of navigable free space is different. Diakité and Zlatanova (2018) defined a Flexible Space Subdivision framework where the types and functions (such as movable capability) of indoor objects could be classified to partition a set of subspaces with dedicated properties. However, the popularity of BIM is not very high, which may restrict the widespread application of the BIM-based methods.

* Corresponding author. E-mail: jtyang@sdust.edu.cn

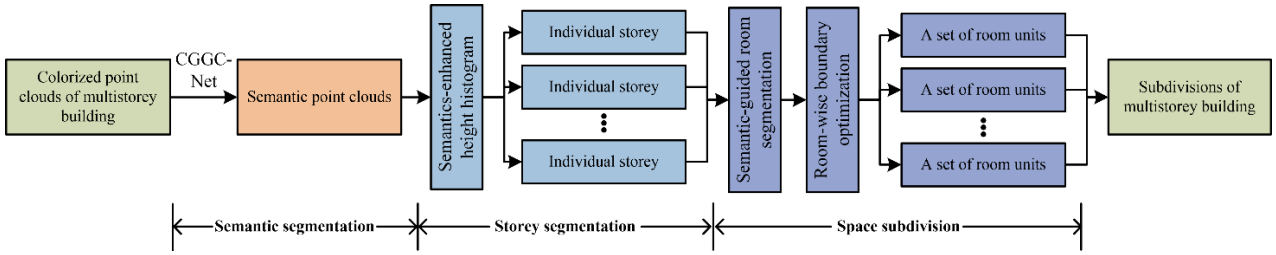


Figure 1. the overview of the developed method

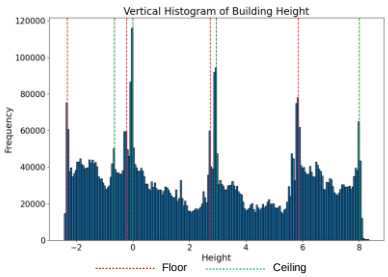
Although Computer-Aided Design (CAD) is used for the design task in the field of architecture both domestically and internationally and a large number of floorplans have been accumulated. Thus, some studies have also devoted to generate the reasonable indoor navigation network from floorplans. Shang et al., (2015) integrated semantics with geometric information to create a fine-grained navigation network from 2D floor plans, where the physical structures are used to define the indoor subspace and semantics to further partition the subspaces. Wu et al., (2021) interpreted the floorplan using the Mask R-CNN to simplify the building elements into rectangles with different shapes and then repaired their inconsistent topological relationships. It might be insufficient for the route-planning task if the semantics and geometric information of indoor spaces are only introduced without taking the corridor structures into consideration. Pang et al., (2020) mainly focused on the visibility of corridor spaces to reconstruct the relationships between different types of corridors, where the key locations of corridor structures are found. Nevertheless, the up-to-date of the floorplans is poor and the description of spatial topology is not comprehensive, making it challenging to meet users' concerns for accurate navigation and personalized location services within the indoor spaces.

In recent years, the development of Light Detection and Ranging (LiDAR) and photogrammetry technology has made real-time and efficient collection of indoor 3D spatial information possible (Meyer et al., 2023). Both can obtain massive (tens of millions) of discrete 3D point clouds, which can achieve the true expression of indoor 3D scenes and become important data sources for generating the spatial data for indoor map (Luo et al., 2023). Nevertheless, the massive amount of discrete 3D point clouds is unstructured, and the presence of noises makes it difficult to accurately partition indoor spatial structures. Although the size and shape of indoor navigation spaces may vary, they are all composed of building components such as floors, ceilings, walls, stairs, and openings. This prior knowledge is also the basis for users to obtain reliable path guidance between these independent navigation spaces (Yang et al., 2021).

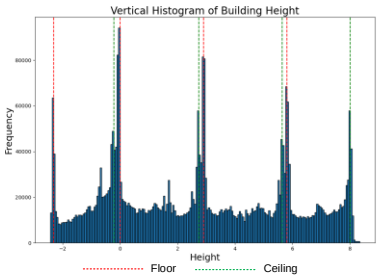
Although there have been lots of methods developed to produce the spatial data for indoor map from 3D point clouds, it is still greatly challenging due to the following aspects. Firstly, the existing methods primarily concentrate on mapping the spatial data of small-scale indoor scenes (Chen et al., 2019; Luo and Huang, 2022; Shabani et al., 2023). It is difficult to extend them into the entire storey of large-scale indoor scenes, which would restrict the usability of these existing methods. Furthermore, the complex situations, such as occlusions, the placement of furniture, different states of doors, are common, which also poses a challenge to the subdivision and reconstruction of indoor spaces.

To address these challenges, therefore, this paper develops a semantics-guided spatial data generation for complex large-scale indoor map from 3D colorized point clouds. The main contributions are as follows: (1) We developed a semantics-guided generation method of indoor spatial data for mapping indoor spaces, where the roles of semantics are investigated for the subdivision and reconstruction of indoor spaces. (2) By defining multiple energy terms within Markov Random Field (MRF) framework for solving both over segmentation and room boundary optimization problems, the refinement of subdivision and reconstruction can be cast as an energy minimization issue. Consequently, the spatial structures of indoor spaces could be recovered properly and effectively under semantic constraints. (3) Both quantitative and qualitative experiments are carried out on the public benchmark-Stanford Large-Scale 3D Indoor Spaces (S3DIS) dataset (Armeni et al., 2016) and Matterport3D dataset

(a)



(b)



(c)

Figure 2. Comparison between (b) conventional and (c) semantics-enhanced height histogram for the same (a) indoor point clouds.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-961-2025 | © Author(s) 2025. CC BY 4.0 License.

