remote sensing

MDPI

Article

IAMS (Interior-Anchored Mean-Shift) Algorithm for Supervoxel Segmentation of Airborne LiDAR Roof Points

Hanyu Zhou [1,2], Liang Zhang [1,2,*], Zhiyue Zhang [1,2], Haiqiong Yang [3], Xiongfei Tang [3], Hongchao Ma [4] and Chunjing Yao [4]

[1] Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China; zhouhanyu@stu.hubu.edu.cn (H.Z.); zhangzhiyue@stu.hubu.edu.cn (Z.Z.)
[2] Hubei Key Laboratory of Regional Development and Environmental Response, Hubei University, Wuhan 430062, China
[3] Yunnan Remote Sensing Center, Kunming 650034, China; ynhcyg@163.com (H.Y.); txiongfei@163.com (X.T.)
[4] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China; hchma@whu.edu.cn (H.M.); yaocj@whu.edu.cn (C.Y.)
* Correspondence: zhangliang@hubu.edu.cn

Highlights

What are the main findings?

- This study introduces an Interior-Anchored Mean-Shift (IAMS) algorithm—a paradigm shift in supervoxel initialization that anchors seeds reliably within object interiors, eliminating cross-boundary segmentation errors at their source.
- By unifying geometric consistency, point density, and spatial correlation into a single interior-awareness criterion, IAMS generates highly homogeneous supervoxels without heuristic thresholds or post-refinement.

What are the implications of the main findings?

- This interior-anchored seeding optimization directly translates to state-of-the-art roof classification accuracy on airborne LiDAR, significantly outperforming VCCS and other baselines across diverse urban scenes.
- The approach demonstrates exceptional robustness under real-world challenges—including sparse sampling and heavy vegetation occlusion—enabling more reliable building extraction for urban remote sensing applications.

Abstract

Accurate building roof classification from airborne LiDAR point clouds is fundamental to reliable three-dimensional (3D) urban reconstruction. While supervoxel-based methods offer efficiency and resilience to uneven point density, their performance is critically undermined by cross-boundary segmentation errors—a direct consequence of random seed initialization that merges geometrically similar yet semantically distinct objects. To address this root cause, this study proposes Interior-Anchored Mean-Shift (IAMS), a novel supervoxel segmentation framework that rethinks seed placement as a geometry-aware interior localization problem. By integrating local geometric consistency point density, and spatial correlation into a unified kernel density estimator, supplemented by density-adaptive voxel weighting and a semi-variogram-driven bandwidth, IAMS reliably anchors seeds within object interiors, yielding highly homogeneous supervoxels without post-processing. Extensive experiments on three diverse airborne LiDAR datasets demonstrated that IAMS consistently outperformed state-of-the-art baselines. On the International Society for Photogrammetry and Remote Sensing (ISPRS) Vaihingen benchmark, our approach improved roof classification completeness, correctness, and quality by up to 7.1% (per-object) over the

Check for updates

Academic Editor: Massimiliano Pepe

Received: 11 February 2026
Revised: 12 March 2026
Accepted: 18 March 2026
Published: 23 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

conventional Voxel Cloud Connectivity Segmentation (VCCS) algorithm while being significantly faster than recent boundary-preserving alternatives. Critically, IAMS maintains robust performance under challenging conditions, including sparse sampling and dense vegetation occlusion, making it a practical solution for real-world urban remote sensing.

Keywords: LiDAR; building roof classification; Mean-Shift; supervoxel segmentation; adaptive bandwidth; kernel density estimation

1. Introduction

Three-dimensional (3D) digital building models provide valuable source data for the data infrastructure in urban planning and residential environments [1]. Over the decades, airborne light detection and ranging (LiDAR) technology has become one of the most efficient techniques for acquiring detailed, accurate 3D measurements of buildings [2]. With this technique, massive laser point clouds reflected from building roofs and facades can be collected with high accuracy and density. Therefore, it considerably simplifies the data acquisition process [3].

However, the 3D building reconstruction from airborne point clouds remains a challenge, where effective roof point labelling is an imperative and critical preprocessing step [4]. From a pattern recognition perspective, this constitutes a classification problem. Consequently, the classification of building components (e.g., roofs and facades) from LiDAR point clouds has been widely addressed in the literature through machine learning [5–7] and deep learning [4,8–10] approaches. While the development of novel deep learning approaches for point cloud classification is valuable, the manual semantic labeling of large training datasets remains time-consuming and labor-intensive. Furthermore, such models typically suffer from limited generalizability across different urban environments or point cloud datasets [11]. Machine learning classifiers, such as Random Forest (RF), can rapidly and accurately assign semantic labels to point clouds using only a small training set [7,12]. Moreover, extracting the discriminative features of ground objects from noisy, sparse, and unstructured LiDAR data is a key step in achieving accurate classification results [11,13]. Conventional features for point cloud classification are primarily geometry-based and are typically computed from a local 3D neighborhood (e.g., points within a 3D sphere or k-nearest neighbors). Hence, determining the optimal neighborhood size for geometry-based feature generation—and thereby improving classification accuracy—poses a challenge for both the point-cloud-processing and machine learning communities, especially when classifying urban point clouds with non-uniform densities.

In recent years, supervoxel-based classification methods have been widely used for 3D point cloud classification. Supervoxel is a 3D analogue of superpixel in 2D image processing [14,15]. It not only significantly enhances data processing efficiency but also provides a natural and compact representation of point clouds. By simplifying the data, it enables downstream operations, such as clustering, to be executed voxel-wise rather than point-wise [16]. Recent literature shows that supervoxel-based classifications can effectively address the challenges posed by non-uniform point densities and reduce computational complexity [11,17,18]. To ensure classification accuracy, supervoxels must avoid violating object boundaries; otherwise, they may contain more than one class [14]. However, due to the random seed initialization and the fixed resolutions (e.g., seed and voxel resolutions), seed voxels tend to merge geometrically similar but semantically distinct voxels, leading to cross-boundary segmentation errors. Such errors may propagate to the subsequent classification processes. To preserve object boundaries, several effective algorithms

have been proposed, including Boundary-Enhanced Supervoxel Segmentation (BESS) [19], Boundary Preserved Supervoxel Segmentation (BPSS) [16], VCCS-kNN [20], and subset selection methods [21,22]. Most of them utilize individual points as processing units, which significantly increases the computational complexity.

In light of the issues outlined above, this paper proposes a robust and effective roof point classification approach based on supervoxel segmentation. By refining the Mean-Shift framework to explicitly anchor seeds within the interior regions of individual roof objects, the method generates more homogeneous supervoxels and effectively suppresses cross-boundary segmentation errors at their source, thereby improving downstream classification accuracy. To ensure reliable interior localization of seeds, three key refinement strategies are introduced:

(1) The kernel density estimation in Mean-Shift is enhanced by jointly incorporating voxel-wise point density and geometric features derived from Fast Point Feature Histograms (FPFH). This joint formulation restricts seed voxel selection to those exhibiting both the highest local point density and strong geometric consistency within the search radius, thereby improving the fidelity of interior region localization.
(2) Each voxel is assigned a weight defined as the ratio of its internal point count to the total number of points within the search neighborhood. This density-proportional weighting preserves fine-scale spatial variations in LiDAR point distributions and effectively guides seed voxels toward true interior density maxima.
(3) The spatial correlation among voxels is quantified using the semi-variogram, and the bandwidth of the Mean-Shift kernel is dynamically adjusted according to this measure. This spatial-correlation-aware bandwidth adaptation enhances the robustness of seed voxel optimization across heterogeneous urban structures—particularly under sparse sampling or complex occlusion.

Collectively, these three mutually reinforcing refinements enable the reliable anchoring of seed voxels within object interiors, and the generated supervoxels thus exhibit higher homogeneity and stronger boundary adherence—an improvement that directly translates into enhanced accuracy for building roof classification. In stark contrast to existing adaptive seeding algorithms such as Saliency-guided Supervoxels (SSV) [23], which lack the integration of spatial correlation in their modeling framework, Interior-Anchored Mean-Shift (IAMS) uniquely achieves the unified fusion of geometric consistency, point density and spatial correlation features for supervoxel initialization, addressing the core limitations of prior methods.

The remaining sections of this paper are organized as follows. In Section 2, relevant research by other scholars is presented. In Section 3, the classification approach proposed in this paper is detailed. Section 4 describes the experiments that were carried out according to the approach, the experimental data, the final results, and the evaluation of the accuracy. Finally, Section 5 presents the conclusions.

2. Related Works

Urban scenes contain numerous complex objects, including buildings, high vegetation, low vegetation, overpasses, streetlights, vehicles, and pedestrians, posing considerable challenges in accurately classifying and extracting building roofs. To improve classification accuracy, many scholars employed supervised classifiers, such as Random Forest (RF) [24], Artificial Neural Networks (ANNs) [25], Support Vector Machines (SVMs) [26], PointNet++ [10], Convolutional Neural Networks (CNNs) [8], and RandLa-Net [4]. These supervised classifiers have demonstrated strong performance in point cloud classification tasks. Furthermore, some researchers explored the multimodal fusion of 2D texture features with 3D geometric features derived from LiDAR data. For example, Du et al. [27] calculated

the gray-level co-occurrence matrix (GLCM) of the nDSM extracted from LiDAR as a crucial feature for building classification. Shirowzhan et al. [5] integrated the texture features of the nDSM with the intensity of LiDAR data, and employed SVM for the classification and change detection of buildings. Lai et al. [28] introduced the texture features of the elevation map to distinguish between buildings and trees accurately. Integrating texture features extracted from imagery provides supplementary features that significantly improve the identification performance of building points. However, the accurate registration of imagery and point clouds is a critical prerequisite for effective feature fusion.

Airborne LiDAR data consists of massive laser points with 3D spatial information, imposing significant demands on the efficiency of classification algorithms and the capabilities of computers. Therefore, the object-oriented classification algorithm based on segmentation exhibits distinct advantages. In these algorithms, supervoxels serve as essential primitives to simplify data representation. The supervoxel extends the concept of superpixel—a paradigm originally proposed for unsupervised over-segmentation of 2D images into compact and homogeneous regions. High-quality supervoxels are characterized by two key attributes: they avoid object boundary crossings and have a regular distribution over the area being segmented [14]. In classification tasks, the former is more important, as it preserves semantic homogeneity as much as possible. The VCCS algorithm [14] is a classical supervoxel segmentation algorithm that uses a variant of k-means clustering. It enforces strict 3D spatial connectivity among occupied voxels when selecting voxels for clustering and adopts a flow constrained clustering approach, helping avoid crossing object boundaries. Building upon the VCCS, Setin et al. [29] proposed Locally Convex Connected Patches (LCCP), which further segments supervoxels based on the concavity and convexity features of supervoxel facets. These two methods generate supervoxels that satisfy both desired properties and have therefore been widely adopted in point cloud processing. However, the reliance on uniformly distributed seeds and fixed supervoxel size in 3D space makes it difficult to process point clouds with non-uniform density, which still leads to cross-boundary issues.

In previous research, three types of solutions have been proposed to address the above issues. The first type adheres to boundaries by recognizing boundary points and subsequently utilizing them to delineate segmentation. For example, BESS algorithm [20] was proposed, which excluded boundary points during the construction of the voxel adjacency graph. Xu et al. [30] developed a boundary-refined supervoxelization algorithm that mitigates the "zig-zag" effect by detecting and refining boundary points. The second type of solution departs from the conventional criteria of uniformly distributed seeds and fixed resolution, yielding supervoxels that dynamically conform to object geometry, such as the adaptive seeds guided by saliency (SSV) [23] and normal histogram [17]. In recent years, researchers have focused on the third type of solution, which formalizes supervoxel segmentation as a subset selection problem. This concept was initially proposed by Lin et al. [16] and subsequently refined by Xiao et al. [21] and Su et al. [22]. These algorithms first select some representative points from the input point cloud, each representing an individual supervoxel. The point cloud is then supervoxelized by assigning additional points to these representative points based on an energy function. A defining characteristic of this type of solution is their independence from neighborhood information during the selection phase, which necessitates the employment of approximation strategies for efficient point allocation. Notably, the first and third solution types use individual points as processing elements, resulting in relatively low computational efficiency. In contrast, the second type of solution employs voxels as the fundamental element and demonstrates a substantial advantage. Moreover, this type of solution has received comparatively limited attention in contemporary research, underscoring the need for further exploration.

3. Methodology

3.1. Overview of the Approach

The workflow of the proposed building roof point classification approach is depicted in Figure 1, which comprises three sequential steps:

1. PTD-Based Ground Point Filtering

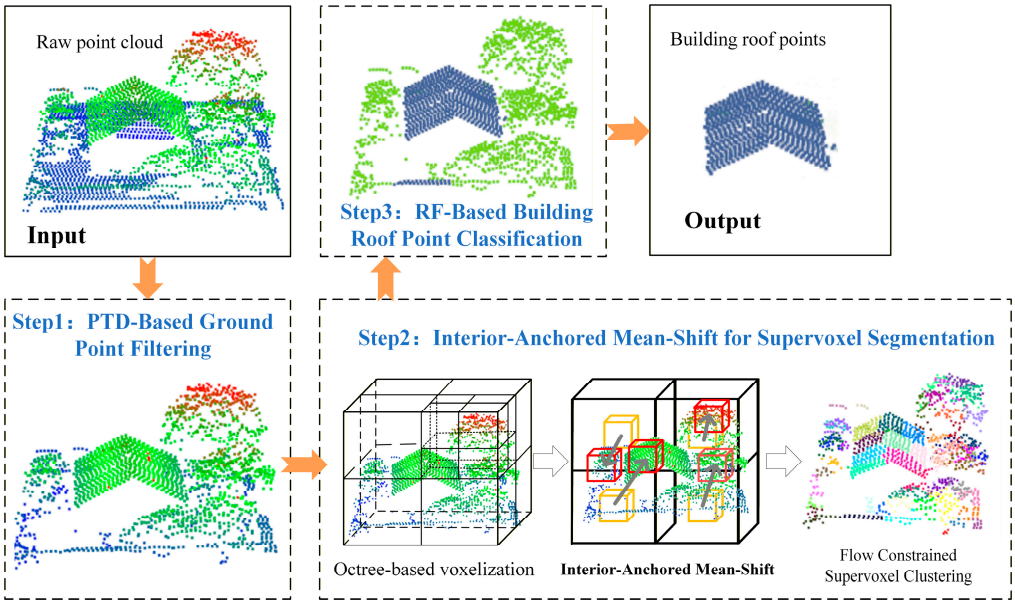


Figure 1. The proposed building roof point classification approach. The raw point cloud is colored by elevation. In Step 2, the yellow and red outlined cubes represent the initial and optimized seed voxels, respectively, while the generated supervoxels are assigned random colors based on their unique IDs. In Step 3, the dark blue and green points represent building roof and vegetation, respectively.

