Coarse-to-Fine Separation of Wood and Leaf From MLS Street Tree Point Clouds Using Branch Tilt Prior and Enhanced Shortest Path Tracing

Yueqian Shen, Shuangshuang Ji, Jinhu Wang, Weidong Liu, Jinguo Wang, Yanming Chen, Zili Deng, Shihan Fu, and Dong Chen, *Member, IEEE*

***Abstract*— Trees play a crucial role in promoting green, ecological, and low-carbon cities, with street trees being essential for urban roadways. Understanding the 3-D structure and biological characteristics of these trees requires accurate separation of wood and leaf. Mobile laser scanning (MLS) technology, known for its high efficiency and resolution, offers significant advantages. MLS data, however, often contain missing or overlapping areas due to occlusions and scanning geometry, complicating precise urban tree modeling. To address these challenges, this article introduces a coarse-to-fine approach for distinguishing wood from leaves in urban street trees. The proposed method begins with a hierarchical workflow that integrates the density-based spatial clustering of applications with noise (DBSCAN) algorithm to identify individual tree nodes. These nodes form the basis for constructing a graph structure for each tree. By leveraging prior knowledge of branch tilt angles, we enhance the shortest path algorithm, facilitating the extraction of features like shortest path frequency and length. This initial step completes a coarse differentiation between wood and leaves. To further refine accuracy, the identified wood and leaf points undergo analysis to extract multiscale geometric features. Integrating these features with the random forest (RF) algorithm results in a more precise separation of wood and leaf points. Our method demonstrates promising segmentation capabilities in MLS-captured roadside trees. Compared to four state-of-the-art methods for wood and leaf separation, our approach shows superior accuracy and efficiency, particularly in accurately identifying trunk points and minor branch points, as well as classifying the outer canopy layer.**

***Index Terms*— Frequency of occurrence, hierarchical separation, mobile laser scanning (MLS), random forest (RF), shortest path feature, wood-leaf separation.**

Received 28 June 2024; revised 1 September 2024; accepted 23 October 2024. Date of publication 1 November 2024; date of current version 18 November 2024. This work was supported in part by the Key Laboratory of Land Satellite Remote Sensing Application, Ministry of Natural Resources of the People's Republic of China under Grant KLSMNR-G202314; and in part by the National Natural Science Foundation of China under Grant 41801379 and Grant 42271450. *(Corresponding author: Dong Chen.)*

Yueqian Shen, Shuangshuang Ji, Jinguo Wang, Yanming Chen, Zili Deng, and Shihan Fu are with the School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China (e-mail: y.shen_lidar@hhu.edu.cn; jss1232021@163.com; wang_jinguo@hhu.edu.cn; chenyanming@hhu.edu.cn; zilideng@126.com; shihanfu@njfu.edu.cn).

Jinhu Wang is with the Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, 1018 WB Amsterdam, The Netherlands (e-mail: jinhu.wang@hotmail.com).

Weidong Liu is with Jiangsu Provincial Surveying and Mapping Engineering Institute, Nanjing 210019, China (e-mail: 442834577@qq.com).

Dong Chen is with the College of Civil Engineering, Nanjing Forestry University, Nanjing 210037, China (e-mail: chendong@njfu.edu.cn).

Digital Object Identifier 10.1109/TGRS.2024.3488696

I. Introduction

THE forest environment stands out as the most significant carbon sink in terrestrial ecosystems. To achieve carbon neutrality, it is essential to focus on enhancing forest carbon sinks [1]. Trees play a critical role in these ecosystems, and increasing their numbers is key to creating green, low-carbon urban landscapes. This strategy not only helps combat climate change but also improves urban environmental quality [2]. Above-ground biomass (AGB) is the main carbon storage in trees, including the total organic matter in both living and deceased forms. This comprises trunks, branches, and leaves, which are crucial for maintaining the equilibrium of the carbon cycle. Trunks and branches act as conduits for water and nutrients, facilitating the transport of essential substances. Leaves, on the other hand, are responsible for key physiological processes such as photosynthesis, respiration, transpiration, and carbon sequestration. The separation of branches and leaves is foundational for computing various metrics that reflect the physiological functions of trees. This includes the calculation of leaf area index and density, crown volume and width, diameter at breast height (DBH), branch length, and branch volume. This process is key for extracting individual tree structures and modeling trees [3], [4], [5], [6].

To bridge the gap between 2-D and 3-D studies of forest ecosystems, Laser Scanning has emerged as a crucial technology for wood and leaf separation. LiDAR is an active remote sensing technology capable of precisely collecting 3-D data for forest monitoring and assessment. Based on the data acquisition platform, LiDAR is typically categorized into three main types: airborne laser scanning (ALS), mobile laser scanning (MLS), and terrestrial laser scanning (TLS) [7]. ALS is renowned for its ability to scan large areas and penetrate the spaces between plant leaves; however, its efficiency is limited in capturing detailed information on individual trees, such as their position at breast height and DBH, as demonstrated by [8] and [9]. In contrast, TLS is well-suited for medium-size forest monitoring and investigation but is often time-consuming. MLS technique involves mounting a LiDAR sensor on a mobile platform, as described [10]. These typical platforms include human-based (handheld or backpack-mounted), vehicle-based, vessel-based, and sled-based platforms. Human-based platforms are carried by individuals, while wheel-based platforms include push carts,

rail-mounted vehicles, motorcycles, bicycles, and motor vehicles. Platforms based on boats and sleds are limited to specific fields and environments, as highlighted in [11], [12], and [13].

Recent advances in using point clouds for wood and leaf separation have opened new possibilities, particularly for carbon sequestration applications. Numerous studies have established benchmarks in this area, primarily focusing on TLS point clouds. TLS research has emphasized excellent data integrity and high accuracy. It is, however, crucial to recognize that MLS data presents notable variations in accuracy and quality compared to TLS data. While TLS has been extensively studied, understanding and addressing the unique challenges posed by MLS data is essential for its effective application in wood and leaf separation methodologies. Significant challenges still exist for wood and leaf separation when using MLS point clouds. These challenges include point cloud variability, limited resolution, noise or outliers, and scanning geometry issues.

In this work, we propose a coarse-to-fine method for wood and leaf separation using MLS point cloud data. The remainder of this article is organized as follows: Section I provides an overview of recent advances in wood and leaf separation using point clouds. Section II gives the details of the proposed methodology, including the tree node acquisition, rough leaf and wood separation, and precise leaf and wood separation. Section III presents experiments and analyses, while Sections IV–V cover the discussion and conclusion.

A. Related Work

Existing wood and leaf separation methods using LiDAR point clouds can be broadly classified into four groups: radiometric feature-based, geometric feature-based, graph-based, and machine learning-based methods.

1) Radiometric Feature-Based Methods: Since most of the LiDAR sensors record the backscattered intensity of the emitted laser pulses, the reflectance differences between wood and leaf can be used for detailed delineation of wood and leaf points [14], [15], [16]. For instance, Beland et al. [17] used the fact that TLS can penetrate the space of leaves, resulting in more echoes from leaves compared to wood areas. Variables like distance, incidence angle, partial occlusion, and varying reflectance qualities of canopy materials can, however, affect intensity values. Danson et al. [18] and Strahler et al. [19] suggested using two different wavelength laser scanners to improve the accuracy of wood and leaf separation. As highlighted in [20] and [21], achieving precise intensity calibration, nevertheless, remains challenging. This difficulty is exacerbated by factors such as instrument characteristics, beam properties, target surface attributes, and incidence angles. The complexity of ensuring accurate system calibration and the diverse equipment configurations further complicates the widespread implementation of this technology, as denoted by [22] and [23].

2) Geometric Feature-Based Methods: Geometric feature-based methods primarily use point location, point distribution, and neighbor relationships to separate wood and leaf points, which usually outperform in terms of accuracy and applicability across various LiDAR platforms [24]. These methods, however, also have limitations, particularly in how features can be influenced by local scale considerations [25], [26]. Current research employs two main approaches to obtain geometric features: computing point-to-neighborhood relationships using either spherical or k-nearest neighbors (KNNs) and subsequently categorizing wood points based on their distinctive geometric properties. Notably, the choice of neighborhood size significantly impacts feature representation. A relatively large and fixed radius neighborhood approach was used by [27], which reveals scattered distribution patterns for leaf points and linear patterns for wood points. Conversely, Li et al. [28] found planar distribution patterns for leaf points and curved surface distribution patterns for wood points with a smaller neighborhood search radius. These variations underscore the significance of neighborhood size in describing and classifying geometric features in tree structures. Moorthy et al. [21] proposed a multiscale neighborhood computation approach to enhance the distinction between wood and leaf points in the tree. This approach improves feature classification across different scales, highlighting the need for sophisticated computational techniques in characterizing complex tree structures. An alternative strategy involves cluster segmentation of point clouds to characterize each cluster's attributes. For example, Ferrara et al. [29] used density-based spatial clustering of applications with noise (DBSCAN) to differentiate tree wood points from the canopy, leveraging local density variations in voxelized oak trees.

Efforts have been made to enhance classification accuracy by combining geometric and radiometric features, which offer more powerful discriminative information between wood and leaf separation. For instance, Sun et al. [6] introduced a three-step framework that integrates intensity and geometric features for the separation of wood and leaf points. Similarly, Tan et al. [30] initially used the intensity-corrected model to preliminarily separate leaf points based on their reflectance differences. Subsequently, they employed density property to further refine the remaining leaf points.

3) Graph-Based Methods: Graph-based methods organize the canopy and wood points into a coherent topological graph and employ graph-based computational geometry methods to distinguish primary wooden branches from smaller leaf structures. For example, Wang [31] used recursive graph partitioning algorithms to differentiate wood and leaf points based on linear features and point density. Another Hui et al. [32] constructed a network graph using shortest path features after applying mean shift clustering to identify tree cluster centers as model points. It is still difficult to use geometric features for segmentation due to the intricate structure of tree canopies. Smaller characteristics of branches might not be immediately apparent, making it challenging to accurately describe the canopy. To produce trustworthy and accurate results, further exploration and refinement of the geometric tree segmentation process are thus necessary. A low frequency of occurrence of leaf nodes at the end of the path and a high path frequency on the main wood nodes are frequently used to estimate the shortest path from the graph. The shortest path well describes the tree skeleton and its affiliated leaves. Vicari et al. [23] established empirical thresholds to separate

wood and leaf points using path tracking and frequency. The work initially constructed a wood skeleton using a shortest path analysis algorithm and then extracted wood points by selecting nearby skeleton points. Additionally, Tian and Li [33] proposed a graph-based leaf-wood separation (GBS) method using only the X-Y-Z coordinates from individual tree points to distinguish wood and leaf points.

4) Machine Learning-Based Methods: More recently, many scholars have attempted to apply machine learning methods to realize wood and leaf point segmentation based on the above-mentioned features; however, the performance of machine learning methods is limited by the lack of training samples, varied classifiers, and feature selection. Some researchers have tried to combine some radiometric features to enhance classification performance, with intensity values proving particularly useful in these efforts [34]. Zhu et al. [35] used seven radiometric features and six geometric features for the task of separation. Their results underscored the greater importance of geometric features over radiometric ones in machine learning-based classification tasks. Windrim and Bryson [36] introduced a novel point-based deep learning architecture for stem segmentation, incorporating intensity information into the learning process. Their experimental findings indicated that the voxel-based deep learning approach, using the intensity information for each point, yielded superior segmentation results. Another Moorthy et al. [21] presented a wood and leaf separation method that combined geometrical features derived from radially bounded nearest neighbors across multiple spatial scales within a machine learning framework. They evaluated the performance of the proposed method using random forest (RF), XGBoost, and other classifiers.

Existing methods for wood and leaf separation using MLS encounter challenges arising from factors such as occlusion, positioning issues, and incomplete scanning in urban street surveys. These issues result in missing, incomplete, and overlapping points, particularly around roadside trees. Despite numerous proposed approaches for wood and leaf separation, several knowledge gaps remain in the MLS domain:

1) *TLS Data-Dominant Wood and Leaf Separation:* Most researches about wood and leaf separation primarily focus on TLS data, which is less challenging compared to MLS point set because TLS yields more uniform and dense point clouds.
2) *Corrupted Separation:* Segmented results are often corrupted by uniformly distributed omission or commission errors in wood and leaf separation tasks.
3) *Scanning Limitations:* MLS scanning mechanisms and angle constraints pose challenges in capturing the complete tree structure. Incomplete point clouds from MLS surveys complicate data processing and algorithm applicability.
4) *Underused Geometric Priors:* For the inventory of street trees, numerous works overlook the diverse geometric priors, which vary in age and species even within the same urban setting.

