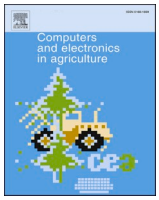
Computers and Electronics in Agriculture 226 (2024) 109378

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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



Towards automatic urban tree inventory: Enhancing tree instance segmentation via moving object removal and a chord length-based DBH estimation approach

Wai Yi Chau [a], Jun Kang Chow [a], Tun Jian Tan [a], Jimmy WU [a], Mei Ling Leung [a], Pin Siang Tan [a], Siu Wai Chiu [a,b], Billy Chi Hang Hau [c], Hok Chuen Cheng [d], Yu-Hsing Wang [a,*]

[a] Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, HKSAR, China
[b] School of Life Sciences, Faculty of Science, The Chinese University of Hong Kong, Hong Kong HKSAR, China
[c] School of Biological Sciences, Faculty of Science, The University of Hong Kong, HKSAR, China
[d] Department of Physics, Faculty of Science, The Chinese University of Hong Kong, Hong Kong HKSAR, China

ARTICLE INFO

Keywords:
Urban tree inventory
Instance tree segmentation
Data collection system
360° camera
Multi-beam flash LiDAR
Deep learning

ABSTRACT

To enhance urban forestry efficacy in Hong Kong, implementing a paradigm shift towards an automated urban tree inventory that utilizes advanced sensing technologies and artificial intelligence is essential for streamlined data collection and analysis. This study advances this objective by creating a comprehensive framework for estimating diameter at breast height (DBH) and extracting tree images. This framework encompasses five key stages: (1) data acquisition utilizing StructXray, a mobile mapping system equipped with a 360° camera and a multi-beam flash LiDAR sensor; (2) vegetation point clouds extraction using deep learning techniques; (3) individual tree segmentation through machine learning algorithms; (4) DBH estimation; and (5) tree image extraction. Six datasets were collected, yielding tree detection precision, recall and F1 score of 0.88, 0.95 and 0.91 respectively. The presence of moving objects within the 3D point cloud map, exhibiting diverse geometric structures, hinders precise vegetation point cloud segmentation by the pointwise neural network. To tackle this challenge, SalsaNext was employed to rectify the predictions of a pointwise neural network, specifically RandLA-Net in this study, eliminating 91 % of misclassified moving object point clouds and completely removing them from 47 % of affected individual tree point clouds. Additionally, a chord length-based method was proposed to enhance DBH estimation accuracy by dividing the point cloud slice into sectors and summing the chord lengths to estimate the tree trunk perimeter. Compared to the ellipse least squares fitting method, this approach reduced the root-mean-square error of the estimated DBH by 1.31 cm.

1. Introduction

Since the Industrial Revolution, greenhouse gas emissions have risen, resulting in an increasing number of extreme weather events. Urban forestry mitigates climate change impacts by cooling the environment through evapotranspiration, reflecting radiation, and providing shade, as well as reducing stormwater runoff by intercepting precipitation with their canopy and absorbing water through their soil and root systems. Although the increase in urban tree carbon sequestration only offset 3 % of urban carbon emissions annually, accurately recording the spatio-temporal dynamics of the urban forest is crucial for better urban planning and maximizing the benefits of urban greening (Pataki et al., 2021). However, urban tree inventory is often incomplete, rare, and infrequently updated due to the time-consuming, costly, and labour-intensive nature of traditional field surveys (Zhang et al., 2015, Chen et al., 2019).

Recent methods utilizing satellite and aerial imagery offer viable alternatives to traditional field surveys (Nielsen et al., 2014) by enabling large-scale tree detection and detailed tree information. Street-level images provide a relatively affordable option for street tree inventories (Choi et al., 2022). However, these image-based methods face challenges, including illumination and occlusion (Anagnostis et al., 2021; Singh et al., 2021; Ye et al., 2024), making LiDAR a promising solution. LiDAR technology utilizes laser beams to precisely measure distances, reducing reliance on external lighting conditions and

* Corresponding author.
E-mail address: ceyhwang@ust.hk (Y.-H. Wang).

https://doi.org/10.1016/j.compag.2024.109378
Received 2 April 2024; Received in revised form 13 August 2024; Accepted 22 August 2024
Available online 29 August 2024

generating comprehensive 3D point cloud maps. These maps serve as a valuable tool for accurately estimating tree geometric parameters, including diameter at breast height (DBH) and tree height. Three types of LiDAR systems are applied in recent studies: terrestrial laser scanning (TLS), airborne laser scanning (ALS), and mobile laser scanning (MLS). TLS captures 3D point clouds from a fixed position, making it an ideal method for accurate tree mapping (Chen et al., 2024), though it requires multiple scans to reduce occlusion errors, resulting in a longer data collection and processing time (Bauwens et al., 2016). MLS, on the other hand, by mounting the LiDAR sensor on a moving platform (Chen et al., 2019; Tsuchiya et al., 2023; Wang et al., 2023a), enables efficient tree surveys over large areas with reduced occlusion effects. The point cloud data acquired via MLS is processed through the simultaneous localization and mapping (SLAM) algorithm, integrating inputs from LiDAR measurements, inertial measurement units (IMU), and Global Navigation Satellite Systems (GNSS) to accurately estimate scanning positions and generate 3D point cloud maps (Bauwens et al., 2016). However, both TLS and MLS have limitations in collecting tree canopy data due to the occlusion of lower canopy layers (Chiappini et al., 2022; Chen et al., 2024). Conversely, ALS offers a better perspective from above the canopy for tree height estimation, but it yields lower point density in the stem area (Chen et al., 2024). It is particularly useful for capturing data in inaccessible or hazardous areas. Overall, LiDAR provides numerous advantages for tree surveys, enabling efficient data collection and accurate estimation of tree geometric parameters.

Due to the numerous benefits offered by LiDAR, there has been a growing focus on utilizing 3D point clouds for individual tree segmentation in urban tree inventory research. Yao & Fan (2013) removed man-made objects which were identified by the Canny operator in a footprint map and used a spectral clustering method to partition the remaining point clouds into individual trees. Weinmann et al. (2017) detected tree point clouds by using random forest classification on the geometric 2D and 3D features and then performed individual tree segmentation by mean-shift clustering. Fan et al. (2020) employed a similar approach, whereby they first extracted pole-like structures based on the roundness of the nearest neighbour of points, random forest soft classification and min-cut method. Subsequently, individual trees were identified by random forest classification. Li et al. (2021) extracted the volumetric geometric structure of supervoxels, which were assumed to be tree crowns, and performed columnar clustering and "up-hill" clustering to accomplish tree instance segmentation. Ning et al. (2023) separated overlapped tree crowns based on the point density and centroid declination angle. However, the complexity of objects in urban areas, the spatial heterogeneity of urban forests, and the diverse structure and shape of urban trees pose difficulties for rule-based individual tree segmentation (Zhang et al., 2015). The advancement of Graphics Processing Units (GPUs) has led to the increasing popularity of deep learning algorithms for urban tree inventory applications in recent years. Chen et al. (2021b) proposed the PointNLM network for vegetation segmentation. The network consists of a local branch and a non-local branch for extracting low-level, high-level, and long-distance feature. Luo et al. (2021) extracted pointwise direction vectors using pointwise direction embedding network (PDE-net). The tree centres are further estimated through pointwise direction aggregation. Based on the predicted direction vectors and tree centres, the tree point clouds are further separated into instance-level by using voxel-based region growing. Wang et al. (2023a) proposed a deep learning network that includes a feature extraction block, two parallel decoders, and a feature fusion block for individual tree segmentation in tree point clouds. In general, the deep learning approach for individual tree segmentation has proven to be more robust in complex urban scenarios. However, to the best of the authors' knowledge, the impact of moving objects on the segmentation of individual trees in point cloud data has not yet been addressed in current research. These moving objects, commonly found in urban forests, include vehicles, pedestrians, cyclists, and animals, may manifest as elongated trails in the 3D point cloud map (Chen et al., 2021a). Their varying shapes, sizes, and point structures, influenced by their trajectories and velocities, present a significant challenge for pointwise neural network to accurately detect and classify. This impacts the precision of tree instance segmentation and the accuracy of tree structural parameter estimation, further complicating the task. Hence, it is crucial to eliminate these dynamic objects from each LiDAR scan for precise data analysis.