A fast, progressive TIN densification filtering (PTD) algorithm [31] is used to extract non-ground points from the raw LiDAR point cloud. Thus, the redundant ground points are removed for the subsequent segmentation.

2. Interior-Anchored Mean-Shift for Supervoxel Segmentation

Large-scale non-ground points are partitioned into a set of simplified supervoxels, each of which represents a homogeneous category of ground objects. Specifically, this process includes three sub-steps: (i) Octree-based voxelization of the preprocessed non-ground points; (ii) Anchor seed voxels within object interiors to eliminate cross-boundary segmentation errors caused by random seed initialization; and (iii) implementation of flow constrained clustering for the generation of refined supervoxel clusters.

3. RF-Based Building Roof Point Classification

Supervoxel-level features (point-distribution features and contextual features) are extracted and fed into an RF classifier for the building roof point classification, establishing a direct mapping between supervoxels and building roof categories.

3.2. Mean-Shift-Based Seed Optimization for Supervoxel Segmentation

The conventional supervoxel segmentation workflow involves two steps [14]: (1) converting the unstructured raw point cloud into a structured voxel model via octree-based voxelization, and (2) aggregating proximal voxels with similar geometric features to generate supervoxels primitives. Seed initialization for the clustering process is generally achieved through uniform spatial partition at a fixed seed resolution R_{seed}. As shown in

Figure 2, this strategy ensures a geometrically regular supervoxel distribution but induces segmentation errors at object boundaries, where seeds tend to miscluster adjacent voxels with similar geometric features [32]. Such cross-boundary segmentation errors propagate downstream, thereby degrading the accuracy of building roof classification. To address this issue, seed voxels should be initialized within object interiors, where local point density is higher and geometric features are more stable compared to boundary regions (Figure 2). The Mean-Shift algorithm, a non-parametric clustering method, is highly effective at detecting local density maxima in datasets with unknown distributions [33], making it particularly suitable for anchoring interior regions of typical ground objects (e.g., buildings, vegetation). Thus, this study leverages the Mean-Shift algorithm to optimize seeds. The detailed implementation procedure is presented as follows.

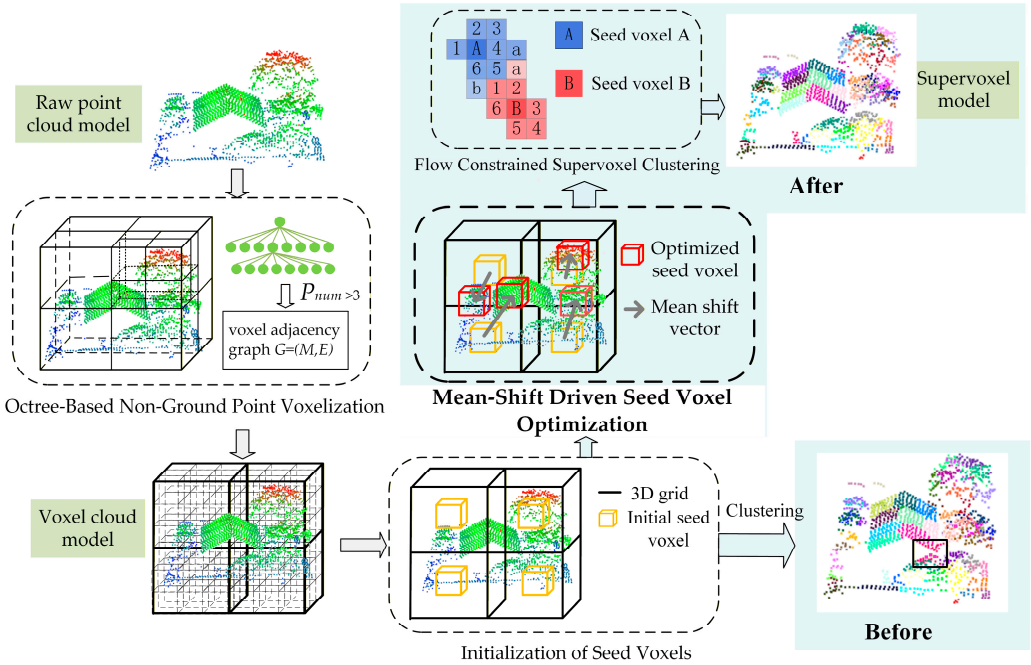


Figure 2. Workflow of the supervoxel segmentation via Mean-Shift-driven seed optimization. In the step of Flow Constrained Clustering, numbers represent the local indices of voxels within each supervoxels, and letters denote the indices of seed voxels.

(1) Octree-Based Non-Ground Point Voxelization

After PTD-based ground-point filtering, an octree structure [34] is employed to partition non-ground points. Empty nodes ($\leq$3 points) are removed, and the remaining non-empty nodes are further subdivided until the termination criteria ($R \leq R_{voxel}$) is satisfied. Here, R represents the voxel size, and R_{voxel} denotes the voxel resolution. Subsequently, a voxel adjacency graph, which offers spatial adjacency relationships information among voxels, is constructed as a prerequisite for the subsequent supervoxel clustering.

(2) Mean-Shift Driven Seed Voxel Optimization

Employing its capability to detect local density maxima, the Mean-Shift algorithm anchors seed voxels within object interiors. Figure 3 illustrates this optimization process. As shown in Figure 2, the initial seeds, which serve as the starting cluster centers for Mean-Shift, are initialized at the centroid of each grid cell.

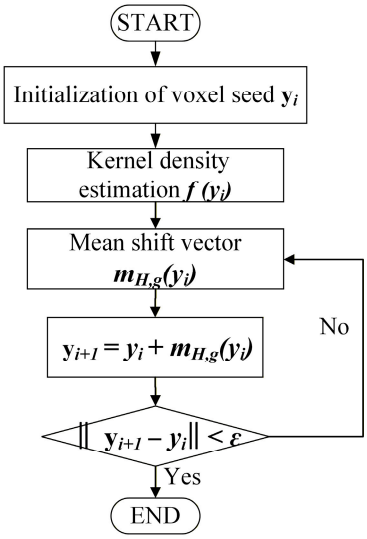


Figure 3. Detailed implementation of Mean-Shift-driven seed optimization process.

Given a point cloud dataset $\{x_i\}$ of n points, $i = 1, \ldots, n$, the 3D kernel density estimation (KDE) is defined as:

$$\hat{f}(X) = \frac{1}{n}\sum_{i=1}^{n} K_H(x - x_i) \tag{1}$$

where K_H is the kernel function. In this study, the Gaussian kernel is adopted to estimate the probability density distribution of discrete points. Its nonlinear weight decay with spatial distance effectively mitigates the interference from the non-uniform density of LiDAR point clouds. The $H \in \mathbb{R}^{3\times3}$ represents the bandwidth matrix, which is a symmetric positive definite matrix.

Starting from the geometric center of an initial seed voxel y_i, the mean shift vector relative to its neighboring voxels $\{x_j\}$ within the search radius is calculated as follows:

$$m_{H,g}(y_i) = \frac{\sum_{j=1}^{n} G\left(\left\|\frac{y_i - x_j}{H}\right\|^2\right)\cdot y_i}{\sum_{j=1}^{n} G\left(\left\|\frac{y_i - x_j}{H}\right\|^2\right)} - y_i \tag{2}$$

The seed voxel sequence is then iteratively updated by:

$$y_{i+1} = y_i + m_{H,g}(y_i), i = 0, 1, 2 \ldots\ldots \tag{3}$$

This iterative process continues until the termination criterion $\|y_{i+1} - y_i\| < \varepsilon$ is satisfied, where ε is a predefined threshold.

(3) Flow Constrained Supervoxel Clustering

Finally, the flow constrained clustering algorithm [14] is implemented to generate supervoxels. This process produces highly homogeneous supervoxels, ensuring unambiguous object attribution for the subsequent building roof classification.

3.3. Supervoxel Segmentation Based on Interior-Anchored Mean-Shift (IAMS)

By iteratively shifting seed voxels toward local density maxima via the Mean-Shift algorithm, the cross-boundary segmentation errors are effectively eliminated. Nevertheless, the performance of the algorithm is highly sensitive to the choice of core hyperparameters. Fixed hyperparameter configurations fail to accommodate the urban point cloud data characterized by non-uniform density and complex scene topologies, ultimately leading to suboptimal segmentation outcomes. To overcome this limitation, an IAMS algorithm is

proposed (Figure 4), which integrates local geometric consistency, point density, and spatial correlation into a unified kernel density estimator, supplemented by density-adaptive voxel weighting and a semi-variogram-driven bandwidth. These strategic refinements are tailored to ensure that the seed voxels are consistently anchored within object interiors.

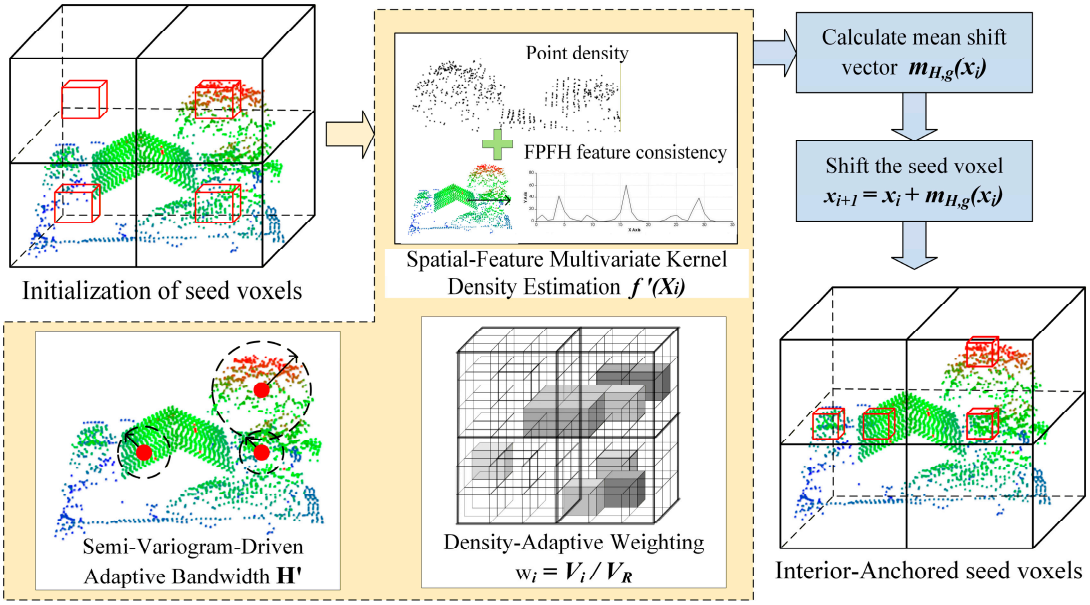


Figure 4. Detailed implementation of the proposed Interior-Anchored Mean-Shift process. This process corresponds to the "Mean-Shift Driven seed voxel optimization" stage depicted in Figure 2. The red outlined cubes represent the seed voxels. The red dots represent the geometric centers of the seed voxels, while the black dashed circles denote search radius.

3.3.1. Spatial-Feature Multivariate Kernel Density Estimation

Conventional kernel density estimation for Mean-Shift relies solely on 3D spatial coordinates in Euclidean space [34], making it difficult to distinguish voxels with the highest local point density but distinct geometric features. The literature [35] has demonstrated new geometric similarity properties of the kernel distance to infer topological characteristics, which suggests the feasibility of incorporating geometric features into kernel density estimators. Therefore, to overcome the limitations of purely spatial analysis, this study extends kernel density estimation to the feature space. By augmenting the 3-dimentional spatial coordinates of geometric center point with 33-dimensional FPFH feature vector, a 36-dimensional joint vector X_i for each voxel is constructed:

$$X_i = (x_i, y_i, z_i, f_{1i}, f_{2i}, f_{3i} \ldots, f_{33i})^T \in \mathbb{R}^{36} \quad (4)$$

This integration ensures that only voxels exhibiting both density maxima and local geometric consistency are selected as seeds, thereby refining the accuracy of interior region localization.

Since spatial coordinates and feature values possess disparate physical units and scales, a direct Euclidean distance calculation would be biased. To resolve this, all dimensions are mapped into [0, 1] through Min-Max Normalization. For each dimension $d \in \{1, 2, \ldots, 36\}$, the normalized value is computed as:

$$\hat{X}_{i,d} = \frac{X_{i,d} - \min(X_d)}{\max(X_d) - \min(X_d)} \quad (5)$$

Then, the spatial-feature multivariate kernel density estimation in this 36-dimensional space is given by:

$$\hat{f}\prime(\hat{X}) = \frac{1}{nH}\sum_{i=1}^{n} K\left(\frac{\hat{X}-\hat{X}_i}{H}\right) \quad (6)$$

where H represents the initial bandwidth, which is set to the seed resolution R_{seed}. Given the prior normalization step, the bandwidth must be mapped into the normalized space utilizing the equation $H = R_{\text{seed}}/L_{\text{max}}$, where L_{max} is the scaling factor, and $L_{\text{max}} = \{L_{\text{x}}, L_{\text{y}}, L_{\text{z}}\}$. L_{x}, L_{y}, L_{z} represents the maximum spatial extents along the corresponding axes in the global 3D space.

