

RESEARCH ARTICLE

Boundary-Aware Supervoxel Segmentation for Indoor 3D Point Clouds

FEI SU¹, YU LIU², KAIXIAO NIE³, YAOHUI LIU¹, JINGXUE BI¹,
RAN ZHANG¹, AND GUOQIANG ZHENG¹

¹School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan 250101, China

²Institute of Environment and Development, Guangdong Academy of Social Sciences, Guangzhou 510635, China

³School of Basic Medicine, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan 250000, China

Corresponding author: Guoqiang Zheng (gqzheng@sdjzu.edu.cn)

This work was supported in part by the Natural Science Foundation of Shandong Province under Grant ZR2021QD105 and Grant ZR2021QD074, in part by the National Natural Science Foundation of China under Grant 42201077 and Grant 42177453, and in part by the Guangdong Office of Philosophy and Social Science under Grant GD21CTS01.

ABSTRACT Supervoxels provide a natural and compact representation of 3D point clouds that enables operations to be performed on regions rather than on scattered points. However, most supervoxel segmentation methods generate supervoxels based on geometric dissimilarity among points and cannot preserve accurate structure boundaries when noise and outliers are present due to the uncertainty in the random sampling of representative points (RPs). This paper formulates indoor supervoxel segmentation as selecting RPs and matching each point with its RP simultaneously. The RPs are selected from nonboundary areas instead of randomly sampled points, guaranteeing the correctness of sampling from the same structure. The best point-to-RP matching is achieved through iterative refinement/clustering via an energy descent method, which ensures the optimal overall segmentation. Experimental tests on five publicly available datasets demonstrate that our method preserves indoor structure boundaries and small structures more effectively than other state-of-the-art methods, resulting in performance scores of approximately 0.65 for boundary recall.

INDEX TERMS Point cloud, indoor segmentation, indoor structure, supervoxel.

I. INTRODUCTION

Fast and stable oversegmentation of 3D point clouds into supervoxels is an important preprocessing step in many applications, such as object detection [1], [2], [3], [4], semantic segmentation [5], [6], and line extraction [7], [8], because of its ability to reduce complexity while fully utilizing the available spatial information. In contrast to the supervoxels in videos and images, which correspond to 3D voxels in the spatiotemporal domain [9], a supervoxel in a 3D point cloud is defined as a cluster of 3D points exhibiting coherence in both appearance and features [9], [10], [11].

Indoor supervoxel segmentation is a crucial preprocessing technique, but it remains in the developmental stage due to the complexity of indoor environments and data restrictions. Despite recent research efforts, satisfactory solutions for segmenting indoor supervoxels have not yet been developed. The

pioneering voxel cloud connectivity segmentation (VCCS) method generates supervoxels based on the neighbor relationships of voxels at fixed resolutions [12]. However, this method relies on the initialization of seed points and may not accurately preserve the boundaries of structures, especially when the density is nonuniform [9]. Later works formalized supervoxel segmentation as subset selection equations to improve the adherence of supervoxels to structure boundaries [9], [13], [14], [15]. However, these methods may not be suitable for indoor scenes containing noise and outliers because of the uncertainty in the random sampling of representative points (RPs). Therefore, there is a critical need to develop a method that can accurately segment indoor supervoxels with awareness of structures, which remains a significant challenge for indoor supervoxel segmentation.

This paper proposes a novel approach to segment indoor supervoxels that accounts for the boundaries of structures. Unlike previous methods that randomly sample points or rely on fixed-resolution voxel connectivity, our method selects

The associate editor coordinating the review of this manuscript and approving it for publication was Kegen Yu¹.

RPs from nonboundary areas and iteratively refines their matching with input points using an energy descent method that minimizes the point-to-RP fitting residuals with awareness of structure boundaries. Furthermore, the point-to-RP fitting residuals are evenly divided among the supervoxels, ensuring optimal overall segmentation.

The rest of this paper is organized as follows. Related works are introduced in Section II. The proposed supervoxel segmentation approach is formulated and implemented as described in Section III. Experiments performed on five real-world datasets are reported in Section IV. Finally, Section V outlines the conclusions and future directions of research.

II. RELATED WORKS

Supervoxel is a natural extension of the concept of superpixels in image processing [16], [17], [18], [19], [20]. Despite the thoroughly research and widespread application of superpixel segmentation in electronics fields, methods for segmenting supervoxels from 3D point clouds are still in an early stage of development.

In video and 3D image segmentation, a supervoxel represents a stack of 2D image regions, and the related segmentation methods include energy minimization methods [21], feature clustering methods [10], [22], graph-based methods [23], and contour optimization methods [24]. However, these methods are designed for data with regular structures, for which the primitives are uniformly distributed [8], and thus cannot be directly applied to unordered and unstructured 3D point clouds.

The VCCS method [12] was proposed as a pioneering approach to address this issue by voxelizing a point cloud via octree clustering based on the geometric and contextual similarity among k-nearest neighbor (kNN) voxels. Although VCCS is efficient and yields satisfactory results depending on the voxel resolution, it may violate structure boundaries due to the fixed resolution in the seed selection, especially in point clouds with nonuniform densities [9], [14]. To overcome this limitation, the boundary-enhanced supervoxel segmentation (BESS) [15] method treats supervoxel segmentation as a boundary detection issue and solves it in two steps. In the first step, boundary points are detected based on the discontinuity of consecutive points along the scanner's scan-line. In the second step, nonboundary points are clustered by means of a generated neighborhood graph. However, the practical applicability of BESS is limited to datasets with sequential point ordering along the same scan line, which restricts its applicability to general point cloud datasets.

To improve the practicality and effectiveness of supervoxel segmentation, Lin et al. [9], [25] and Xiao et al. [26] formulated supervoxel segmentation as a subset selection problem and solved it within a heuristic optimization framework. In their approaches, the RPs for each supervoxel are randomly selected based on a dissimilarity metric among points, thus achieving superior effectiveness in both indoor and outdoor environments. However, these methods still suffer from

missing structural details due to the uncertainty in the random selection of RPs based on geometric deviation.

III. METHODS

A. OVERVIEW

Inspired by previous works [6] and [9], the task of segmenting a raw indoor point cloud $P = \{p_1, p_2, \dots, p_N\}$, consisting of N points, into supervoxels $V = \{v_1, v_2, \dots, v_K\}$ can be formulated as seeking representative points RP and finding optimal correspondences between nonrepresentative point $p_i \in P$ and its corresponding $r_{li} \in RP$ simultaneously. This can be achieved via an energy minimization algorithm (Eq. 1) based on a binary indicator matrix $x = \{0, 1\}_{N \times K}$.

$$x^* = \underset{x}{\operatorname{argmin}} \left(\sum_{li=1}^{\operatorname{size}(RP)} \sum_{i=1}^N x * D(p_i, r_{li}) + \lambda(\operatorname{size}(RP) - K) \right) \quad (1)$$

where $D(p_i, r_{li})$ represents the dissimilarity metric between p_i and r_{li} and K represents the number of supervoxels. As RP is closely related to the supervoxels, in this paper, we first distinguish boundary points from all points via a contour detection method and then select the points in RP from nonboundary areas.

The entire workflow consists of three main steps: boundary detection, initial supervoxel generation and boundary refinement, as depicted in Fig. 1. First, the boundary areas are extracted from the raw point cloud via a 2D contour detection method. Then, every point in RP is selected from the nonboundary areas, and the supervoxel corresponding to each RP is constructed via an energy descent algorithm. Finally, the boundaries of the supervoxels are further refined via structure boundary enhancement and refinement processing.

B. BOUNDARY DETECTION

According to the definition of supervoxels, the RPs should not be selected from the boundary areas of indoor structures. Therefore, the first step of our method is to detect the structure boundaries and filter out these points. Each segmented patch is converted into a binary image via 3D plane projection, and its contours are extracted via the iterative reweighted least-squares (IRLS) algorithm. The structure boundary in each patch is obtained via a 2D contour reprojection method. Fig. 2 presents an example of the boundary detection results obtained in this way.