Precise estimation of tree structural characteristics such as DBH, tree height, and crown width holds significant importance for strategic planning and effective management of urban forestry. Various methods have been employed to approximate DBH using 3D point cloud data, but their accuracy is influenced by several factors. Common methods include least squares fitting (Cabo et al., 2018; Fan et al., 2020) and the random sample consensus (RANSAC) fitting (Chen et al., 2019) which assume the tree trunk has a specific shape, usually in a circle, an ellipse or a cylinder (Balenović et al., 2021). This assumption might result in the inaccurate estimations of DBH for trunks with irregular shapes (Bauwens et al., 2016). Furthermore, bark roughness can also influence DBH estimation. LiDAR scans both the fissure and outer surface of the tree trunk resulting in estimated DBH values that do not align with tape measure results (Tsuchiya et al., 2023). Moreover, rough bark structures can introduce additional noise in point cloud data (Zeybek and Vatandaşlar, 2021), making DBH estimation more challenging. Additionally, errors in LiDAR measurement (Forsman et al., 2018) and the influence of point cloud density (Balenović et al., 2021) can introduce additional inaccuracies in DBH estimation. To improve the accuracy of DBH estimation, it is imperative to develop a novel methodology.

Given the preceding discussions, the primary objectives of this study are as follows:

(1) Develop a comprehensive framework utilizing a mobile mapping system and artificial intelligence to advance the automation of urban tree inventory processes.
(2) Investigate the significance of integrating moving object segmentation into the process of individual tree segmentation within urban settings, ultimately supporting the advancement of automated estimation for tree geometric parameters.
(3) Propose an alternative approach that could potentially offer a more efficient estimation of DBH in comparison to the least squares fitting method.

Fig. 1 illustrates the framework employed to attain the aforementioned objectives of this study. This framework employs a mobile data collection platform equipped with advanced sensors, including a 360° camera and a multi-beam flash LiDAR sensor, to gather image and point cloud data (Stage I). The collected point cloud data is processed to generate 3D point cloud maps. Vegetation point clouds are identified using deep learning algorithms (Stage II) and further segmented at the instance level through machine learning techniques (Stage III). The segmented tree point clouds are then used for estimating DBH (Stage IV) and extracting tree images (Stage V). This approach was validated with data collected from the Hong Kong University of Science and Technology (HKUST) campus and an urban area on Mei Yuen Street, Sai Kung, Hong Kong, demonstrating its potential to enhance urban tree inventory processes.

This manuscript starts by introducing the proposed framework in the Materials and Methods section. Subsequent sections of the paper present case studies that demonstrate the efficacy of this proposed framework in the Results section. The Discussion and Conclusions sections are provided at the end of this manuscript.

2. Materials and methods

The proposed framework, as depicted in Fig. 1, provides an overview of its functionalities. Specifically, a mobile data collection platform named StructXray (Chow et al., 2021), equipped with a 360° camera and

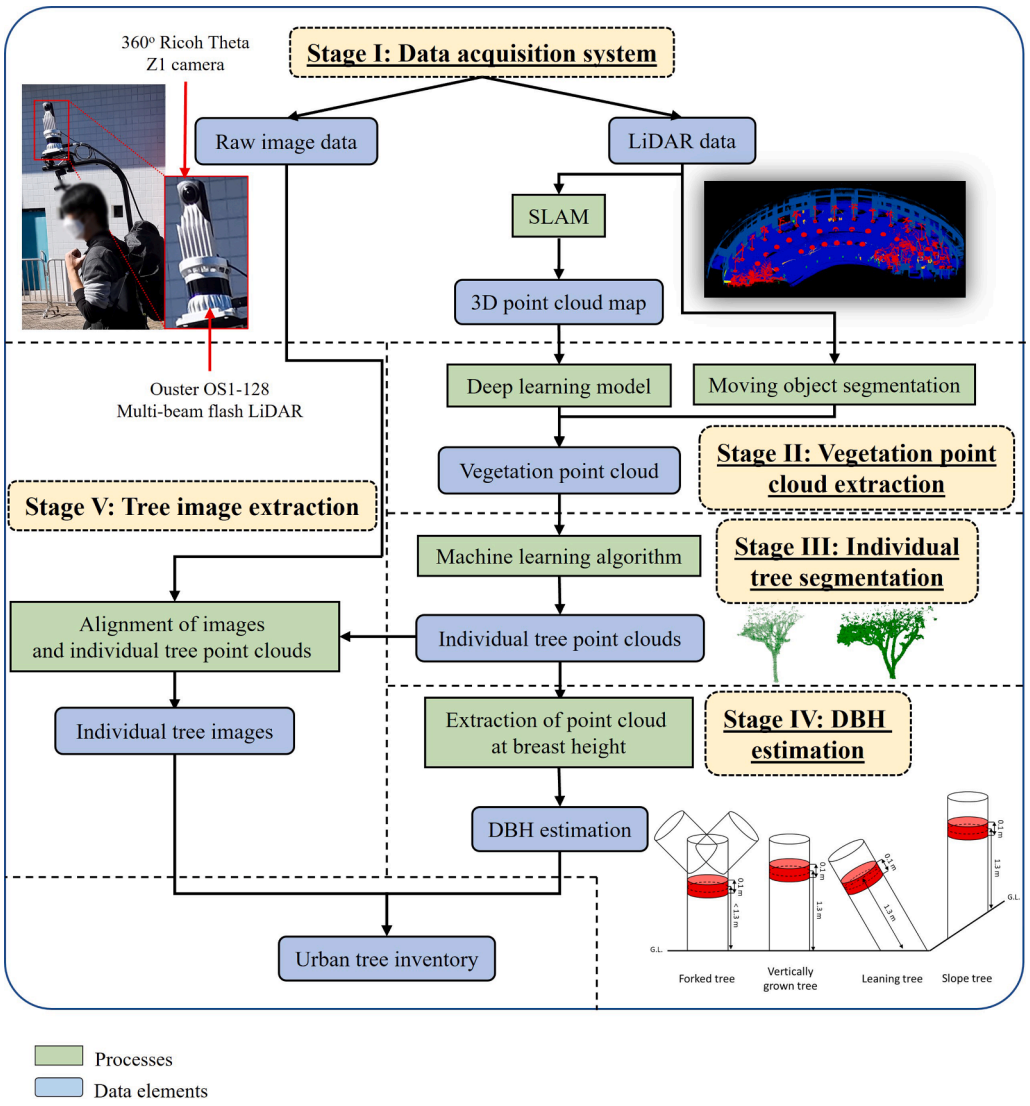


Fig. 1. Overview of the proposed framework for estimating DBH and extracting tree images in this study which consists of five key components: data acquisition system, vegetation point cloud extraction, individual tree segmentation, DBH estimation and tree image extraction. Noted that in Stage IV, the point cloud within the red slice is extracted for DBH estimation.

a multi-beam flash LiDAR sensor, is adopted for data collection. The raw point cloud frames collected are fed into the SLAM algorithm, FAST-LIO2 (Xu et al., 2022), to generate a 3D point cloud map. The map is then analysed using the lightweight pointwise neural network RandLA-Net (Hu et al., 2020) to achieve semantic segmentation. The prediction is subsequently rectified by a projection-based neural network SalsaNext (Cortinhal et al., 2020) which identifies moving objects. Points predicted as vegetation are further segmented at the instance-level by the HDBSCAN algorithm (Campello et al., 2013) and using Dijkstra's algorithm (Dijkstra, 1959). Following this, DBH estimation is conducted by a chord length-based methodology. The integration of 360° camera and LiDAR data allows overlaying the point clouds onto the panoramic images, facilitating the extraction of individual tree images. The details of the processing are elaborated in the following subsections.