This work introduces a pipeline for wood and leaf separation using MLS cloud data to address these challenges. The proposed approach incorporates an RF and an enhanced shortest-path algorithm. This combination leverages the growth patterns and specific characteristics of street trees to improve accuracy and robustness in the separation process.

B. Contributions

The main contributions of the proposed method are outlined as follows:

1) *Coarse-to-Fine Wood and Leaf Separation Approach:* We propose a coarse-to-fine method for wood and leaf separation. Initially extracting wood and leaf nodes at a coarse level allows us to emphasize multiscale features. Integration with an RF further refines the separation process, ensuring accurate differentiation of wood and leaf points.
2) *Probability Density-Based Weighted Shortest Path Algorithm:* We present a novel algorithm that uses probability density for weighted shortest paths, enhancing the stability and reliability of the separation results.
3) *Holistic Framework for Wood-Leaf Separation:* We introduce a comprehensive framework that maximizes the use of tree growth characteristics for wood-leaf separation. This approach has demonstrated effectiveness across various scenarios, providing robust probability density functions for each scene.

II. Methodology

A. Overall Framework

This work proposes a method for separating trunks from leaves based on shortest path features and RF classification. The method mainly comprises two parts.

1) *Tree Node Extraction and Graph Creation:* The DBSCAN clustering method is employed to identify cluster center points of trees as tree nodes. Considering the growth angles of wood points, a graph of tree nodes is created. Initially, wood and leaf nodes are roughly extracted by adaptively setting path length and frequency thresholds.
2) *Multiscale Feature Calculation and Classification:* Quantitative features of initial wood and leaf nodes are computed within the remaining nodes to be classified. This process involves multiscale neighborhood analysis to distinguish between wood and leaf points. Fig. 1 illustrates the pipeline of the proposed method, depicting the sequential steps involved in tree node extraction, graph creation, feature calculation, and classification using an RF classifier.

B. Individual Tree Node Acquisition

1) Point Cloud Slicing: As stated by [37], 3-D point clouds of trees reveal distinctive structural variations at different height levels, where leaves often display irregular shapes and wood manifests in circular or semicircular distributions on the projection plane of the xoy. To facilitate clustering of wood and leaf points, we slice the tree points at a fixed height level of dh, and project these sliced points onto the yoz. Assuming

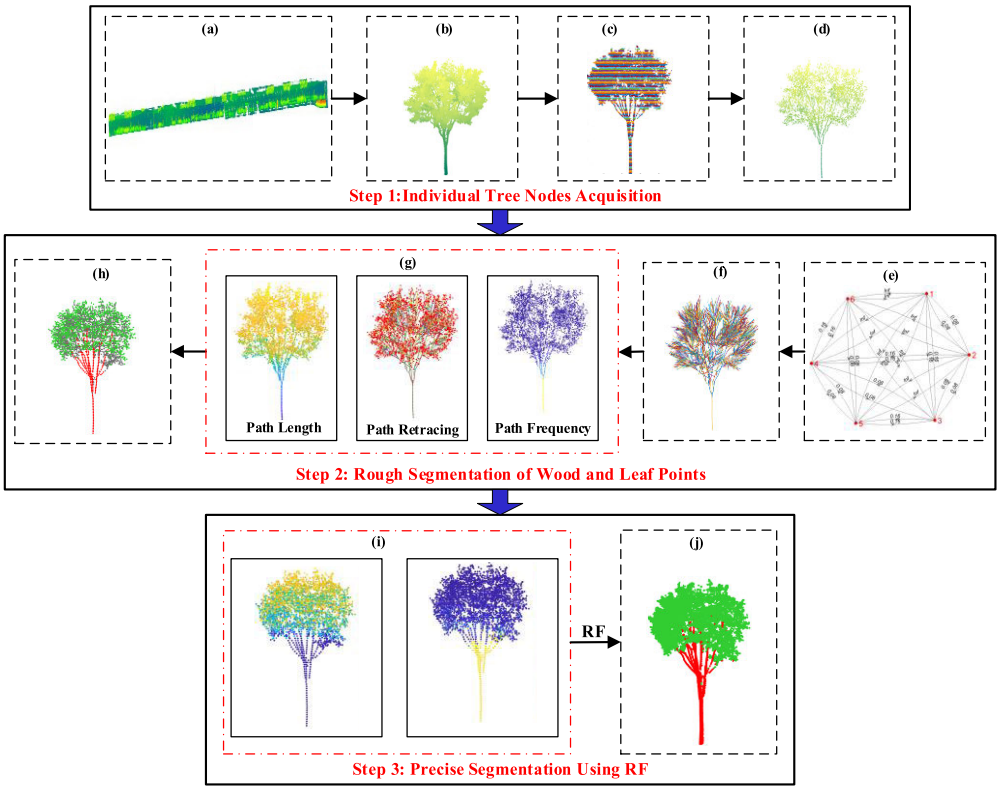


Fig. 1. Overall workflow of the proposed method. (a) Street point cloud acquisition. (b) Single tree segmentation. (c) Point cloud slicing. (d) Trees nodes acquisition. (e) Graph construction. (f) Shortest paths generation. (g) Shortest test path feature. (h) Rough wood and leaf nodes. (i) Multiscale feature calculation.

that the point cloud of one tree contains n points with corresponding coordinates $P_n = \{P_i|(x_i, y_i, z_i), i = 1, 2, \ldots, n\}$. The layer sequence number of any point P_i is determined by

$$L_i = \text{ceil}\left(\frac{Z_i - Z_{\min}}{dh}\right) \tag{1}$$

where L_i is the level number for the point P_i; Z_i refers to the vertical coordinate of the point P_i. $Z_{\min}$ refers to the minimum value of the vertical coordinate of the current tree. dh represents the slice height and ceil$(\cdot)$ means rounding up operation.

2) Nodes Acquisition: As noted in [32], the use of mode points proves effective in the segregation of wood and leaf. This allows the transition from a point-wise separation of wood and leaf to a cluster-wise approach, significantly enhancing the computational efficiency. To achieve this, mode points, referred to as nodes are initially used for the rough separation of wood and leaf. These nodes are identified using the DBSCAN clustering algorithm [38]. Given the variation in point cloud density between the tree trunk and crown, specific neighborhood radii are established accordingly. If the tree point cloud is clustered into n_c clusters, the clustering center M_i of each cluster is estimated as follows:

$$M_i = \left(\sum_{j=1}^{m_c} x_j/m_c, \sum_{j=1}^{m_c} y_j/m_c, \sum_{j=1}^{m_c} z_j/m_c\right) \tag{2}$$

where $i = 1, 2, \ldots, n_c$, and m_c is the total number of points within the ith cluster.

C. Rough Segmentation of Wood and Leaf Points

1) Constructing a Weighted Directed Graph With Consideration of Tilt Angle: When the nodes are acquired, the next step is to construct the graph using these node points. The subsequent task is to generate the shortest path for all nodes to the base node with the lowest elevation. The Dijkstra algorithm is commonly used for generating the shortest path in single source-weighted problems [39]. Here we introduce the Dijkstra algorithm to implement the graph structure based on the tree nodes obtained in Section II-B. Given that the nodes in graphs are typically interconnected, the shortest paths for the notes could be expressed by

$$\text{SP}(G, B, M_C) = \{M_C, \ldots, B\} \tag{3}$$

where G represents the constructed weighted directed graph, B represents the base node, M_C represents a node in the graph, and SP$(\cdot)$ represents the shortest path from the node in the graph to the base node.

To quantitatively assess the growth status and morphological characteristics of roadside trees, a total of 2577 branches from 180 trees were segmented using the CloudCompare open-source tool.[1] The tilt angles of each branch were then calculated by

$$\text{angle}\left(M_{ij}, M_z\right) = a\cos\left(\frac{\text{dot}(M_{ij}, M_z)}{|M_{ij}||M_z|}\right) \tag{4}$$

[1]https://www.danielgm.net/cc/

where M_{ij} represents the node vector from M_i to M_j, and M_z is the 3-D zenith direction vector, denoted as $M_z = (0, 0, 1)$. The symbol dot$(\cdot)$ denotes the inner product of the two vectors, $a\cos(\cdot)$ represents the angle between the two vectors, and angle$(\cdot)$ is the angle value between the edge M_{ij} and the M_z. The probability density function based on branch tilt is constructed in Section IV-B.

As the sample size increases, the tilt angles of the branches more accurately reflect the actual growth conditions of the trees. In this article, a total of 180 roadside trees were selected as samples. The tilt angles of all branches were calculated using (4), and a branch angle frequency histogram was generated at an interval of 1°. Through extensive sample analysis, it has been observed that the distribution of the branch tilt angles approximately follows a double Gaussian distribution. The probability density of any given title angle can, therefore, be estimated as follows:

$$F\left(x_{\text{angle}}\right) = a_1 \cdot e^{-\left(\frac{x_{\text{angle}}-b_1}{c_1}\right)^2} + a_2 \cdot e^{-\left(\frac{x_{\text{angle}}-b_2}{c_2}\right)^2} \tag{5}$$

where x_{angle} is the tilt angle value and $F(\cdot)$ is the probability value corresponding to that angle. a_1 and a_2 are amplitude coefficients, b_1 and b_2 are mean parameters, and c_1 and c_2 are standard deviation parameters.

The procedures for generating the shortest path using tilt angle probability weighting are summarized as follows:

1) The Euclidean distance between two adjacent nodes, M_i and M_j is given by $E(M_i, M_j)$.
2) The tilt angle angle(M_{ij}, M_j) is calculated using (4).
3) The tilt angle is weighed by the angle probability value F, which is calculated as follows:

$$\text{angle}^1\left(M_{ij}, M_z\right) = F \cdot \text{angle}\left(M_{ij}, M_z\right) \tag{6}$$

where angle$^1(\cdot)$ represents the weighted angle value, and angle$(\cdot)$ represents the angle between the edge M_{ij} and the zenith direction M_z.

4) The Euclidean distance and the weighted angle value between two nodes are normalized as follows:

$$\begin{cases} E_{\text{norm}}\left(M_i, M_j\right) = \dfrac{E\left(M_i, M_j\right)}{\max\left(E\left(M_i, M_j\right)\right)} \\ \text{angle}^1_{\text{norm}}\left(M_{ij}, M_z\right) = \dfrac{\text{angle}^1\left(M_{ij}, M_z\right)}{\max\left(\text{angle}^1\left(M_{ij}, M_z\right)\right)} \end{cases} \tag{7}$$

where $E_{\text{norm}}(\cdot)$ represents the normalized distance weight, $\text{angle}^1_{\text{norm}}(\cdot)$ represents the normalized angle weight, and $\max(E(\cdot))$ is the maximum distance weight, and $\max(\text{angle}^1(\cdot))$ is the maximum angle.

5) Generate weights by a combination of normalized Euclidean distance and normalized tilt angle probability. To avoid negative weights, an additional coefficient μ,which generates the smallest normalized Euclidean weight, is introduced as follows:

$$\begin{cases} w\left(M_i, M_j\right) = E_{\text{norm}}\left(M_i, M_j\right) \\ \qquad\qquad - \mu \cdot \text{angle}^1_{\text{norm}}\left(M_{ij}, M_z\right) \\ \text{s.t. } \mu = \min\left(E_{\text{norm}}\left(M_i, M_j\right)\right). \end{cases} \tag{8}$$

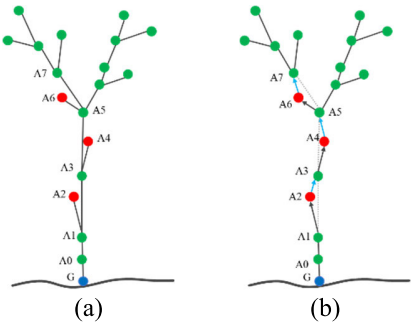


Fig. 2. Effectiveness of hybrid weights. When only normalized Euclidean weights are considered, twig artifacts, marked by red circle places subfigure (a) appear; however, using the proposed hybrid weights in subfigure (b) helps remedy these artifacts. Note that the circle symbols represent the generated wood and leaf nodes. The red circles indicate areas where edge connection errors occur, while the blue circles denote the base nodes of the tree trunk.