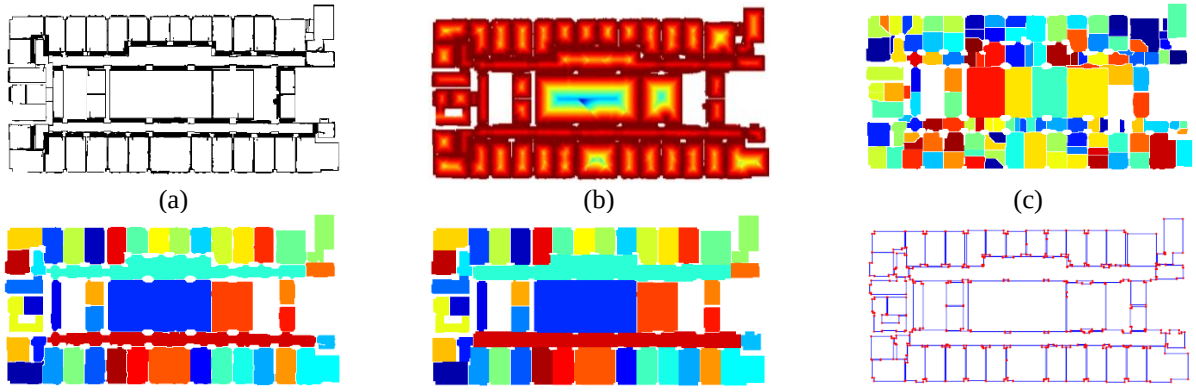


Figure 3. An example of floorplan generation and selection on Area 1 on S3DIS dataset. (a) is the 2D grid projection map of vertical elements point cloud. (b) is the distance transformation map. (c) is the result of watershed operation. (d) is the result of room segmentation optimization. (e) is the result of room-wise boundary optimization. (f) is the vector-graphic floorplan.

(Chang et al., 2017), which would confirm that the developed method outperforms the state-of-the art.

The rest of this paper is organized as follows. Section 2 describes the proposed method in detail. Following this, experimentation and analysis is conducted in Section 3. Finally, a brief summary with the outlook is given in Section 4.

2. Methodology

The overview of our proposed method can be shown in Figure 1. Figure 1 illustrates the flow chart of the proposed method, which consists of the following four parts: (1) Semantic segmentation of 3D indoor scene; (2) Storey segmentation using semantics-enhanced height histogram; (3) Semantics-guided room segmentation based on building physical structures; (4) Room-wise boundary optimization using semantics-aware Recursive Search. Key algorithms are given in detail below.

2.1 Semantic segmentation of 3D indoor scene

As described in Section 1, vertical elements of indoor physical structures mainly composed of walls, beams, doors, columns, separate the internal space of the room from the external space and divide adjacent rooms, which would offer the semantic guidance for reasoning the indoor spatial relationships. Therefore, we utilize our previous work CGGC-Net (Wang et al., 2023) to effectively segment the original point clouds into six categories, including floors, ceilings, doors, windows, vertical elements of building structures and others.

2.2 Storey segmentation using semantics-enhanced height histogram

Accurately distinguishing and separating floors is a key step in building modeling, structural analysis, and data understanding, which is crucial for modeling multi-story buildings. Traditional floor segmentation methods rely on vertical histograms, using height information to identify floors and ceilings (Pexman et al., 2021). While this approach is effective to some extent, it struggles when dealing with complex building layouts, particularly when the heights of adjacent floors and ceilings are similar, or when various furniture and equipment are placed in the interior. In such cases, the performance of vertical histograms for separation is often compromised, as shown in Figure 2(a). Existing methods typically fail to account for the structural and functional differences of the building, leading to less accurate segmentation results.

To address this issue, we propose an innovative floor segmentation method based on semantic-enhanced height histograms, as shown in Figure 2(b). Unlike traditional methods, our approach significantly improves floor separation accuracy by introducing semantic guidance. Specifically, we enhance the point cloud data with semantic information and remove points unrelated to floor separation (such as furniture, indoor equipment, hanging objects, etc.) from the point cloud. This ensures that the height histogram is generated based solely on the physical structural features of the building (e.g., floors, ceilings, columns, walls). This method not only improves the quality of segmentation but also prevents interference from furniture and other objects in complex environments.

By using this semantic-guided segmentation technique, we achieve more accurate floor separation, especially when dealing with structurally complex buildings and diverse interior layouts. Traditional geometry-based methods often struggle with these complexities, whereas our approach combines semantic information with geometric features, improving segmentation accuracy while ensuring that the resulting building model aligns more closely with the actual structure. This innovative method has a unique advantage over existing technologies and pushes forward the development of floor segmentation techniques, offering a smarter and more robust solution.

2.3 Semantic-guided room segmentation based on building physical structures

The purpose of this subsection is to divide indoor spaces into a set of individual units, which consists of initial room segmentation based on watershed algorithm and room segmentation optimization under MRF framework.