2.1. Description of the study areas

A total of six datasets were collected, with four obtained from the HKUST campus and two from Mei Yuen Street on 25th October 2023. The decision to gather four datasets from the HKUST campus, as opposed to collecting all datasets from urban areas, was influenced by the prevalent monoculture planting practices in Hong Kong (Kwong, 2022). Roadside trees in Hong Kong urban areas are predominantly composed of a limited number of species, which does not offer an appropriate environment for evaluating the framework's effectiveness in detecting trees of various forms. In contrast, the HKUST campus, situated on a hillside in Sai Kung, contains over ten thousand trees, including a variety of urban tree species found in Hong Kong. The selected sites feature an average of five different tree species. The primary tree species present at the sites are enumerated in Table 1. Notable examples include *Aleurites moluccana* and *Delonix regia*, known for their broad, spreading canopies; *Araucaria heterophylla*, distinguished by its columnar structure; *Melaleuca viminalis*, recognized for its weeping branches; and the palm trees *Archontophoenix alexandrae* and *Phoenix roebelenii*. The diverse array of tree species and forms on campus provides ample opportunities to test the robustness of the proposed framework. Although the scenes on the HKUST campus might not be as complex as those in typical urban areas, the four study areas as shown in Figs. A1a - d are located adjacent to main roads and infrastructure and contain various urban furniture such as poles, barriers, lamp posts, and traffic signs. Therefore, the settings of the selected study areas on the HKUST campus are considered to be similar to those found in urban areas. To further extend the validation of the proposed framework, data were also collected at Mei Yuen Street, which provides scenes with more

Table 1
Primary tree species across various sites.

Dataset No.	Location	Primary tree species
1	HKUST campus	*Araucaria heterophylla*, *Archontophoenix alexandrae*, *Delonix regia*
2	HKUST campus	*Lagerstroemia speciosa*, *Melaleuca viminalis*, *Phoenix roebelenii*
3	HKUST campus	*Aleurites moluccana*, *Bischofia javanica*, *Peltophorum tonkinense*
4	HKUST campus	*Bischofia javanica*, *Ficus variegata* var. *chlorocarpa*, *Koelreuteria bipinnata*, *Liquidambar formosana*
5	Mei Yuen Street	*Cinnamomum camphora*
6	Mei Yuen Street	*Aleurites moluccana*, *Cinnamomum camphora*

complex urban environments compared to the HKUST campus. Mei Yuen Street is an L-shaped road, lined with trees, predominantly *Cinnamomum camphora*, and minimally with *Aleurites moluccana*, *Livistona chinensis*, *Sterculia lanceolata*, and *Pongamia pinnata*. A portion of Mei Yuen Street is currently designated as a pedestrian-only area, and as such, vehicular traffic is restricted from entering. This study, therefore, divided Mei Yuen Street into two separate study areas, with one located within the pedestrian-only area and the other outside of it, as shown in Figs. A1e - f.

2.2. Data acquisition (Stage I)

The inset of Fig. 1 illustrates the mobile mapping system, called StructXray (Chow et al. 2021) developed by our team, which consists of a 360° camera and a multi-beam flash LiDAR sensor, for data collection. StructXray was chosen for tree surveying in this study due to its mobility, lightweight design, and integration of high-performance sensors. These features enable efficient and effective data acquisition, making it an ideal choice for the proposed framework.

StructXray is outfitted with a Ricoh Theta Z1 camera, which has a symmetrical dual-fisheye lens. This camera is lightweight (182 g) with its camera body made of magnesium alloy. The sensor is capable of capturing 360° still images at a resolution of 7 K (6720 x 3360, 23 million pixels) and recording videos at 30 fps (frames per second) in 4 K resolution (3840 x 1920, 7 million pixels). The advantage of using a 360° camera lies in its ability to capture images over a wide area, thereby minimizing the need for additional sensors and reducing both battery usage and system complexity. In addition, StructXray is equipped with the Ouster OS1 LiDAR sensor, which features 128 channels. This advanced multi-beam flash LiDAR sensor functions by emitting precise laser beams with a wavelength of 865 nm, effectively illuminating the scene for optimal data collection. Simultaneously, each pixel of the sensor captures reflection from the scanned objects, allowing for the efficient capture of a comprehensive and detailed 3D representation of the surrounding environment. With a remarkable capacity to capture 5,242,880 points per second, the sensor provides an ample quantity of points to establish a dense 3D point cloud for further processing within the framework. Its exceptional performance is further enhanced by its maximum measurement range of ~ 200 m and a vertical field of view of 45°, allowing for superior coverage and promising opportunities to capture tree data. The sensor demonstrates the standard deviation of precision ranging between 0.5 cm and 3 cm, and achieves an accuracy of approximately 2.5 cm when measuring objects with Lambertian surfaces measured from distances up to 90 m. Thus, StructXray is capable of precise estimation of tree geometric parameters. According to the datasheet of the Ouster OS1 LiDAR sensor, there is an observable increase in the standard deviation of precision when target objects are ~ 10 m away from the sensor. To ensure high precision measurements, point clouds measured at distances greater than 10 m in the plan view from the sensor are excluded from the analysis. The LiDAR sensor also outputs IMU data, including 3-axis gyroscope and 3-axis accelerometer measurements. The integration of LiDAR measurement and IMU data as inputs for SLAM enables the generation of a 3D point cloud map and facilitates estimation of the sensor's position and orientation within the point cloud map. FAST-LIO2 (Xu et al., 2022) was adopted for SLAM in this study. It fuses LiDAR and IMU measurements using a tightly coupled iterated Kalman filter, allowing for accurate and precise state estimation and mapping. To validate the performance of the FAST-LIO2 algorithm, the reconstructed 3D point cloud map was compared with measurements obtained from the GOWIN TKS-200 Serial Electronic Total Station with prism sets, as shown in Fig. A2. This instrument was selected for its extensive measurement range (over 1000 m) and high accuracy (±2 mm + 2 ppm * distance for distance measurement, and 2″ for angle measurement). Since the 360° camera was mounted on top of the LiDAR sensor (see, inset in Fig. 1), the relative distance and azimuth of both sensors were fixed during the survey. Translation and rotation need to be applied only once to align both panoramic images and the spherical projection of point clouds into the same coordinate system. For more details, please refer to Chow et al. (2021).

2.3. Vegetation point cloud extraction (Stage II)

Building on the established alignment of the panoramic images and spherical projection of point clouds, tree instance segmentation can be conducted on a single data type, either the image data or the LiDAR data, thereby obtaining segmentation results for both data types simultaneously. After thorough consideration, this research focused on conducting tree instance segmentation on point cloud data and subsequently utilized the obtained results to extract tree images. This approach was chosen due to the prevalent occurrence of segmentation errors in image data at object boundaries (Xiao et al., 2023). These misclassified pixels at object boundaries may correspond to points considerably distant from the tree point clouds. Consequently, relying exclusively on image segmentation results for point cloud segmentation may lead to a degradation in the quality of individual tree point clouds, hindering the advancement of automated estimation of tree geometric parameters.