1) RANSAC SEGMENTATION

Because indoor structures are generally man-made and can be approximately represented by various planes, the points in the indoor scene are first segmented via the random sampling consensus (RANSAC) method [27]. Given a point cloud $P = \{p_1, p_2, \dots, p_N\}$ with associated normals $N = \{n_1, n_2, \dots, n_N\}$, the output of our algorithm is a set of primitive shapes $M = \{m_1, m_2, \dots, m_m\}$ with corresponding disjoint point sets $DP = \{dp_1, dp_2, \dots, dp_m\}$. Each $dp_i \in PC$ is associated with one corresponding primitive shape $m_i \in M$,


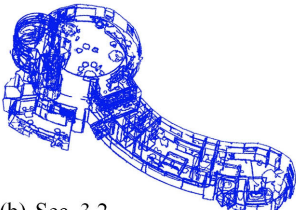
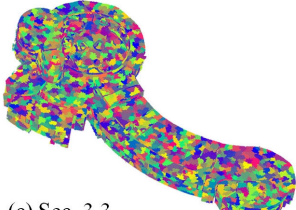
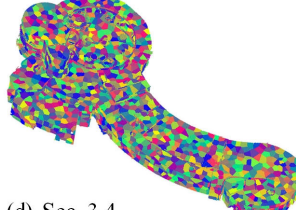
Raw Point Cloud	Boundary Detection	Initial supervoxel generation	Boundary refinement
 (a)	 (b) Sec. 3.2	 (c) Sec. 3.3	 (d) Sec. 3.4

FIGURE 1. The pipeline of the proposed boundary-aware supervoxel segmentation method.

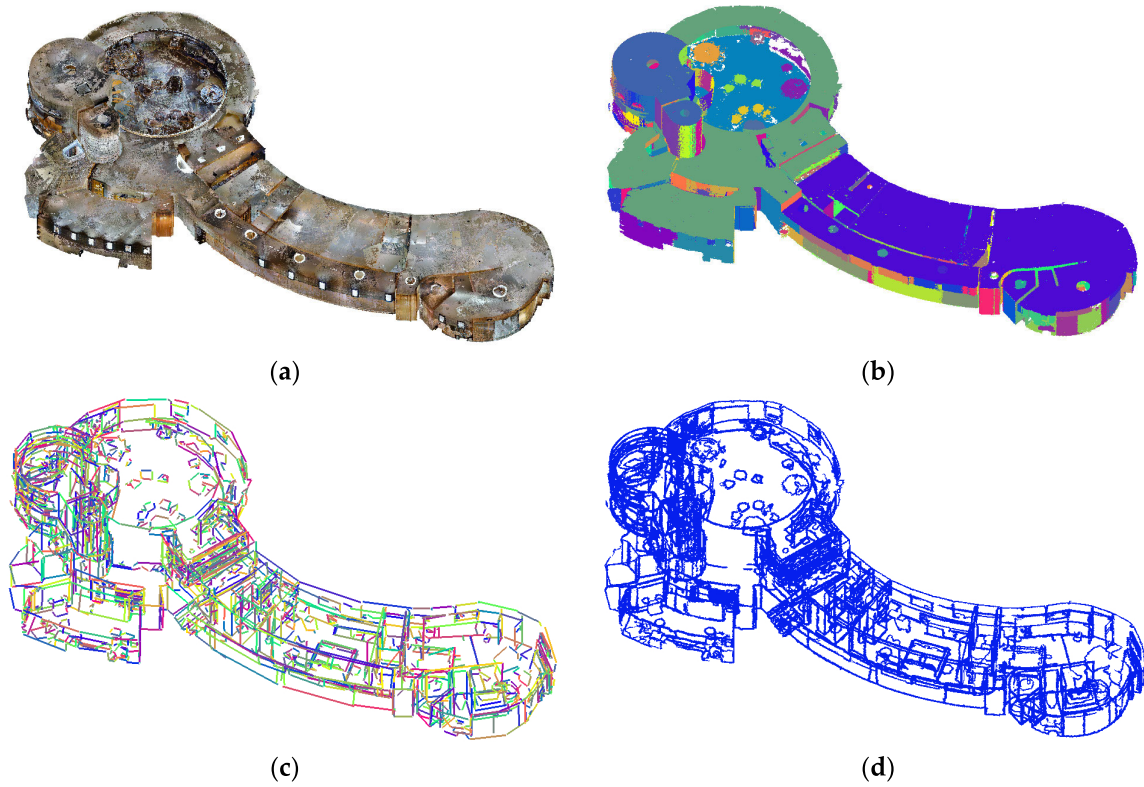


FIGURE 2. Boundary detection results: (a) original point cloud, (b) segmentation results, (c) line detection results and (d) detected boundary points after buffer analysis.

where m_i is a planar shape representing a tangent plane. m_i is parameterized as $\{p_c(x_c, y_c, z_c), n_c(nx_c, ny_c, nz_c)\}$, where p_i and n_i of m_i represent the center point and normal, respectively, of the fitted plane.

The most important parameters are the maximum point-to-distance threshold, α_1 ; the normal angle threshold, α_2 ; the minimum number of supporting points per plane, α_3 ; and the probability of missing the next largest plane candidate in each iteration, α_4 . These parameters are chosen based on the characteristics of the dataset, as described in Section IV-A.3. Given that the density of the point cloud plays an important role in the RANSAC segmentation algorithm, we subsampled all point clouds in our experiments to a minimum point

distance of 1 cm during preprocessing to ensure reliable results.

2) CONTOUR DETECTION

Following segmentation, the points in each segmented region are projected onto occupancy bitmaps on their respective planes. Each point $p_j \in dp_i$ is projected onto its associated planar shape m_i via Eq. 2, as shown in Fig. 3.

$$\begin{cases} x_j = \overrightarrow{p_j p_c} \cdot v_x \\ y_j = \overrightarrow{p_j p_c} \cdot v_y \end{cases} \quad (2)$$

$$\overrightarrow{p_j' p_c} = \overrightarrow{p_j p_c} - (\overrightarrow{p_j p_c} \cdot n_c) \cdot n_c \quad (3)$$

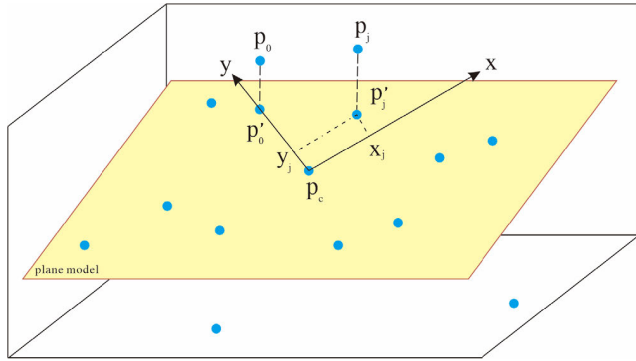


FIGURE 3. Illustration of the procedure for converting a 3D point into 2D.

where p'_j represents the point after p_j is projected onto its shape m_i . $\overrightarrow{p'_j p_c}$ is the x-axis, denoted by v_x ; thus, the y-axis v_y can be calculated as $v_x \times n_c$.

C. INITIAL SUPERVOXEL GENERATION

As described in the previous section, the indoor scene is transformed into boundary points BP and nonboundary points NP , and the RPs are selected from NP through iterative refinement/clustering via an energy descent method. All non-boundary points are initially assigned to supervoxels such that each RP corresponds to its own supervoxel, and then the points are sorted in ascending order according to the curvature of each RP. Then, inspired by [9], when merging two adjacent supervoxels s_i and s_j would reduce the overall energy, these two supervoxels are merged into one supervoxel. This process is repeated until the energy no longer decreases or the supervoxel size approaches T_{min} . The details are shown in Algorithm 1, and the supervoxel merging process and point-to-supervoxel similarity computation are described in Sections III-C.1 and III-C.2, respectively. The settings of R and λ are discussed in Section IV. The initial supervoxels and RPs are shown in Fig. 4.