2.3.1 initial room segmentation based on watershed algorithm: The building layout is divided into a set of room units based on the extracted vertical elements in this subsection. To accomplish this task, we can transform 3D space partition into 2D image segmentation by projecting vertical elements into a 2D grid map to generate a binary image (as shown in Figure 3(a)). More specifically, the value of the corresponding pixel in the grid map is assigned a value of 1 if the grid contains point clouds, otherwise it is assigned a value of 0. The pixel size of this grid

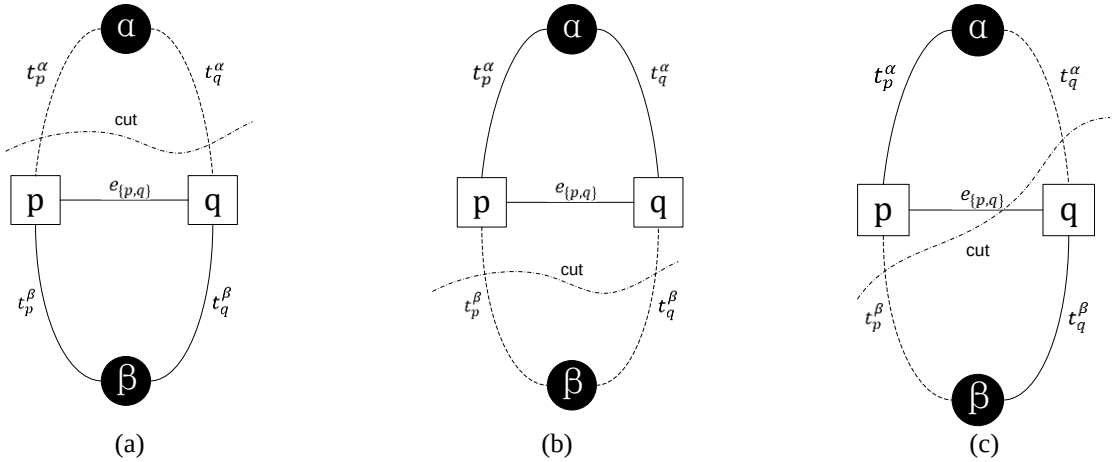


Figure 4. Examples of α-β-swap algorithm. (a), (b) and (c) represent three situations in graph cutting algorithms. Dotted lines indicate the edges cut while solid lines indicate the edges remaining in the original graph.

map derived from the projection transformation is experimentally set to 8 cm.

Following this, distance transformation maps (Diosi et al., 2005) can be obtained on the generated grid map, where the grayscale value of each pixel in the distance transformation map represents the distance from that point to the nearest border. Thus, the local maximum usually exists at the centre of each room, which is the basis for subsequent spatial partitioning.

Finally, after reversing the distance transformation map (as shown in Figure 3(b)), a watershed algorithm can segment the indoor environment into a set of room units (as shown in Figure 3(c)) since the watershed algorithm is an effective image segmentation method that views images as terrain and landforms, where the grayscale values of pixels represent altitude. In this end, the indoor spaces are initially partitioned into a set of individual units.

2.3.2 Room segmentation optimization under MRF framework: As explained in (Yang et al., 2021), over segmentation may occurs even if there is no building architecture inside the room, and it is more likely to emerge in more complicated large-scale indoor scenes. However, in contrast to component connectivity strategy adopted in (Yang et al., 2021), an optimization applies MRF model to merge over-segmented rooms has been proposed.

Markov Random Field is a graphical model used to model joint probability distributions, commonly used to describe relationships between multiple variables and to handle data with complex spatial structures. A Markov model can be represented as an undirected graph $G =< V, E >$, where V represents the various spatial units after initial room segmentation, and E denotes a set of undirected edges. The set of undirected edges consists of two parts: edges connecting source and sink points, and edges connecting spatial units pairwise. Then the merge of over-segmented rooms can be described as a multi-class label optimization problem, given a series of spatial partition units $F = \{f_k | 1 \le k \le m\}$, assign room instance labels $L = \{l_k | 1 \le k \le m\}$ to F. In the field of computer vision, this problem can be transformed into minimizing the energy objective function, as formulated in Eq. (1):

$$D(L) = \sum_{k=1}^{m} D_{data}(l_k) + \lambda * \sum_{(j,k)\in G} D_{smooth}(l_j, l_k) \qquad (1)$$

where $D_{data}(l_k)$ denotes data term and $D_{smooth}(l_j, l_k)$ represents pairwise term. λ represents the weight coefficient that balances data term and pairwise term.

Data term. Date term is designed to measure the degree of adaptation of a certain room unit to its corresponding category. Specifically, it can be expressed as Eq. (2):

$$E_{data}(l_k) = 1 - \frac{A_k}{B_k} \qquad (2)$$

where B_k represents the number of boundary points for a specific room unit f_k, and A_k represents the number of adjacent points between the room unit f_k and other room units.

Pairwise term. Pairwise term is used to encourage two adjacent spatial units to separate or merge. When there are no building vertical elements between two adjacent spatial units, they are more likely to have the same label value. Similarly, when there exists building vertical elements between them, they are more likely to possess different label values. Therefore, the pairwise term can be defined:

$$E_{smooth}(l_k, l_j) = \begin{cases} 0 & \text{if } l_k = l_j \\ |e_{jk}| \cdot (1 - \frac{|e_{jk} \cap v|}{|e_{jk}|}) & \text{if } l_k \neq l_j \end{cases} \qquad (3)$$

where e_{jk} denotes the incident edge between room unit f_j and f_k, v represents the vertical elements, and $|e|$ is the length of edge e.

Ultimately, our proposed MRF model is solved by the α-β-swap algorithm (as shown in Figure 4). The result of the optimized room segmentation can be seen in Figure 3(d).