To ensure the high quality of individual tree point cloud data for accurate tree geometric estimation, semantic segmentation was performed as a data preprocessing step. This process eliminates non-vegetation elements in urban environments, thereby establishing a forest-like setting that aids in the subsequent individual tree segmentation process. RandLA-Net (Hu et al., 2020), including both the original version (Wang et al., 2023b) and the improved versions (Lei et al., 2022, Xia et al., 2023), were commonly used for tree semantic segmentation because of its highly efficient computation time and resource consumption. It is a lightweight, pointwise neural network designed for semantic segmentation on 3D point cloud maps. The network architecture follows a U shape encoder-decoder design, incorporating skip connections. To reduce computational complexity, random sampling is employed. Additionally, a local feature aggregation module is implemented to maintain geometric details. This module comprises local spatial encoding units (LocSE) and aggregation pooling units, which effectively capture spatial information and generate valuable feature vectors. By stacking multiple LocSE and aggregation pooling units, the model's receptive field is expanded, ensuring the preservation of important features. This study followed a methodology similar to that of the study of Hu et al. (2020) in implementing the RandLA-Net architecture for semantic segmentation, with a necessary modification that the model's depth was increased from four to six encoding and decoding layers. The first five layers reduce the number of points by a factor of 4, while the final layer reduces them by a factor of 2. The output feature dimensions for each layer were set to 16, 64, 128, 256, 512, and 1024, respectively. This adjustment enables the model to capture more complex geometric structures in the point cloud, thereby improve its performance. Additionally, the learning rate and the number of training

epochs were modified to 0.002 and 1000, respectively, to prevent overshooting and ensure a smoother descent along the loss surface. Note that the RandLA-Net model was trained using the Open 3D library (Zhou et al., 2018) and an NVIDIA GeForce GTX 1080 Ti GPU.

To address the inherent computational complexity associated with model training and prediction, the Voxel Grid Filtering approach was adopted in this study (Hacinecipoglu et al., 2020). This technique involves downsampling the point clouds by computing the average coordinates of points within a cubic voxel measuring 5 cm x 5 cm x 5 cm. Note that the choice of a cubic voxel with a 5 cm edge length was made to preserve the local geometric structure of the tree trunk, as this information is crucial for the RandLA-Net model to understand the relative spatial positions of neighbouring points. For further details on the rationale behind choosing a 5 cm voxel size for downsampling, please refer to the supplementary material. Furthermore, for the purpose of mitigating noise within the 3D point cloud map to optimize model's performance, points were removed if their corresponding voxel contained less than ten points.

The process of labelling data is a crucial step in training deep learning models. In this study, all datasets were manually labelled using CloudCompare, with meticulous attention to detail for both model training and testing, serving as ground truth. Reference was also made to the images gathered during Stage I for annotation purpose. However, the complexity of urban scenes introduces a challenge as the point clouds can potentially be classified into numerous classes, further complicating the tasks of model training and prediction. Consequently, the decision regarding the number of classes to label becomes a significant consideration in the training of deep learning models. In this study, it was assumed that objects in the urban setting could be grouped into the same nine classes as those annotated in the Paris-Lille 3D dataset (Roynard et al., 2018): ground, buildings, poles, bollards, trash cans, barriers, pedestrians, cars and natural vegetation. The Paris-Lille 3D dataset is a point cloud dataset collected by an MLS, which is equipped with GPS, IMU, and a Velodyne HDL-32E LiDAR sensor, on the urban streets in Paris and Lille, France. Thus, this dataset was also applied for model training in this study, enhancing both the model's performance and its generalization capabilities. Note that moving objects were annotated as pedestrians to represent their dynamic presence within the 3D point cloud maps.

Table A1 provides a summary of the training, validation, and testing datasets used for training the RandLA-Net models. Four RandLA-Net models were trained by different combinations of datasets and performed semantic segmentation on a single dataset to make optimal use of the collected datasets. The datasets collected from HKUST Sites 2 to 4 were used to train the RandLA-Net model (Model 1). This trained model was subsequently utilized for performing semantic segmentation on the datasets collected from Mei Yuen Street Sites 5 to 6. Additionally, three RandLA-Net models (Models 2 to 4) were developed to test the robustness of the deep learning model in semantic segmentation on datasets collected from HKUST Sites 2 to 4. This was done to ensure that the testing datasets were not incorporated into the training datasets. The dataset collected from HKUST Site 1 is representative, not skewed towards any particular class compared to other datasets. As a result, this dataset was specifically chosen as a validation dataset to evaluate the potential occurrence of overfitting during the training of the RandLA-Net models.

Utilizing 3D point cloud maps for training a pointwise neural network is superior to relying solely on individual point cloud frames. The fundamental reason behind this lies in the consistent and comprehensive representation offered by the 3D point cloud map. By encompassing multiple frames, it mitigates the impact of occlusion, ensuring that objects are captured holistically. Consequently, the spatial relationships between objects are accurately preserved, furnishing invaluable contextual information for precise segmentation outcomes. Thus, this study utilizes 3D point cloud maps to train RandLA-Net, aiming to optimize its performance for semantic segmentation. However, it is crucial to acknowledge the inherent difficulty faced by pointwise neural networks in accurately identifying moving entities like pedestrians and vehicles. In the 3D point cloud map, these objects are represented by clusters of points with varying structures, shapes, and point distributions, which are influenced by their velocities and trajectories. Consequently, there may be similarities between the point cloud data associated with these objects and that of vegetation, leading to a potential decrease in the accuracy of semantic segmentation performed by RandLA-Net. To tackle this concern, the application of a deep learning model called SalsaNext (Cortinhal et al., 2020) was employed for moving object segmentation. SalsaNext is a projection-based convolutional neural network with a standard encoder-decoder architecture. In order to enhance the results of moving object segmentation, this study adopted the methodology proposed by Chen et al. (2021a), which involves the fusion of range images and residual images as input. The range images are created by mapping point cloud frames onto a 2D grid, where each pixel represents a specific point within the point cloud frame and stores information such as x, y, z coordinates, range, and intensity. On the other hand, the residual images are created by computing the disparity in range for each pixel between the current scan and previous scans. By concatenating the range and residual images, the resulting composite image contains both spatial and temporal information, enabling the model to effectively analyse sequential data and accurately distinguish between pixels associated with moving objects and those that belong to the background. This study used residual images generated from five previous scans, based on the findings of Chen et al. (2021a), who demonstrated that the performance of moving object segmentation becomes consistent and reliable when using residual images generated from five or more previous scans. The SalsaNext models were trained using the Keras library and four NVIDIA GeForce GTX 1080 Ti GPUs. The training configurations were based on the study by Cortinhal et al. (2020), with modifications: the learning rate, the number of epochs and the batch size were adjusted to 0.0005, 500 and 8, respectively, to enhance model convergence, prevent overshooting optimal parameters and improve the model's generalization to unseen data. Table A2 provides an overview of the training, validation, and testing datasets used for training SalsaNext models. The training and validation datasets maintained a 9:1 ratio, with 90 % of the data allocated for training purposes, while the remaining 10 % reserved for validation. To mitigate potential bias in the original data order, the data was shuffled before model training.

Following the implementation of SalsaNext for moving object segmentation, the results obtained were merged with the semantic segmentation outputs generated by RandLA-Net. It is important to acknowledge that RandLA-Net and SalsaNext process different types of data. RandLA-Net operates on a downsampled 3D point cloud map, while SalsaNext utilizes a concatenated range-residual image. Indeed, as previously mentioned, the concatenated range-residual images and the downsampled 3D point cloud maps were both derived from the individual point cloud frames, establishing an inherent connection between the pixels in the concatenated range-residual image and the points in the downsampled 3D point cloud map. Therefore, to integrate the results of RandLA-Net and SalsaNext, this study determined that if over 50 % of the pixels corresponding to a specific point in the downsampled 3D point cloud map were classified as moving objects, the corresponding point would be labelled as a moving object, irrespective of RandLA-Net's prediction. It is important to note that, in theory, initially removing moving objects using SalsaNext could aid RandLA-Net in better comprehending the static scene, thereby potentially improving its performance in semantic segmentation. However, it is practically infeasible for SalsaNext to accurately identify all moving objects. This limitation could complicate the spatial distribution of points and might diminish the accuracy of semantic segmentation by RandLA-Net. Consequently, the study opted to conduct moving object segmentation using SalsaNext and semantic segmentation using RandLA-Net independently, and subsequently integrate the results from both methods. Upon the successful

execution of semantic segmentation, the proposed methodology was evaluated utilizing the Intersection over Union (IoU) metric, calculated as follows:

$$IoU = \frac{TP}{TP + FP + FN} \tag{2.1}$$

where TP, FP, and FN represent the numbers of points that correspond to true positive, false positive, and false negative for a given class, respectively.