3.3.2. Density-Adaptive Voxel Weighting

Conventional supervoxel clustering algorithms (e.g., VCCS [14]) employ a uniform weighting for all voxels, ignoring the variation in point count among voxels. This constitutes a significant drawback when processing LiDAR datasets characterized by non-uniform density distributions. To resolve this, weights are assigned to voxels based on their point density ratio (Equation (7)), guiding seeds to anchor high-density regions (representative of object interiors) rather than sparse, boundary-like areas.

$$w_i = \frac{V_i}{V_R} \quad (7)$$

where V_i represents the number of points within voxel X_{i}, and V_R represents the total number of points within the search radius (R_{seed}).

The weight factor W_i is incorporated into the kernel density estimation and the weighted kernel density estimation $\hat{f}_w(X)$ can be defined as:

$$\hat{f}_w(X) = \frac{1}{n}\sum_{i=1}^{n} w_i K_H(x - x_i) \quad (8)$$

By introducing the weight factor, voxels with higher local point density exert greater influence on the kernel density estimation, thereby reinforcing the algorithm's convergence toward high-density regions.

3.3.3. Semi-Variogram-Driven Adaptive Bandwidth

In kernel density estimation, the bandwidth H determines the extent of the Region of Interest (ROI). Traditional Mean-Shift often employs a fixed bandwidth [33], which fails to accommodate the spatial heterogeneity of LiDAR point clouds, leading to over-segmentation in dense clusters or under-segmentation in sparse areas. To address this, an adaptive bandwidth mechanism is proposed by introducing the semi-variogram function, a core geostatistical tool that quantifies spatial correlation by distance-dependent similarity decay [36]. This function has been widely applied to handle heterogeneity in remote sensing and environmental modeling [37–39]. It aligns perfectly with the intrinsic property of LiDAR data, where spatially proximate points (e.g., within a single roof) exhibit higher density and geometric consistency.

In this study, the voxel density (point density of each voxel) is defined as a regionalized variable Z(P), with the seed resolution R_{seed} serving as the initial bandwidth H. For a given seed voxel, the semi-variogram $\gamma(h)$ is calculated as:

$$\gamma(h) = \frac{1}{2m}\sum_{i=1}^{m} [Z(P) - Z(P+h)]^2 \quad (9)$$

where h denotes the lag distance. The semi-variogram effectively captures the spatial correlation of two voxels. A low $\gamma(h)$ indicates a homogeneous, high-density region (e.g., roof centers), while a high $\gamma(h)$ signifies object boundaries or sparse scanning areas.

The relationship between variogram value $\gamma(h)$ and covariance value *C(h)* is illustrated in Figure 5. As the lag distance h increases, the $\gamma(h)$ progressively rises until it reaches its peak (the Sill), while the *C(h)* continuously declines to 0, demonstrating an inverse correlation between these two functions. Under this condition, the equation $C(h) = \gamma(\infty) - \gamma(h)$ can be concluded.

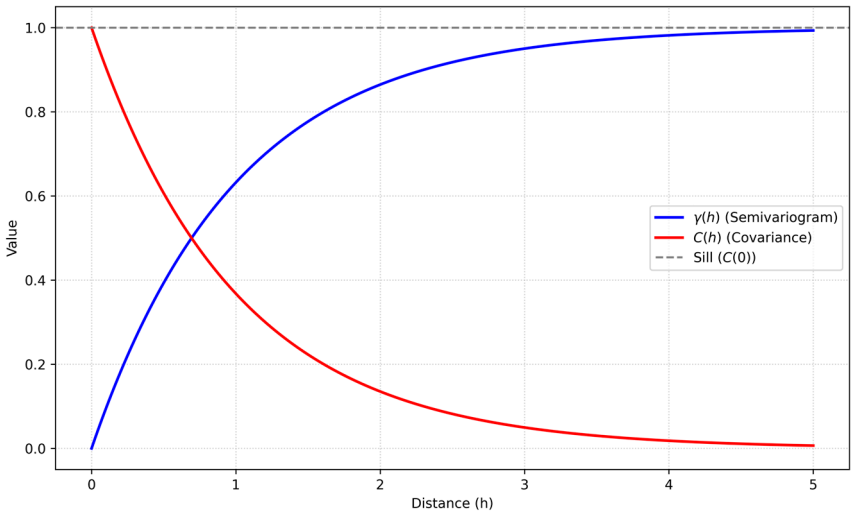


Figure 5. The relationship between variogram and covariance.

To dynamically adjust bandwidth, we derive a spatial correlation weight *W(h)* from the Sill ($\gamma(\infty)$):

$$W(h) = \frac{C(h)}{\gamma(\infty)} = \frac{\gamma(\infty) - \gamma(h)}{\gamma(\infty)} \tag{10}$$

Then, the initial bandwidth H can be adjusted by the following formula:

$$H' = H \times W(h) \tag{11}$$

The weight $W(h)$ ranges from 0 to 1, representing the strength of local point density consistency. A higher value of $W(h)$ signifies superior spatial consistency in point density ($W(h) \approx 1$), which is typically characteristic of object interiors. In such regions, the bandwidth remains substantially constant. Conversely, a lower value of $W(h)$ indicates weaker spatial correlation and greater density fluctuations, suggesting the presence of object boundaries or sparse regions. Under these conditions, the bandwidth is automatically adjusted to maintain the kernel function's high sensitivity.

Integrating the point density weight factor w_i (Equation (7)) and adaptive bandwidth H', the mean shift vector of seed voxel X relative to its neighboring voxels *Xi* is defined by:

$$m_{H',w}(X) = \frac{\sum_{i=1}^{n} w_i G\left(\left\|\frac{\hat{X}-\hat{X}_i}{H'}\right\|^2\right) X_i}{\sum_{i=1}^{n} w_i G\left(\left\|\frac{\hat{X}-\hat{X}_i}{H'}\right\|^2\right)} - X \tag{12}$$

Then, seed voxel *Xi* shifts along the mean shift vector until it converges to the region with the highest local density and feature consistency. This iterative procedure terminates when $\left\|m_{H',w}(X)\right\| < \mathrm{R}_{\mathrm{voxel}}$, signifying that the minimum voxel granularity is achieved.

3.4. Supervoxel-Level Feature Extraction and Random Forest-Based Building Roof Classification

Through the IAMS algorithm, massive and unorganized point clouds are partitioned into highly homogeneous supervoxels, which serves as simplified, high-fidelity geometric primitives. Since these supervoxels maintain superior geometric consistency and adhere to object boundaries, the complexity of the subsequent classification is substantially reduced. Consequently, an RF classifier is capable of achieving high-precision results through the learning of discriminative features.

To classify building roofs, both point-distribution and contextual features are extracted. Specifically, five point-distribution features (*VOL*, *STDV*, *CLUS*, *Nm*, *Nr*) are extracted to capture the local geometric properties of supervoxels, thereby enabling the effective discrimination of building roofs from non-target objects such as noise, facades, and vegetation. However, the supervoxelization itself is an over-segmentation process, thus using merely point-distribution features can hardly represent the geometric characteristics of an object [40]. To address this limitation, four contextual features (*RE*, *NEBI*, *RP*, *RD*) are integrated to depict the spatial relationships among supervoxels. These features, detailed in Table 1, are then fed into the RF classifier to perform the building roof identification.

Table 1. Feature vectors set for roof point classification.

Category	Feature	Abbr.	Definition
Point-distribution features	Supervoxel volume	VOL	Number of Points
	Standard deviation [7]	STDV	$\sqrt{\frac{\left(\bar{x}-x_i\right)^2+\left(\bar{y}-y_i\right)^2+\left(\bar{z}-z_i\right))^2}{n}}$
	Clustered points [7]	CLUS	n/N
	Mean of normal vectors [41]	Nm_x	$\frac{1}{n}\sum_{i=1}^{n} Nx_i$
		Nm_y	$\frac{1}{n}\sum_{i=1}^{n} Ny_i$
		Nm_z	$\frac{1}{n}\sum_{i=1}^{n} Nz_i$
	Range of normal vectors [41]	Nr_x	$Nx_{max}-Nx_{min}$
		Nr_y	$Ny_{max}-Ny_{min}$
		Nr_z	$Nz_{max}-Nz_{min}$
Contextual features	Relative elevation [7]	RE	$\frac{\lvert a\bar{x}+b\bar{y}+c\bar{z}+d\rvert}{\sqrt{a^2+b^2+c^2}}$
	Number of neighbors [7]	NEBI	/
	Relative position [30]	RP	$\frac{1}{n}\sum_{i=1}^{n} doi$
	Relative direction [30]	RD	$\frac{1}{n}\sum_{i=1}^{n} noi$

4. Experiment and Analysis

4.1. Data Description

To evaluate the effectiveness and robustness of the proposed building roof classification approach, experiments were conducted across three datasets with varying point densities and diverse scenarios (Table 2). The first dataset is the Vaihingen dataset from International Society for Photogrammetry and Remote Sensing (ISPRS) [42], which exhibits a point density of 4–7 pts/m^2, characterized by buildings surrounded by trees. The second dataset is the Hainan dataset, which was independently acquired by our research group in Hainan Province, China, using a Livox Avia (Livox Technology Co., Ltd., Shenzhen, China) laser scanner. This dataset features a markedly lower point density of 0.3 pts/m^2 and a sparser vegetation and building distribution compared to the former dataset. The third dataset is the Porirua City dataset in New Zealand from Open Topography [43], which possesses a significantly higher point density, reaching 27 pts/m^2. Compared to Vaihingen, it features a denser vegetation distribution, with many building roofs obscured by trees. This distinct-dataset evaluation strategy is essential for further validating the robustness of our supervoxel segmentation and building roof point classification algorithm.

Table 2. Experimental datasets.

Dataset	Source	Point Density	Distribution of Trees and Buildings
Hainan	ALS	0.3 pts/m^2	Sparse distribution of trees and buildings
Vaihingen	ISPRS	4–7 pts/m^2	Buildings are surrounded by trees
Porirua City	Open Topography	27 pts/m^2	Dense distribution of trees with obscured roofs

The Vaihingen dataset [42] was captured using a Leica ALS50 system (Leica Geosystems, Heerbrugg, Switzerland) over Vaihingen, Germany, and comprises three test areas with diverse building architectures (Figure 6). Area 1 features complex historical buildings in the city center; Area 2 is characterized by high-rise residential buildings amidst dense vegetation; and Area 3 consists of typical detached small houses.

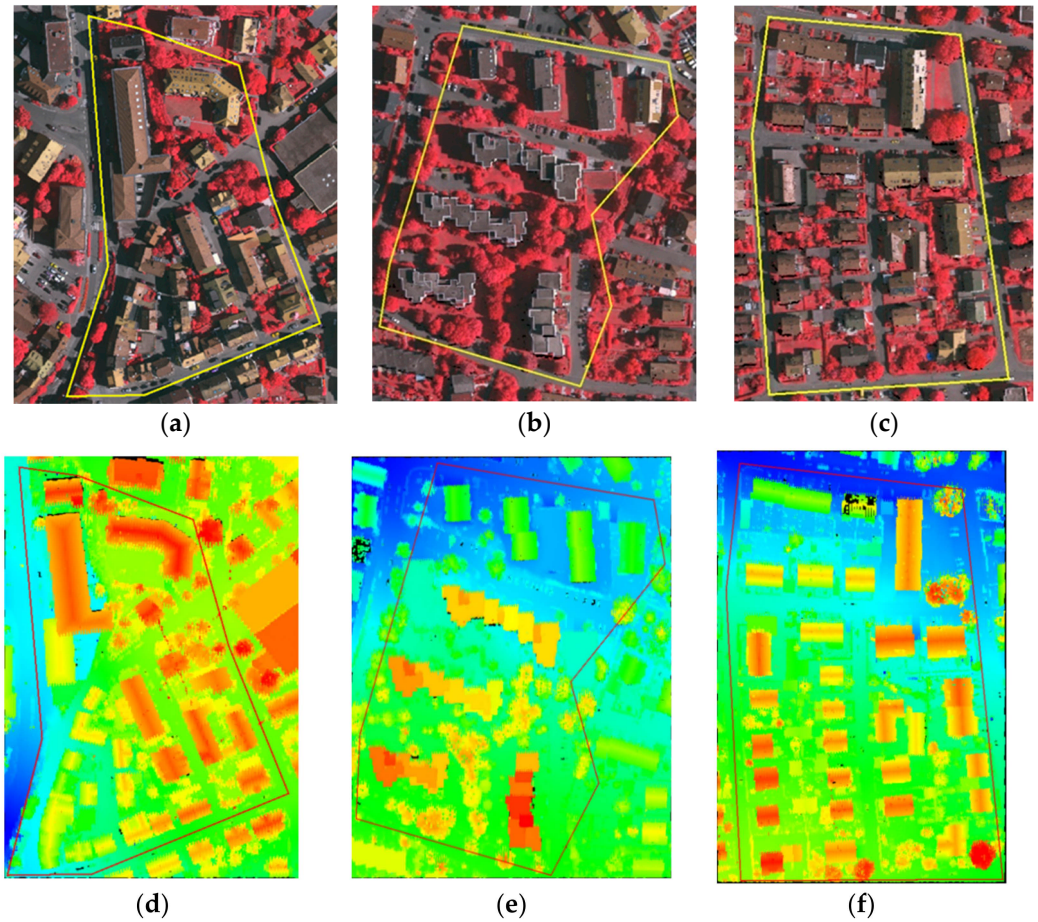
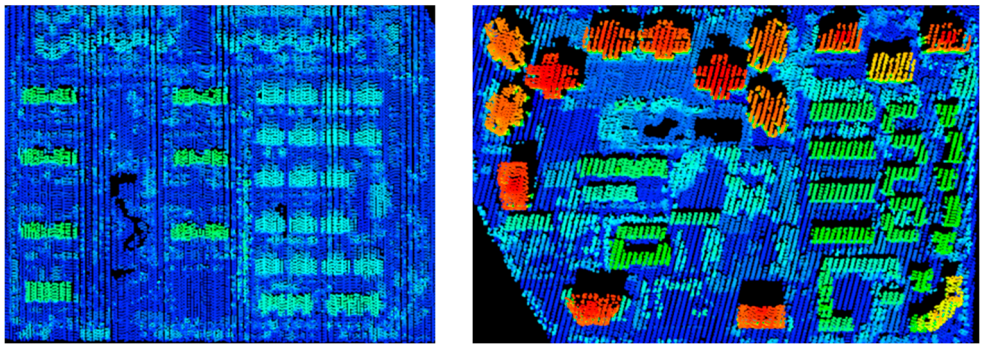


Figure 6. ISPRS Vaihingen test areas. (**a**–**c**) Aerial images of Areas 1–3; (**d**–**f**) Raw LiDAR data of Areas 1–3. Points in (**d**–**f**) are colored by height. The yellow and red outlines indicate the extent of the corresponding test areas.