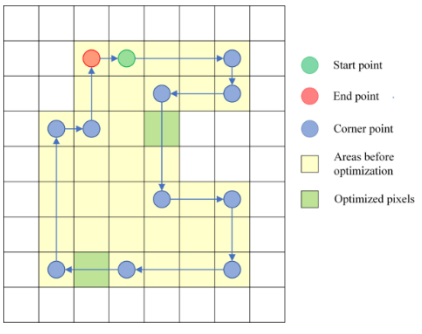


Figure 5. Illustration of room-wise boundary optimization algorithm.

Algorithm 1: Semantics-aware Recursive Search

Input: Current room boundary loop: $CurBoundary$,
Current room area of Curboundary: $CurArea$,
Original room boundary loop: R_i',
Original room area of R_i': P_i',
The rectangular area where the room is located: $RectArea$,
Start position in searching: $StartPos$,
Current position in searching: $CurPos$,
First stored searching direction: $FDirection$,
Second stored searching direction: $SDirection$,
Optimal penalty in searching: $OpitmPenalty$,
Current penalty in searching: $CurPenalty$
Output: Refined room boundary loop: R_i,

if $CurPos == StartPos$
  $CurPenalty \leftarrow U$(Curboundary);
  if $CurPenalty < OpitmPenalty$
    $R_i \leftarrow CurBoundary$;
    $OpitmPenalty \leftarrow CurPenalty$;
  return;
$CurPos \leftarrow CurPos + FDirection$;
if $CurPos \in RectArea$
  if $CurPos \in R_i'$ && $P_i'(CurPos)! = 0$
  $RecursiveSearch$;
else if $P_i'(CurPos) == 0$
  $SDirection \leftarrow CalculateDirection\ (\ CurPos\ ,\ R_i'$);
  $RecursiveSearch$;
  $FDirection \leftarrow SDirection$;
  $RecursiveSearch$;
else
  $FDirection \leftarrow SDirection$;
  $CurPos \leftarrow CurPos + FDirection$;
  $RecursiveSearch$;

2.4 Room-wise boundary optimization using semantics-aware Recursive Search

Due to the presence of noise points, the boundaries of the rooms segmented in the previous step are irregular, thus the objective of this subsection is to obtain room borders of regular while maintaining its original shapes as much as possible. The inspiration of our work is to transform the room boundary optimization into an energy objective formulation minimization, which is formulated in Eq. (4):

$$\mathrm{U}(\mathrm{R_i}) = \mathrm{U_{complexity}}(\mathrm{R_i}) + \beta * \mathrm{U_{freedom}}(\mathrm{R_i}) \quad (4)$$

where $\mathrm{U_{complexity}}(\mathrm{R_i})$ is complexity term, and $\mathrm{U_{freedom}}(\mathrm{R_i})$ represents the freedom term. β is a weight parameter balancing the two terms. $\mathrm{R_i}$ represents the loop of room boundary, and specifically it is defined as a polygonal curve with a loop topology, consisting of a sequence of pixels at integer coordinates.

Complexity term. $\mathrm{U_{complexity}}(\mathrm{R_i})$ is the model penalty describing the complexity of the refined room boundary, which can be quantified by the number of corners within each loop (as defined in Eq. (5)).

$$\mathrm{U_{complexity}}(\mathrm{R_i}) = \{\text{number of corners of } \mathrm{R_i}\} \quad (5)$$

Freedom term. $\mathrm{U_{freedom}}(\mathrm{R_i})$ describes the penalty of degrees of freedom in boundary refinement, which can be formulated in Eq. (6).

$$\mathrm{U_{freedom}}(\mathrm{R_i}) = \sum_{\mathrm{p}\in \mathrm{P_i}, \mathrm{p}\notin \mathrm{P_i}'} \mathrm{V(p)} \quad (6)$$

where p represents the pixels of loop $\mathrm{R_i}$, $\mathrm{P_i}'$ and $\mathrm{P_i}$ denotes the room areas before and after optimization, respectively. V(p) is used to measure the penalty value of pixels at different positions.

$$V(p) = \begin{cases} \mu_1 & \text{if labelmap}(p) = 0 \\ \mu_2 & \text{if labelmap}(p) \neq 0 \end{cases} \quad (7)$$

where labelmap(p) refers to the 2D grid map generated by projecting the original point cloud onto the X-Y plane. When there is no point cloud within a certain grid, its value is set to 0. Considering the presence of noise points or point cloud deficiency caused by occlusion, the generated 2D grid map labelmap cannot accurately reflect the indoor building structure. Therefore, the areas with value of 0 within labelmap are also required to be considered, but the penalty term for its corresponding pixel should be larger. In our experiments, μ_1 is set to 1 while μ_2 is set to 3.

We search for the optimal room boundary by minimizing the energy formulation through a designed semantics-aware recursively searching strategy which is described in **Algorithm 1**. Figure 5 provides a more intuitive representation of our proposed algorithm. The purpose of our method is to search within the boundaries of the original room, filling in some blank pixels during the search process to minimize the energy formulation. Afterwards, the boundary points of the 8-connected domain are extracted to connect adjacent points into a line and merge the line segments on the same line. Ultimately, as shown in Figure 3(e), the refined room boundary can be obtained after converting the rasterized image of the above results into vector-graphic data (as shown in Figure 3(f)).