Based on (8), the final weight matrix for any two nodes is generated as follows:

$$\begin{cases} W_{\text{edge}}\left(M_i, M_j\right) = \begin{cases} w\left(M_i, M_j\right), & \text{if } w\left(M_i, M_j\right) \le r \\ \text{inf}, & \text{otherwise} \end{cases} \\ \text{s.t. } r = \text{mean}\left(w\left(M_i, M_j\right)\right) - 2 \cdot \text{std}\left(w\left(M_i, M_j\right)\right) \end{cases} \tag{9}$$

where $W_{\text{edge}}(\cdot)$ represents the final edge weight between two nodes M_i and M_j, mean$(\cdot)$ represents the average weight value, std$(\cdot)$ represents the standard deviation of the weight, and r is the edge weight threshold. Once the weight matrix between nodes has been constructed using (9), a single-source weighted directed graph can be easily created using the topological connection relationships preserved in the weight matrix.

We provide a detailed schematic in Fig. 2 to illustrate the effectiveness of our hybrid node weight $W_{\text{edge}}(\cdot)$, which combines normalized Euclidean and tilt angle weights as described in (9). As shown in Fig. 2(a), when only the normalized Euclidean weight is used and the normalized tilt angle weight is ignored, the shortest path from node A7, for example, follows the sequence A7-A5-A3-A1-A0-G. This path bypasses nodes A6, A4, and A2, leading to a significant deviation from the true tree skeleton by inaccurately representing the growth of twigs (edges from A3 to A4 and from A1 to A2) originating from the trunk. Incorporating the normalized tilt angle weight, however, helps avoid these twig artifacts and results in a more accurate representation of the trunk shapes, as evident in Fig. 2(b). Our hybrid weight representation thus captures better the intrinsic growth patterns of tree branches and aligns the tree skeleton more closely with its actual shape.

The algorithm calculates the mean and standard deviation of all edge weights. Subsequently, the edge weight threshold is determined by subtracting twice the standard deviation from the mean value. The effects of different standard deviations on graph construction are discussed in Section IV-C. For a detailed representation of the algorithm, refer to the pseudocode displayed in Table I.

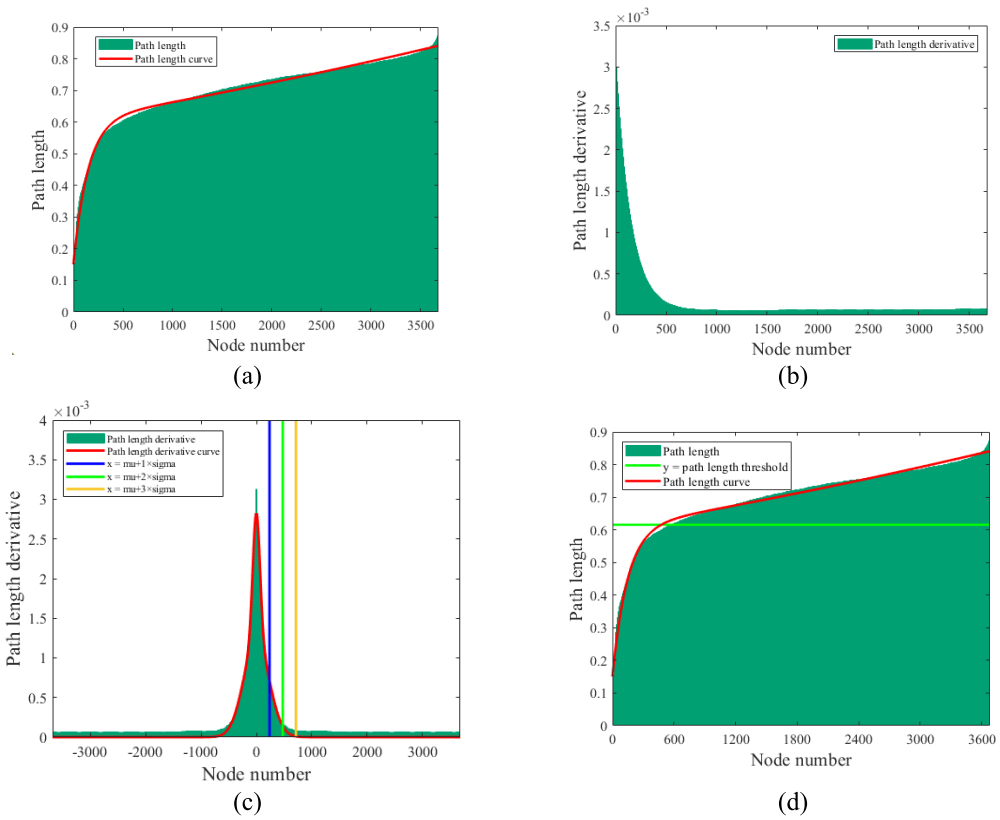


Fig. 3. Determine the path length threshold extracted by the initial wood node. (a) Histogram of path length. (b) Histogram of first-order derivatives of path length. (c) Symmetrical treatment along ordinates. (d) Path length threshold.

TABLE I
EVALUATION INDEX CALCULATION RESULTS OF THE PROPOSED METHOD (UNIT: %)

Accuracy	F1-score for wood	F1-score for leaf	Kappa
97.08	89.53	98.29	87.82

2) Rough Detection of Leaf Nodes: As shown in Fig. 2, leaf nodes are generally located at the end of each path. This implies that most paths must traverse through wood nodes to reach the base node, meaning fewer paths pass through leaf nodes compared to wood nodes. To detect leaf nodes, an idea called pass retracing proposed in [32] is employed, which involves continuously deleting nodes from the end node of each path to the base node. Given the leaf nodes are typically end nodes, we retrace one step for each path. Assuming the nodes in the shortest path are represented by $\{M_n, M_{n-1}, \ldots, M_1, B\}$, after one step path retracing, the nodes will be $\{M_{n-1}, \ldots, M_1, B\}$, efficiently removing the end node $\{M_n\}$. This operation is implemented to all nodes, allowing the leaf nodes to be identified after a single step of shortest path tracing.

$$\{\text{leaf}\} = \text{diffset}\{M_c', V\} \tag{10}$$

where {leaf} represents the obtained preliminary leaf node, V is the set of tree nodes, M_c' is the node set remaining after the last node is deleted for all paths, and diffset$\{\cdot\}$ represents the point belonging to node V but not to node M_c'.

By backtracking one step, the resulting nodes include not only the leaf nodes but also the branch nodes. Notably, the leaf nodes obtained through the aforementioned method are rather coarse, necessitating further refinement to enhance their accuracy. It should be noted that nodes located beneath the canopy may be incorrectly classified as leaf nodes due to the absence of leaves in that area.

Recognizing that paths of nodes beneath the canopy are anticipated to be generally short, we propose setting a path length threshold after tracing the path back one step to capture the leaf nodes. The procedures for this detection are summarized as follows:

1) After generating the shortest paths for all nodes, the path lengths can be estimated using histogram analysis. As shown in Fig. 3(a), the histogram exhibits the frequency of path lengths among the tree nodes. The red curve representing the path length frequency is fit using a least squares regression algorithm [40]. We can clearly see that the curve initially monotonically increases before flattening out.
2) It is a fact that the path length of the leaf nodes is greater than that of the wood nodes; therefore, it needs to determine the elbow point on the red curve in Fig. 3(a) to effectively separate the leaf nodes from the wood nodes. To achieve this, we differentiate the distribution function curve to obtain the probability density distribution, as illustrated in Fig. 3(b).
3) It is apparent that the rate of frequency change in Fig. 3(b) obeys a half-normal distribution. Symmetric

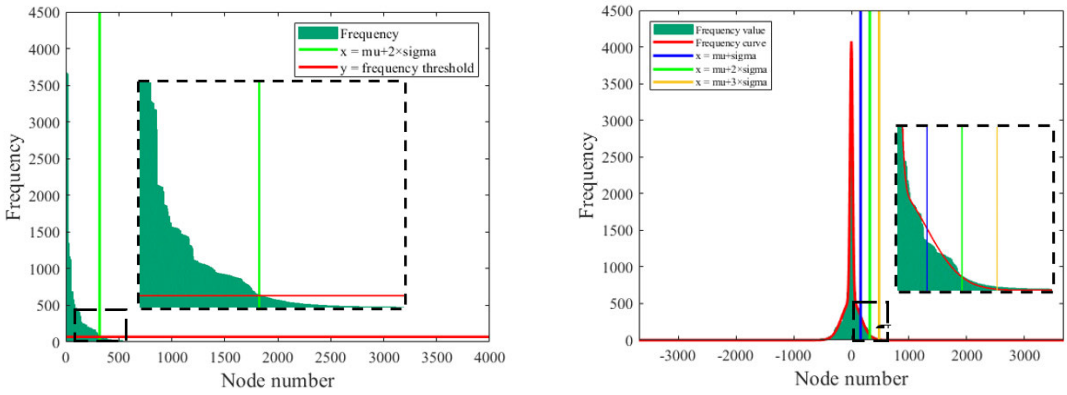


Fig. 4. Frequency of occurrence threshold determination. (a) Histogram of frequency. (b) Gaussian fitting acquisition threshold.

multiprocessing along the y-axis is applied to transform the half-normal distribution into a symmetric normal distribution, as depicted in Fig. 3(c). The Gaussian distribution is then employed to fit the symmetric normal distribution (mirrored frequency change rate curve), and the corresponding μ and σ are estimated. As shown in Fig. 3(c), the blue, green and yellow lines represent the positions of $\mu+\sigma\,\mu+2\sigma$, and $\mu+3\sigma$, respectively. As the frequency change rate stabilizes the position of $\mu+2\sigma$, the value is chosen as the elbow value for separating the wood nodes and leaf nodes.

4) When the elbow point value is determined, the corresponding path length frequency can be estimated using the Gaussian probability density function [see Fig. 3(c)]. The resulting blue line in Fig. 3(d) represents the elbow line used to separate the leaf nodes from the wood nodes.

As such, the path length threshold is determined by eliminating all remaining duplicate nodes from the path. Nodes exceeding this threshold are identified as the leaf nodes initially removed and then restored to obtain the first set of leaf points.

3) Rough Detection of Wood Nodes: Since the paths of leaf nodes always pass through wood nodes, the access frequency of wood nodes is higher compared to that of leaf nodes. A frequency threshold can, therefore, be set to separate leaf nodes from wood nodes. The specific procedures are outlined as follows.

1) The frequency of occurrence of each node is calculated, and the corresponding histogram is generated.
2) The distribution of frequency of occurrence appears to follow a half-normal distribution. Similar to the path length frequency, symmetric multiprocessing along the y-axis is implemented, as illustrated in Fig. 4. Next, the Gaussian distribution is applied to fit the mirrored frequency of occurrence of the change rate curve, and the corresponding μ and σ are estimated. As the frequency of occurrence tends to be stable from the position of $\mu+2\sigma$, this value is chosen as the threshold for distinguishing between the wood and leaf nodes.

The nodes that surpass the frequency threshold are considered the initially extracted wood nodes, leveraging the characteristic that wood nodes typically have a significantly higher access frequency than certain leaf nodes. These nodes are then restored to obtain the initial set of wood points.

D. Precise Segmentation Using RF

Once the wood and leaf nodes are roughly acquired, they serve as the basis for obtaining local geometric features of the leaf and wood nodes, respectively. These features are subsequently fed into the RF classifier for accurate segmentation of leaf nodes and stem nodes. The process involves two main steps: multiscale features calculation and selection, and RF classification, which are described below.

1) Multiscale feature calculation and selection. Wood nodes are linear with a low density, and their geometric features are not well described by spherical or cylindrical neighborhoods. As illustrated in [41], KNN shows more flexibility with large variations in point density. To capture the multiscale characteristics of the data, we perform KNN searches across multiple spatial scales, using K values of 20, 40, 60, and 80, respectively. This approach allows us to comprehensively describe local geometric features by analyzing them at different scales. Due to the relatively low density of tree nodes, features are less distinguishable when the K value is small. As the K value increases, the variation in features diminishes. Among these, the number of tree nodes within the neighborhood of each tree node is calculated as the feature, including the initial wood and leaf nodes. Note that we consider the neighborhoods of nodes rather than the neighborhoods of points.
2) RF classification. RF has been wildly applied in various fields, including classification, regression, and other tasks, which shows better performance compared to other machine learning algorithms such as naive Bayes and neural networks [21]. Using the multiscale features constructed as described above, and combining them with RF classifier, the wood and leaf nodes are segmented precisely.