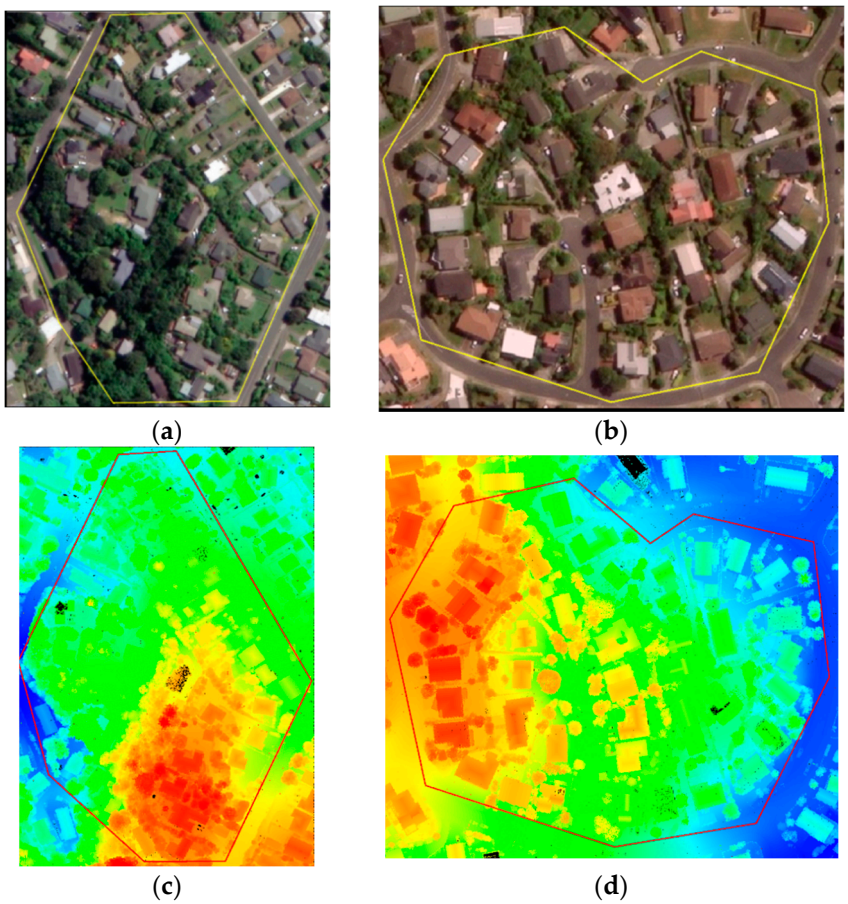
The Hainan dataset comprises two test areas (Figure 7). Area 1 is characterized by many systematically arranged residential buildings, while Area 2 exhibits greater architectural diversity, encompassing both high-rise buildings and low-rise dwellings.



(**a**) (**b**)

Figure 7. Hainan test areas. (**a**) Raw LiDAR data of Area 1; (**b**) Raw LiDAR data of Area 2. Points are colored by height.

The Porirua City dataset was retrieved from the Open Topography database [43]. Compared to the Vaihingen dataset, it exhibits significantly denser vegetation, with many building roofs occluded by trees, which poses a substantial challenge for accurate super-voxel segmentation. Two test areas from this dataset were selected for the experimental evaluation (Figure 8). In Area 1, the buildings are surrounded by dense vegetation, with some rooftops occluded by overhanging tree canopies. Additionally, the scene features extensive areas of lush greenery. In Area 2, although the vegetation is less dense than in Area 1, the tree canopies are at a similar elevation and in close spatial proximity to the rooftops.



(**a**) (**b**)

(**c**) (**d**)

Figure 8. Porirua City test areas. (**a**,**b**) Aerial images of Areas 1–2; (**c**,**d**) Raw LiDAR of Areas 1–2. Points in (**c**,**d**) are colored by height. The yellow and red outlines indicate the extent of the corresponding test areas.

4.2. Experiments on the ISPRS Vaihingen Dataset

4.2.1. Evaluation of Supervoxel Segmentation Result

Two quantitative metrics, boundary recall (BR) and under-segmentation error (UE), were employed to evaluate the accuracy of the generated supervoxels.

BR quantifies the proportion of ground-truth roof boundaries that align with the generated supervoxel boundaries [21], which is defined as:

$$BR(\mathcal{T},\mathcal{S}) = \frac{\sum_{p\in\partial\mathcal{T}}\left(min_{q\in\partial\mathcal{S}}\|p-q\|<\varepsilon\right)}{|\partial\mathcal{T}|} \tag{13}$$

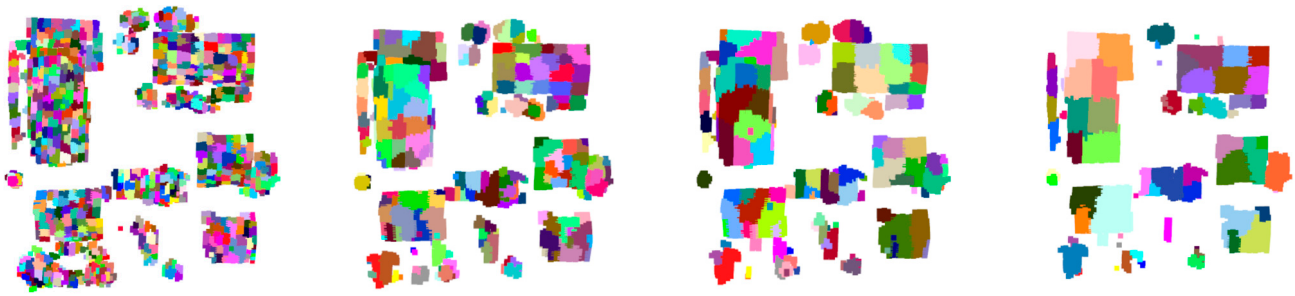
where $\partial\mathcal{T}$ and $\partial\mathcal{S}$ represent the sets of boundary points in the ground-truth roof and the supervoxel segmentation results, respectively. $(\cdot)$ is an indicator function that deems a ground-truth roof point p successfully covered by the supervoxel boundaries if its Euclidean distance to the nearest supervoxel boundary point q is less than a predefined threshold ϵ. Following [21], the threshold ϵ is set to 0.03 m in this study.

UE measures the degree of supervoxels across the ground-truth roof boundaries [21], which is defined as:

$$UE(\mathcal{T},\mathcal{S}) = \frac{1}{n}\left(\sum_{i=1}^{M}\left(\sum_{s_j|s_j\cap T_i\neq\emptyset}|s_j|\right)-n\right) \tag{14}$$

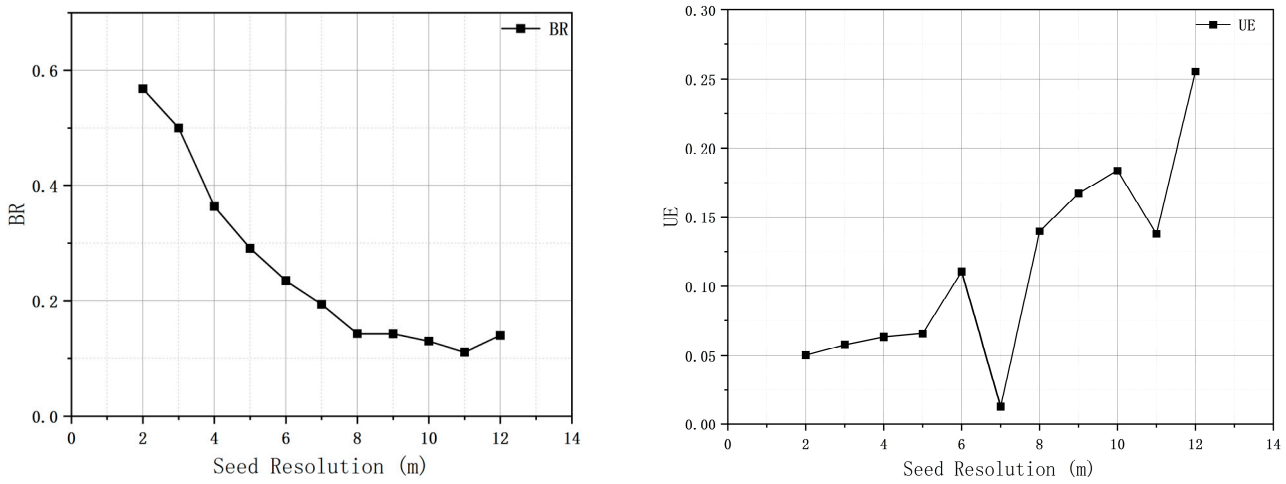
where $T = \{T_1, T_2, \ldots, T_M\}$ is the set of ground truth roof regions, and n is the total number of points in T. For each T_i, the corresponding set of overlapping supervoxels is identified, i.e., $\{s_j|s_j\cap T_i\neq\varnothing\}$. The under-segmentation error is then computed by counting the points of these supervoxels that leak outside T_i, divided by the normalization factor n.

The appropriateness of the seed resolution R_{seed} and voxel resolution R_{voxel} is critical to the performance of supervoxel segmentation. According to studies by [16], the R_{seed} must be approximately 10 times R_{voxel}. Given that the R_{seed} directly dictates supervoxel size, its selection is the primary focus of our analysis. As shown in Figure 9, when R_{seed} is set too small, it results in over-segmentation, producing redundant segments within flat roof interiors. Conversely, when R_{seed} exceeds the minimum building length (>5 m), specific structural details are eliminated, and the *UE* increases significantly. This is because the seeds are too sparse to capture small structures. Figure 10 illustrates the sensitivity analysis of the impact of R_{seed} on *BR* and *UE* metrics. It is observed that the most stable segmentation performance is achieved when the seed resolution is approximately equal to the minimum roof length. Our findings suggest that R_{seed} should be set to the minimum roof length L_{min} and not exceed $2 \times L_{min}$ to ensure robustness of the algorithm.



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

Figure 9. Visual comparison of supervoxel segmentation results under different R_{seed}. Each supervoxel is assigned a random color based on its unique ID. (**a**) R_{seed} = 2 m, *UE* = 0.05; (**b**) R_{seed} = 5 m, *UE* = 0.06; (**c**) R_{seed} = 7 m, *UE* = 0.12; (**d**) R_{seed} = 10 m, *UE* = 0.18. The minimum roof length is 5 m.



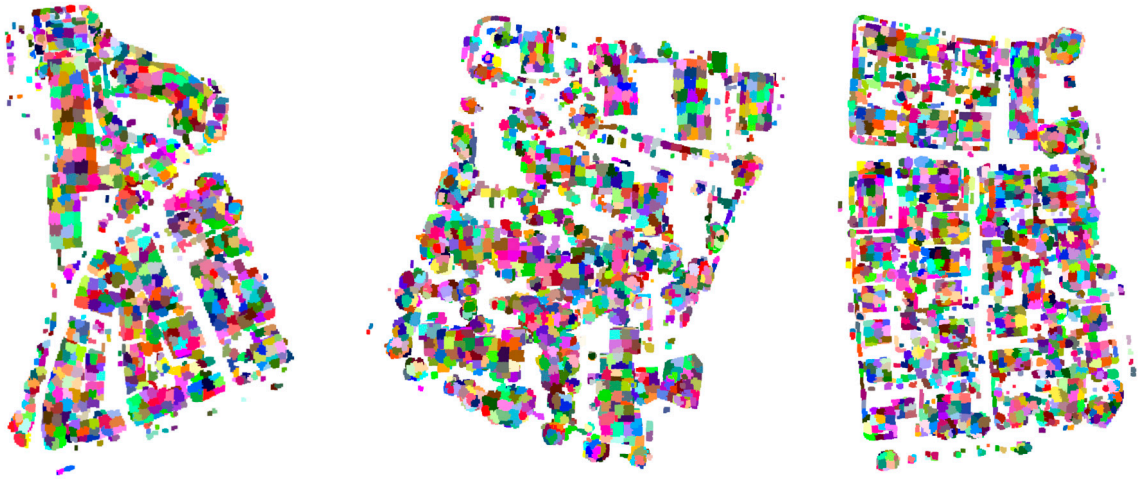
(**a**) (**b**)

Figure 10. The sensitivity analysis of the impact of R_{seed} on segmentation accuracy. (**a**) Impact of R_{seed} on *BR*; (**b**) Impact of R_{seed} on *UE*.