3. Experimentation and analysis

3.1 Evaluation on room segmentation

Four metrics are adopted to evaluate the performance of room segmentation, including, precision, recall and $\mathrm{F_{1\text{-}score}}$. IoU measures the ratio of the overlapping area between the predicted rooms and the ground truth. The precision, recall and F1 score are formulated in Eqs. (7-10).

$$precision = \frac{TP}{TP + FP} \quad (7)$$

$$recall = \frac{TP}{TP + FN} \quad (8)$$

$$F_{1-\mathrm{score}} = \frac{TP}{TP + FP} \quad (9)$$

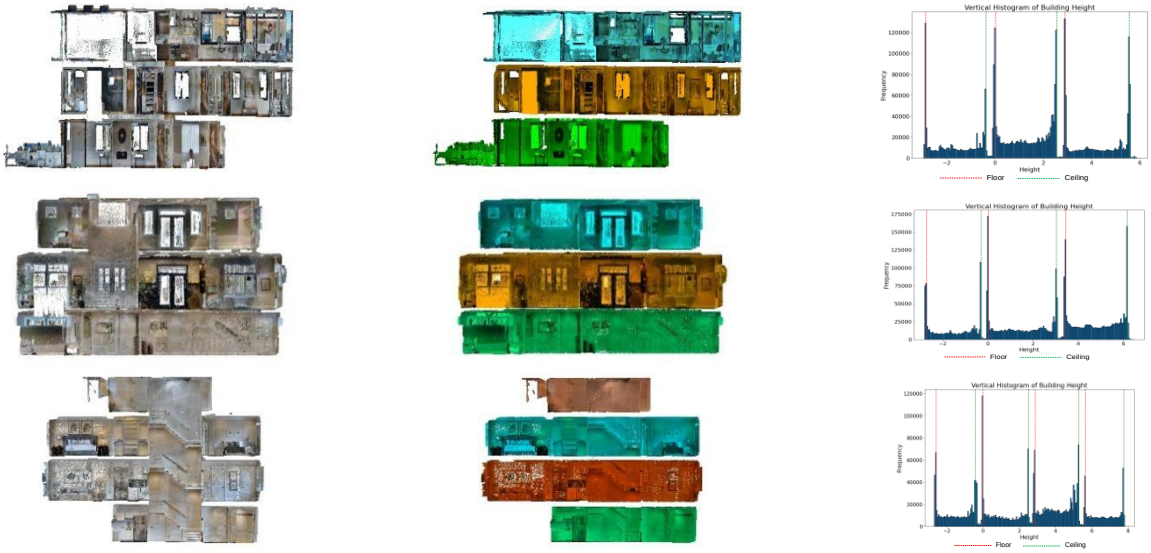
$$\mathrm{IoU} = \frac{TP_i}{GT_i + FP_i} \quad (10)$$

where TP represents the number of true positives, TN represents the number of true negatives, FN represents the number of false negatives, and FP represents the number of false positives. Additionally, $\mathrm{TP_i}$, $\mathrm{GT_i}$, and $\mathrm{FP_i}$ represent the number of true positives, ground truth, and false positives in class i, respectively.

3.2 Storey segmentation results

In the section of the storey separation, the semantic point clouds are used to analyse the height characteristics of the building and calculated the semantics-enhanced height histogram along the vertical direction. Based on this histogram, peak detection algorithm was applied to successfully identify the heights of multiple floors and ceilings, and Figure 6 shows significant height differences, which are in line with the characteristics of building structures.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-961-2025 | © Author(s) 2025. CC BY 4.0 License.



(a) (b) (c)

Figure 6. Storey segmentation results:(a)Colorized point clouds, (b)Segmented Storeys, (c)Semantics-enhanced height histogram

3.3 Room segmentation results

3.3.1 Room segmentation on S3DIS dataset: Figure 8 shows the generation results of indoor subspaces on S3DIS dataset at different steps, including watershed segmentation, room segmentation through MRF optimization and room-wise boundary refinement. There are many over segmentation phenomena in the indoor space partition based on the building physical structures, which is caused by the limitations of the watershed algorithm. Based on room optimization module under the Markov random field framework, the problem of merging over segmented units is transformed into an energy minimization problem and solved using graph cuts. The third column of Figure 8 illustrates the corresponding results. It shows that room segmentation optimization based on Markov random field model has achieved the merging of over-segmented indoor spatial units, especially in rooms such as some long corridors. Moreover, due to the influence of point cloud noises, the boundaries of the room units are not regular. The room-wise boundary optimization combined with topological information transforms the building boundary optimization problem into an energy minimization problem. From the forth column in Figure 8, the room boundary has become more regular while retaining finer details.

Type of room	Size(m^2)	Recall (%)	Precision (%)	$F_{1\text{-score}}$ (%)
ConferenceRoom_1	24.2	88.8	100.0	94.1
ConferenceRoom_2	35.7	88.3	99.8	93.7
CopyRoom_1	10.2	89.5	100.0	94.5
Hallway_1	6.7	97.9	93.3	95.5
Hallway_2	104.6	92.7	97.0	94.8
Office_1	13.0	91.5	100.0	95.5
Office_2	30.8	93.8	100.0	96.8
Office_3	17.2	92.5	100.0	96.1
Office_4	17.2	94.3	99.8	97.0
WC_1	2.2	83.9	99.9	91.2

Table 1 Quantitative evaluations on Area 1 of S3DIS dataset.