1) SUPERVOXEL MERGING

Given two adjacent supervoxels s_i and s_j , if these two supervoxels were to be merged into one supervoxel s'_i , the merging energy E_{Δ} would equal

$$E_{\Delta} = E_{new} - E_{old} = \left(\sum_{p_n \in s'_i} D(p_n, r_i) + \sum_{p_m \in s_j} D(p_m, r_j) - \sum_{p_t \in s'_i} D(p_t, r_i) \right) - \lambda \quad (4)$$

To reduce the time complexity of $O(size(s_j) * 2)$, we assume that the dissimilarity D between s_i and s_j is a metric that satisfies the triangle inequality. Therefore, Eq. 4 can be written as

$$\begin{aligned} E_{\Delta} &\leq \sum_{p_m \in s_j} D(p_m, r_j) \\ &\quad - \left(\sum_{p_m \in s_j} D(p_m, r_i) + \sum_{p_m \in s_j} D(r_j, r_i) \right) - \lambda \\ &= size(s_j) * D(r_j, r_i) - \lambda \end{aligned} \quad (5)$$

Algorithm 1 Initial Supervoxel Generation

```

1 Input: nonboundary points NP
2 Output: supervoxel set S and RP set RP
3 Set R and λ
4 Initialize: S ← NP; RP ← NP
5 Compute the RP number size(RP) by R
6 while size(S) ≥ size(RP)
7   Sort S by curvature
8   for each s_i ∈ S
9     for each s_j ∈ neighbor(s_i)
10      if λ > size(s_i) * D(r_j, r_i) //Section III-C.1
11        Merge s_j into s_i
12        Delete r_j from RP
13      end if
14    end for
15  end while
16  λ ← λ * 2
17 end while
18 Return S and RP

```

Thus, if $size(s_j) * D(r_j, r_i) - \lambda < 0$, then $E_{\Delta} < 0$, which means that merging s_i and s_j will reduce the total energy, and therefore, these two supervoxels should be merged.

2) SUPERVOXEL SIMILARITY METRIC $D(r_i, r_j)$

Given two supervoxels s_i and s_j with respective representative points r_i and r_j , the supervoxel similarity can be computed as shown in Eq. 6.

$$D(r_i, p_j) = w_1 D_{geo}(i, j) + w_2 D_{spt}(i, j) + w_3 D_{bou}(i, j) \quad (6)$$

where $D_{geo}(i, j)$ is the geometric deviation between r_i and r_j and is computed as the normalized distance from r_j and p_i defined in Eq. 7; $D_{spt}(i, j)$ is the normal deviation, which measures the normal similarity between these two points, as shown in Eq. 8; $D_{bou}(i, j)$ penalizes supervoxels containing too many boundary points with varying normals and is computed as the ratio between the number of boundary points and the total number of points in these two supervoxels, which is proposed to control the segmentation conform to the structure boundaries and ensure that points in the same supervoxel cannot cross shape structures; and w_1 , w_2 and w_3 are corresponding scalar weights, which are introduced in Section IV-A.3.

$$D_{geo}(i, j) = -\ln\left(\frac{1}{\sqrt{2\pi}R} \exp\left(-\left(\frac{\|r_j - r_i\|^2}{2R^2}\right)\right)\right) \quad (7)$$

$$D_{spt}(i, j) = 1 - (n_j \cdot n_i)^2 \quad (8)$$

$$D_{bou}(i, j) = \frac{size(s_i \cup s_j \cap BP)}{size(s_i) + size(s_j)} \quad (9)$$

where n_i and n_j represent the normals of r_i and r_j , respectively, and $size(\cdot)$ is the function for computing the number of points in a set.

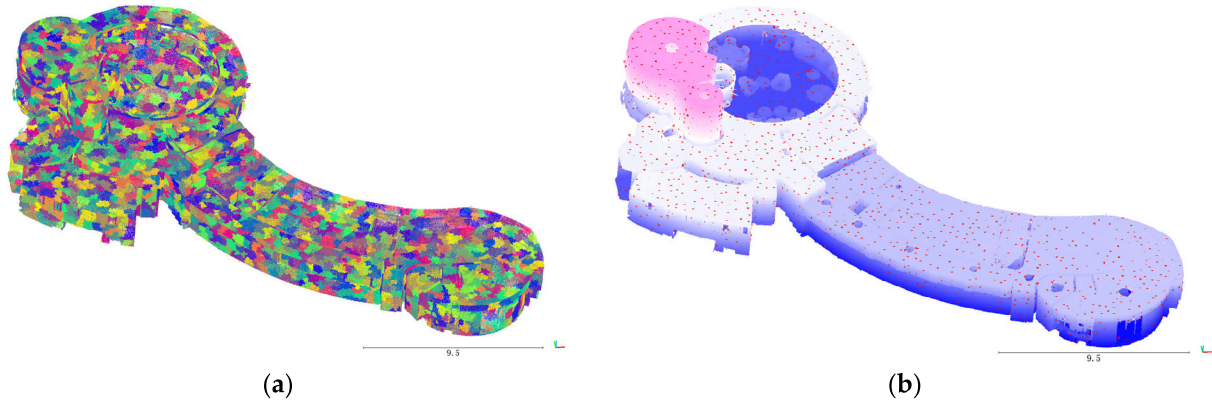


FIGURE 4. Initial supervoxel generation results: (a) each supervoxel is colored according to its label, and (b) the RPs are colored red.

D. BOUNDARY REFINEMENT

As described in the previous section, the indoor scene is segmented into various supervoxels with noncompact shape patterns, as shown in Fig. 1c, since the energy descent algorithm does not consider the spatial coherence among supervoxels. Therefore, it is necessary to apply boundary enhancement and refinement to smooth the supervoxel boundaries while preserving sharp indoor structures. For each point $p_i \in NP$, we extract its RP r_i as shown in Algorithm 2.

Algorithm 2 Structure Boundary Enhancement

```

1 Input:  $S$ ,  $RP$  and  $NP$ 
2 Initialize:  $Q \leftarrow \emptyset$  3 for each  $p_i \in NP$ 
4   for each neighboring supervoxel  $s_j$  of  $p_i$ 
5     Compute  $D(i, j)$  for each supervoxel  $s_j$ 
6   Find a supervoxel  $s_t$  with minimal  $D$ 
7   Assign  $p_i$  to  $s_t$  8 end for

```

Then, if $p_i \in P$ and its neighboring point p_j satisfy $D(p_i, r_i) > D(p_j, r_j)$, where r_i and r_j are the RPs of p_i and p_j , respectively, the boundary of s_j can be further compacted by assigning p_i to s_j . This process is iterated until no additional such enhancements can be made. The example illustrated in Fig. 1d highlights that exchanging supervoxel boundaries in this way can yield improved boundaries and more regular shapes. Further details are provided in Algorithm 3.

IV. RESULTS

A. EXPERIMENTAL SETUP

To evaluate the feasibility and robustness of the proposed supervoxel segmentation method, we conducted both qualitative and quantitative evaluations on various point cloud datasets, as depicted in Fig. 5a. The algorithm was entirely implemented using Point Cloud Library (PCL) and CloudCompare. We performed all experimental studies on an Intel(R) Core(TM) i7-12700H @ 3.50 GHz processor with 64 GB of RAM.

Algorithm 3 Supervoxel Boundary Exchange

```

1 Input:  $S$  and  $RP$ 
2 Initialize:  $Q \leftarrow \emptyset$  3 for each  $r_i \in RP$ 
4   for each  $p_i \in neighbor(r_i)$ 
5     if  $p_i$  does not belong to  $s_i$ 
6       Add  $p_i$  and  $r_i$  into  $Q$  7 while  $Q \neq \emptyset$ 
8     for each  $p_j \in neighbor(p_i)$ 
9       if  $size(s_{p_i}) * D(s_{p_i}, p_j) > size(s_i) * D(s_i, p_j)$ 
10        Assign  $p_j$  to  $s_i$ 
11        if  $p_i \notin Q$ 
12          Add  $p_i$  into  $Q$ 
13        end if
14      end if 15 end for 16 end while

```

1) DATA SPECIFICATIONS

The performance of the proposed method was evaluated using five benchmark indoor datasets, which were colored and visualized in CloudCompare, as shown in Fig. 6a. The statistics and parameters for these datasets are provided in Tables 1 and 2. Datasets 1, 2 and 3 were obtained from the Matterport3D benchmark [29], [30] and captured using a Matterport3D scanner. Dataset 4 was captured by a Zeb-Revo sensor using an MLS device [31]. Dataset 5 was obtained from the Boston Semantic Interpretation Challenge (Indoor) Benchmark [32] and captured by a common RGBD sensor. Dataset 4 is a TLS dataset and has relatively low accuracy compared to the other four datasets.