With R_{voxel} = 0.5 and R_{seed} = 5.4, the segmentation produced 959, 1711, and 1701 supervoxels for Areas 1–3, respectively (Table 3). The segmentation results are presented in Figure 11.

Table 3. The results from supervoxel segmentation.

Test area	Point Density (pts/m^2)	Total Number of Points	Total Number of Supervoxels
Area 1	3.5	92858	959
Area 2	7.9	290,958	1711
Area 3	7.9	256,718	1701



(**a**) (**b**) (**c**)

Figure 11. Supervoxel segmentation results of the ISPRS Vaihingen dataset. Each supervoxel is assigned a random color based on its unique ID. (**a**) Area 1, (**b**) Area 2, (**c**) Area 3.

To validate the effectiveness of our supervoxel segmentation algorithm in preserving object boundaries and improving segmentation accuracy, our algorithm was compared to the VCCS and two state-of-the-art algorithms. To ensure a fair comparison, the supervoxel resolution for all algorithms was consistently set to 5.4 m. As seen in Table 4, our results show higher *BR* and lower *UE* compared to the VCCS algorithm, which is facilitated by

the proposed seed voxel optimization strategy. In comparison with other two advanced algorithm, IAMS demonstrates competitive performance relative to VCCS_kNN and BPSS. A visual comparison of the segmentation performance is presented in Figure 12. It can be observed that, the VCCS algorithm is highly susceptible to cross-boundary segmentation errors, especially in regions where tree crowns and building roofs are in close proximity. Even for the two state-of-the-art algorithms, there remains a marginal risk of such errors. Conversely, IAMS effectively preserves the distinct boundaries between vegetation and roofs in these challenging areas. This enhanced performance is critical for mitigating errors in the subsequent building roof classification stage.

Table 4. Evaluation of the supervoxel segmentation results of the ISPRS Vaihingen dataset.

Test Area	VCCS		VCCS_kNN		BPSS		IAMS	
	BR	UE	BR	UE	BR	UE	BR	UE
Area 1	0.52	0.37	0.77	0.33	0.73	0.34	0.76	0.33
Area 2	0.65	0.24	0.80	0.21	0.78	0.22	0.83	0.20
Area 3	0.63	0.26	0.82	0.22	0.79	0.23	0.81	0.23
Average	0.60	0.29	0.80	0.25	0.77	0.26	0.80	0.25

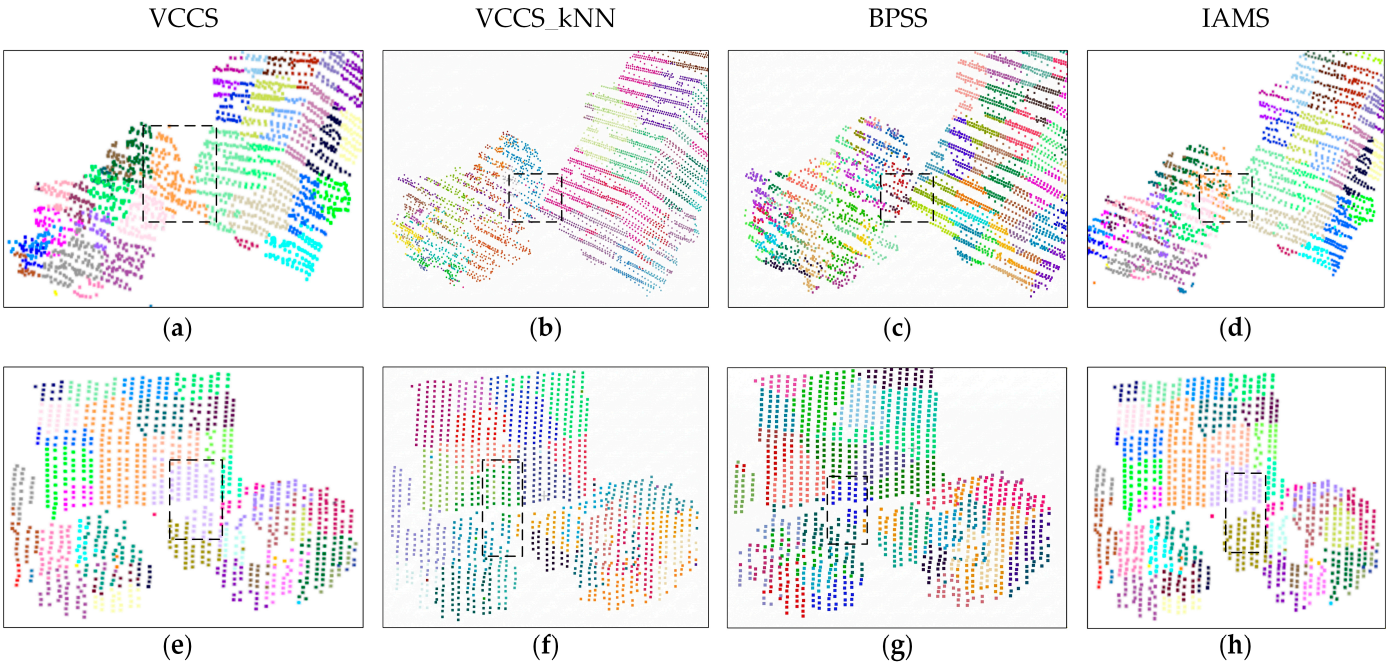


Figure 12. Performance of our supervoxel segmentation algorithm and other algorithms. Each supervoxel is assigned a random color based on its unique ID. (**a**–**d**) The detailed visualization of example 1; (**e**–**h**) The detailed visualization of example 2.

To further evaluate the contribution of each module in the IAMS algorithm, an ablation experiment was conducted using the traditional Mean-Shift (MS) algorithm as the baseline. As shown in Table 5, integrating the MKDE module significantly enhanced segmentation accuracy, with the BR improving from 0.66 to 0.73 and the *UE* decreasing from 0.38 to 0.32, demonstrating that fusing FPFH geometric features with spatial coordinates effectively improves the algorithm's ability to distinguish object boundaries. With the addition of the DW module, the *BR* was further elevated to 0.78, indicating that point density weighting can optimize the weight distribution of voxels and reduce the likelihood that boundary voxels are erroneously identified as seeds. When the AB module was fully integrated, the optimal performance was achieved, attaining a *BR* of 0.80 and a *UE* of 0.25. This consistent

improvement in performance validates the effectiveness of each module within the IAMS algorithm in enhancing the accuracy of supervoxel segmentation.

Table 5. Ablation study for each module of the IAMS algorithm. MKDE: spatial-feature multivariate kernel density estimation; DW: density-adaptive voxel weighting; AB: semi-variogram-driven adaptive bandwidth.

Variant	BR	UE
MS	0.66	0.38
MS + MKDE	0.73	0.32
MS + MKDE + DW	0.78	0.27
MS + JKDE + DW + AB (IAMS)	0.80	0.25

4.2.2. Evaluation of Classification Accuracy

To quantitatively analyze the classification results, the accuracy evaluation system provided by the ISPRS [42] was adopted. The completeness, the correctness, and the quality metrics were determined on a per-pixel level ($Cm_{pix}/Cr_{pix}/Q_{pix}$) and on a per-object level ($Cm_{ob}/Cr_{ob}/Q_{ob}$). In this study, per-pixel is defined as the level of laser points, and per-object is defined as the level of building roof objects. These metrics can be calculated using Equations (15)–(17), respectively.

$$C_m = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{15}$$

$$C_r = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{16}$$

$$Q = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{17}$$

where TP, FN, and FP represent the number of correctly classified, omitted, and misclassified building points or objects, respectively.

Figure 13 shows the building roof classification results compared against the ISPRS benchmark ground truth. The results demonstrate that the proposed approach achieved accurate, stable performance, correctly identifying the majority of large and medium-sized buildings. The quantitative evaluation (Table 6) indicates that the method performs well on both the per-pixel and per-object levels, achieving an average Cm_{pix} of 94.9% and Cr_{pix} of 93.5%, and Cm_{ob} of 87.8% and Cr_{ob} of 95.6%, respectively. Notably, the lower per-object completeness in Area 2 is mainly attributed to the prevalence of high-rise residential complexes, where roofs are characterized by numerous parapets and small structures. The pronounced geometric discrepancy between these parapets and the roofs, coupled with the insufficient point density of small structures, makes it difficult to extract stable geometric features and leads to a higher rate of missed detections.

Table 6. Evaluation of the building classification results of the ISPRS Vaihingen dataset.

Test Site	Per-Pixel Level (%)			Per-Object Level (%)		
	Cm_{pix}	Cr_{pix}	Q_{pix}	Cm_{ob}	Cr_{ob}	Q_{ob}
Area 1	94.7	92.7	88.1	97.3	92.3	90.0
Area 2	95.9	96.1	91.4.	71.4	100.0	71.4
Area 3	94.0	91.8	86.7	94.6	94.6	89.8
Average	94.9	93.5	88.7	87.8	95.6	83.7

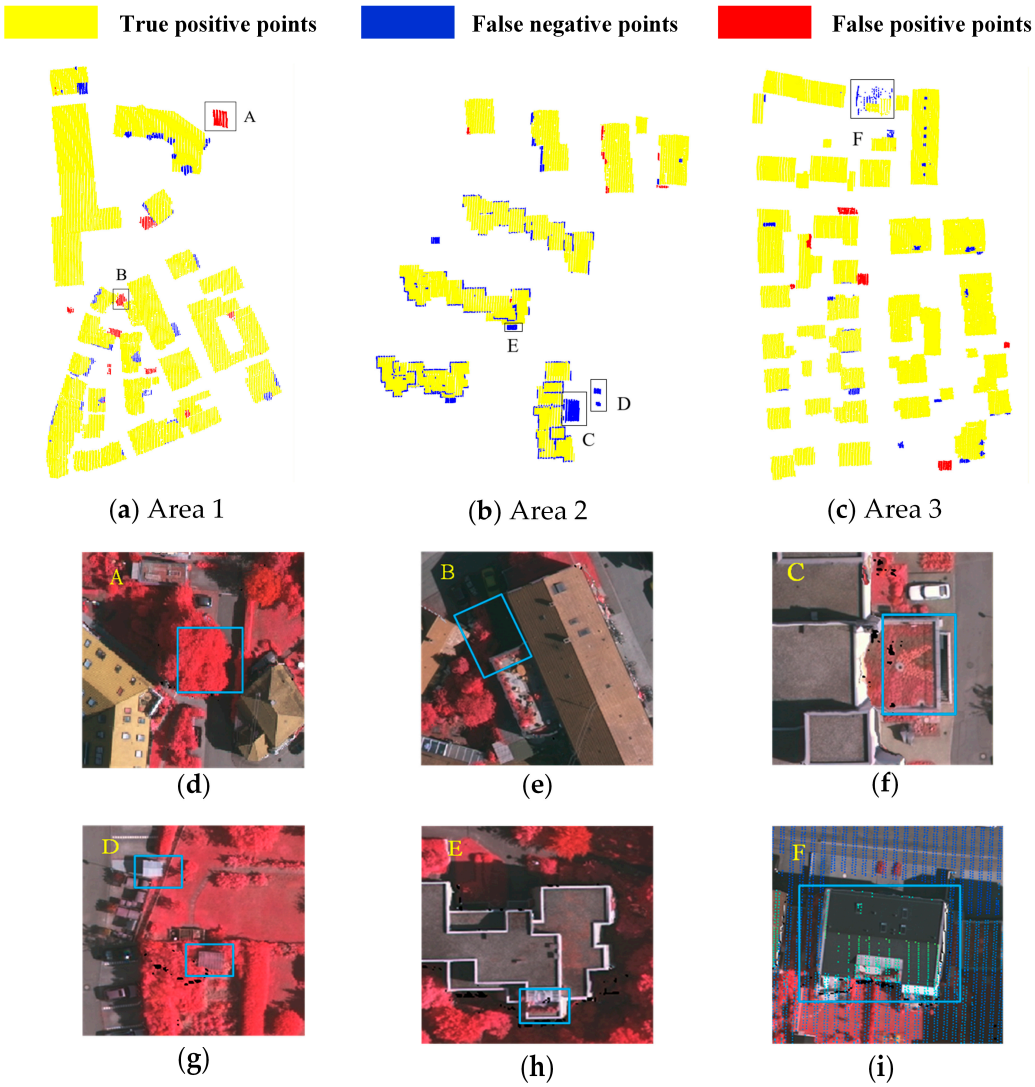


Figure 13. Visualization of extracted buildings for the Vaihingen dataset and exposed error factors. A–F—The local details of the classification results corresponding to the respective ground truth sub-figures displayed; (**a**–**c**) Building roof classification results of Areas 1–3; (**d**) details of A in (**a**); (**e**) details of B in (**a**); (**f**) details of C in (**b**); (**g**) details of D in (**b**); (**h**) details of E in (**b**); (**i**) details of F in (**c**).