2.4. Individual tree segmentation (Stage III)

Upon completing Stage II, only points classified as natural vegetation and ground were preserved for further processing. In Stage III, the vegetation point cloud was further divided into instance tree levels using a combination of machine-learning and rule-based algorithms. To accomplish this, the point clouds of natural vegetation were initially split into two halves. The lower half focused on segmenting individual tree trunks, while the upper half concentrated on tree crown segmentation. The decision to employ a two-stage approach for individual tree segmentation was based on the observation that tree trunks typically exhibit more distinct separations compared to the tree crown, which often overlap with each other, posing challenges in segmentation when relying on clustering methods. Consequently, this study initially partitioned the tree trunk point cloud into individual clusters using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm (Campello et al., 2013). Based on the results of individual tree trunk segmentation, a root node was established for each cluster, serving as a reference for individual tree crown segmentation through the utilization of Dijkstra's algorithm (Dijkstra, 1959). A detailed description of the procedures involved in each step is provided in the subsequent paragraphs.

2.4.1. Construction of digital terrain model

To partition natural vegetation point clouds into their respective top and bottom halves, the most straightforward method was to construct a digital terrain model (DTM) using points classified as ground by RandLA-Net in Stage II, which serves as the basis for determining the height of each point. However, relying solely on these points for the direct establishment of the DTM was vulnerable to various possible errors. Measurement errors from the LiDAR sensor, coupled with mapping inaccuracies inherent to the SLAM algorithm, resulted in the apparent thickness in the ground point clouds. To mitigate this issue, the spatial averaging of the ground points within a 5 cm x 5 cm grid in the x-y plane was calculated and considered as the digital terrain model. This grid size was consistent with the voxel size used for downsampling, as mentioned in Stage II. Outliers resulting from misclassification were corrected by adjusting the z-coordinate of the grid to the average z-coordinate of neighbouring grids, provided their average height difference was 0.2 m or greater, based on the standard requirement for a stair's riser height. Additionally, point clouds labelled as natural vegetation that were outside the DTM boundary were removed to prevent errors in individual tree segmentation.

2.4.2. Individual tree trunk segmentation

The natural vegetation point cloud was first divided into two halves at a height of 1.35 m. This specific height was chosen based on the criterion that a plant is categorized as a tree when its trunk diameter measures 9.5 cm or greater at a height of 1.3 m above ground level in Hong Kong. Vegetation below 1.3 m in height does not meet the defined criteria for being classified as a tree. As detailed in Section 2.3, a voxel size of 5 cm was utilized for the Voxel Grid Filtering approach. Therefore, an additional 5 cm is necessary to determine if a plant qualifies as a tree. In this study, the lower half of the natural vegetation point cloud was subjected to clustering via the application of the HDBSCAN algorithm (Campello et al., 2013) for individual tree trunk segmentation. HDBSCAN is a density-based clustering algorithm, an extension of the DBSCAN algorithm by incorporating the hierarchical representation. As a result, HDBSCAN can effectively separate point clouds into clusters with varying densities and shapes. A minimum cluster size parameter of 108 points was chosen for HDBSCAN. This selection is based on the requirement for tree dimensions in Hong Kong and the voxel size discussed earlier. Specifically, the minimum number of points needed to represent the tree trunk was calculated by multiplying the horizontal and vertical point requirements (4 x 27), resulting in a minimum cluster size of 108 points.

After the HDBSCAN operation, the resulting clusters can be classified into three main types: those consisting of an individual tree trunk, those comprising a single group of bushes, and those that are a combination of both bushes and tree trunks. Shrubs in urban areas are typically pruned to maintain a height of less than 1.3 m, which allows them to be easily managed. Thus, clusters that did not reach a height of 1.3 m were considered to be bushes and subsequently excluded from the subsequent analysis.

For the clusters that comprise both bushes and tree trunks, it is necessary to separate such clusters into distinct groups, one consisting of bushes and a single trunk, and subsequently remove the bushes before DBH estimation. To ascertain clusters that comprise both bushes and tree trunks, an examination of the width of the cluster at various height levels was conducted. It is observed that the width of bushes is typically greater than that of tree trunks. Thus, in this study, a cluster was assumed as a combination of shrubs and trunks if the ratio of its width below a height of 1.25 m to its width between 1.25 m and 1.35 m was greater than 2. Figs. 2 and 3 illustrate the processing of such clusters. Points at a height of 1.25 m to 1.35 m above ground level were presumed to be the part of the tree trunks and were further grouped into different clusters by HDBSCAN where the selected minimum number of points in a cluster was 8 as shown in Fig. 2b. The spatial average of points within each cluster was subsequently considered as the root node. Next, a weighted undirected graph was constructed, where the points below 1.25 m served as the vertices. Each vertex was connected to its three closest neighbouring points, and the weight of the connection, or edge, was determined by the square of the Euclidean distance. Due to the non-uniform distribution of points, multiple graphs were typically formed. Consequently, edges were added to connect the closest vertices between two graphs, resulting in the creation of a single graph. The shortest paths between each unclassified point and every root node were then calculated by Dijkstra's algorithm. In cases where the shortest path between an unclassified point and a certain root node was found to be the minimum, the unclassified point was subsequently assigned to the cluster corresponding to that root node, as illustrated in Fig. 2b & 2c. The subsequent stage involves the removal of the bushes from the point cloud cluster, as depicted in Fig. 3. Three underlying assumptions were being made: firstly, that there were no points located within the tree trunk; secondly, that the noise level of the point cloud at the surface of the tree trunk was uniformly distributed; and lastly points at 1.25 m to 1.35 m were assumed to belong to be part of the tree trunk. The point cloud clusters, previously downsampled, were restored to their original resolution and subsequently divided into vertical segments of 10 cm each. Next, the Hough transform (Hough, 1962) algorithm was applied to detect the centre of the tree trunk on the slice between 1.25 m and 1.35 m, as illustrated in Fig. 3a. Subsequently, the noise level, indicated by the thickness of the point cloud at the surface of the tree trunk on the same slice, was quantified by measuring the difference in distance between the farthest and closest points relative to the center. Then points that fall outside the range of the tree trunk thickness were removed from the point cloud for slices positioned below 1.25 m, as shown in Fig. 3b. Fig. 4 presents one of the results after bushes removal at the x-z plane and y-z plane.