2) EVALUATION CRITERIA

The boundary recall (BR) is a commonly used metric for evaluating supervoxel segmentation performance [9], [12]. It is defined as the number of correctly detected boundary points divided by the total number of detected boundary points, as shown in Eq. 10.

$$BR = \frac{|B_T \cap B_G|}{|B_G|} \quad (10)$$

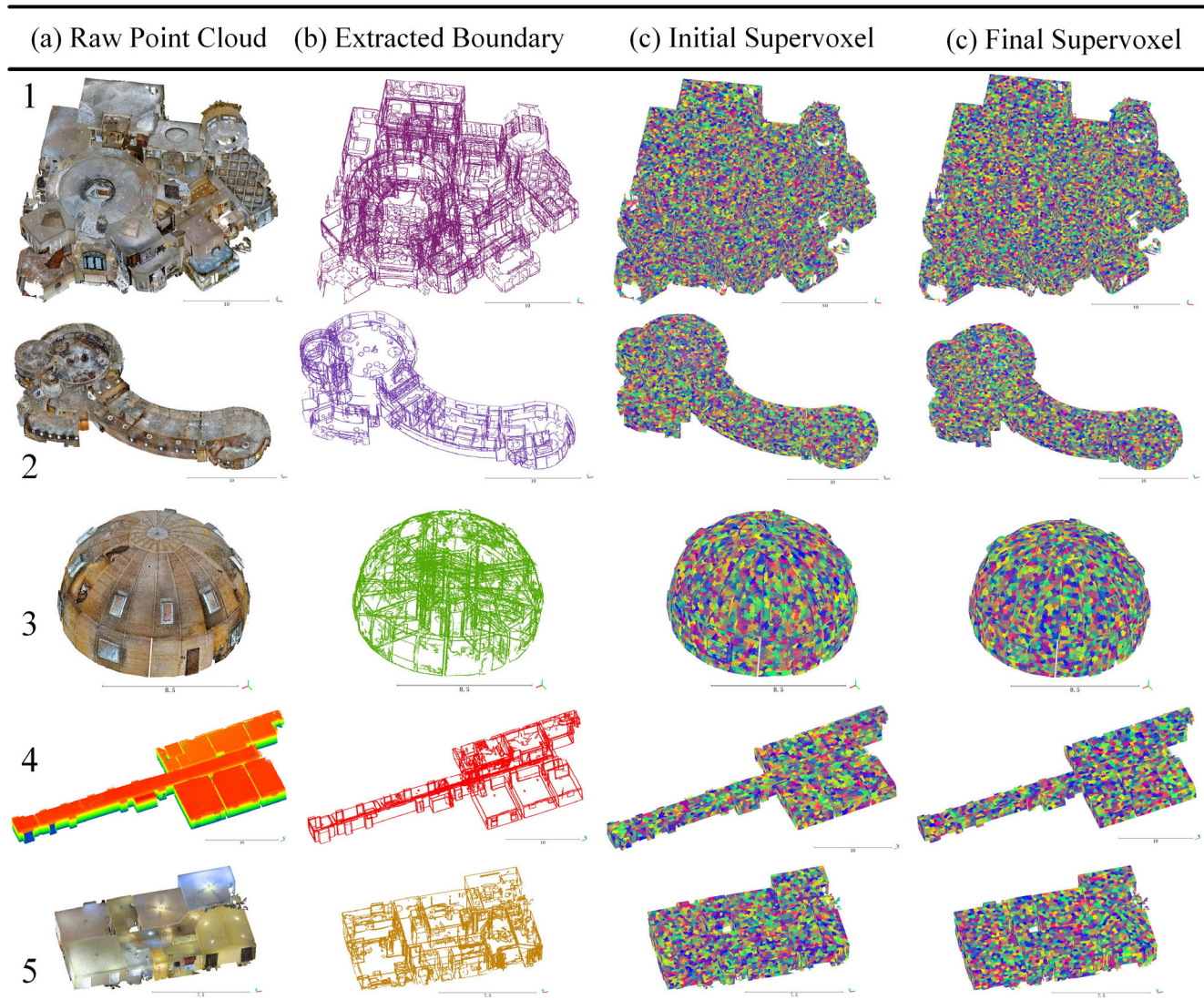


FIGURE 5. Qualitative results of indoor supervoxel segmentation: (a) original data; (b) the extracted boundaries for each point cloud; (c) the generated initial supervoxels, represented by different colors; and (d) the final results after structure refinement.

TABLE 1. Descriptions of the datasets.

Test Site	Points (million)	Area (m ²)	Height (m)	Relative Accuracy	Density (points/m ³)
Dataset 1	10.9	500	12	1–5 cm	1831
Dataset 2	3.1	650	5.6	1–5 cm	863
Dataset 3	4.3	150	7.5	3–15 cm	3792
Dataset 4	4.9	600	2.5	2–3 cm	3270
Dataset 5	4.5	120	3	3–15 cm	13206

where B_T and B_G are the sets of boundary points in the supervoxel segmentation results and the ground truth, respectively, while $|B_T|$ and $|B_G|$ are the numbers of boundary points in the supervoxel segmentation results and the ground truth, respectively. A higher BR indicates better performance in

capturing the boundaries of the indoor structures in the point cloud.

3) PARAMETER SETTINGS

The input parameters for our method are presented in Table 2. Unless otherwise specified, all parameters are applicable for all experiments reported in this paper.

The RANSAC-based boundary detection process relies on four critical parameters, namely, α_1 , α_2 , α_3 , and α_4 , which are determined based on the characteristics of the point cloud, including the scanner accuracy, density, and noise level. α_1 specifies the maximum distance to a compatible point and is typically set to 0.5% of the width of the input point cloud’s bounding box. α_2 restricts the deviation of a point’s normal vector from its fitting model [32] and is typically set to 20°. α_3 denotes the minimum number of points in each segmented shape, which varies based on the density of the input point

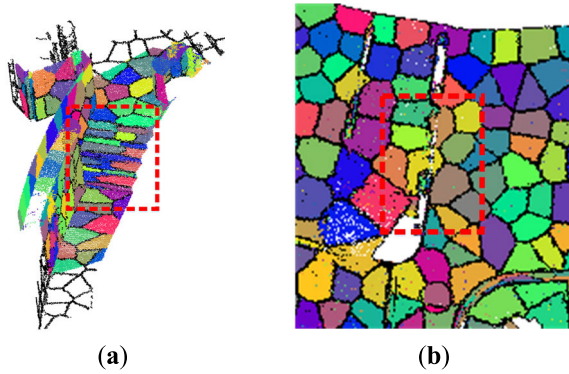


FIGURE 6. In cases where our method failed, the boundaries of the supervoxels are colored in black. (a) Undersegmentation on a set of stairs due to insufficient boundary detection results. (b) Undersegmented cases violating the boundaries of a thin wall.

TABLE 2. Parameters of the proposed method.

PM	Value	Description
α_1	*	maximum point-to-distance threshold
α_2	5–25°	normal angle threshold
α_3	≥ 200	minimum number of supporting points per primitive
α_4	0.001	probability of missing the next largest primitive candidate in each iteration
R	*	desired resolution for each supervoxel
λ	*	median of the lowest dissimilarity distances
w_1	0.4	weight of $D_{geo}(i, j)$, which measures the geometric similarity among supervoxels
w_2	1	weight of $D_{spt}(i, j)$, which measures the normal similarity among supervoxels
w_3	0.3	weight of $D_{bou}(i, j)$, which represents the penalty for boundary splitting/merging among supervoxels

The symbol ‘*’ indicates that the corresponding parameter is determined for each dataset individually.

cloud. Finally, α_4 is the probability of missing a better candidate during sampling and is generally set to 0.001.

The supervoxel generation process depends on two parameters, R and λ . R specifies the desired resolution for each supervoxel and is used to determine the expected number of supervoxels, as elaborated in Section IV-D. On the other hand, λ is a regularization parameter that determines whether two supervoxels should be merged into one, and it can be set to the median of the lowest dissimilarity distances between each point and its neighboring points.

The dissimilarity distance computation algorithm involves three important parameters, w_1 , w_2 and w_3 . These weights control the similarity among supervoxels and are associated with the geometric similarity, normal similarity and boundary penalty, respectively. The default weights are $w_1 = 0.4$, $w_2 = 1$, and $w_3 = 0.3$.