To further quantify the impact of different error sources, a detailed error statistic analysis was performed (Table 7). Among the FP errors, 65% were attributed to 'Regular-shaped vegetation' (as seen in A of Area 1), where vegetation exhibits a regular surface, like buildings. The classifier is prone to misclassifying supervoxels in these regions as building roofs. This is primarily because the classifier relies solely on geometric features, and these supervoxels exhibit planarity attributes highly similar to those of actual roof structures. The other 27% of FP stemmed from 'Vegetation overspreading' (as seen in B of Area 1), where the dense canopy adjacent to the roofs shared similar geometric features, leading to characterization errors. The remaining 8% of FP resulted from man-made structures with regular geometries, such as large trucks. Consistent with the prior categories, the intrinsic limitation of a geometry-only approach prevents the classifier from effectively discriminating between these non-roof supervoxels and true building structures. To mitigate these errors, integrating high-resolution multispectral imagery or LiDAR multi-echo information may be a viable strategy. The spectral indices (e.g., NDVI) provided by multispectral imagery offer crucial spectral discriminative power by exploiting the high reflectance of

chlorophyll, effectively decoupling flat-topped vegetation from roofs. Furthermore, multi-echo information captures distinct physical properties. Unlike solid roofs that generate a single strong return, vegetation acts as a sparse and porous structure producing multiple returns. Incorporating these complementary attributes will be instrumental in avoiding such misclassification errors in future research. Regarding FN, 47% were attributed to ancillary roof structures, such as dormers and chimneys. These features possess smaller volumes and distinct geometric shapes compared to primary roof planes, making them susceptible to misclassification as vegetation. Small buildings, including garden rooms and garages (e.g., C and D of Area 2), constituted the other primary error source, accounting for 43% of errors. A size-dependent sensitivity analysis (Figure 14) reveals that completeness metrics at both the per-pixel and per-object levels exceed 90% for buildings > 50 m^2, whereas performance drops significantly for smaller structures. This quantitative gap confirms that building size is a significant factor influencing classification completeness. For structures < 50 m^2, the relatively low sampling density results in supervoxels that fail to provide distinctive geometric features, thereby rendering them highly susceptible to being omitted during the classification process. To address these limitations, future research will focus on studying multi-scale analysis methods. By extracting and fusing the geometric features of finer-resolution supervoxels, the geometric features of small roofs can be precisely captured, thereby mitigating omission errors.

Table 7. Statistics of the error source of the building classification results.

Error Type	Specific Source	Points	Contribution (%)
FP	Regular-shaped vegetation	1215	65
	Vegetation overspreading	495	27
	Other man-made structures	148	8
FN	Ancillary structures of roofs	1434	47
	Small buildings	1312	43
	Others	305	10

To validate the effectiveness of the seed voxel anchoring strategy, a comparative experiment was performed using the VCCS algorithm as the baseline. As shown in Table 8, our algorithm outperforms the VCCS algorithm in completeness, correctness, and quality metrics on both the per-pixel and per-object levels. In particular, the correctness metric shows exhibited significant improvements, with gains of 5.2% on the per-pixel level and 7.1% on the per-object level, indicating that the seed voxel anchoring strategy can effectively enhance the accuracy of building roof classification.

Table 8. Comparative experiment for seed voxel anchoring strategy.

Variant	Per-Pixel Level (%)			Per-Object Level (%)		
	Cm_{pix}	Cr_{pix}	Q_{pix}	Cm_{ob}	Cr_{ob}	Q_{ob}
VCCS	92.6	88.3	82.5	85.7	88.5	77.1
IAMS	94.9	93.5	88.7	87.8	95.6	83.7

In addition, the performance of the proposed algorithm was evaluated against six traditional geometric-based approaches that are available on the ISPRS benchmark website [44] and three advanced deep learning models (Table 9). It can be observed that our algorithm yielded results comparable to those of advanced deep learning approaches, despite relying solely on point distribution and contextual features within an RF framework. This performance is primarily attributed to the interior-anchored seeding mechanism that ensures the

generation of boundary-adherent and complete roof supervoxels. By yielding high-quality supervoxel primitives during the segmentation phase, the proposed algorithm simplifies the complexity of the subsequent classification of building roofs. These results validate the effectiveness of the proposed supervoxel segmentation algorithm for building classification and demonstrate its significant potential for further improvement.

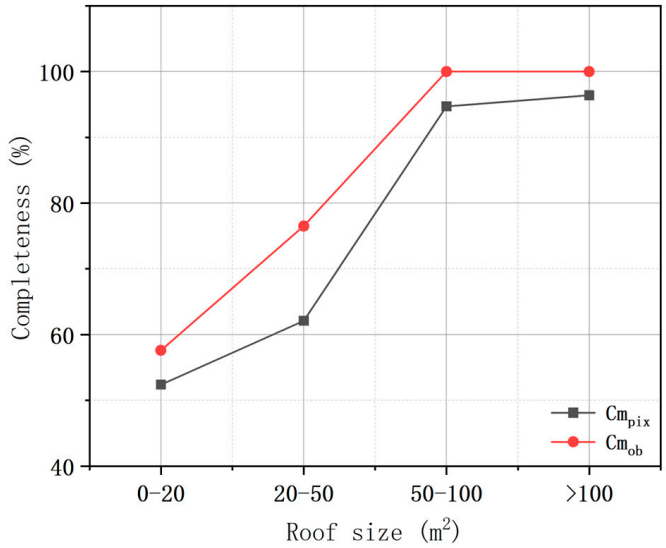


Figure 14. Sensitivity analysis of classification completeness across different roof scales.

Table 9. Comparison of our method with other methods. The best values per column are shown in bold font. '×' represents the absence of related experimental results in the source publication.

Method	Per-Pixel Level (%)			Per-Object Level (%)		
	Cm_{pix}	Cr_{pix}	Q_{pix}	Cm_{ob}	Cr_{ob}	Q_{ob}
WHUY2	89.7	90.9	82.3	83.0	97.5	81.3
MAR2	89.7	95.2	85.5	80.6	93.7	76.5
UMTA	92.3	87.5	81.5	80.0	98.6	79.1
HANC3	91.3	95.9	87.8	85.4	82.2	71.7
WHU_QC	85.8	**98.7**	84.8	80.9	**99.0**	80.3
MON4	94.3	82.9	79.0	83.9	93.8	79.3
PointNet [45]	73.0	82.2	63.0	×	×	×
PointNet++ [45]	92.8	81.0	76.2	×	×	×
Point Transformer [46]	**96.4**	96.8	**93.3**	86.3	78.6	70.3
IAMS	94.9	93.5	88.7	**87.8**	95.6	**83.7**

4.2.3. Runtime Analysis

Apart from classification accuracy, computational efficiency is a critical evaluation metric for large-scale LiDAR processing. A comparative analysis of the running times of our algorithm and several others for supervoxel segmentation and building roof classification is presented in Figure 15. All experiments were conducted on a PC equipped with an Intel Core i7-14650HX CPU @ 2.20 GHz and 16 GB RAM running Windows 11 (64-bit). To ensure the stability of the results, average runtimes were calculated from 10 trials.

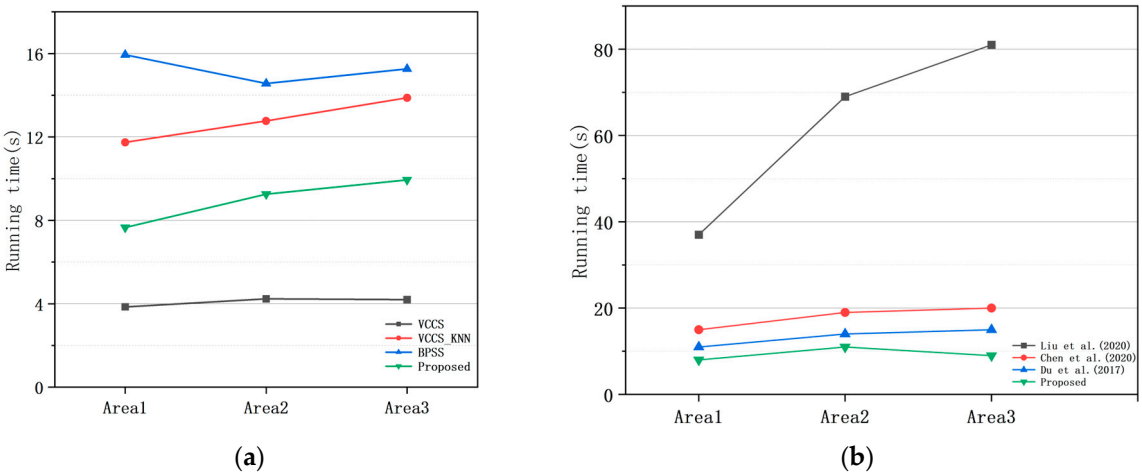


Figure 15. Runtime comparison of different algorithms. (**a**) Runtime comparison during supervoxel segmentation stage; (**b**) Runtime comparison during building roof classification stages, and the proposed algorithm is compared with Liu et al. [47], Chen et al. [48], and Du et al. [27].

The results indicate that in the supervoxel segmentation stage, although VCCS emerged as the most efficient due to its simpler architecture, it often fails to respect object boundaries. Our algorithm, despite an average runtime increase of 3.5 s to 5.1 s due to seed voxel optimization (Figure 16), consistently outperformed the other boundary-improvement algorithms. This efficiency is primarily achieved by operating on a voxel-based framework rather than processing individual laser points, thereby significantly reducing the search space complexity. As a result, our algorithm achieves a corresponding reduction in runtime for roof classification compared to algorithms proposed in recent literature [27,47,48].

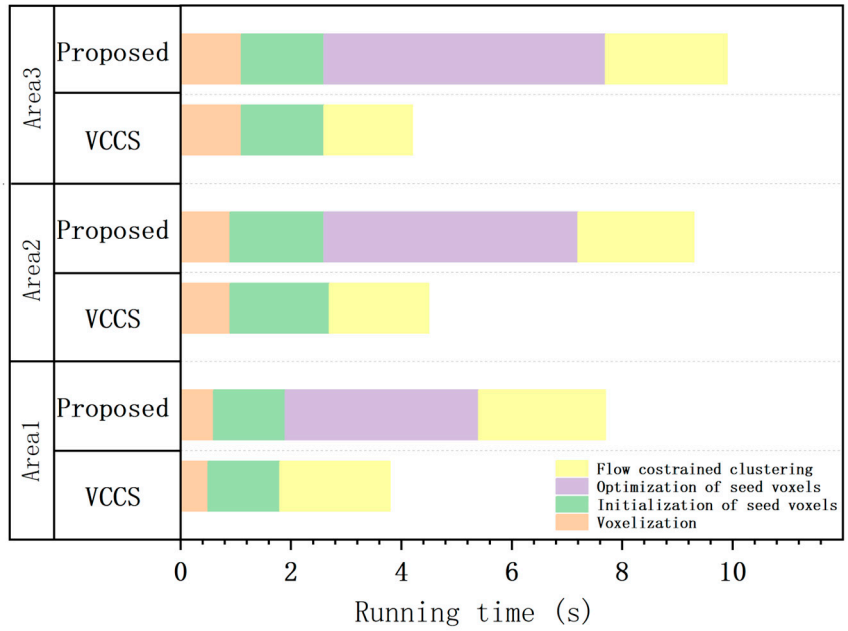


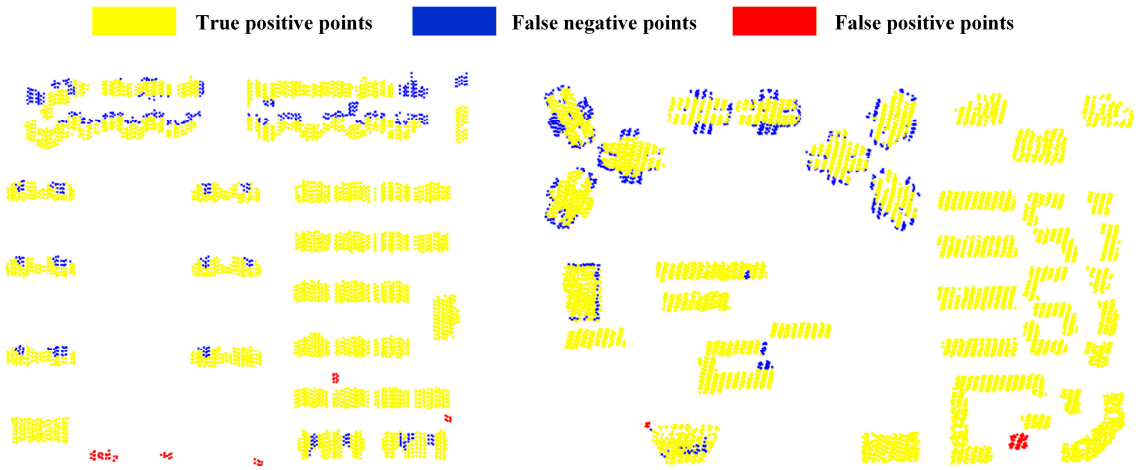
Figure 16. Runtime comparison between VCCS and our supervoxel segmentation algorithm.

4.3. Experiments on the Other Two Datasets

4.3.1. Results for the Hainan Dataset

The classification results for the Hainan dataset are shown in Figure 17. Although our algorithm effectively identified the vast majority of building roofs, a limited number of FN errors were observed. This is corroborated by the quantitative evaluation in Table 10,

where the results for the Cm_{pix} and Q_{pix} metrics were significantly inferior to those for the Cr_{pix} metric. These FN errors are primarily due to sparse laser points failing to adequately describe the geometric features of the roof supervoxels.



(**a**) Area 1 (**b**) Area 2

Figure 17. Visualization of extracted buildings at the per-pixel level for the Hainan dataset and exposed error factors.

Table 10. Evaluation of the building roof classification accuracy using the Hainan dataset.