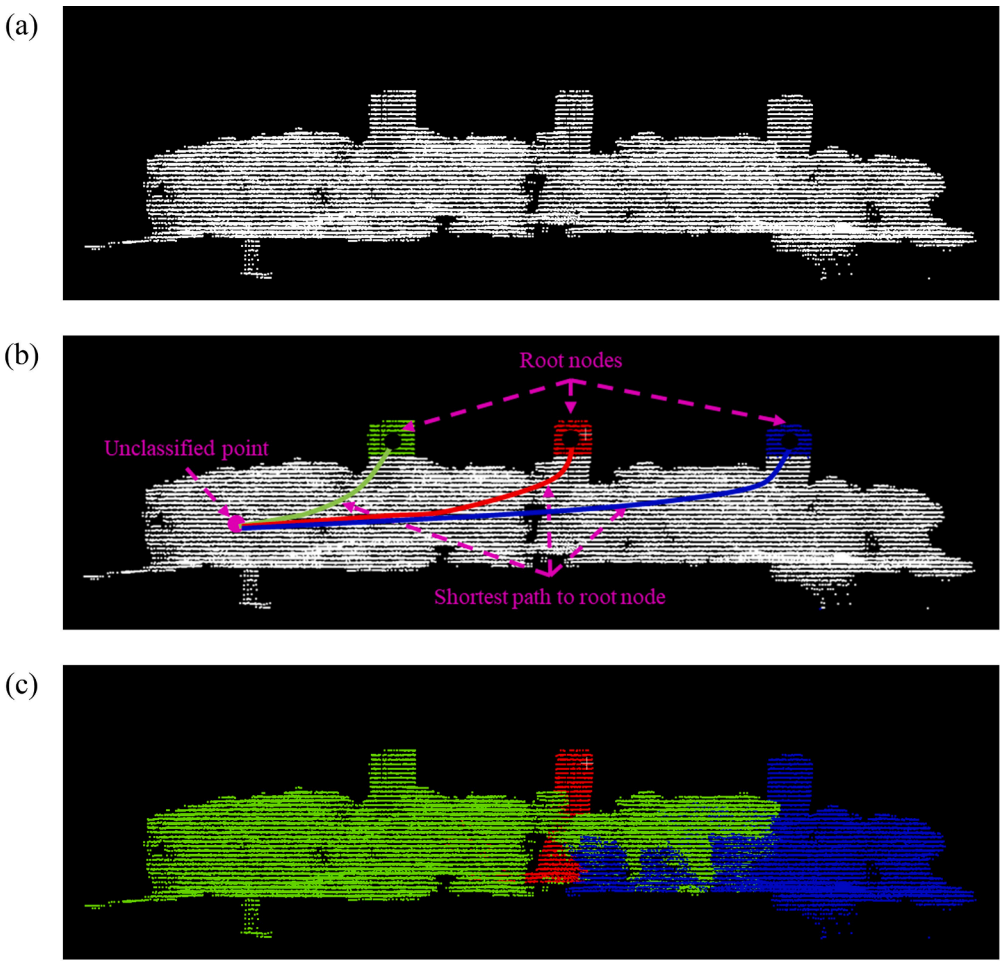


Fig. 2. Illustration of separating a cluster that consists of bushes and multiple trunks into clusters of bushes and a single trunk. (a) After applying HDBSCAN to the lower half of the natural vegetation point cloud, a cluster consists of bushes and three trunks; (b) Points at 1.25 m to 1.35 m in this cluster are extracted and further clustered by HDBSCAN. Three groups of points are obtained. The spatial average of every group of points serves as a root node. Points below 1.25 m in height are used for building a weighted undirected graph and are further separated by using Dijkstra's algorithm. The shortest paths (i.e. green, red and blue lines) from an unclassified point (i.e. pink circle) to every root node (i.e. black circles) are calculated. If the shortest path between an unclassified point and a root node is discovered to be the minimum (i.e. green line), the unclassified point is then assigned to the cluster that corresponds to that root node; (c) This cluster is finally separated into three groups, each comprising bushes and a single tree trunk. Note that the unclassified points are shown in white color as shown in (a) & (b) while these points are separated into three groups which are highlighted in red, green and blue color as shown in (b) & (c).

2.4.3. Individual tree crown segmentation

The process of individual tree crown segmentation involved a methodology similar to that used for separating clusters containing both bushes and multiple tree trunks into clusters with a single trunk and bushes, as described in Section 2.4.2 and Fig. 2. The sole difference is that the weighted undirected graph was constructed using the upper half of the vegetation point cloud, as this process focuses on individual tree crown segmentation. After completing the individual tree crown segmentation, the individual tree trunk cluster can be combined with the individual tree crown cluster to create a comprehensive individual tree point cloud. The effectiveness of this tree detection process was subsequently evaluated using precision, recall, and F1 score. The equations of evaluation metric are as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2.2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2.3}$$

$$\text{F1 score} = \frac{2TP}{2TP + FP + FN} \tag{2.4}$$

where TP, FP, and FN represent the number of true positives, false positives, and false negatives for detected trees, respectively.

2.5. DBH estimation (Stage IV)

Upon acquiring the individual tree point clouds, which were previously downsampled, meticulous efforts were made to upsample them to their original resolution, thereby enhancing the accuracy of DBH estimation. The definition of DBH is the diameter of the tree trunk at 1.3 m above ground. However, for some special circumstances, the position for DBH estimation is not straightforward. In this study, a procedure for automatically locating the position for DBH estimation was established for three special cases: (1) vertically grown trees located at the sloping ground, (2) trees with a fork below 1.3 m height, and (3) leaning trees. For case (1), the DBH is measured at a height of 1.3 m above ground at the upper side of the slope. For case (2), the DBH is measured below the tree fork. For case (3), the DBH is measured at 1.3 m in length along the leaning trunk. In order to obtain a precise estimation of the DBH, it is necessary to accurately locate the position for DBH estimation. First, the ground level was determined by identifying the nearest ground data points with the highest z-coordinate value. Then the tree point cloud below 1.3 m in height was extracted and segmented into different slices

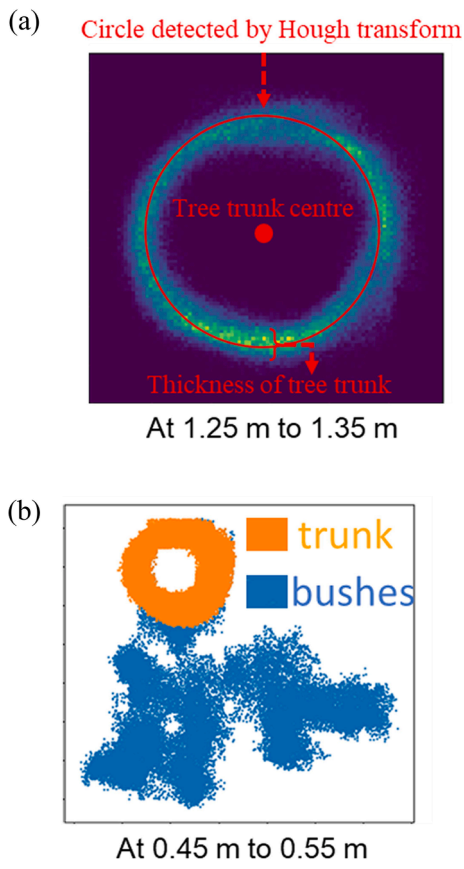


Fig. 3. Illustration of removing bushes from a cluster of bushes and a single trunk. (a) Apply Hough transform for detecting the tree trunk centre on the point cloud slice at 1.25 m to 1.35 m in height and calculate the noise level of the point cloud (i.e., thickness) at the tree trunk surface; (b) remove points that fall outside the range of the tree trunk thickness.

in the vertical direction. If the width of consecutive slices increased from bottom to top, it indicated the presence of a tree fork. After that the leaning angle of the tree trunk was calculated and the trunk point cloud was further rotated in the vertical direction. Finally, the position for DBH estimation would be at either 1.3 m in height or at the tree fork, depending on the circumstances. This aligns with the standard position for in-situ DBH measurement. In order to ensure the inclusion of an adequate number of points for precise estimation of DBH, points were extracted specifically at the location designated for DBH estimation with a predefined slice thickness of 10 cm, as illustrated in the inset of Fig. 1. This selection aligns with the approach employed by Huang et al. (2011) for vertically grown tree and Proudman et al. (2021) for leaning tree. Note that numerous research efforts, such as the study by Koreň et al. (2020), have focused on examining the accuracy of DBH estimation across various slice thicknesses. Despite these investigations, a consensus on the most effective slice thickness for precise DBH estimation has yet to be established. Therefore, this study adopts the assumption that a slice thickness of 10 cm is a suitable choice.