Type of room	Size(m^2)	Recall (%)	Precision (%)	$F_{1\text{-score}}$ (%)
Auditorium_1	213.9	98.0	98.0	98.0
Auditorium_2	282.7	99.7	100.0	99.8
ConferenceRoom_1	66.8	87.4	100.0	93.3
Hallway_1	5.8	97.2	93.0	95.0
Hallway_2	21.6	99.3	98.5	98.9
Hallway_3	24.3	99.6	96.8	98.2
Office_1	9.2	97.8	96.6	97.2
Office_2	12.2	95.3	100.0	97.6
WC_1	12.8	99.2	99.9	99.5
Storage_1	3.4	99.5	100.0	99.7

Table 2 Quantitative evaluations on Area 2 of S3DIS dataset.

Type of room	Size(m^2)	Recall (%)	Precision (%)	$F_{1\text{-score}}$ (%)
Region1	13.0	98.1	97.1	97.6
Region2	16.3	94.4	93.8	94.1
Region3	7.16	97.3	92.3	94.7
Region4	50.5	98.2	95.2	96.6
Region5	7.1	91.8	98.5	95.0
Region6	25.9	94.4	97.5	95.9

Table 3 Quantitative evaluations on Area 1 of Matterport3D dataset.

Type of room	Size(m^2)	Recall (%)	Precision (%)	$F_{1\text{-score}}$ (%)
Region1	31.9	95.3	100.0	97.6
Region2	55.4	91.3	99.9	95.4
Region3	30.2	94.4	94.2	94.3
Region4	8.25	97.0	99.2	98.1
Region5	30.1	94.6	96.2	95.4
Region6	8.6	94.2	98.5	96.3

Table 4 Quantitative evaluations on Area 2 of Matterport3D dataset.

(a) (b) (c)

Figure 7. Examples of the door in different states: (a) open, (b) closed, (c) partially open.

To quantitatively evaluate the accuracy of room segmentation, we cannot directly report the recall, precision and $F_{1\text{-score}}$ for the entire scenario since compared with previous researches (ISPRS 2021), we focus on more complex and challenging large-scale indoor scenes, which are prone to over segmentation and under segmentation. As a result, the predicted number of segmented rooms is different from ground truth. Thus, in this subsection, we display some selected representative rooms in each scene for the quantitative evaluation.

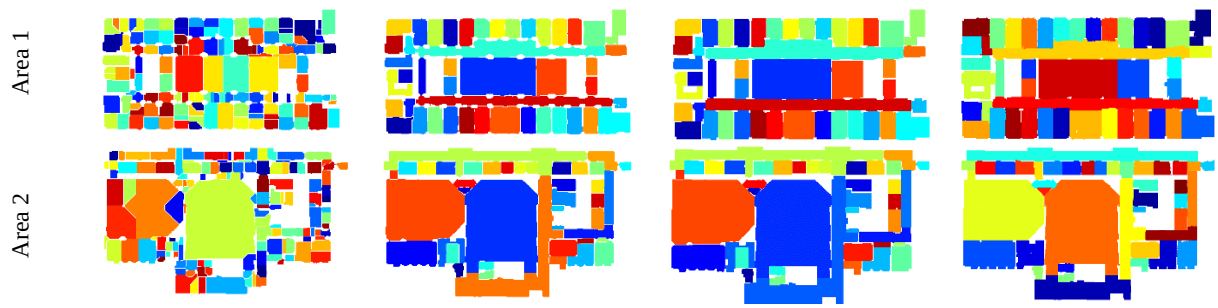
Tables 1 and 2 present the results of room segmentation from the quantitative perspective in different scenarios on S3DIS dataset. For most of room segmentation results, the recall is above 90% and the precision is above 95%. It can be concluded that the semantic guided indoor room segmentation method is effective for complex layout indoor spaces in large scenes. In addition, the vertical physical structures of buildings obtained through semantic information and subsequent spatial partitioning using distance transformation maps and watershed algorithms can provide prior information for subsequent room segmentation, ensuring a high recall rate of segmentation results. Based on Markov random fields, optimizing and merging the over segmented rooms ensures high accuracy of the segmentation results. It is worth noting that for some large rooms over 200 square meters, the room segmentation can still achieve satisfying results, with its precision, recall, and $F_{1\text{-score}}$ all exceeding 98%.

3.3.2 Room segmentation on Matterport3D dataset: Compared with S3DIS dataset, although the number of rooms in Matterport3D dataset is relatively small, around 20, the sparse point clouds and missing point clouds in some scenes make the room segmentation even more challenging. We select three scenarios with different complexity and characteristics for evaluation.

The results of room segmentation at different stages are listed in Figure 9. The ground truth is created by projecting the room labels provided by the dataset onto a two-dimensional grid map, so the ground truth may not be accurate. But intuitively, our segmentation results can restore the spatial distribution of the original scene. Like the previous subsection, quantitative evaluation for selected rooms is reported in Tables 3 and 4. We can observe that most of rooms can get a satisfying performance, with recall, precision exceeding 88.1%, 92.3% and 93.2% respectively. The relatively low recall rate is mainly due to inaccurate ground truth.