Test Sites	Per-Pixel Level (%)			Per-Object Level (%)		
	Cm_{pix}	Cr_{pix}	Q_{pix}	Cm_{ob}	Cr_{ob}	Q_{ob}
Area 1	87.8	98.6	86.7	94.7	91.5	87.1
Area 2	90.0	99.2	89.4	100.0	95.1	95.1
Average	88.9	98.9	88.1	97.4	93.3	91.1

4.3.2. Results on the Porirua City Dataset

Figure 18 illustrates the building roof classification results of the Porirua City dataset. It is evident that our algorithm maintained robustness even in environments with dense vegetation. Similar to the Vaihingen dataset, the FP errors primarily arise from vegetation with regular geometric shapes being misidentified as building structures, as shown in D and F in Figure 18. As for FN errors, due to obstruction from tree canopies, such as B and C, the laser points reflected from roofs were insufficient to accurately depict their geometric characteristics, leading to the misclassification of the occlusion sections as vegetation. The quantitative evaluation in Table 11 shows that our algorithm achieved superior performance on the per-pixel level, with average Cm_{pix}, Cr_{pix}, and Q_{pix} values exceeding 90%. In contrast, the metrics on per-object-level were less than ideal, which is attributed to the aforementioned vegetation interference. Potential enhancements could involve integrating echo features into the RF classifier.

Table 11. Evaluation of the building roof classification accuracy using the Porirua City datasets.

Test Site	Per-Pixel Level (%)			Per-Object Level (%)		
	Cm_{pix}	Cr_{pix}	Q_{pix}	Cm_{ob}	Cr_{ob}	Q_{ob}
Area 1	96.3	97.5	94.0	90.6	74.4	70.0
Area 2	96.1	96.5	92.8	97.7	84.6	83.0
Average	96.2	97.0	93.4	94.2	79.5	76.5

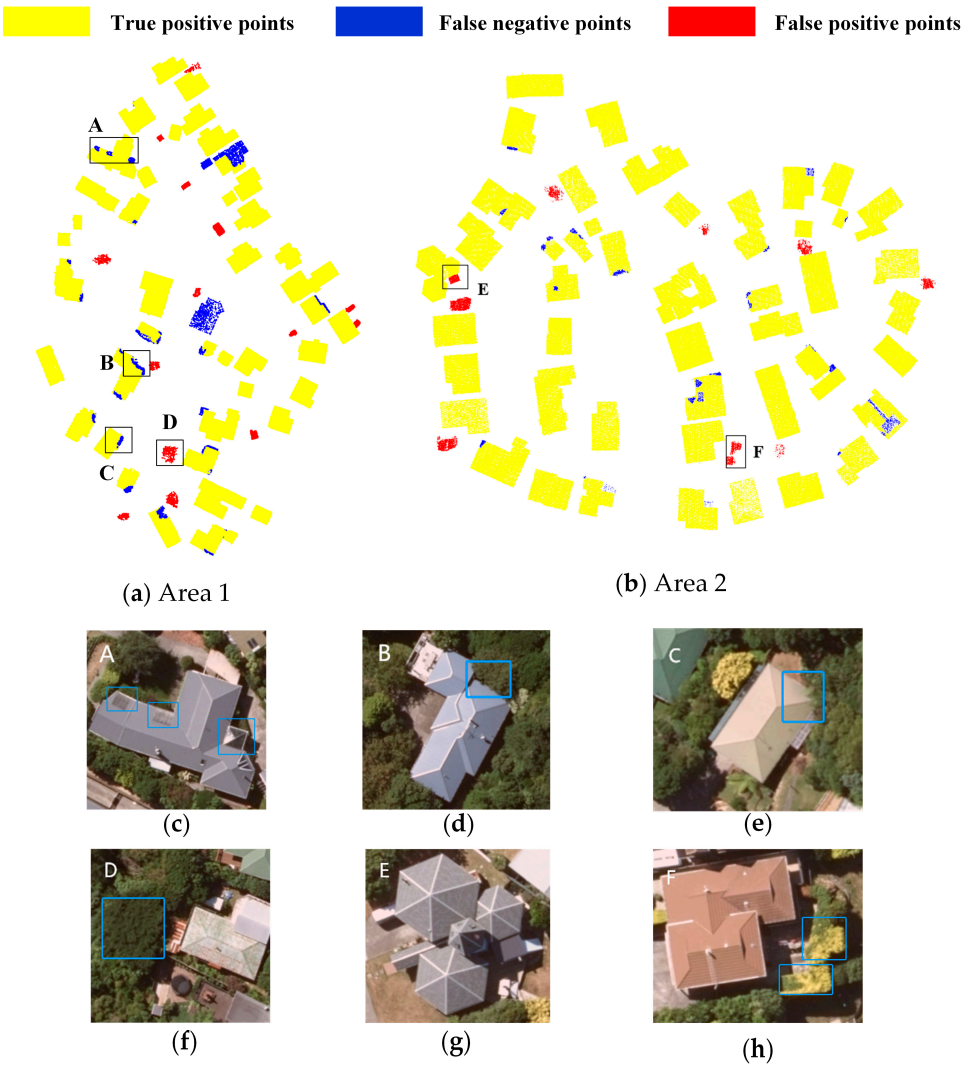


Figure 18. Visualization of extracted buildings for the Porirua City dataset and exposed error factors. A–F—The local details of the classification results corresponding to the respective ground truth sub-figures displayed; (**a**,**b**) Building roof classification results of Areas 1–2; (**c**) details of A in (**a**); (**d**) details of B in (**a**); (**e**) details of C in (**a**); (**f**) details of D in (**a**); (**g**) details of E in (**b**); (**h**) details of F in (**b**).

To further validate the effectiveness of our supervoxel segmentation algorithm on datasets with denser vegetation distributions, its classification results were compared with those of the VCCS algorithm. Table 12 presents the comparison of quantitative evaluation, which indicates that our algorithm achieved evident improvements. The average Cm_{ob}, Cr_{ob}, and Q_{ob} increased by 5.9%, 2.5%, and 5.6% on a per-pixel level, and 5.3%, 0.7%, and 4.3% on a per-object level, respectively. Two instances of buildings are presented in Figure 19. It can be observed that, with the VCCS algorithm, supervoxels frequently crossed the boundaries between adjoined roofs and tree canopies. In contrast, our algorithm effectively adhered to object boundaries, thereby mitigating both FN and FP errors. This comparative analysis demonstrates that, despite the challenges posed by dense vegetation, our algorithm is robust and effectively reduces classification errors compared to the conventional supervoxel segmentation algorithm.

Table 12. Comparison of our algorithm with the VCCS-based method.

Method	Per-Pixel Level (%)			Per-Object Level (%)		
	Cm_{pix}	Cr_{pix}	Q_{pix}	Cm_{ob}	Cr_{ob}	Q_{ob}
VCCS	90.3	94.5	87.8	82.1	81.3	73.3
IAMS	96.2	97.0	93.4	87.4	82.0	77.6

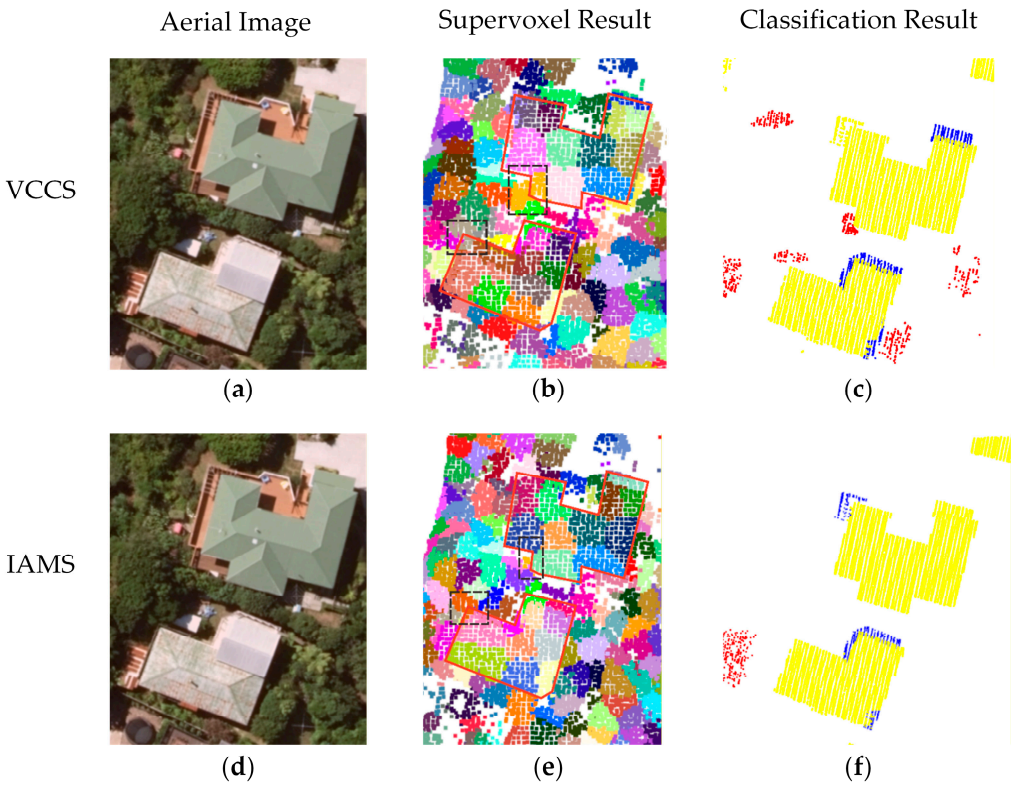


Figure 19. The effectiveness of our supervoxel segmentation algorithm on a dataset with a denser vegetation distribution. (**a**–**c**) The classification process based on the VCCS algorithm; (**d**–**f**) The classification process based on our algorithm. In (**b**,**e**), each supervoxel is assigned a random color based on its unique ID. In (**c**) and (**f**), yellow points represent TP, while blue and red represent FN and FP, respectively.

4.3.3. Robustness Analysis Across Diverse Roof Types

Figure 20 illustrates detailed views of roof classification results across diverse building types, encompassing high-rise buildings (Figure 20c–j) and buildings with hexagonal pyramid roofs (Figure 20k). For high-rise buildings, although some minor architectural details were occasionally omitted, the vast majority of roof points were accurately classified. Additionally, our algorithm exhibited precise identification of hexagonal pyramid roofs, as exemplified by K. These results validated the robustness of the proposed building roof classification framework, demonstrating its effectiveness in accommodating both complex geometries and specialized roof forms.

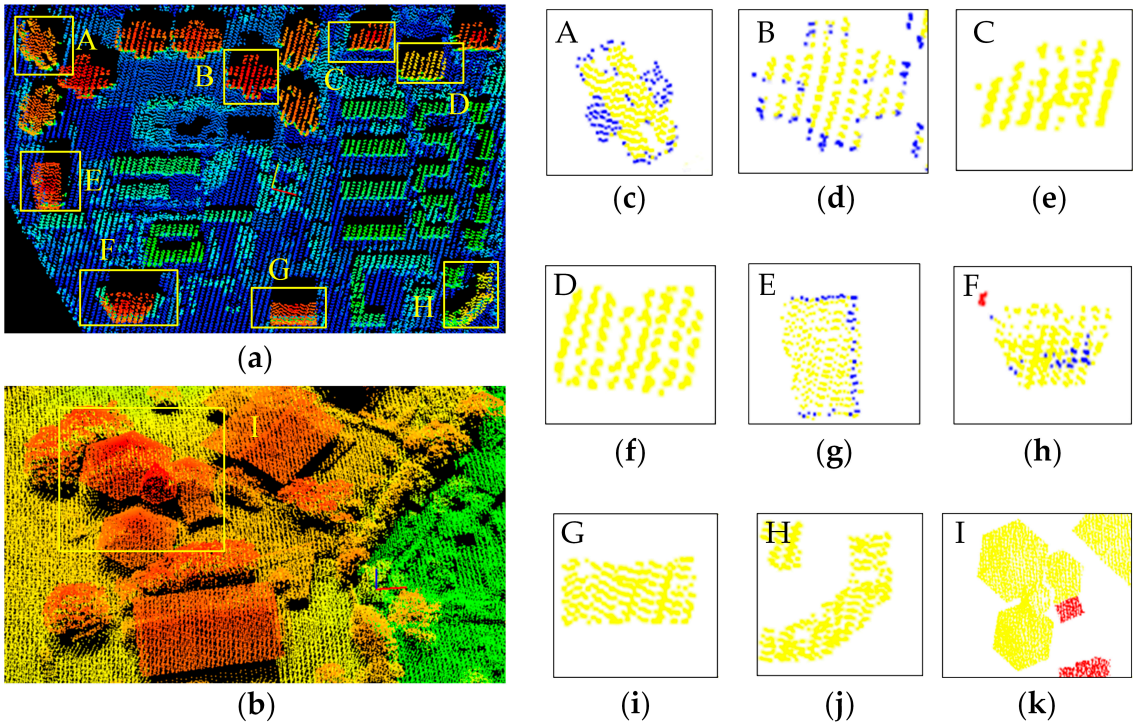


Figure 20. Detailed views of the classification results of different types of building roofs in the Hainan and Porirua City dataset. (**a**) The raw point cloud of the Hainan dataset in Area 2; (**b**) The raw point cloud of the local scene in Porirua City dataset; (**c**–**k**) The detailed views of roof classification results: yellow points represent TP, while blue and red represent FN and FP, respectively. A–I—The local details of the classification results corresponding to the raw point cloud displayed.

5. Conclusions

This study proposes Interior-Anchored Mean-Shift (IAMS), which optimizes supervoxel initialization to alleviate cross-boundary segmentation issues caused by random seed initialization in conventional supervoxel segmentation algorithms. By unifying geometric consistency, point density, and spatial correlation into a single interior-awareness criterion, the approach reliably anchors seeds within object interiors. This strategy effectively suppresses segmentation errors at their source, generating highly homogeneous supervoxels without the need for heuristic thresholds or intensive post-refinement. Experimental results across the Vaihingen, Hainan, and Porirua datasets demonstrate that IAMS consistently outperforms state-of-the-art algorithms, including VCCS. The proposed method significantly enhances building roof classification, yielding improvements in per-object quality by up to 7.1%. Notably, the algorithm maintains exceptional robustness under challenging real-world conditions, such as sparse sampling and heavy vegetation occlusion, while ensuring high computational efficiency. The high fidelity of the resulting classification provides a more reliable data foundation for downstream applications, such as 3D city modeling and urban planning.