As highlighted in the Introduction, numerous approaches have been employed to estimate DBH using LiDAR point cloud data. However, the accuracy of these estimations has been influenced by various factors, leading to potential errors. In light of this, a chord length-based method was proposed to mitigate the errors associated with DBH estimation. Drawing inspiration from the study by Fernández-Sarría et al. (2013) on estimating crown diameter, the proposed method employs a similar approach to estimate DBH. The proposed method initially extracted the points located at the position designated for DBH estimation and subsequently divided them into different sectors of 5°. It was assumed that

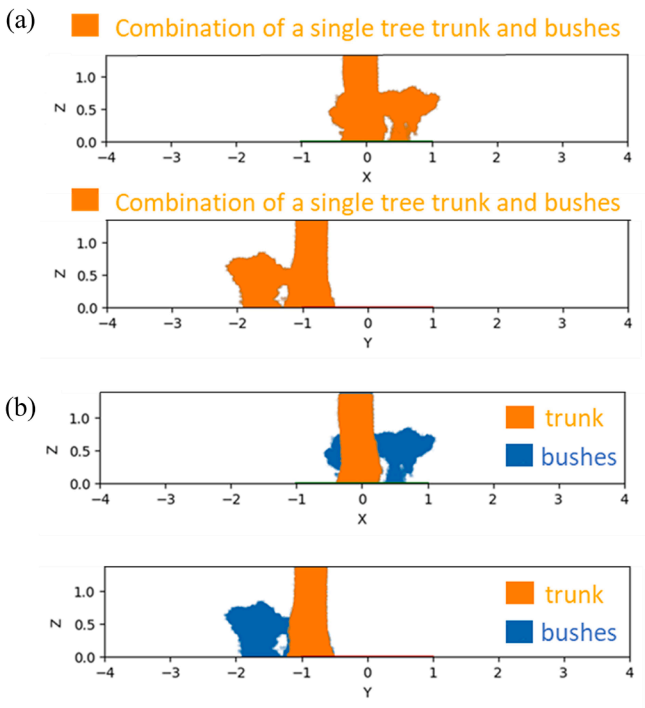


Fig. 4. Results of bushes removed from a cluster that contains both bushes and a single tree trunk at the x-z plane and the y-z plane. (a) the original point cloud cluster; (b) the point cloud cluster after bushes removal.

the outer surface of the tree trunk was located at a certain percentile of the distribution of the distance between the points and the tree trunk centre in that sector which requires calibration for every LiDAR sensor. To ensure dependable calibration results in this study, it is crucial to collect a varied range of DBH values that are uniformly distributed. This approach prevents the calibration process from disproportionately favoring any specific subset of the data, thereby resulting in a more generalized and robust calibration model. Consequently, a Chi-squared goodness-of-fit test was performed to evaluate whether the DBH values measured by tape measure at each site conform to a uniform distribution. Based on the test results, the trees at HKUST Site 3 and Mei Yuen Street Site 5 exhibit a more uniform distribution of DBH values compared to the other three sites. Thus, it was decided that trees at HKUST Site 3 and Mei Yuen Street Site 5 were selected for calibration. Once calibrated, the perimeter of the tree trunk could be estimated by summing of the chord length of the sectors. Assuming the trunk at breast height exhibits a circular cross-section, the estimated perimeter can be converted into DBH using the standard equation for the circumference of a circle. Fig. 5 shows the illustration for the DBH estimation proposed in this study. Kindly note that the choice to segment the point cloud data into 5° sectors has been made in this study to strike an optimal balance among the accuracy of DBH estimation, the effectiveness in precisely detecting the outer surface of trees, and computational efficiency. The discrepancies between the LiDAR measurements and field measurements, obtained using a tape measure, were assessed using Root Mean Square Error (RMSE) and relative RMSE (rRMSE) where their equations are shown as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(DBH_{estimate} - DBH_{tape\ measure}\right)^2} \tag{2.5}$$

$$rRMSE = \frac{RMSE}{\overline{DBH_{tape\ measure}}} \times 100\% \tag{2.6}$$

where n is the number of trees, $DBH_{estimate}$ is the DBH value estimated by the proposed method, $DBH_{tape\ measure}$ is the DBH value measured by the

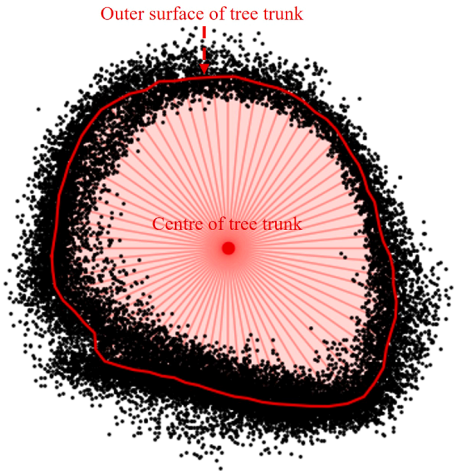


Fig. 5. An illustration of DBH estimation by the proposed method in this study. The point cloud at the position for DBH estimation is divided into multiple circular sectors with varying diameters. The outer surface of the tree trunk is assumed to be located at a certain percentile of the distribution of the distance between the points and the tree trunk centre in a sector. Using this information, the chord length of each sector can be calculated. Summing these chord lengths provides an estimate of the tree trunk's perimeter, which can be then converted into DBH using the standard circumference formula for a circle.

tape measure and $\overline{DBH_{tape\ measure}}$ is the average of DBH value measured by the tape measure.

2.6. Tree image extraction (Stage V)

Despite potential challenges such as occlusion and lighting variations, images remain valuable for tree monitoring and management objectives, including tasks like tree species classification (Liu et al., 2019; Choi et al., 2022) and tree health assessment. Therefore, it becomes essential to extract individual tree images from the images collected from 360° camara. As discussed in Sections 2.2 & 2.3, individual tree images could be extracted by projecting the point cloud frames onto the panoramic images and subsequently leveraging the outcomes of individual tree segmentation on the 3D point cloud map. However, it is important to acknowledge that the vertical field of view of the LiDAR sensor employed in this study is constrained to a maximum of 45°, as mentioned in Section 2.2. This limited range may not offer adequate coverage for capturing a comprehensive representation of the entire tree. Consequently, the act of superimposing the spherical projection of point cloud frames could result in incomplete extraction of individual tree images. To alleviate this constraint, the study opted to overlay the spherical projection of the 3D point cloud map onto the panoramic images. By utilizing the surveying route inferred by the SLAM algorithm, knowledge of the sensors' pose in the 3D point cloud map at the time of the image captured was obtained. This enabled the projection of the 3D point cloud map onto the identical image scene. Using the results of individual tree segmentation in Stage III, labels were assigned to the points within a 3D point cloud map, distinguishing them as either individual trees or non-tree objects. This process facilitated the identification of the precise position of each tree within the image, enabling the subsequent generation of a rectangular bounding box for each tree. Hence, the pixels within these bounding boxes were subsequently extracted to obtain individual tree images. However, it is essential to recognize the potential limitation of this approach. The 3D point cloud map comprises the individual tree point clouds of all the detected trees within the site. When mapping the 3D point cloud map into a spherical projection, it is possible that many trees, particularly those located far from the sensor's position, may become obstructed, similar to what is observed in the panoramic image. Thus, extracting the image of these trees might not yield practical insights for urban forestry applications. To ensure the practicality of the extracted tree images for urban forestry applications, the number of points corresponding to each tree in the point cloud frame captured simultaneously with the corresponding panoramic image was quantified. If the count of points associated with a tree exceeded a predefined threshold value, the corresponding image was considered appropriate for extraction.