B. QUALITATIVE EVALUATION

The qualitative evaluation of the results of indoor supervoxel segmentation on these datasets is presented in Fig. 5. Fig. 5a displays the five raw point clouds, where the color of each point is determined by its RGB value. Fig. 5b presents the extracted boundaries in the selected point clouds. Fig. 5c illustrates the initial partitioned supervoxels with noncompact shapes, where each supervoxel is represented by a different color. Fig. 5d shows the final results after boundary refinement.

For structural planes, most supervoxels conform to the structure boundaries because of the boundary penalty. When a supervoxel crosses a boundary, causing the number of boundary points it contains to increase, the boundary penalty increases significantly. For curved areas, most supervoxels have compact shape patterns due to the boundary refinement processing; examples include the cylindrical walls in Dataset 1, the long curved walls of the long corridor in Dataset 2, and the dome in Dataset 3. However, the proposed method failed to detect relatively small-scale regions with small numbers of points due to insufficient boundary detection results, as exemplified by the area in the red box in Fig. 6a, which depicts a few supervoxels that cross the boundaries of a set of stairs. Undersegmentation mainly occurred in areas where the structures were not clearly distinguished. Despite the structure boundary splitting/merging penalty, as shown in the case of the thin wall in Fig. 6b, some supervoxels still cross boundaries due to their geometric similarity.

In general, most supervoxels adhere closely to the indoor structure boundaries and exhibit compact shapes, as illustrated in Fig. 5d, with the exception of the dome surface in Dataset 3. The experimental results indicate that the proposed method achieves outstanding performance in the segmentation of both a TLS point cloud (Dataset 4) and RGBD point clouds (Datasets 1, 2 and 3 and 5).

C. PERFORMANCE COMPARISON

We further discuss the effectiveness of our method in comparison with other state-of-the-art segmentation algorithms on the same datasets. The results are shown in Table 4 and Figs. 7–10. VCCS [12] is the most popular method at present and has proven to be useful for both indoor and outdoor scenes. KNNVCCS is a modified version of VCCS in which VCCS’s voxel-based neighborhood is replaced with K-nearest neighbor points. KBBS [9] also formulates the supervoxel segmentation problem as a subset selection problem, making it the most similar to our method. Accordingly, these advanced methods of supervoxel segmentation are selected for performance evaluation. The key parameters of all these methods were set as suggested in their original papers. Figs. 7–9 visualize the results of these four methods on three selected datasets (Datasets 2, 3, and 5) for comparison. Fig. 10 shows the corresponding metrics on the selected datasets.

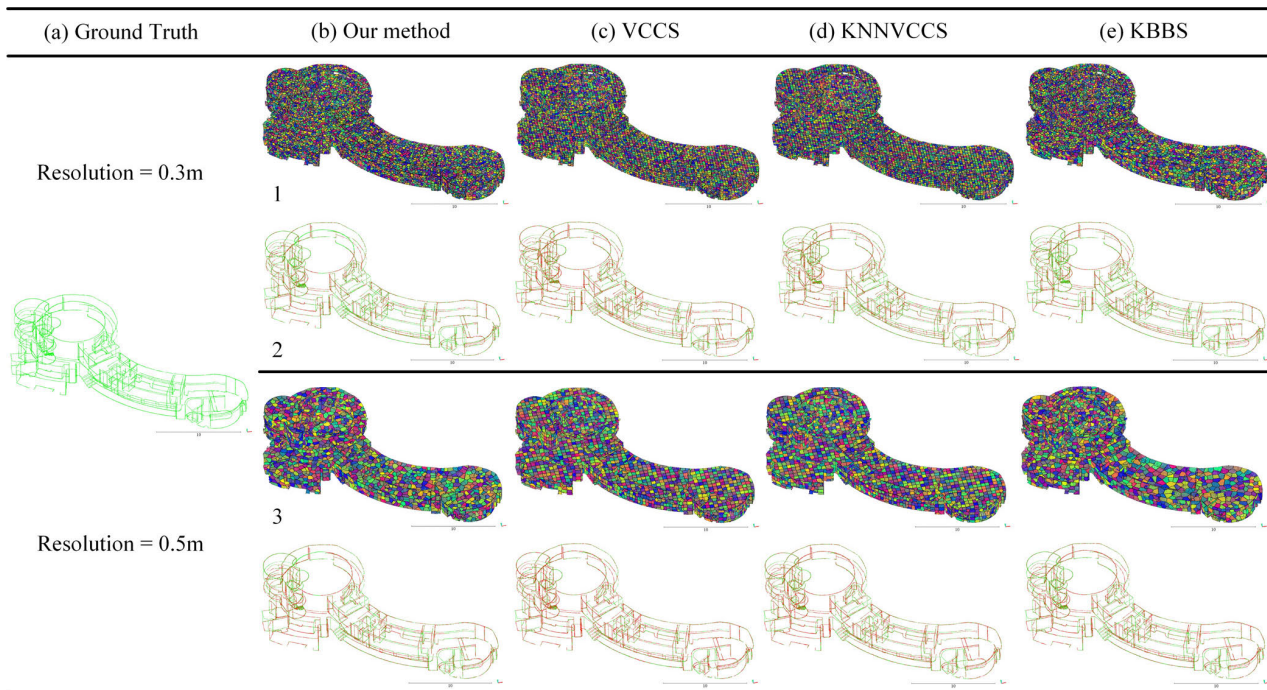


FIGURE 7. Supervoxel segmentation results on Dataset 2, where each supervoxel is colored differently and the supervoxel boundaries are displayed in black. The resolution R of a supervoxel is 0.3 m in rows 1 and 2, while it is 0.5 m in rows 3 and 4. The differences between the boundaries of the supervoxels and the ground-truth boundaries are highlighted in rows 2 and 4, where the detected true boundaries and incorrect boundaries are depicted as green and red lines, respectively.

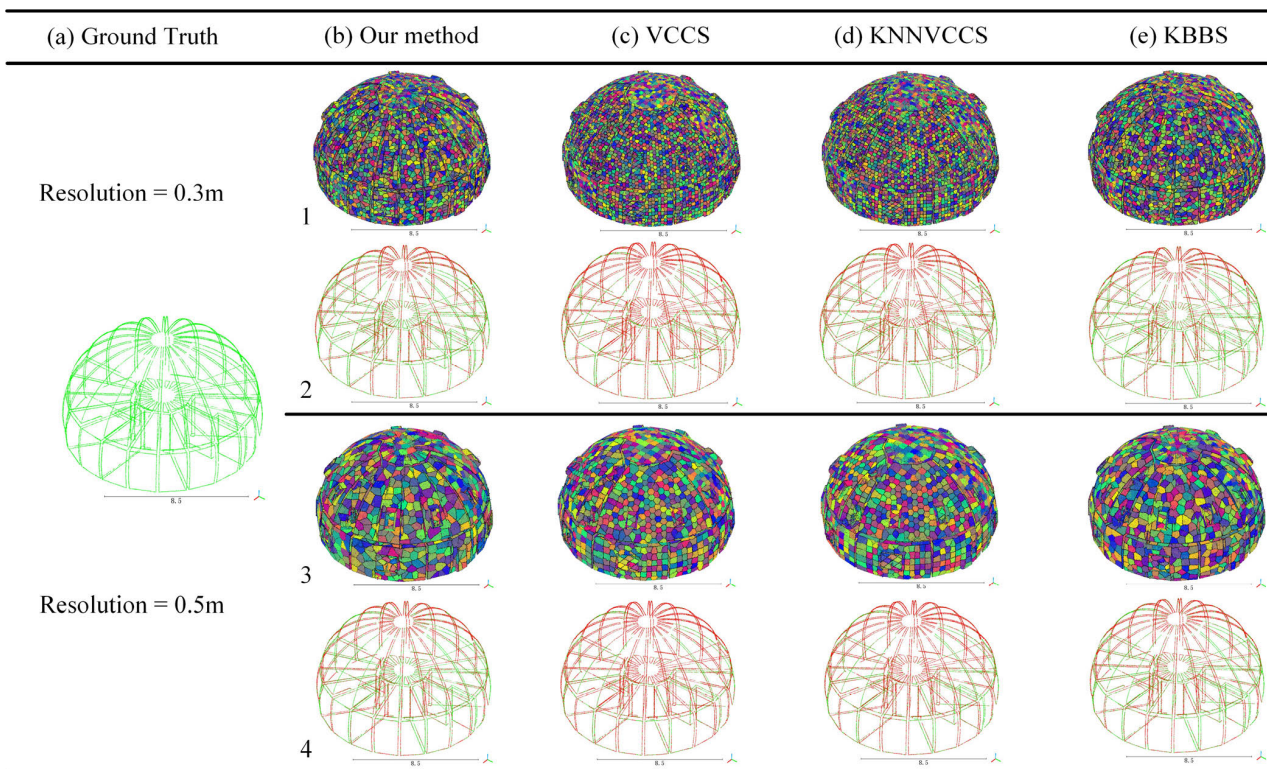


FIGURE 8. Supervoxel segmentation results on Dataset 3, where each supervoxel is colored differently and the supervoxel boundaries are displayed in black. The resolution R of a supervoxel is 0.3 m in rows 1 and 2, while it is 0.5 m in rows 3 and 4. The differences between the boundaries of the supervoxels and the ground-truth boundaries are highlighted in rows 2 and 4, where the detected true boundaries and incorrect boundaries are depicted as green and red lines, respectively.