III. Results and Validations

In this section, extensive experiments are performed to evaluate the proposed method, and the performance of key components is analyzed in detail.

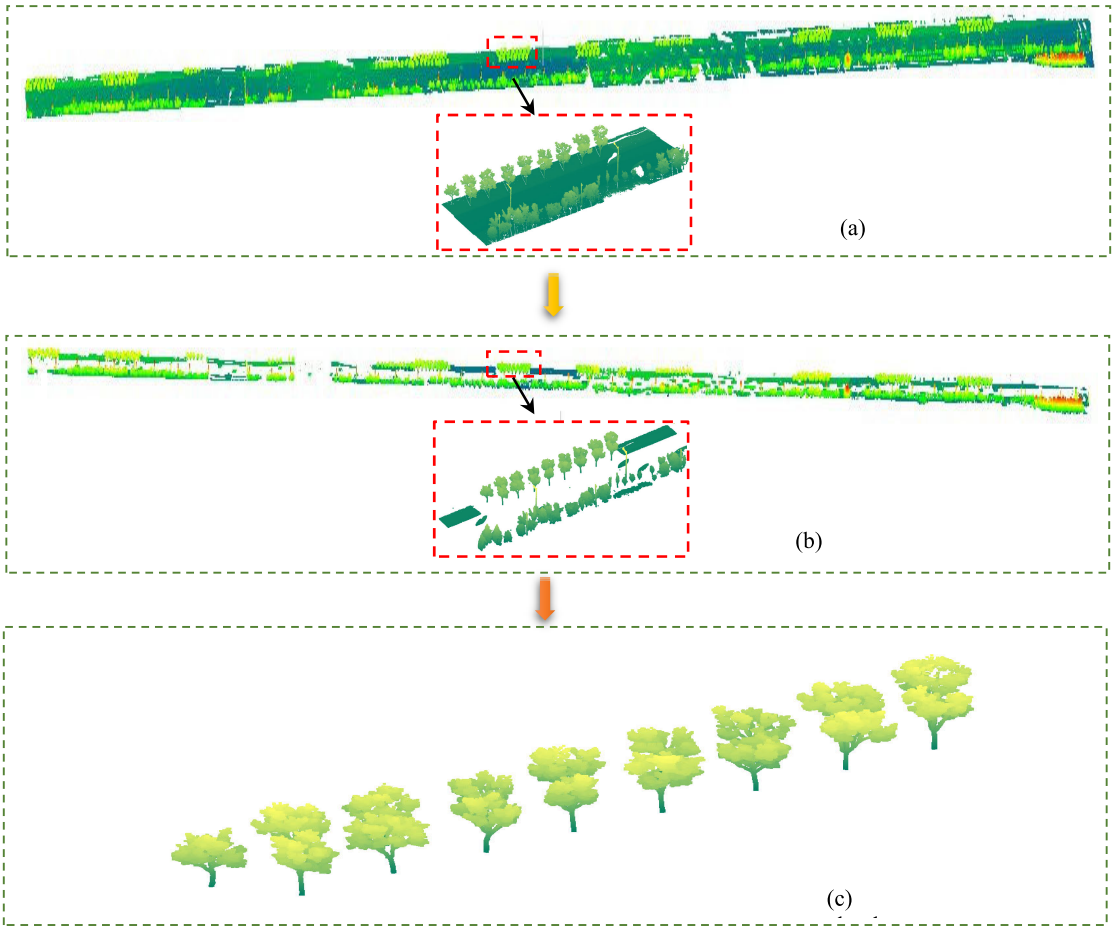


Fig. 5. Instance-level tree segmentation. (a) Street point. (b) Then nonground point cloud. (c) Single tree point.

A. Data Description

The proposed methodology was validated using point cloud data acquired by MLS. The scan was conducted in 2020 in Beijing. The data has a volume point density of approximately 600 pts/m^3, totaling 125 457 788 points over a length of 2104 m. Nonground points were extracted using the cloth simulation filter (CSF) [42], and point clouds for 180 trees were obtained using CloudCompare software. Fig. 5 illustrates the process of extracting the individual tree points.

Ground truth points representing wood and leaf for each tree were also manually segmented which can be used for evaluating the performance of the proposed methodology. The heights of 180 tree samples range from 4.29 to 9.49 m, with the number of points in each tree varying from 4786 to 97 683. These characteristics are crucial for assessing the stability across various point densities and its effectiveness in handling various tree heights.

B. Validation Metrics

In this work, three key metrics, accuracy, $F1$-score, and Kappa coefficient, were calculated to provide a quantitative evaluation of the proposed methodology.

Accuracy demonstrates the proportion of correctly classified points relative to the total number of points, which can be calculated using the following equation:

$$\text{Accuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{11}$$

where T_P represents the number of correctly extracted wood points, F_P represents the number of wrongly extracted wood points, i.e., leaf points are extracted as wood points, T_N is the number of correctly extracted leaf points, and F_N is the number of wrongly extracted leaf points, i.e., wood points are extracted as leaf points.

$F1$-score can be calculated using precision (P) and recall (R) according to (12). Precision (P) represents the ratio of the correctly classified true positive samples to the total classified positive samples, while recall (R) represents the ratio of correctly classified true positive samples to the reference positive class

$$\begin{cases} F_1\text{-score} = 2 \times \dfrac{P \times R}{P + R} \\ P = \dfrac{T_P}{T_P + F_N} \\ R = \dfrac{T_P}{T_P + F_P} \end{cases} \tag{12}$$

where P is the accuracy rate, which indicates the proportion of points that are correctly classified as wood points or leaf points. R is the recall, which indicates the proportion of tree

point cloud data that is finally correctly classified as wood points or leaf points.

The Kappa coefficient is a crucial metric for assessing the accuracy and dependability of classification results according to (13), the values of which are typically between 0 and 1, where 0∼0.2 denotes extremely low consistency, 0.21∼0.40 denotes general consistency, 0.41∼0.60 denotes medium consistency, 0.61∼0.80 denotes high consistency, and 0.81∼1 denotes almost complete consistency.

$$\begin{cases} \text{Kappa} = \dfrac{(\text{po-pe})}{(1-\text{pe})} \\ \text{po} = \dfrac{T_P + T_N}{N} \\ \text{pe} = \dfrac{(T_P + T_N) \times (T_P + F_P) + (F_P + T_N) \times (F_N + T_N)}{N \times N} \end{cases} \tag{13}$$

where, po is proportional consistency, pe is random consistency, and N is the number of points.

To further analyze the performance of the proposed methodology in achieving wood and leaf separation, the Type I error (T_1) and the Type II error (T_2) are used to evaluate the segmentation results. Type I error is also known as omission error, which refers to the proportion of wood points wrongly identified as leaf points. Type II error is also known as redundancy error, which refers to the proportion of leaf points wrongly identified as wood points [43]. These errors can be calculated using the following equation:

$$\begin{cases} T_1 = \dfrac{F_N}{F_N + T_P} \\ T_2 = \dfrac{F_P}{F_P + T_P}. \end{cases} \tag{14}$$

C. Parameters Illustration

Four main parameters are involved in this article: the slice height dh as described in (1), the DBSCAN clustering radius, the minimum number of points, and the edge weight threshold r as described in (9). Notably, all tests using the proposed method implemented the same fixed parameter values. The slice height and clustering radius are critical for identifying tree-mode points, with the slice height being smaller than the tree height. This value determines both the quantity and quality of the tree nodes obtained. Specifically, if the slice height is too large, it results in fewer tree nodes when applying DBSCAN clustering. Conversely, if the threshold is set too low, the number of clusters will be excessively high, yielding a result similar to processing the points directly; therefore, the slice height has been empirically determined to be 0.08 m. Given the presence of isolated points and sparse point clouds in the canopy area, the minimum number of points in each cluster is set to 2. Specific details regarding the selection of clustering radius and edge weight threshold are provided in Sections IV-A and IV-C, respectively.

D. Quantitative Analysis

Generally, the number of leaf points is significantly greater than that of wood points for a single tree. A smaller number of mis-segmentation points will not, therefore, significantly affect the $F1$-score and accuracy of leaf points, but it will greatly reduce the $F1$-score and Kappa coefficient of wood points; hence, the $F1$-score and Kappa coefficient of wood points can better reflect the performance of the wood-leaf separation. The four accuracy indicators, accuracy, $F1$-score for wood, $F1$-score for leaf, and Kappa, of the proposed methodology are calculated, the results of which are shown in Table I. It can be observed that all the accuracies of the 30 tree samples are above 94%, with a maximum value of 98.92% and a minimum value of 94.74%. The $F1$-score for wood points and leaf points ranged from 80.75% to 94.65% and 96.68% to 99.42%, respectively. The Kappa coefficients varied from 78.09% to 93.67%. This demonstrates that the proposed method achieves consistent results for wood and leaf separation. The mean accuracy, $F1$-score for wood points, $F1$-score for leaf points, and Kappa are approximately 97.08%, 89.53%, 98.29%, and 87.82%, respectively. Considering the significant variation in point density among the 30 trees, the findings indicate that the proposed method demonstrates a favorable ability to separate the wood and leaf points in the street trees with different densities.

To further evaluate the performance of the proposed method, we compared it against four other state-of-the-art approaches for wood and leaf separation: Graph-based GBS, Leaf-wood separation (LeWos), CANUPO, and RF modelRF. GBS, proposed in 2022 by [33], relies solely on xyz information of individual trees and effectively uses shortest path characteristics for point cloud segmentation, cluster identification, and regional growth across various tree species and sizes. LeWoS, introduced in 2020 by [44], is an unsupervised classification method based on geometric features and effectively uses shortest path characteristics for point cloud segmentation, cluster identification, and regional growth across various tree species and sizes. The CANUPO model [45] is another prominent machine learning technique that captures multiscale geometric features for accurate wood and leaf point segmentation. The improved RF model [21] uses multiscale local geometric features as predictive variables. This model was trained using TLS data from five plots and two single trees located in two different tropical forest locations. These four methods were tested on the same 30 tree samples to benchmark their performance against the proposed methodology.

The accuracy indicators, i.e., accuracy, $F1$-score for wood, $F1$-score for leaf, and Kappa, for all 30 trees using the abovementioned four methods were calculated. Fig. 6(a)–(d) presents the results for accuracy, $F1$-score for wood, $F1$-score for leaf, and Kappa, respectively, for these methods. From Fig. 6(a) and (b), it is evident that the RF model exhibits the lowest stability, displaying the largest fluctuation range among the accuracy indicators. Conversely, the proposed method demonstrates not only a small fluctuation range but also achieves the highest accuracy across all metrics.

The average values of the four accuracy indicators, i.e., accuracy, $F1$-score for wood, $F1$-score for leaf, and Kappa, for the abovementioned four methods, are shown in Table II. It is evident that the proposed method is 97.08%, slightly surpassing the GBS method, which has an accuracy of 96.61%.