Future research will focus on three key directions: (1) Exploring the synergy between IAMS and advanced deep learning architectures for point cloud processing, where high-homogeneity, boundary-adherent supervoxels generated by IAMS serve as structured primitives to not only reduce the computational overhead of deep learning models but also enhance the stability and semantic consistency of 3D scene recognition; (2) Incorporating multi-echo LiDAR and spectral information into the IAMS feature set to enrich structural and spectral discriminative cues, thereby improving the separability of geometrically similar yet semantically distinct objects in complex urban scenes; and (3) Refining the computational efficiency of the overall framework to maintain or surpass current high-

precision per-object segmentation performance while achieving more efficient and robust urban point cloud analysis.

Author Contributions: Conceptualization, H.Z. and L.Z.; methodology, H.Z. and L.Z.; software, H.Z.; validation, Z.Z., H.M. and C.Y.; investigation, H.Z.; resources, L.Z., H.Y. and X.T.; data curation, H.Z.; writing—original draft preparation, H.Z.; writing—review and editing, L.Z. and Z.Z.; funding acquisition, L.Z., H.Y. and X.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Natural Resources Provincial-Ministerial Cooperation Project (2023ZRBSHZ033) and Science and Technology Research and Development Program of the China State Railway Group (L2023G016).

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Acknowledgments: The authors would like to thank Open Topography for providing the Porirua City dataset used in the study. During the preparation of this manuscript, the authors used GenAI for the purposes of language editing and stylistic enhancement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare that this study received funding from China State Railway Group. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. Ghasemian Sorboni, N.; Wang, J.; Najafi, M.R. Fusion of Google Street View, LiDAR, and Orthophoto Classifications Using Ranking Classes Based on F1 Score for Building Land-Use Type Detection. *Remote Sens.* **2024**, *16*, 2011. [CrossRef]
2. Tarsha Kurdi, F.; Lewandowicz, E.; Gharineiat, Z.; Shan, J. High-Resolution Building Indicator Mapping Using Airborne LiDAR Data. *Electronics* **2025**, *14*, 1821. [CrossRef]
3. Wang, F.; Zhou, G.; Xie, J.; Fu, B.; You, H.; Chen, J.; Shi, X.; Zhou, B. An Automatic Hierarchical Clustering Method for the LiDAR Point Cloud Segmentation of Buildings via Shape Classification and Outliers Reassignment. *Remote Sens.* **2023**, *15*, 2432. [CrossRef]
4. Sun, X.; Guo, B.; Li, C.; Sun, N.; Wang, Y.; Yao, Y. Semantic Segmentation and Roof Reconstruction of Urban Buildings Based on LiDAR Point Clouds. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 19. [CrossRef]
5. Shirowzhan, S.; Trinder, J. Building Classification from Lidar Data for Spatio-temporal Assessment of 3D Urban Developments. *Procedia Eng.* **2017**, *180*, 1453–1461. [CrossRef]
6. Park, Y.; Guldmann, J.-M. Creating 3D city models with building footprints and LIDAR point cloud classification: A machine learning approach. *Comput. Environ. Urban Syst.* **2019**, *75*, 76–89. [CrossRef]
7. Aljumaily, H.; Laefer, D.F.; Cuadra, D.; Velasco, M. Point cloud voxel classification of aerial urban LiDAR using voxel attributes and random forest approach. *Int. J. Appl. Earth Obs. Geoinf.* **2023**, *118*, 3208. [CrossRef]
8. Maltezos, E.; Doulamis, A.; Doulamis, N.; Ioannidis, C. Building Extraction From LiDAR Data Applying Deep Convolutional Neural Networks. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 155–159. [CrossRef]
9. Soilán, M.; Lindenbergh, R.; Riveiro, B.; Sánchez-Rodríguez, A. Pointnet for the Automatic Classification of Aerial Point Clouds. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *IV-2/W5*, 445–452. [CrossRef]
10. Shin, Y.-H.; Son, K.-W.; Lee, D.-C. Semantic Segmentation and Building Extraction from Airborne LiDAR Data with Multiple Return Using PointNet++. *Appl. Sci.* **2022**, *12*, 1975. [CrossRef]
11. Liao, L.; Tang, S.; Liao, J.; Li, X.; Wang, W.; Li, Y.; Guo, R. A Supervoxel-Based Random Forest Method for Robust and Effective Airborne LiDAR Point Cloud Classification. *Remote Sens.* **2022**, *14*, 1516. [CrossRef]
12. Wang, Y.; Jiang, T.; Yu, M.; Tao, S.; Sun, J.; Liu, S. Semantic-Based Building Extraction from LiDAR Point Clouds Using Contexts and Optimization in Complex Environment. *Sensors* **2020**, *20*, 3386. [CrossRef]
13. Meng, Y.; Chen, S.; Liu, Y.; Li, L.; Zhang, Z.; Ke, T.; Hu, X. Unsupervised Building Extraction from Multimodal Aerial Data Based on Accurate Vegetation Removal and Image Feature Consistency Constraint. *Remote Sens.* **2022**, *14*, 1912. [CrossRef]
14. Papon, J.; Abramov, A.; Schoeler, M.; Worgotter, F. Voxel Cloud Connectivity Segmentation—Supervoxels for Point Clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Portland, OR, USA, 23–28 June 2013*; IEEE: Piscataway, NJ, USA, 2013; pp. 2027–2034.

15. Xu, Y.; Tong, X.; Stilla, U. Voxel-based representation of 3D point clouds: Methods, applications, and its potential use in the construction industry. *Autom. Constr.* **2021**, *126*, 3675. [CrossRef]
16. Lin, Y.; Wang, C.; Zhai, D.; Li, W.; Li, J. Toward better boundary preserved supervoxel segmentation for 3D point clouds. *ISPRS J. Photogramm. Remote Sens.* **2018**, *143*, 39–47. [CrossRef]
17. Puligandla, V.A.; Loncaric, S. A Supervoxel Segmentation Method With Adaptive Centroid Initialization for Point Clouds. *IEEE Access* **2022**, *10*, 98525–98534. [CrossRef]
18. Yue, L.; Yun, M.; Han, D. A Novel Roof Detection Algorithm Based SLSC Supervoxel for Airborne LiDAR Point Cloud Data. In *Proceedings of the 2025 IEEE 5th International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 17–19 January 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 186–190.
19. Song, S.; Lee, H.; Jo, S. Boundary-enhanced supervoxel segmentation for sparse outdoor LiDAR data. *Electron. Lett.* **2014**, *50*, 1917–1919. [CrossRef]
20. Sha, Z.; Zhu, Q.; Chen, Y.; Wang, C.; Nurunnabi, A.; Li, J. A Boundary-Enhanced Supervoxel Method for 3D Point Clouds. In *Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–28 September 2017*; IEEE: Piscataway, NJ, USA, 2020; pp. 2771–2774.
21. Xiao, Y.; Chen, Z.; Lin, Z.; Cao, J.; Zhang, Y.J.; Lin, Y.; Wang, C. Merge-Swap Optimization Framework for Supervoxel Generation from Three-Dimensional Point Clouds. *Remote Sens.* **2020**, *12*, 473. [CrossRef]
22. Su, F.; Liu, Y.; Nie, K.; Liu, Y.; Bi, J.; Zhang, R.; Zheng, G. Boundary-Aware Supervoxel Segmentation for Indoor 3D Point Clouds. *IEEE Access* **2023**, *11*, 73478–73489. [CrossRef]
23. Gao, G.; Lauri, M.; Zhang, J.; Frintrop, S. Saliency-guided Adaptive Seeding for Supervoxel Segmentation. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24-28 September 2017*; IEEE: Piscataway, NJ, USA, 2017.
24. Feng, H.; Chen, Y.; Luo, Z.; Sun, W.; Li, W.; Li, J. Automated extraction of building instances from dual-channel airborne LiDAR point clouds. *Int. J. Appl. Earth Obs. Geoinf.* **2022**, *114*, 3042. [CrossRef]
25. Özdemir, E.; Remondino, F. Classification of Aerial Point Clouds with Deep Learning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 103–110. [CrossRef]
26. Duran, Z.; Ozcan, K.; Atik, M.E. Classification of Photogrammetric and Airborne LiDAR Point Clouds Using Machine Learning Algorithms. *Drones* **2021**, *5*, 104. [CrossRef]
27. Du, S.; Zhang, Y.; Zou, Z.; Xu, S.; He, X.; Chen, S. Automatic building extraction from LiDAR data fusion of point and grid-based features. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 294–307. [CrossRef]
28. Lai, X.; Yang, J.; Li, Y.; Wang, M. A Building Extraction Approach Based on the Fusion of LiDAR Point Cloud and Elevation Map Texture Features. *Remote Sens.* **2019**, *11*, 1636. [CrossRef]
29. Stein, S.C.; Schoeler, M.; Papon, J.; Wörgötter, F. Object Partitioning using Local Convexity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23-28 June 2014*; IEEE: Piscataway, NJ, USA, 2014; pp. 304–311.
30. Xu, Y.; Ye, Z.; Yao, W.; Huang, R.; Tong, X.; Hoegner, L.; Stilla, U. Classification of LiDAR Point Clouds Using Supervoxel-Based Detrended Feature and Perception-Weighted Graphical Model. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 72–88. [CrossRef]
31. Li, H.; Ye, C.; Guo, Z.; Wei, R.; Wang, L.; Li, J. A Fast Progressive TIN Densification Filtering Algorithm for Airborne LiDAR Data Using Adjacent Surface Information. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 12492–12503. [CrossRef]
32. Ao, W.; Wang, L.; Shan, J. Point Cloud Classification by Fusing Supervoxel Segmentation with Multi-Scale Features. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W13*, 919–925. [CrossRef]
33. Yan, W.; Guan, H.; Cao, L.; Yu, Y.; Gao, S.; Lu, J. An Automated Hierarchical Approach for Three-Dimensional Segmentation of Single Trees Using UAV LiDAR Data. *Remote Sens.* **2018**, *10*, 1999. [CrossRef]
34. Vo, A.-V.; Truong-Hong, L.; Laefer, D.F.; Bertolotto, M. Octree-based region growing for point cloud segmentation. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 88–100. [CrossRef]
35. Phillips, J.M.; Wang, B.; Zheng, Y. Geometric Inference on Kernel Density Estimates. In *Proceedings of the 31st International Symposium on Computational Geometry (SoCG 2015), Eindhoven, The Netherlands, 22–25 June 2015*; Dagstuhl Publishing: Wadern, Germany, 2015; pp. 857–871.
36. Houlding, S. *Practical Geostatistics: Modeling and Spatial Analysis*; Manual; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2000.
37. Delobel, L.A.; Moffat, D.; Tebbs, E.; Baas, A.C.W. Segmenting and characterising ripple patterns on sand dunes using machine learning and 2D semi-variogram. *Remote Sens. Environ.* **2025**, *331*, 5031. [CrossRef]
38. Mingxin, W.; Chao, F.; Bingbo, G.; Zhoupeng, R.; Fadong, L. A spatial random forest interpolation method with semi-variogram. *Chin. J. Eco-Agric.* **2022**, *30*, 451–457. [CrossRef]
39. Seddiki, A.; Dehimi, S. Effect of choosing a variogram model to predict salinity and its impact on the environment and geotechnical structures. *Tech. Soc. Sci. J.* **2023**, *39*, 860. [CrossRef]

40. Sun, Z.; Xu, Y.; Hoegner, L.; Stilla, U. Classification of Mls Point Clouds in Urban Scenes Using Detrended Geometric Features from Supervoxel-Based Local Contexts. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *IV-2*, 271–278. [CrossRef]
41. Ramiya, A.M.; Nidamanuri, R.R.; Krishnan, R. Segmentation based building detection approach from LiDAR point cloud. *Egypt. J. Remote Sens. Space Sci.* **2017**, *20*, 71–77. [CrossRef]
42. Rottensteiner, F.; Sohn, G.; Jung, J.; Gerke, M.; Baillard, C.; Benitez, S.; Breitkopf, U. The ISPRS benchmark on urban object classification and 3D building reconstruction. In Proceedings of the XXII ISPRS Congress, Melbourne, Australia, 25 August–1 September 2012; Copernicus GmbH: Göttingen, Germany, 2012; pp. 293–298.
43. Open Topography. Available online: https://doi.org/10.5069/G94Q7S68 (accessed on 9 February 2026).
44. International Society for Photogrammetry and Remote Sensing (ISPRS). Vaihingen 3D Semantic Labeling Results. Available online: https://www.isprs.org/resources/datasets/benchmarks/UrbanSemLab/results/vaihingen-3d-semantic-labeling.aspx (accessed on 9 February 2026).
45. Su, Z.; Peng, J.; Feng, D.; Li, S.; Yuan, Y.; Zhou, G. A Building Point Cloud Extraction Algorithm in Complex Scenes. *Remote Sens.* **2024**, *16*, 1934. [CrossRef]
46. Zhao, H.; Jiang, L.; Jia, J.; Torr, P.H.; Koltun, V. Point Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 16259–16268.
47. Liu, K.; Ma, H.; Ma, H.; Cai, Z.; Zhang, L. Building Extraction from Airborne LiDAR Data Based on Min-Cut and Improved Post-Processing. *Remote Sens.* **2020**, *12*, 2849. [CrossRef]
48. Chen, S.; Shi, W.; Zhou, M.; Zhang, M.; Chen, P. Automatic Building Extraction via Adaptive Iterative Segmentation with LiDAR Data and High Spatial Resolution Imagery Fusion. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2081–2095. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.