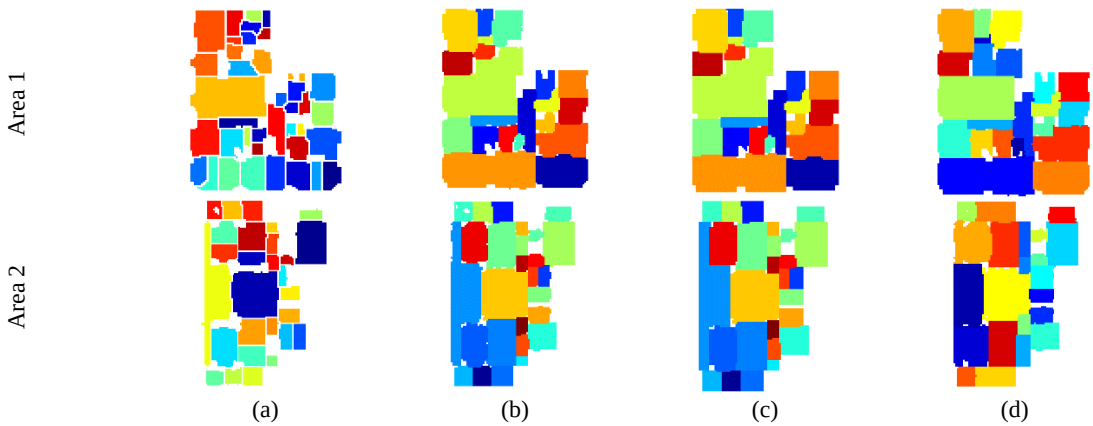
3.4 Impact of Doors on Room Segmentation

Doors are structures that connect adjacent spaces, marking the separation between rooms. Since it is not possible to ensure that



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

Figure 8. The generation of indoor subdivision on S3DIS. Different rooms are rendered with different colors: (a)Watershed result, (b)MRF optimization, (c)Room boundary optimization, (d)Ground Truth.



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

Figure 9. The generation of indoor subdivision on Matterport3D. Different rooms are rendered with different colors: (a)Watershed result, (b)MRF optimization, (c)Room boundary optimization, (d)Ground Truth.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-X-G-2025-961-2025 | © Author(s) 2025. CC BY 4.0 License.

every door remains closed during data collection, the open/closed state of the door needs to be considered in room segmentation. The states of a door can be classified into three types: closed, open, and partially open, as shown in Figure 7.

For closed doors, the model presented in this paper recognizes them as part of the vertical building category for subsequent spatial segmentation, which does not affect the room segmentation results. For open or partially open doors, gaps may exist between adjacent rooms, which theoretically could lead to under-segmentation of rooms. However, the proposed model avoids this issue because doors and door frames always appear together. Regardless of whether the door is open or partially open, the door frame consistently aligns with the wall. Therefore, when the door frame is semantically recognized as part of the vertical building category, it does not affect the room segmentation results. Compared to traditional methods that rely on geometric constraints, the approach based on semantic information guidance handles the impact of door status changes on segmentation results more effectively, ensuring that room segmentation accuracy remains unaffected even when doors are open.

Additionally, the method proposed in this paper utilizes Markov Random Field (MRF) to optimize the over-segmentation issue, considering factors such as the location and length of walls, rather than relying solely on room connectivity. This optimization approach avoids segmentation issues caused by the open/closed state of the door, ensuring the precision of the segmentation results.
Through these improvements, the model demonstrated in this paper shows high stability and accuracy in handling the influence of door status, effectively mitigating the negative impact of door status changes on room segmentation.

4. Summary and outlook

This paper introduces a semantic-guided pipeline for large-scale indoor spatial data generation, focusing on room segmentation and boundary optimization. Unlike methods that rely on geometric constraints for indoor space partitioning, which can be hindered by issues like furniture placement, uneven point cloud distribution, or object occlusion, our approach leverages semantic information. This allows for the extraction of geometric primitives that enable efficient space subdivision across multiple storeys. Through optimization using multiple energy functions, we address over-segmentation and room boundary optimization, with the refined subdivision generated by minimizing these energy functions. Our experiments on the Stanford Large-scale 3D Indoor Spaces Dataset and Matterport3D dataset demonstrate the effectiveness and robustness of the proposed method.

However, despite the strong results in our experiments, challenges remain in handling sparse and noisy data, which may impact the performance of the method in more complex environments. In sparse data situations, when the point cloud density is low—particularly in areas with minimal furniture or simple room layouts—the method may struggle to capture the necessary geometric features of the space. This can lead to inaccurate segmentation results. To address this, future work could incorporate data augmentation techniques like point cloud reconstruction or random sampling to enhance the density of the input data. Multi-scale analysis methods could also be employed to process data at varying resolutions, helping to improve the method’s adaptability and robustness in such conditions.

Expanding the scope of experiments to evaluate the method in more diverse environments would help confirm its generalizability. Testing the method in more complex or non-traditional indoor spaces, such as industrial facilities or unconventional buildings, would help demonstrate its versatility and reliability in a wider range of contexts. Additionally, incorporating deep learning techniques like point cloud deep learning networks could further improve the method's ability to handle large-scale, complex datasets. These techniques are particularly effective at extracting meaningful features from dense point clouds and could lead to better segmentation performance in large and intricate environments. What’s more, we also plan to apply the semantic-guided method to outdoor scenarios, such as UAVs and vehicle-mounted LiDAR systems, for 3D scene reconstruction. This extension has significant potential for real-world 3D modeling and smart city applications.

Acknowledgements

This work is jointly supported by the National Natural Science Foundation of China (No. 42201486, 42371453) and Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-temporal Big Data Technology (No. TKL2024A07).

References

Armeni I, Sener O, Zamir A R, et al., 2016. 3d semantic parsing of large-scale indoor spaces[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 1534-1543.

Baek F, Ha I, Kim H, 2019. Augmented reality system for facility management using image-based indoor localization[J]. Automation in construction, 99: 18-26.