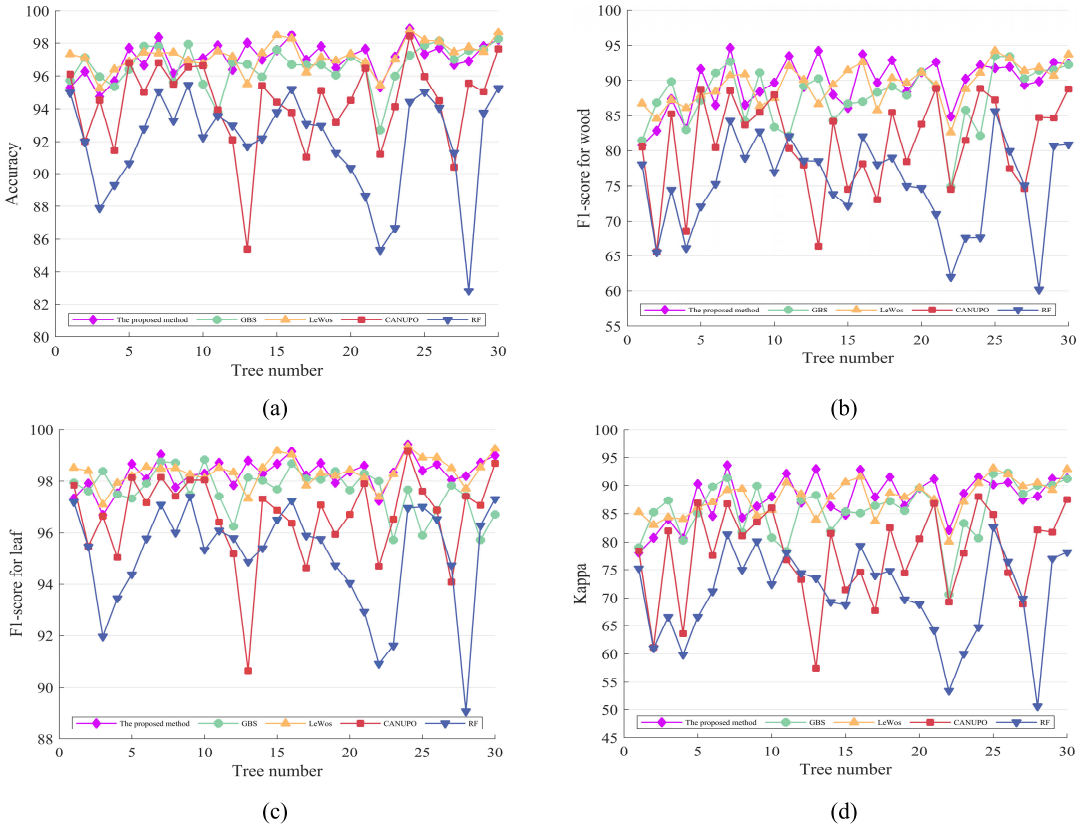


Fig. 6. Accuracy assessments of 30 single trees using different methods. (a) Accuracy. (b) $F1$-score for wood. (c) $F1$-score for leaf. (d) Kappa.

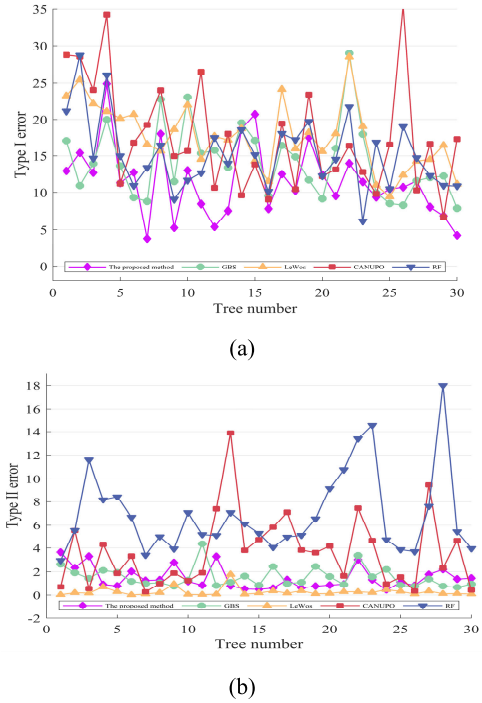


Fig. 7. Errors for 30 trees using four segmentation methods. (a) Type I error. (b) Type II error.

The proposed method, furthermore, significantly outperforms the CANUPO and RF model methods, which exhibit accuracies of 94.32% and 91.93%, respectively. In terms of $F1$-score for wood, the proposed method is significantly superior to the other methods. Noteworthy, it showcases an enhancement of at least nine percentage points compared to CANUPO and RF model methods. The results indicate that the proposed method is more effective in extracting wood points. For the $F1$-score for leaf, the values for the five methods are 98.29%, 98.02%,98.35%, 96.63%, and 95.12%, respectively. This reveals that the proposed method exhibits a slight enhancement in extracting leaf points. LeWoS achieves the highest value for the $F1$-score for leaf points, indicating it is particularly effective in extracting foliage points.

To further analyze the segmentation errors, the Type I and Type II errors were calculated for these four segmentation methods, as shown in Table III. Overall, Type I errors are higher than Type II errors across all methods, which demonstrates that the abovementioned methods exhibit better performance in extracting leaf points than wood points. Clearly, the proposed method achieved the lowest average for Type I error at 11.59%, whereas the other four methods yielded values of 14.25%, 17.62%, 17.53%, and 15.33%, respectively. This indicates that the proposed method is more effective in detecting wood points compared to the other methods. In terms of Type II error, the proposed method, GBS, LeWoS, CANPUO, and RF model exhibit error rates of 1.48%, 1.55%, 0.28%, 3.69%, and 6.91%, respectively. These results indicate that the proposed method performs worse than the LeWoS model in terms of Type II error performance; however, it shows competitive performance with the GBS method. Importantly, the proposed method significantly outperforms CANPUO and the RF model, which exhibit a higher error rate of 3.69% and 6.91%, respectively.

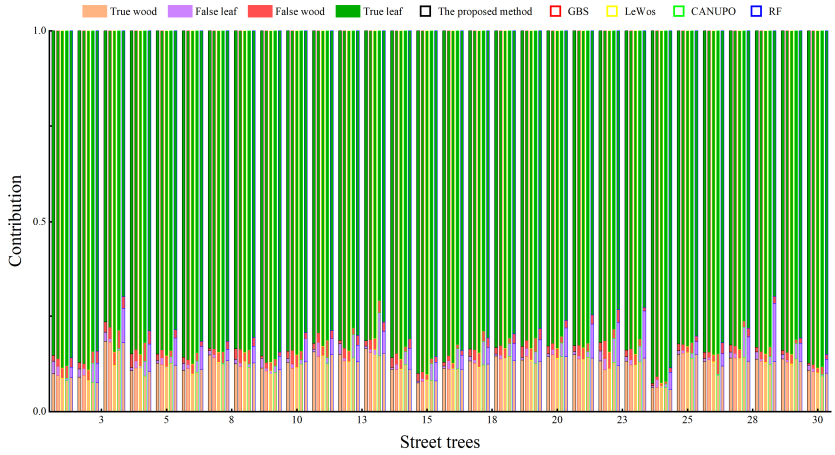


Fig. 8. Distribution of segmenting results of 30 street trees obtained using five methods.

TABLE II

MEAN VALUE OF ASSESSMENT METRICS USING DIFFERENT METHODS (UNIT: %)

Methods	Accuracy	F1-score for wood	F1-score for leaf	Kappa
The proposed method	97.08±2.33	**89.53±8.78**	98.29±1.61	**87.82±9.72**
GBS	96.61±3.91	**87.70±12.83**	98.02±2.29	**85.52±15.10**
LeWoS	96.22±1.97	89.41±6.79	98.05±1.28	87.82±7.74
CANUPO	94.32±8.93	**80.94±15.31**	96.63±5.97	**77.63±20.14**
RF model	91.93±9.10	**75.30±15.13**	95.12±6.06	**70.57±20.00**

TABLE III

COMPARISON OF TYPE I ERROR AND TYPE II ERROR (UNIT: %)

Tree samples	Type I error					Type II error				
	The proposed method	GBS	LeWoS	CANUPO	RF model	The proposed method	GBS	LeWoS	CANUPO	RF model
AVG	11.59	14.25	17.62	17.53	15.33	1.48	1.55	0.28	3.69	6.91

TABLE IV

SAMPLE BRANCHES STRIPS AND ANGLE VALUE INFORMATION (ANGLE VALUE UNIT: °)

Number of sample	**Number of branches**	**Minimum angle value**		**Maximum angle value**		**Mean angle value**		**Maximum frequency angle value**	
Angle interval		**0.5**	**1.0**	**0.5**	**1.0**	**0.5**	**1.0**	**0.5**	**1.0**
30	512	1.0	1.0	74.5	74.0	25	25	14.5	14.0
60	937	1.0	1.0	74.5	74.0	23	23	14.5	17.0
90	1,365	0.0	0.0	74.5	74.0	23	23	14.5	17.0
120	1,737	0.0	0.0	78.0	78.0	24	24	14.5	15.0
150	2,099	0.0	0.0	82.0	82.0	24	24	14.5	17.0
180	2,577	0.0	0.0	82.0	82.0	24	24	17.5	17.0
AVE	**1,538**	**0.3**	**0.3**	**77.6**	**77.3**	**23.8**	**23.8**	**15.0**	**16.2**

The LeWoS method, furthermore, exhibits the highest Type I error, indicating its limited effectiveness in accurately identifying wood points; however, it achieves the lowest average Type II error value, highlighting its proficiency in extracting leaf points. Comparatively, the proposed method shows similar performance to the GBS method, with Type I error rates of 1.483% and 1.55%, respectively. It is, however, notably inferior to the other two methods, which report higher Type I error rates of 3.69% and 6.91%, respectively. The proposed method, furthermore, efficiently maintains a balanced performance between the Type I and Type II errors, reflecting its robustness in segmentation tasks.

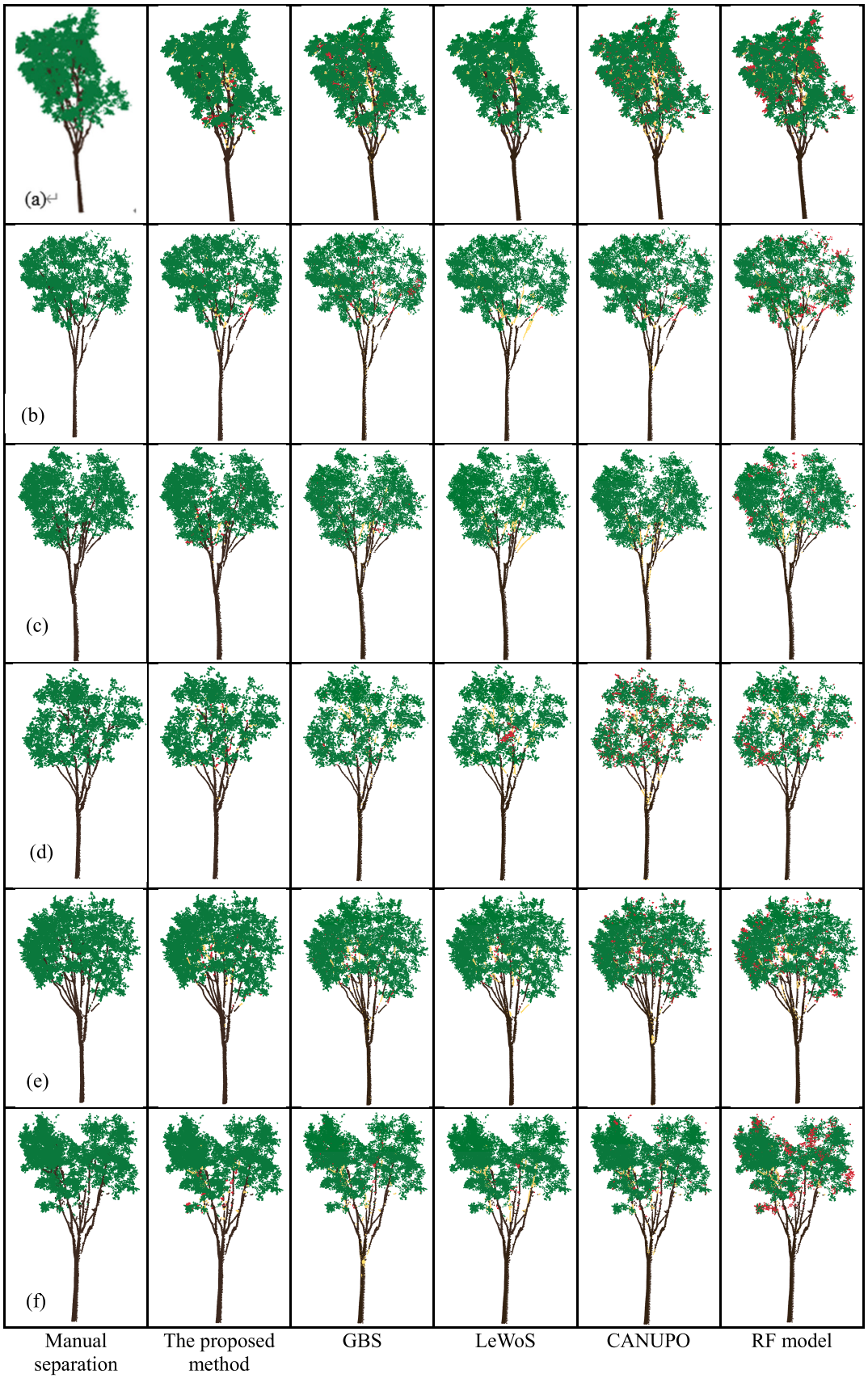


Fig. 9. Woods and leaves separation errors in six street trees. (a) Tree 4. (b) Tree 5. (c) Tree 7. (d) Tree 13. (e) Tree 18. (f) Tree 21.