The VCCS-based algorithms obtained relatively lower BR values than KBBS and our method on the three datasets. This

lower performance can be attributed to the fact that VCCS segments indoor scenes with a fixed resolution, making it

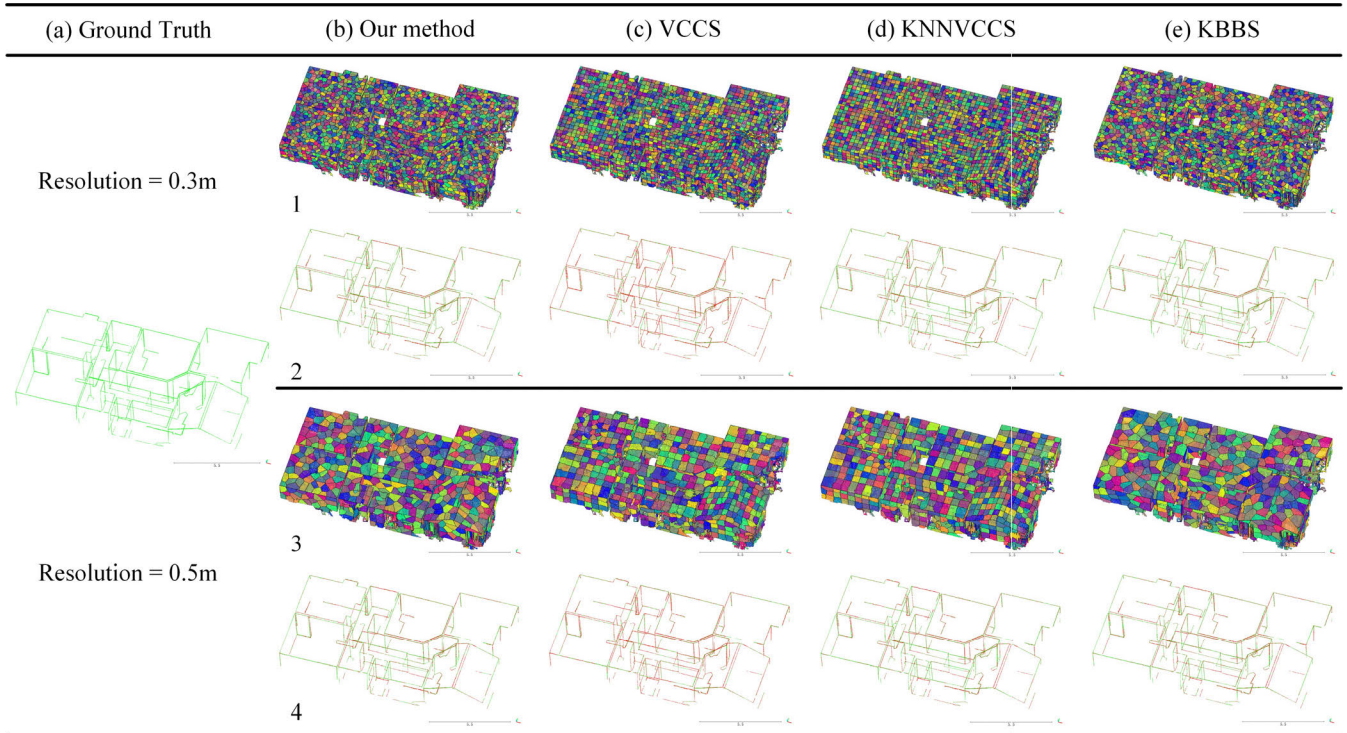


FIGURE 9. Supervoxel segmentation results on Dataset 5, where each supervoxel is colored differently and the superpixel boundaries are displayed in black. The resolution R of a supervoxel is 0.3 m in rows 1 and 2, while it is 0.5 m in rows 3 and 4. The differences between the boundaries of the supervoxels and the ground-truth boundaries are highlighted in rows 2 and 4, where the detected true boundaries and incorrect boundaries are depicted as green and red lines, respectively.

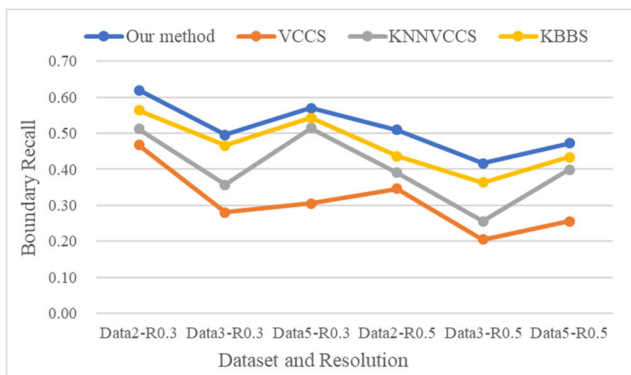


FIGURE 10. Quantitative evaluation of the four methods on Datasets 2, 3 and 5.

less robust to varying densities [9] and high levels of noise, resulting in missing structural details. Supervoxel segmentation may be better formulated as a subset (or RP) selection problem to balance the geometric fitting errors and spatial coherence among supervoxels, which significantly improves performance.

The proposed method also achieved better performance than KBBS, especially for curved structures with small deviations in curvature, as shown by the long corridor in Dataset 2 and the dome in Dataset 3. This is because KBBS randomly selects RPs based on geometric deviation, which significantly

reduces its BR on Datasets 2 and 3. Our method uses the extracted boundaries as constraints and introduces a boundary penalty to select splitting/merging operations for supervoxels in the boundary areas, which in turn improves the performance of indoor supervoxel segmentation.

D. PERFORMANCE COMPARISON

To investigate the effect of the resolution R on supervoxel segmentation, we conducted additional experiments with varying resolutions using our method. The number of generated supervoxels increases with decreasing resolution, as shown in Figs. 11 and 12. The run times in this experiment are shown in Table S1 in the supplementary file. Notably, even when the supervoxel resolution is larger than the width of certain structures, our method can still accurately preserve the boundaries, as illustrated by the regions marked with black boxes in Figure 11. The experimental results demonstrate that the optimal resolution R for supervoxel segmentation depends on the density and accuracy of the point cloud data. Specifically, for input data with high density and accuracy and few occluded areas, as in Dataset 4, $R=0.5$ achieves the best performance. On the other hand, for point clouds with low density or accuracy, such as RGBD point clouds (e.g., Datasets 1, 2, 3 and 5), a lower resolution of $R=0.3$ yields better results. Therefore, the appropriate choice of resolution should be determined in accordance with the characteristics