Boguslawski P, Mahdjoubi L, Zverovich V, et al., 2016. Automated construction of variable density navigable networks in a 3D indoor environment for emergency response[J]. Automation in Construction, 72: 115-128.

Chang A, Dai A, Funkhouser T, et al., 2017. Matterport3d: Learning from rgb-d data in indoor environments[J]. arXiv preprint arXiv:1709.06158.

Chen J, Liu C, Wu J, et al., 2019. Floor-sp: Inverse cad for floorplans by sequential room-wise shortest path[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2661-2670.

Diakité A A, Zlatanova S, 2018. Spatial subdivision of complex indoor environments for 3D indoor navigation[J]. International Journal of Geographical Information Science, 32(2): 213-235.

Diosi A, Taylor G, Kleeman L, 2005. Interactive SLAM using laser and advanced sonar[C]. Proceedings of the 2005 IEEE International Conference on Robotics and Automation. IEEE, 1103-1108.

Farahsari P S, Farahzadi A, Rezazadeh J, et al., 2022. A survey on indoor positioning systems for IoT-based applications[J]. IEEE Internet of Things Journal, 9(10): 7680-7699.

Fathalizadeh A, Moghtadaiee V, Alishahi M, 2023. Indoor geo-indistinguishability: Adopting differential privacy for indoor location data protection[J]. IEEE Transactions on Emerging Topics in Computing, 12(1): 293-306.

Fu M, Liu R, Qi B, et al., 2020. Generating straight skeleton-based navigation networks with Industry Foundation Classes for indoor way-finding[J]. Automation in Construction, 112: 103057.

Gao R, Ye F, Luo G, et al., 2018. Indoor map construction via mobile crowdsensing[J]. Smartphone-Based Indoor Map Construction: Principles and Applications, 3-30.

Isikdag U, Zlatanova S, Underwood J, 2013. A BIM-Oriented Model for supporting indoor navigation requirements[J]. Computers, Environment and Urban Systems, 41: 112-123.

Jeamwatthanachai W, Wald M, Wills G., 2017. Map data representation for indoor navigation by blind people[J]. International Journal of Chaotic Computing, 4(1): 70-78.

Kang H K, Li K J, 2017. A standard indoor spatial data model—OGC IndoorGML and implementation approaches[J]. ISPRS International Journal of Geo-Information, 6(4): 116.

Kang Z, Yang J, Yang Z, et al., 2020. A review of techniques for 3d reconstruction of indoor environments[J]. ISPRS International Journal of Geo-Information, 9(5): 330.

Luo J, Ye Q, Zhang S, et al., 2023. Indoor mapping using low-cost MLS point clouds and architectural skeleton constraints[J]. Automation in Construction, 150: 104837.

Luo Z, Huang W, 2022. FloorplanGAN: Vector residential floorplan adversarial generation[J]. Automation in Construction, 142: 104470.

Meyer T, Brunn A, Stilla U, 2023. Geometric BIM verification of indoor construction sites by photogrammetric point clouds and evidence theory[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 195: 432-445.

Moreno M V, Zamora M A, Skarmeta A F., 2014. User-centric smart buildings for energy sustainable smart cities[J]. Transactions on emerging telecommunications technologies, 25(1): 41-55.

Noureddine H, Ray C, Claramunt C, 2022. A hierarchical indoor and outdoor model for semantic trajectories[J]. Transactions in GIS, 26(1): 214-235.

Otero R, Lagüela S, Garrido I, et al., 2020. Mobile indoor mapping technologies: A review[J]. Automation in Construction, 120: 103399.

Pang Y, Zhou L, Lin B, et al., 2020. Generation of navigation networks for corridor spaces based on indoor visibility map[J]. International Journal of Geographical Information Science, 34(1): 177-201.

Park S, Yu K, Kim J, 2020. Data model for IndoorGML extension to support indoor navigation of people with mobility disabilities[J]. ISPRS International Journal of Geo-Information, 9(2): 66.

Pexman K, Lichti D D, Dawson P, 2021. Automated storey separation and door and window extraction for building models from complete laser scans[J]. Remote Sensing, 13(17): 3384.

Shabani M A, Hosseini S, Furukawa Y, 2023. Housediffusion: Vector floorplan generation via a diffusion model with discrete and continuous denoising[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5466-5475.

Shang J, Tang X, Yu F, et al., 2015. A semantics-based approach of space subdivision for indoor fine-grained navigation[J]. Journal of Computational Information Systems, 11(9): 3419-3430.

Tashakkori H, Rajabifard A, Kalantari M, 2015. A new 3D indoor/outdoor spatial model for indoor emergency response facilitation[J]. Building and Environment, 89: 170-182.

Teo T A, Cho K H, 2016. BIM-oriented indoor network model for indoor and outdoor combined route planning[J]. Advanced engineering informatics, 30(3): 268-282.

Wang X, Yang J, Kang Z, et al., 2023. A category-contrastive guided-graph convolutional network approach for the semantic segmentation of point clouds[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16: 3715-3729.

Wu Y, Shang J, Chen P, et al., 2021. Indoor mapping and modeling by parsing floor plan images[J]. International Journal of Geographical Information Science, 35(6): 1205-1231.

Yang J, Kang Z, Zeng L, et al., 2021. Semantics-guided reconstruction of indoor navigation elements from 3D colorized points[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 173: 238-261.

Zhou X, Xie Q, Guo M, et al., 2020. Accurate and efficient indoor pathfinding based on building information modeling data[J]. IEEE Transactions on Industrial Informatics, 16(12): 7459-74