For a deeper insight into the performance of different methods across various trees, Fig. 7(a) and (b) depict the Type I and Type II errors for the 30 trees, respectively. It is evident from Fig. 7(a) and (b) that the proposed method showcases

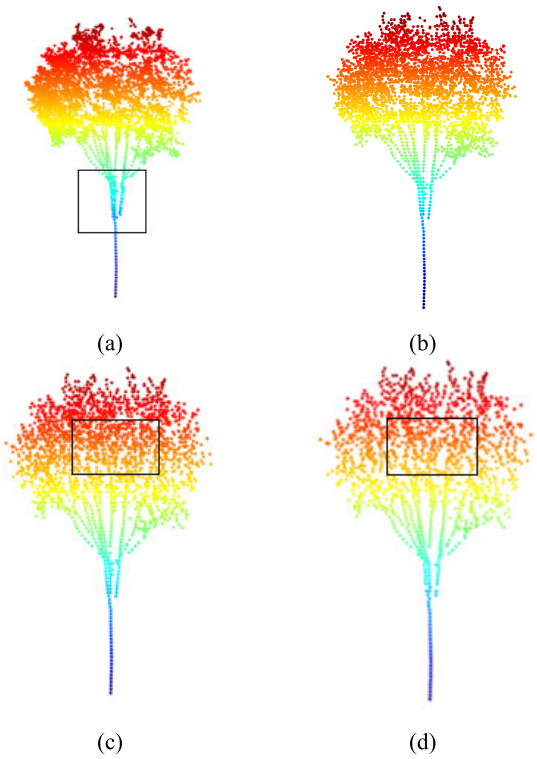


Fig. 10. Tree nodes generated with different coefficient λ. (a) $\lambda = 0.25$. (b) $\lambda = 0.5$. (c) $\lambda = 0.75$. (d) $\lambda = 1$.

higher stability across different trees in terms of both Type I and Type II errors compared to the other four methods. This observation highlights the proposed method's versatility across various densities and tree heights.

E. Qualitative Analysis

Fig. 8 illustrates the distribution of segmenting results for 30 street trees using the five mentioned methods, while Fig. 9 presents the visualization of segmenting results for six trees with varying point densities. These figures reveal certain tendencies among the methods: LeWoS and CANUPO models often misidentify wood points as leaf points, whereas the RF model tends to misclassify leaf points outside the canopy as wood points. In contrast, the outcomes from the proposed method consistently show fewer incorrectly labeled wood or leaf points compared to the GBS method across most cases. When optimizing canopy shape adjustments, the proposed method, furthermore, effectively and thoroughly extracts wood points, resulting in finer and more distinguishable branches.

IV. Sensitivity Analysis

A. Selection of DBSCAN Clustering Radius

The clustering radius chosen for DBSCAN plays a crucial role in determining the number of tree nodes generated. If the clustering neighborhood radius is too small, it can result in an excessive number of tree nodes, leading to increased computational workload and storage requirements during the graph construction process. Additionally, a small clustering radius may result in imprecise branch segmentation, making it challenging to distinguish between different tree branches. Conversely, if the clustering neighborhood radius is too large, it might merge distinct tree branches together, resulting in inaccuracies in representing tree morphology. This approach, moreover, may overlook certain details and small branches, making it difficult to extract comprehensive structural information about the trees. Selecting an appropriate clustering radius for the tree canopy is, therefore, crucial to ensure that the obtained tree nodes accurately reflect the original morphology of the trees. To address this issue, we propose using distinct clustering radii for the main trunk and the tree canopy, respectively. The procedures contain two steps.

1) Main trunk radius acquisition. After completing hierarchical processing on individual trees, the point cloud is projected onto the XOY plane. Subsequently, the Least Squares algorithm is applied to fit circles for different layers within the tree trunk, yielding corresponding radii. Among these radii, the largest one is selected as the clustering neighborhood radius for the trunk.
2) Canopy clustering radius determination. Using the trunk clustering radius as a reference, select a scaling factor and calculate the canopy clustering radius using Formula 18. To illustrate the process of coefficient determination, we use Tree 18 as an example. Scaling factors λ of 0.25, 0.5, 0.75, and 1, were selected, resulting in the acquisition of 5722, 3678, 2234, and 1523 tree nodes, respectively, as shown in Fig. 10. Indeed, it is evident that when λ is set to 0.25, the canopy nodes are noticeably dense, and the segmentation at the main branches is not adequately defined; however, as λ is increased to 0.75 and 1, the segmentation of the main branches improves, albeit causing the canopy nodes to become sparser. When λ is set to 0.5, the segmentation of the main branches is effective, and the canopy nodes maintain a moderate density

$$\text{Eps} = \lambda \times R_{\text{trunk}} \tag{15}$$

where Eps is the cluster neighborhood radius of the crown layer, R_{trunk} is the cluster radius of the trunk layer, and λ is the coefficient.

B. Branches Tilt-Based Probability Density Function Construction

Distribution histograms of branch tilt angles are depicted in Figs. 11 and 12, with angle intervals of 0.5° and 1°, respectively. Both figures reveal that the distribution of tilt angles initially increases and then decreases as the angle grows. Specifically, the blue line denotes the most frequent angle value, typically ranging between 10° and 20°, while the yellow line represents the average angle value, concentrated around 20° and 30°. Table IV presents details regarding the number of branches acquired under various sample sizes, including the minimum angle value, maximum angle value, average angle value, and the angle value, with the highest frequency for angle intervals of 0.5° and 1°, respectively. The table indicates an average number of branches per tree of approximately 14. The minimum angle value is approximately 0°, while the maximum angle values are concentrated between

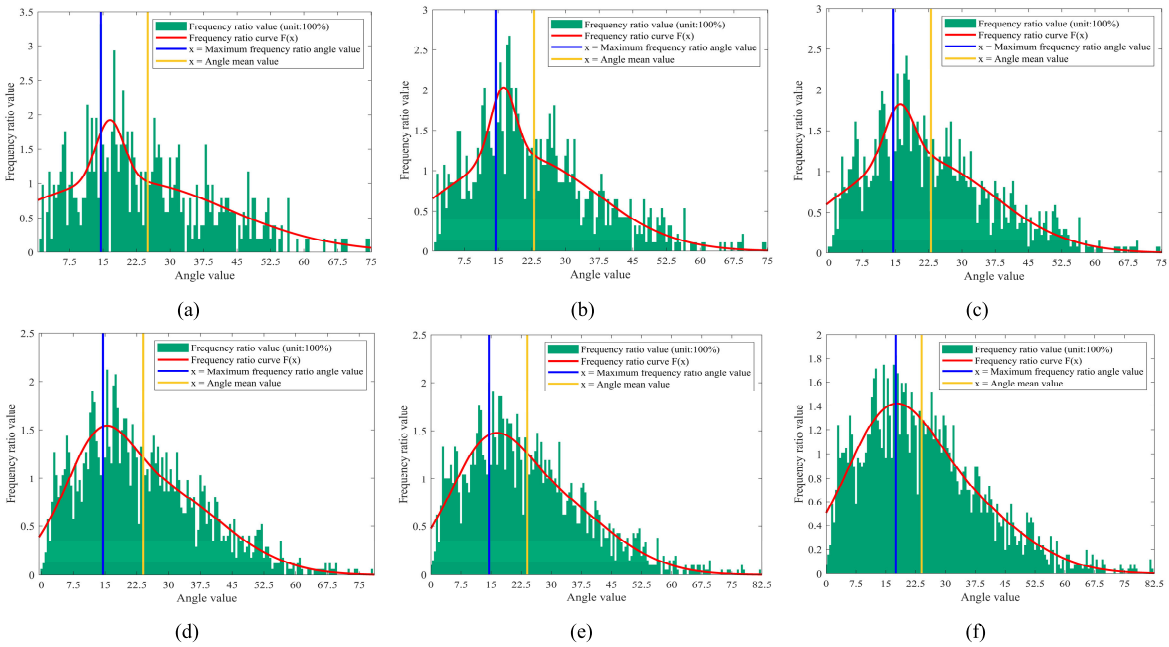


Fig. 11. Frequency distribution of angles between the main direction of tree branches and the direction of zenith angle (angle interval 0.5°). Number of trees. (a) 30. (b) 60. (c) 90. (d) 120. (e) 150. (e) 180.

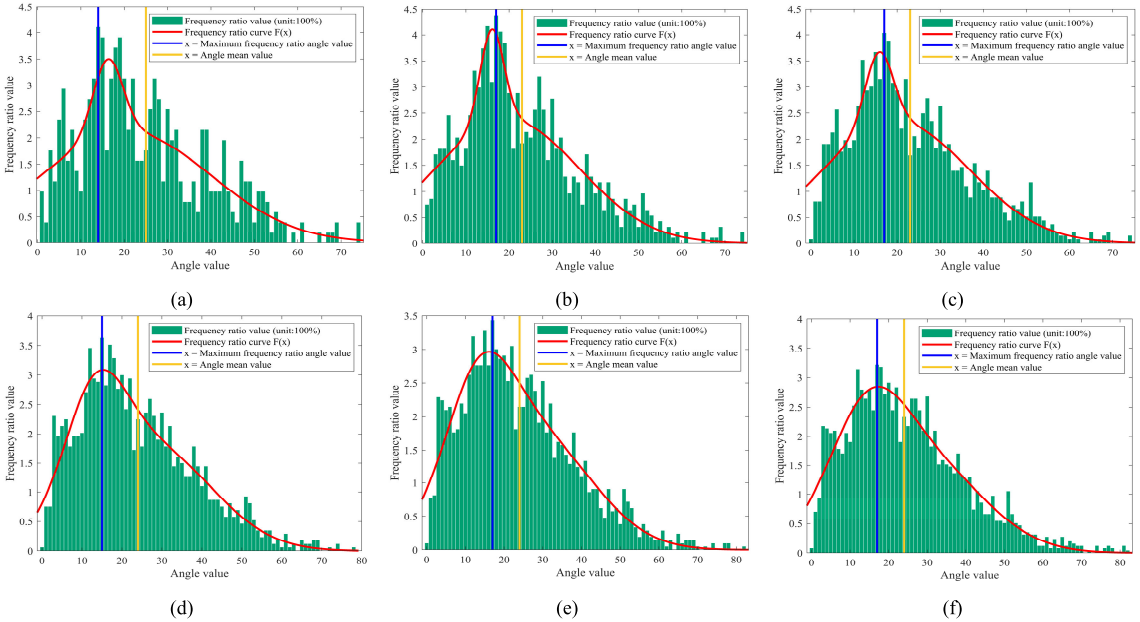


Fig. 12. Frequency distribution of angles between the main direction of tree branches and the direction of zenith angle (angle interval 1°). Number of trees. (a) 30. (b) 60. (c) 90. (d) 120. (e) 150. (e) 180.

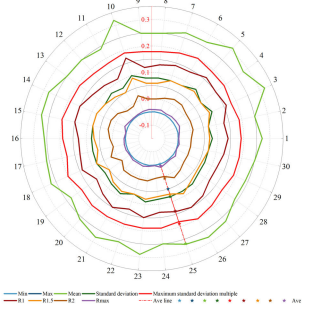


Fig. 13. Edge-weighted information map of 30 street trees during graph construction.

72° and 84°. Remarkably, the minimum angle value, maximum angle value, average angle value, and angle with the highest frequency exhibit relatively consistent fluctuations across various scenarios. Hence, the frequency distribution trends of tree branch tilt angles remain relatively stable regardless of changes in angle intervals or sample sizes. Concerning the distribution of various angle values, it is apparent that the majority of branches exhibit tilt angles ranging from 10° to 20°.