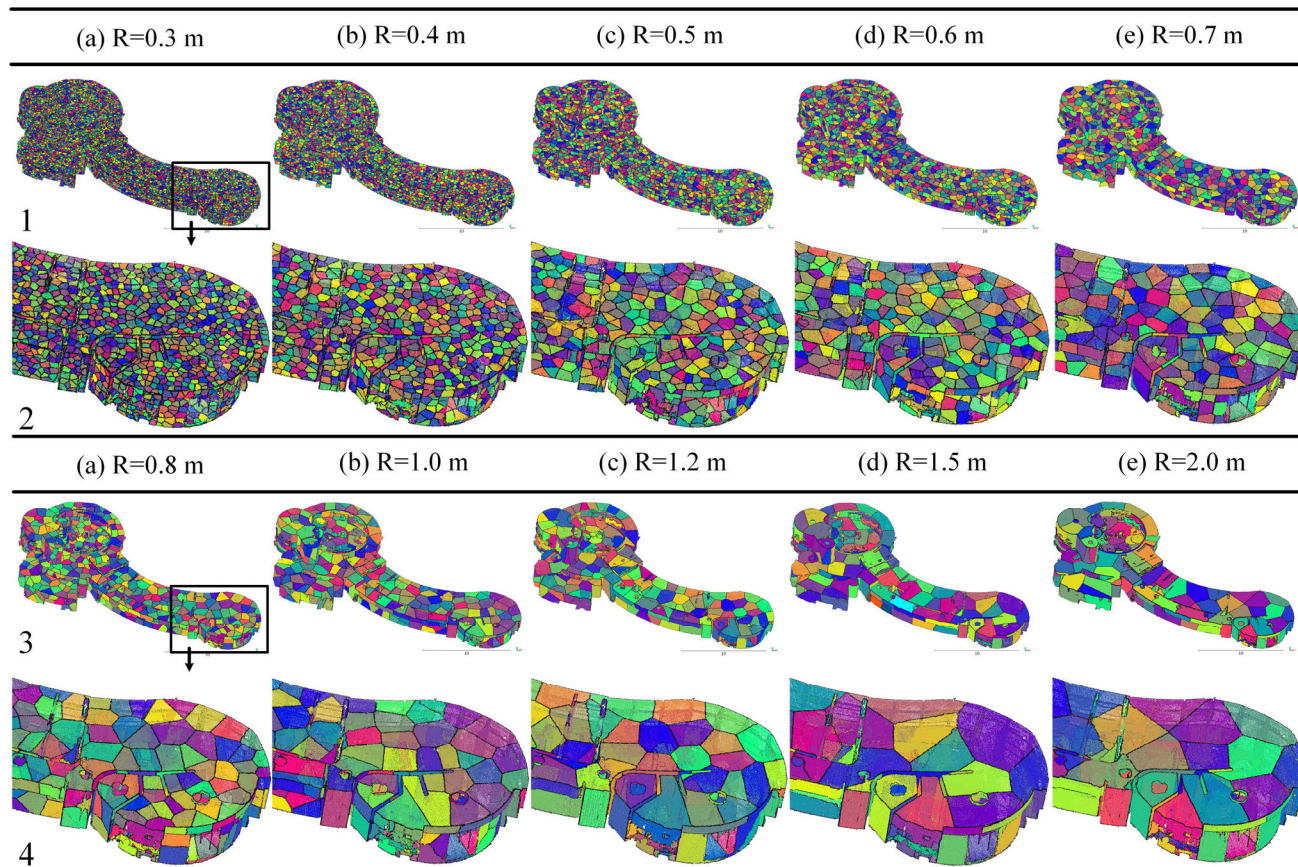


FIGURE 11. Supervoxel segmentation results on Dataset 2 with different resolutions, where each supervoxel is colored differently and the boundaries are colored black. The figures in rows 2 and 4 are magnifications of the point clouds corresponding to the black boxes in rows 1 and 3, respectively.

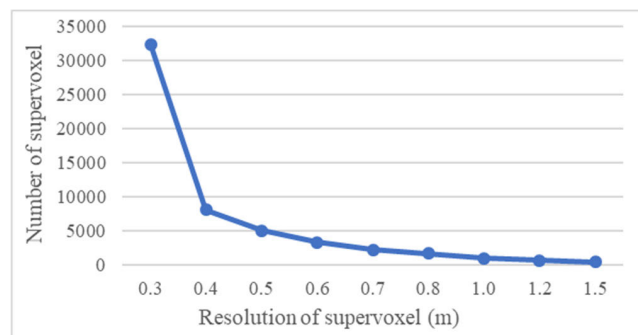


FIGURE 12. Numbers of supervoxels with different resolutions in Dataset 2.

of the input data to ensure the best possible performance of the proposed method.

V. CONCLUSION

Existing methods for the segmentation of supervoxels from indoor point clouds rely on geometric dissimilarity to distinguish the structure boundaries and show obvious defects in the intersection regions of curved structures. To address this issue, we propose a boundary-aware supervoxel segmentation method and evaluate its

performance on five point clouds. Our comprehensive experiments demonstrate that our approach achieves superior performance compared to other state-of-the-art methods across various types of point clouds.

Despite the promising results, our approach also has some limitations. For instance, it may fail in relatively small-scale regions due to insufficient boundary extraction, as illustrated by the region marked with the red box in Fig. 6a. In addition, our method still produces some noncompact supervoxel shapes, as observed in the dome area in Dataset 3. The supervoxel segmentation of domes and orbicular structures will be further considered in future work.

ACKNOWLEDGMENT

The authors would like to gratefully acknowledge Dr. Lin Yangbin for their assistant in code acquisition.

REFERENCES

- [1] J. Li and X. Cheng, "Supervoxel-based extraction and classification of pole-like objects from MLS point cloud data," *Opt. Laser Technol.*, vol. 146, Feb. 2022, Art. no. 107562, doi: 10.1016/j.optlastec.2021.107562.
- [2] F. Su, H. Zhu, T. Chen, L. Li, F. Yang, H. Peng, L. Tang, X. Zuo, Y. Liang, and S. Ying, "An anchor-based graph method for detecting and classifying indoor objects from cluttered 3D point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 172, pp. 114–131, Feb. 2021, doi: 10.1016/j.isprsjrs.2020.12.007.

- [3] Z. Dong, B. Yang, P. Hu, and S. Scherer, "An efficient global energy optimization approach for robust 3D plane segmentation of point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 137, pp. 112–133, Mar. 2018, doi: [10.1016/j.isprsjprs.2018.01.013](https://doi.org/10.1016/j.isprsjprs.2018.01.013).
- [4] X. Chen, H. Wu, D. Lichti, X. Han, Y. Ban, P. Li, and H. Deng, "Extraction of indoor objects based on the exponential function density clustering model," *Inf. Sci.*, vol. 607, pp. 1111–1135, Aug. 2022.
- [5] Z. Kang and J. Yang, "A probabilistic graphical model for the classification of mobile LiDAR point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 143, pp. 108–123, Sep. 2018, doi: [10.1016/j.isprsjprs.2018.04.018](https://doi.org/10.1016/j.isprsjprs.2018.04.018).
- [6] F. Su, H. Zhu, L. Li, G. Zhou, W. Rong, X. Zuo, W. Li, X. Wu, W. Wang, F. Yang, H. Hu, and S. Ying, "Indoor interior segmentation with curved surfaces via global energy optimization," *Autom. Construct.*, vol. 131, Nov. 2021, Art. no. 103886, doi: [10.1016/j.autcon.2021.103886](https://doi.org/10.1016/j.autcon.2021.103886).
- [7] Y. Lin, C. Wang, B. Chen, D. Zai, and J. Li, "Facet segmentation-based line segment extraction for large-scale point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4839–4854, Sep. 2017, doi: [10.1109/TGRS.2016.2639025](https://doi.org/10.1109/TGRS.2016.2639025).
- [8] D. Zai, J. Li, Y. Guo, M. Cheng, Y. Lin, H. Luo, and C. Wang, "3-D road boundary extraction from mobile laser scanning data via supervoxels and graph cuts," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 802–813, Mar. 2018, doi: [10.1109/TITS.2017.2701403](https://doi.org/10.1109/TITS.2017.2701403).
- [9] Y. Lin, C. Wang, D. Zhai, W. Li, and J. Li, "Toward better boundary preserved supervoxel segmentation for 3D point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 143, pp. 39–47, Sep. 2018, doi: [10.1016/j.isprsjprs.2018.05.004](https://doi.org/10.1016/j.isprsjprs.2018.05.004).
- [10] R. Yi, Z. Ye, W. Zhao, M. Yu, Y. Lai, and Y. Liu, "Feature-aware uniform tessellations on video manifold for content-sensitive supervoxels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3183–3195, Sep. 2021, doi: [10.1109/TPAMI.2020.2979714](https://doi.org/10.1109/TPAMI.2020.2979714).
- [11] L. Li, F. Su, F. Yang, H. Zhu, D. Li, X. Zuo, F. Li, Y. Liu, and S. Ying, "Reconstruction of three-dimensional (3D) indoor interiors with multiple stories via comprehensive segmentation," *Remote Sens.*, vol. 10, no. 8, p. 1281, Aug. 2018, doi: [10.3390/rs10081281](https://doi.org/10.3390/rs10081281).
- [12] J. Papon, A. Abramov, M. Schoeler, and F. Worgotter, "Voxel cloud connectivity segmentation—Supervoxels for point clouds," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2027–2034, doi: [10.1109/cvpr.2013.264](https://doi.org/10.1109/cvpr.2013.264).
- [13] Z. Sha, Y. Chen, Y. Lin, C. Wang, J. Marcato, and J. Li, "A supervoxel approach to road boundary enhancement from 3-D LiDAR point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6500405, doi: [10.1109/LGRS.2020.3037484](https://doi.org/10.1109/LGRS.2020.3037484).
- [14] Y. Xiao, Z. Chen, Z. Lin, J. Cao, Y. J. Zhang, Y. Lin, and C. Wang, "Merge-swap optimization framework for supervoxel generation from three-dimensional point clouds," *Remote Sens.*, vol. 12, no. 3, p. 473, Feb. 2020, doi: [10.3390/rs12030473](https://doi.org/10.3390/rs12030473).
- [15] S. Song, H. Lee, and S. Jo, "Boundary-enhanced supervoxel segmentation for sparse outdoor LiDAR data," *Electron. Lett.*, vol. 50, no. 25, pp. 1917–1919, Dec. 2014, doi: [10.1049/el.2014.3249](https://doi.org/10.1049/el.2014.3249).
- [16] H. Zhu, F. Meng, J. Cai, and S. Lu, "Beyond pixels: A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 12–27, Jan. 2016, doi: [10.1016/j.jvcir.2015.10.012](https://doi.org/10.1016/j.jvcir.2015.10.012).
- [17] Y. Cai and X. Guo, "Anisotropic superpixel generation based on Mahalanobis distance," *Comput. Graph. Forum*, vol. 35, no. 7, pp. 199–207, Oct. 2016, doi: [10.1111/cgf.13017](https://doi.org/10.1111/cgf.13017).
- [18] H. Fu, X. Cao, D. Tang, Y. Han, and D. Xu, "Regularity preserved superpixels and supervoxels," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 1165–1175, Jun. 2014, doi: [10.1109/TMM.2014.2305571](https://doi.org/10.1109/TMM.2014.2305571).
- [19] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [20] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art," *Comput. Vis. Image Understand.*, vol. 166, pp. 1–27, Jan. 2018, doi: [10.1016/j.cviu.2017.03.007](https://doi.org/10.1016/j.cviu.2017.03.007).
- [21] O. Veksler, Y. Boykov, and P. Mehrani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. European Conf. Comput. Vis. (ECCV)*, 2010, pp. 211–224.
- [22] L. Zhang, S. Lu, C. Hu, D. Xiang, T. Liu, and Y. Su, "Superpixel generation for SAR imagery based on fast DBSCAN clustering with edge penalty," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 804–819, 2022, doi: [10.1109/jstars.2021.3131187](https://doi.org/10.1109/jstars.2021.3131187).
- [23] C. Xu and J. J. Corso, "LIBSVX: A supervoxel library and benchmark for early video processing," *Int. J. Comput. Vis.*, vol. 119, no. 3, pp. 272–290, Sep. 2016, doi: [10.1007/s11263-016-0906-5](https://doi.org/10.1007/s11263-016-0906-5).
- [24] Y. Liang, J. Shen, X. Dong, H. Sun, and X. Li, "Video supervoxels using partially absorbing random walks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 5, pp. 928–938, May 2016, doi: [10.1109/TCSVT.2015.2406232](https://doi.org/10.1109/TCSVT.2015.2406232).
- [25] Z. Ban, Z. Chen, and J. Liu, "Supervoxel segmentation with voxel-related Gaussian mixture model," *Sensors*, vol. 18, no. 1, p. 128, Jan. 2018, doi: [10.3390/s18010128](https://doi.org/10.3390/s18010128).
- [26] H. Ni and X. Niu, "SVLA: A compact supervoxel segmentation method based on local allocation," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 300–311, May 2020, doi: [10.1016/j.isprsjprs.2020.03.011](https://doi.org/10.1016/j.isprsjprs.2020.03.011).
- [27] R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for point-cloud shape detection," *Comput. Graph. Forum*, vol. 26, no. 2, pp. 214–226, Jun. 2007.
- [28] X. Lu, Y. Liu, and K. Li, "Fast 3D line segment detection from unorganized point cloud," 2019, *arXiv:1901.02532*.
- [29] I. Armeni, S. Sax, A. R. Zamir, and S. Savarese, "Joint 2D-3D-semantic data for indoor scene understanding," 2017, *arXiv:1702.01105*.
- [30] E. Valero, A. Adan, and F. Bosché, "Semantic 3D reconstruction of furnished interiors using laser scanning and RFID technology," *J. Comput. Civil Eng.*, vol. 30, no. 4, 2016, Art. no. 04015053, doi: [10.1061/\(ASCE\)CP.1943-5487.0000525](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000525).
- [31] K. Khoshelham, L. Díaz Vilarinho, M. Peter, Z. Kang, and D. Acharya, "The ISPRS benchmark on indoor modelling," *Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci.*, vol. 42, pp. 367–372, Sep. 2017, doi: [10.5194/isprs-archives-XLII-2-W7-367-2017](https://doi.org/10.5194/isprs-archives-XLII-2-W7-367-2017).
- [32] R. Ambrus, S. Claiici, and A. Wendt, "Automatic room segmentation from unstructured 3-D data of indoor environments," *IEEE Robot. Autom. Lett.*, vol. 2, no. 2, pp. 749–756, Apr. 2017, doi: [10.1109/lra.2017.2651939](https://doi.org/10.1109/lra.2017.2651939).



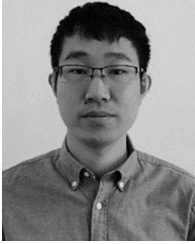
FEI SU received the B.S. degree from Shandong Agricultural University, Taian, China, in 2013, and the Ph.D. degree from Wuhan University, China, in 2020. He is currently with the School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan, China. His current research interests include geographic information science, point cloud processing, and 3-D indoor segmentation and reconstruction.



YU LIU is an Associate Professor with the Institute of Environment and Development, Guangdong Academy of Social Sciences. His research interests include point cloud, spatio-temporal big data analysis, and geographic information mining.



KAIXIAO NIE received the bachelor's degree from the School of Veterinary Medicine, Northwest Agriculture and Forestry University, and the Ph.D. degree from Tsinghua University, China, in 2020. She is currently with the Department of Pathogen Biology, School of Basic Medicine, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China. Her current research interests include simultaneous localization and mapping (SLAM) and point cloud processing.



YAOHUI LIU received the Ph.D. degree in quaternary geology from the Institute of Geology, China Earthquake Administration, China, in 2020. He is currently a Lecturer with Shandong Jianzhu University. His research interests include computer vision, remote sensing, deep learning, and risk management.



RAN ZHANG received the bachelor's degree in surveying and mapping engineering from the Heilongjiang University of Science and Technology, Harbin, China, in 2021. She is currently pursuing the master's degree with Shandong Jianzhu University. Her research interests include geographic information science, point cloud processing, and contour extraction.



JINGXUE BI received the M.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2015, and the Ph.D. degree from the China University of Mining and Technology (CMUT), Xuzhou, China, in 2019. He is currently an Associate Professor with Shandong Jianzhu University. His current research interests include indoor positioning and human activity recognition.



GUOQIANG ZHENG received the B.S. degree from Northeast Forestry University, Harbin, China, in 2001, and the Ph.D. degree from Nanjing University, Nanjing, China, in 2004. He is currently with the School of Surveying and Geo-Informatics, Shandong Jianzhu University, Jinan, China. His current research interests include remote sensing image processing, image classification, point cloud processing, and 3-D reconstruction.

• • •