C. Impact of Varying Standard Deviation Coefficients on Graph Construction

As mentioned earlier, this article employs statistical methods to compute the mean and standard deviation of all edge weights, initially establishing the threshold for edge weights using the techniques. The magnitude of edge weights significantly influences graph construction: excessively large weights

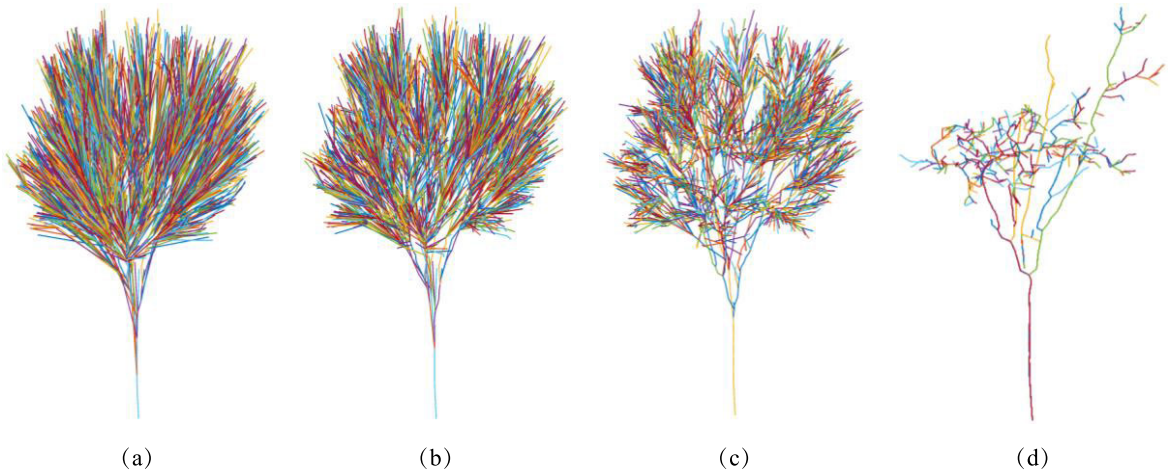


Fig. 14. Shortest path with various standard deviation coefficients. (a) 1. (b) 1.5. (c) 2. (d) 2.4.

TABLE V
INFORMATION ON EDGE WEIGHTS OF 30 SINGLE STREET TREES DURING GRAPH CONSTRUCTION

Tree samples	Min (10^{-4})	Max (10^{-2})	Mean	Standard deviation	Maximum coefficients of standard deviation	R1	R1.5	R2	Rmax (10^{-2})
AVG	7.84	99.78	0.32	0.13	2.40	0.19	0.12	0.06	0.74

can lead to a complex graph, while overly small weights may result in some points lacking the shortest path. The edge weight threshold can be estimated using (16), where the coefficient represents the standard deviation. To ensure nonnegative edge weights, a maximum value for the coefficient is provided by (17). Table V presents the edge weight information during the graph construction process for 30 roadside trees under different standard deviation coefficients with the values of 1.0, 1.5, 2.0, and $S_{\max}$, respectively. The table shows the minimum edge weight is close to zero, while the maximum value is close to one. Additionally, the maximum standard deviation coefficient is around 2.4. The results across different trees show minimal fluctuations, indicating a high level of consistency. Here we use tree 18 as an example to further explain the selection of the standard deviation coefficients. The values of 1.0, 1.5, 2.0 and 2.5, are adopted to calculate the edge weight threshold, and the shortest path is generated, as shown in Fig. 13. As indicated by Fig. 14, the shortest path generated with a standard deviation coefficient of 2.0 effectively reflects the shape of the tree. In contrast, a coefficient of 2.5 failed to generate a valid shortest path

$$r = \text{mean}\big(w\big(M_i, M_j\big)\big) - S \times \text{std}\big(w\big(M_i, M_j\big)\big) \quad (16)$$

$$S_{\max\_\text{mul}} = \frac{\text{floor}\left(\frac{\text{mean}\left(w\left(M_i, M_j\right)\right)}{\text{std}\left(w\left(M_i, M_j\right)\right)} \times 10\right)}{10} \quad (17)$$

where $S_{\max\_\text{mul}}$ is the maximum standard deviation coefficient, floor$(\cdot)$ represents rounding toward negative infinity, $w(\cdot)$ is the edge weight value, mean$(\cdot)$ represents the mean value, std$(\cdot)$ is standard deviation, r is the edge weight threshold, and S is the coefficient of the standard deviation.

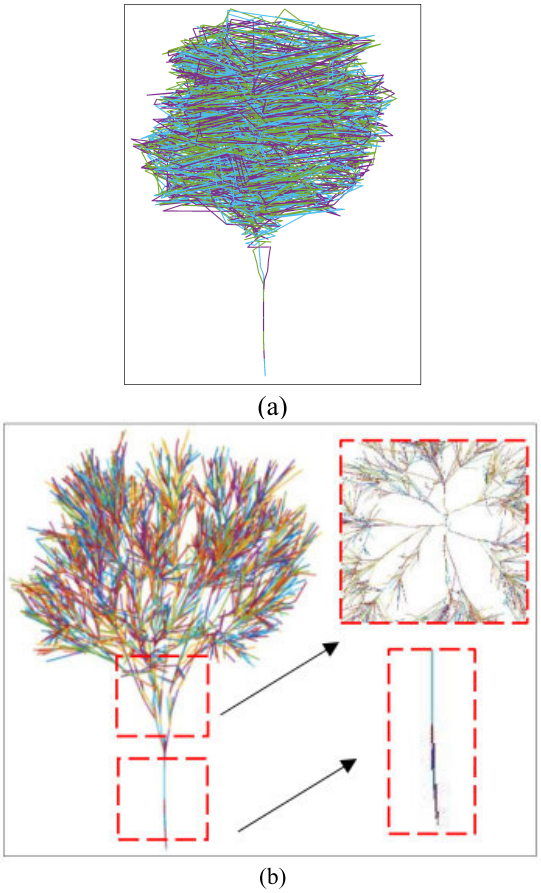


Fig. 15. Single-source shortest path diagram with each path randomly colored. (a) Shortest path to the formation of a directed weightless graph. (b) Shortest path formed by a directed weighted graph constructed with Euclidean distance as the edge weight.

D. Comparison of Shortest Path Generation

Fig. 15(a) illustrates the shortest path graph for all nodes to the base node, generated by a directed unweighted graph with all edge weights set to 1. The graph reveals that the shortest paths are disorganized, and neighboring branches fail

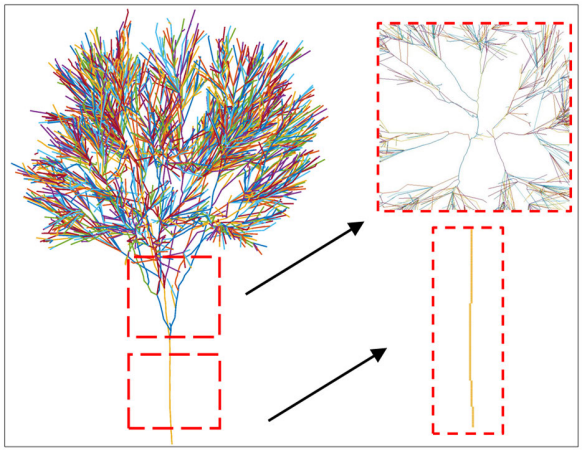


Fig. 16. Shortest path is formed by a directed weighted graph that considers the edge weights corresponding to the tilt angle.

to connect in an ordered manner and missing the shape features of the tree canopy. In contrast, Fig. 15(b) displays the directed weighted graph where the Euclidean distance serves as the edge weight. While this approach reflects the shape features of the trees, the resulting shortest path tree appears excessively rigid.

In Fig. 16, the directed weighted graph is constructed considering the tilt angle. The resulting shortest path not only ensures a well-organized connection between adjacent branch points but also enhances the flexibility of the path. This improved path representation better captures the growth shape of trees, providing a solid foundation for future wood and leaf separation processes.

V. Conclusion

Wood and leaf separation using laser scanning point cloud data has gathered considerable attention and is a critical procedure for diverse applications in fields such as ecology research, urban planning and agricultural management. MLS, known for its efficiency in comparison to ground-based laser scanning, is particularly advantageous. The point cloud obtained through MLS, nevertheless, often suffers from incompleteness, a challenge attributed to scanning geometry, which complicates the task of differentiating between wood and leaves, especially for street trees. This article proposes a new alternative method for separating wood and leaf points of individual trees in the urban street using MLS data. The prior semantic information gathered from the branches is employed for generating a probability density function. This function is subsequently used to optimize the generation of the weighted shortest paths. Next, the local features related to the shortest paths for the nodes within a tree are analyzed and then used to roughly differentiate between wood and leaf points. Subsequently, the approximately separated wood and leaf points are used to derive various geometric features. These features are integrated with the RF algorithm to achieve a more refined separation of wood and leaf points. 30 street tree samples with varying heights and densities are used to assess the performance of the proposed methodology. Experimental results demonstrate that the proposed method effectively distinguishes between wood and leaves. The average accuracy, $F1$-score for wood, $F1$-score for leaf, and Kappa are 97.08%, 89.53%, 98.29%, and 87.82%, respectively. Compared with the other four commonly used wood-leaf separation methods, namely GBS, LeWoS, CANUPO, and RF model, the proposed method outperforms four accuracy metrics. This highlights the effectiveness and reliability of the proposed approach in distinguishing between wood and leaf points of roadside trees across various density conditions. More importantly, this study further analyzes the type I and type II errors of the four methods, and the results show that the type I and type II errors are significantly reduced; further qualitative analysis also shows that the proposed method can detect the trunk points and small branches more completely and effectively when the canopy is well classified, proving that the proposed method is superior to the other three methods; further work is recommended to improve the proposed method to separate wood and leaf points at the forest level. A potential enhancement could, moreover, involve acquiring tree species information beforehand, enabling the use of specific species characteristics to generate a probability density function based on branch tilt angles for the weighted shortest paths. This refinement would further elevate the efficiency and reliability of the proposed method.

References

[1] J. Fang, A. Chen, C. Peng, S. Zhao, and L. Ci, "Changes in forest biomass carbon storage in China between 1949 and 1998," *Science*, vol. 292, no. 5525, pp. 2320–2322, Jun. 2001.

[2] Y. Bao, M. Gao, D. Luo, and X. Zhou, "The influence of plant community characteristics in urban parks on the microclimate," *Forests*, vol. 13, no. 9, p. 1342, Aug. 2022.

[3] D. B. Lindenmayer et al., "New policies for old trees: Averting a global crisis in a keystone ecological structure," *Conservation Lett.*, vol. 7, no. 1, pp. 61–69, Jan. 2014.

[4] Y. Dou, Y. Yang, and S. An, "Above-ground biomass models of *Caragana korshinskii* and *Sophora viciifolia* in the Loess Plateau, China," *Sustainability*, vol. 11, no. 6, p. 1674, Mar. 2019.

[5] Y. Li et al., "Retrieval of tree branch architecture attributes from terrestrial laser scan data using a Laplacian algorithm," *Agricult. Forest Meteorol.*, vol. 284, Apr. 2020, Art. no. 107874.

[6] J. Sun et al., "Wood–leaf classification of tree point cloud based on intensity and geometric information," *Remote Sens.*, vol. 13, no. 20, p. 4050, Oct. 2021.

[7] Q. Guo et al., "Application status and prospect of LiDAR in forest ecosystem monitoring and simulation," *Chin. Sci. Bull.*, vol. 59, no. 6, pp. 459–478, 2014.

[8] M. Maltamo, O. M. Bollandsas, E. Naesset, T. Gobakken, and P. Packalen, "Different plot selection strategies for field training data in ALS-assisted forest inventory," *Forestry*, vol. 84, no. 1, pp. 23–31, Jan. 2011.

[9] X. Liang, V. Kankare, X. Yu, J. Hyyppä, and M. Holopainen, "Automated stem curve measurement using terrestrial laser scanning," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1739–1748, Mar. 2014.

[10] C. Gollob, T. Ritter, and A. Nothdurft, "Forest inventory with long range and high-speed personal laser scanning (PLS) and simultaneous localization and mapping (SLAM) technology," *Remote Sens.*, vol. 12, no. 9, p. 1509, May 2020.

[11] S. Kaasalainen, H. Kaartinen, A. Kukko, K. Anttila, and A. Krooks, "Application of mobile laser scanning in snow cover profiling," *Cryosphere*, vol. 5, no. 1, pp. 135–138, 2011.

[12] M. Vaaja, J. Hyyppä, A. Kukko, H. Kaartinen, H. Hyyppä, and P. Alho, "Mapping topography changes and elevation accuracies using a mobile laser scanner," *Remote Sens.*, vol. 3, no. 3, pp. 587–600, Mar. 2011.

[13] F. Di Stefano, S. Chiappini, A. Gorreja, M. Balestra, and R. Pierdicca, "Mobile 3D scan LiDAR: A literature review," *Geomatics, Natural Hazards Risk*, vol. 12, no. 1, pp. 2387–2429, Jan. 2021.

[14] M. Béland, D. D. Baldocchi, J.-L. Widlowski, R. A. Fournier, and M. M. Verstraete, "On seeing the wood from the leaves and the role of voxel size in determining leaf area distribution of forests with terrestrial LiDAR," *Agricult. Forest Meteorol.*, vol. 184, pp. 82–97, Jan. 2014.

[15] X. Zhao et al., "Active 3D imaging of vegetation based on multi-wavelength fluorescence LiDAR," *Sensors*, vol. 20, no. 3, p. 935, Feb. 2020.

[16] E. S. Douglas et al., "Finding leaves in the forest: The dual-wavelength echidna LiDAR," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 776–780, Apr. 2015.

[17] M. Beland, J.-L. Widlowski, R. A. Fournier, J.-F. Cote, and M. M. Verstraete, "Estimating leaf area distribution in savanna trees from terrestrial LiDAR measurements," *Agricult. Forest Meteorol.*, vol. 151, no. 9, pp. 1252–1266, Sep. 2011.

[18] F. M. Danson et al., "Developing a dual-wavelength full-waveform terrestrial laser scanner to characterize forest canopy structure," *Agricult. Forest Meteorol.*, vol. 198, pp. 7–14, Nov. 2014.

[19] A. H. Strahler et al., "Retrieval of forest structural parameters using a ground-based LiDAR instrument (Echidna®)," *Can. J. Remote Sens.*, vol. 34, pp. 426–440, Nov. 2008.

[20] K. Calders et al., "Evaluation of the range accuracy and the radiometric calibration of multiple terrestrial laser scanning instruments for data interoperability," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2716–2724, May 2017.

[21] S. M. K. Moorthy, K. Calders, M. B. Vicari, and H. Verbeeck, "Improved supervised learning-based approach for leaf and wood classification from LiDAR point clouds of forests," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3057–3070, May 2020.

[22] S. Kaasalainen, A. Krooks, A. Kukko, and H. Kaartinen, "Radiometric calibration of terrestrial laser scanners with external reference targets," *Remote Sens.*, vol. 1, no. 3, pp. 144–158, Jul. 2009.

[23] M. B. Vicari, M. Disney, P. Wilkes, A. Burt, K. Calders, and W. Woodgate, "Leaf and wood classification framework for terrestrial LiDAR point clouds," *Methods Ecol. Evol.*, vol. 10, no. 5, pp. 680–694, May 2019.

[24] M. Wang and M. S. Wong, "A novel geometric feature-based wood-leaf separation method for large and crown-heavy tropical trees using handheld laser scanning point cloud," *Int. J. Remote Sens.*, vol. 44, no. 10, pp. 3227–3258, May 2023.

[25] J. Zhou, H. Wei, G. Zhou, and L. Song, "Separating leaf and wood points in terrestrial laser scanning data using multiple optimal scales," *Sensors*, vol. 19, no. 8, p. 1852, Apr. 2019.

[26] Z. Wang et al., "A multiscale and hierarchical feature extraction method for terrestrial laser scanning point cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2409–2425, May 2015.

[27] L. Ma, G. Zheng, J. U. H. Eitel, L. M. Moskal, W. He, and H. Huang, "Improved salient feature-based approach for automatically separating photosynthetic and nonphotosynthetic components within terrestrial LiDAR point cloud data of forest canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 679–696, Feb. 2016.

[28] S. Li, L. Dai, H. Wang, Y. Wang, Z. He, and S. Lin, "Estimating leaf area density of individual trees using the point cloud segmentation of terrestrial LiDAR data and a voxel-based model," *Remote Sens.*, vol. 9, no. 11, p. 1202, Nov. 2017.

[29] R. Ferrara, S. G. P. Virdis, A. Ventura, T. Ghisu, P. Duce, and G. Pellizzaro, "An automated approach for wood-leaf separation from terrestrial LiDAR point clouds using the density based clustering algorithm DBSCAN," *Agricult. Forest Meteorol.*, vol. 262, pp. 434–444, Nov. 2018.

[30] K. Tan, W. Zhang, Z. Dong, X. Cheng, and X. Cheng, "Leaf and wood separation for individual trees using the intensity and density data of terrestrial laser scanners," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 7038–7050, Aug. 2021.

[31] X. Wang, L. Wang, and T. Zhang, "Geometry-based assessment of levee stability and overtopping using airborne LiDAR altimetry: A case study in the Pearl River Delta, Southern China," *Water*, vol. 12, no. 2, p. 403, Feb. 2020.

[32] Z. Hui, S. Jin, Y. Xia, L. Wang, Y. Yevenyo Ziggah, and P. Cheng, "Wood and leaf separation from terrestrial LiDAR point clouds based on mode points evolution," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 219–239, Aug. 2021.

[33] Z. Tian and S. Li, "Graph-based leaf–wood separation method for individual trees using terrestrial LiDAR point clouds," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5705111.

[34] L. Penasa, M. Franceschi, N. Preto, G. Teza, and V. Polito, "Integration of intensity textures and local geometry descriptors from terrestrial laser scanning to map chert in outcrops," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 88–97, Jul. 2014.

[35] X. Zhu et al., "Improving leaf area index (LAI) estimation by correcting for clumping and woody effects using terrestrial laser scanning," *Agricult. Forest Meteorol.*, vol. 263, pp. 276–286, Dec. 2018.

[36] L. Windrim and M. Bryson, "Detection, segmentation, and model fitting of individual tree stems from airborne laser scanning of forests using deep learning," *Remote Sens.*, vol. 12, no. 9, p. 1469, Jan. 2020.

[37] Z. Su, S. Li, H. Liu, and Y. Liu, "Extracting wood point cloud of individual trees based on geometric features," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1294–1298, Aug. 2019.

[38] H. Zhang, Z. Duan, N. Zheng, Y. Li, Y. Zeng, and W. Shi, "An efficient class-constrained DBSCAN approach for large-scale point cloud clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7323–7332, 2022.

[39] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Math.*, vol. 1, no. 1, pp. 269–271, Dec. 1959.

[40] A. Gruen and D. Akca, "Least squares 3D surface and curve matching," *ISPRS J. Photogramm. Remote Sens.*, vol. 59, no. 3, pp. 151–174, May 2005.

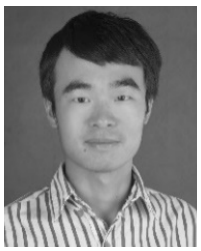
[41] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 286–304, Jul. 2015.

[42] W. Zhang et al., "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," *Remote Sens.*, vol. 8, no. 6, p. 501, Jun. 2016.

[43] X. Lin, A. Li, J. Bian, Z. Zhang, and X. Nan, "Separation method of point cloud branches and leaves of complex trees based on network diagram for ground-based LiDAR," *Remote Sens. Technol. Appl.*, vol. 37, no. 1, pp. 161–172, 2022.

[44] D. Wang, S. Momo Takoudjou, and E. Casella, "LeWoS: A universal leaf-wood classification method to facilitate the 3D modelling of large tropical trees using terrestrial LiDAR," *Methods Ecology Evol.*, vol. 11, no. 3, pp. 376–389, Mar. 2020.

[45] N. Brodu and D. Lague, "3D terrestrial LiDAR data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology," *ISPRS J. Photogramm. Remote Sens.*, vol. 68, pp. 121–134, Mar. 2012.



Yueqian Shen was born in Zhejiang Province, China, in 1988. He received the B.S. degree in surveying engineering, the M.S. degree in geodesy and surveying engineering, and the Ph.D. degree in geodesy and surveying engineering from Hohai University, Nanjing, China, in 2011, 2014, and 2017, respectively.

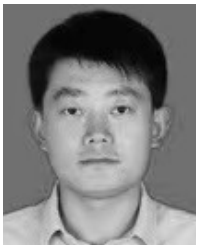
He was a Joint Ph.D. Candidate with Delft University of Technology, Delft, The Netherlands, from September 2015 to October 2016. He is currently an Associate Professor at the School of Earth Sciences and Engineering, Hohai University. His research interests include point cloud data analysis, change detection, and deformation analysis using dense point cloud data and signal processing.

Shuangshuang Ji received the bachelor's degree in geomatics from Suzhou University, Suzhou, China, in 2021, and the M.S. degree in geomatics engineering from Hohai University, Nanjing, China, in 2024.

Her research interests include wood and leaf segmentation using point cloud data, point cloud registration, and signal processing.

Jinhu Wang received the M.S. degree in signal and information processing from the University of Chinese Academy of Sciences, Beijing, China, in 2012, and the Ph.D. degree in optical and laser remote sensing from Delft University of Technology, Delft, The Netherlands, in 2017.

He is currently a Researcher at the University of Amsterdam, Amsterdam, The Netherlands. His research interests focus on object detection and recognition from point cloud data obtained by laser scanning.



Weidong Liu Senior Engineer, was born in Shandong Province, China, in 1985. He received the B.S. degree in surveying engineering from Changsha University of Science and Technology, Changsha, China, in 2008, and the M.S. degree in geodesy and surveying engineering from Hohai University, Nanjing, China, in 2011.

He is currently the Deputy Director of the Center for Surveying and Mapping Reference and Ocean Technology, Jiangsu Provincial Surveying and Mapping Engineering Institute, Nanjing, China. His research interests include the construction and maintenance of surveying and mapping reference systems, hydrographic survey and charting, and water resources investigation.

Jinguo Wang received the B.S. degree in geological engineering and the M.S. and Ph.D. degrees in hydrology and water resources from Hohai University, Nanjing, China, in 1996, 1999, and 2002, respectively.

He is currently a Professor at the School of Earth Sciences and Engineering, Hohai University. He is mainly engaged in teaching and scientific research in geological engineering, hydrogeology, and groundwater pollution mechanism and evaluation. His research interests include the applications of new optical sensors in hydrology, geology and traffic engineering.

Yanming Chen received the B.S. degree from Hohai University, Nanjing, China, in 2007, the M.S. and Ph.D. degrees from Nanjing University, Nanjing, in 2010 and 2015, respectively.

He worked at the School of Geography and Ocean Science, Nanjing University, in 2015, and joined the School of Earth Sciences and Engineering, Hohai University, Nanjing, in 2021. He is engaged in research in the field of photogrammetry and remote sensing technology, focusing on LiDAR remote sensing technology and its geomorphological applications.

Zili Deng was born in Henan Province, China, in 2002. He received the B.S. degree in geomatics engineering from Hohai University, Nanjing, China, in 2024, where he is currently pursuing the M.S. degree in surveying and mapping science and technology.

His research interests include LiDAR point cloud data analysis and object detection.

Shihan Fu was born in Henan Province, China, in 2002. She received the B.S. degree in surveying and mapping engineering from Nanjing Forestry University, Nanjing, China, in 2024. She currently pursuing the M.S. degree in surveying and mapping science and technology with Hohai University, Nanjing.

Her research interests include LiDAR point cloud analysis and object detection using AI techniques.

Dong Chen (Member, IEEE) received the bachelor's degree in computer science from Qingdao University of Science and Technology, Qingdao, China, in 2005, the master's degree in cartography and geographical information engineering from Xi'an University of Science and Technology, Xi'an, China, in 2009, and the Ph.D. degree in geographical information sciences from Beijing Normal University, Beijing, China, in 2013.

He completed a Postdoctoral Fellowship with the Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada. He is currently an Associate Professor at Nanjing Forestry University, Nanjing, China. His research interests include computer vision, remote sensing, computational geometry, and AI-driven data processing, particularly focusing on intelligent point cloud processing and its applications in forestry inventory and large-scale 3-D real-scene reconstruction for smart cities. His work aims to enhance the ability of computers to better understand the physical world.