3. Results

3.1. 3D point cloud map reconstruction and validation

The LiDAR data collected from the six study areas were input into the FAST-LIO2 algorithm (Xu et al., 2022) which successfully generated 3D point cloud maps, as depicted in Fig. 6. In this study, the four datasets collected within the campus of HKUST, i.e., Sites 1 to 4, were used for validation using a total station with prism sets. The relative distances between the corners of man-made structures and the center of the pole-like objects were calculated using the reconstructed 3D point cloud maps and compared to the measurements by the total station. Fig. A3 displays the selected locations in the 3D point cloud maps for comparison, while Tables A3-A6 provide an overview of the accuracy of the FAST-LIO2 algorithm for 3D point cloud reconstruction of Sites 1 to 4. According to the data, the mean absolute error varied between 0.048 m and 0.086 m, while the standard deviation of the absolute error ranged from 0.024 m to 0.090 m across the four sites. Overall, the adopted FAST-LIO2 algorithm demonstrated a high degree of accuracy in reconstructing the 3D point cloud maps, with a RMSE of 0.081 m and a relative RMSE of 0.354 %. These results suggest that the FAST-LIO2 algorithm is promising for 3D point cloud reconstruction in various locations.

3.2. Results of semantic segmentation

As mentioned in Section 2.3, Stage II involved the implementation of two deep learning models, RandLA-Net and SalsaNext, for semantic segmentation in order to extract the vegetation point cloud. The results of the proposed approach for semantic segmentation are presented in Fig. 7. Table 2 summarizes the IoU values of the semantic segmentation results. Since the focus of this study was not on urban elements, they were grouped into a single class labelled as 'Others' in Table 2. Overall, the IoU of the ground, vegetation and others are 0.97, 0.94 and 0.90 respectively with an average IoU of 0.93. It is noteworthy that RandLA-Net misclassified 18,419 moving object points. In contrast, SalsaNext effectively corrected a significant portion of these misclassifications, specifically 16,783 points, achieving a 91 % correction rate. This indicates that SalsaNext successfully removed the moving object point cloud, thereby facilitating further individual tree segmentation and automatic DBH estimation. Fig. 8 provides an illustrative example. In Fig. 8a, the long trail (depicted in yellow) between the urban elements (highlighted in red) and the trees (shown in green) indicates the existence of moving objects. When relying solely on RandLA-Net, a segment of this long trail was predicted to be urban elements, while another segment was classified as vegetation, as depicted in Fig. 8b. However, SalsaNext accurately identified the entire long trail as a moving object, as demonstrated in Fig. 8c. This precise identification allowed for the effective removal of moving objects from the vegetation point cloud, thus enhancing the segmentation of individual trees. Note that the proposed approach resulted in a marginal increase of 0.0011 in the IoU score for vegetation classification. This slight improvement can be attributed to the fact that moving objects comprised only 0.5 % of the point cloud datasets, while vegetation accounted for 36 % of the datasets. The significant disparity in the number of points representing moving objects and vegetation diminishes the statistical significance of the observed increase in IoU.

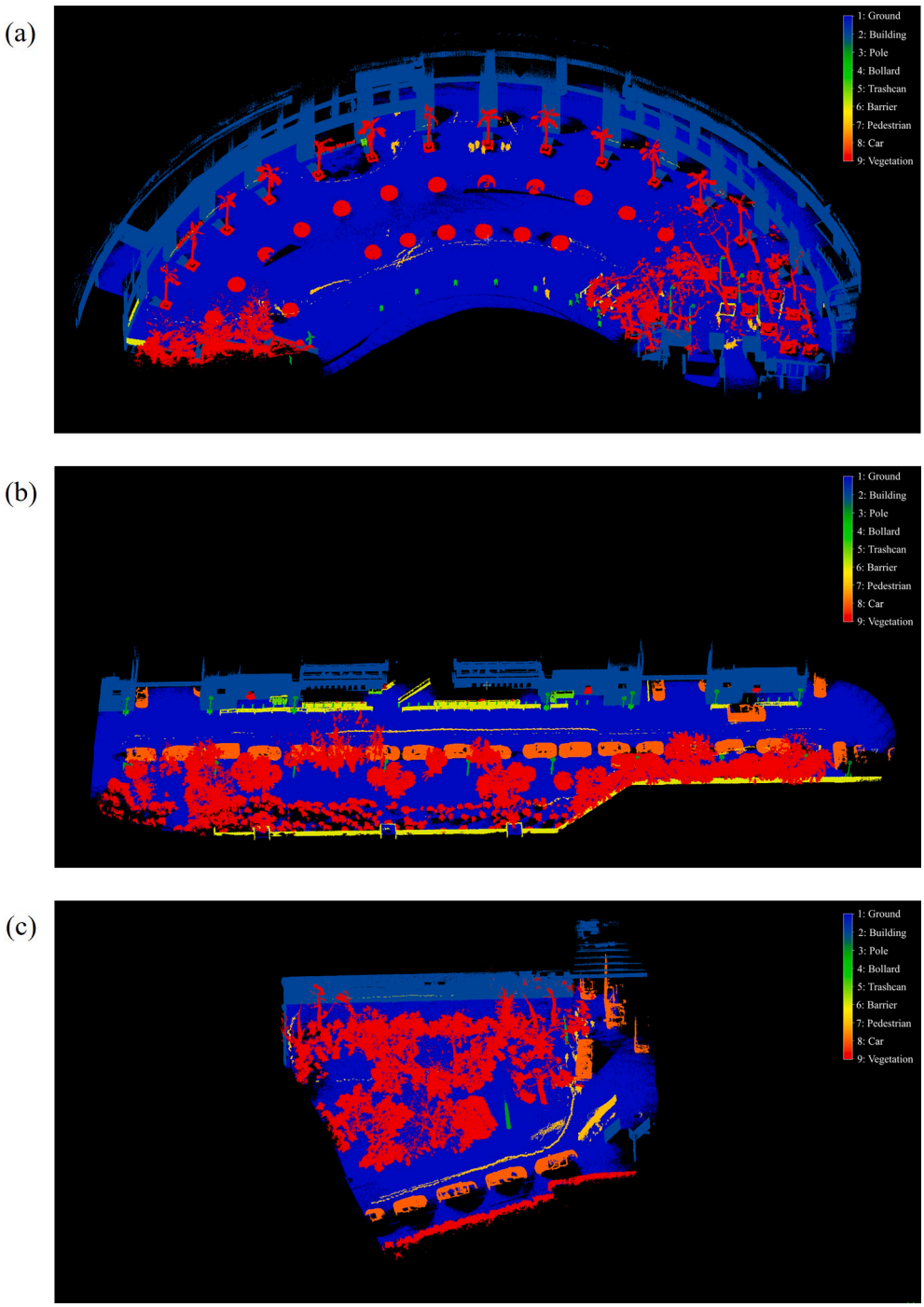


Fig. 6. The reconstructed 3D point cloud maps collected from six different sites: (a) Site 1, (b) Site 2, (c) Site 3, and (d) Site 4 within the campus of HKUST, as well as (e) Site 5 and (f) Site 6 at Mei Yuen Street in Sai Kung, Hong Kong.

3.3. Detection of tree trunk and delineation of tree crown

Following semantic segmentation, only the data points labelled as ground and vegetation were preserved for the purpose of individual tree segmentation in Stage III. The vegetation point cloud was subsequently partitioned into its lower and upper halves, which were then segmented using HDBSCAN and Dijkstra's algorithm, respectively. The outcomes of the individual tree segmentation are depicted in Fig. 9. Additionally, the precision, recall, and F1 score are summarized in Table 3. It is important to note that the success of tree detection is determined by achieving an IoU value exceeding 0.5 for points within the tree point cloud whose height is less than 1.35 m. This criterion is set due to the subjectivity and challenges involved in labeling point clouds of trees with overlapping tree crowns at an individual level. Out of the 262 trees surveyed across the five study areas, 248 trees were successfully identified, resulting in an overall precision, recall, and F1 score of 0.88, 0.95, and 0.91, respectively.

3.4. DBH calibration and estimation

After obtaining the point clouds of individual trees, the DBH of the surveyed trees was estimated in Stage IV. As mentioned in Section 2.5,

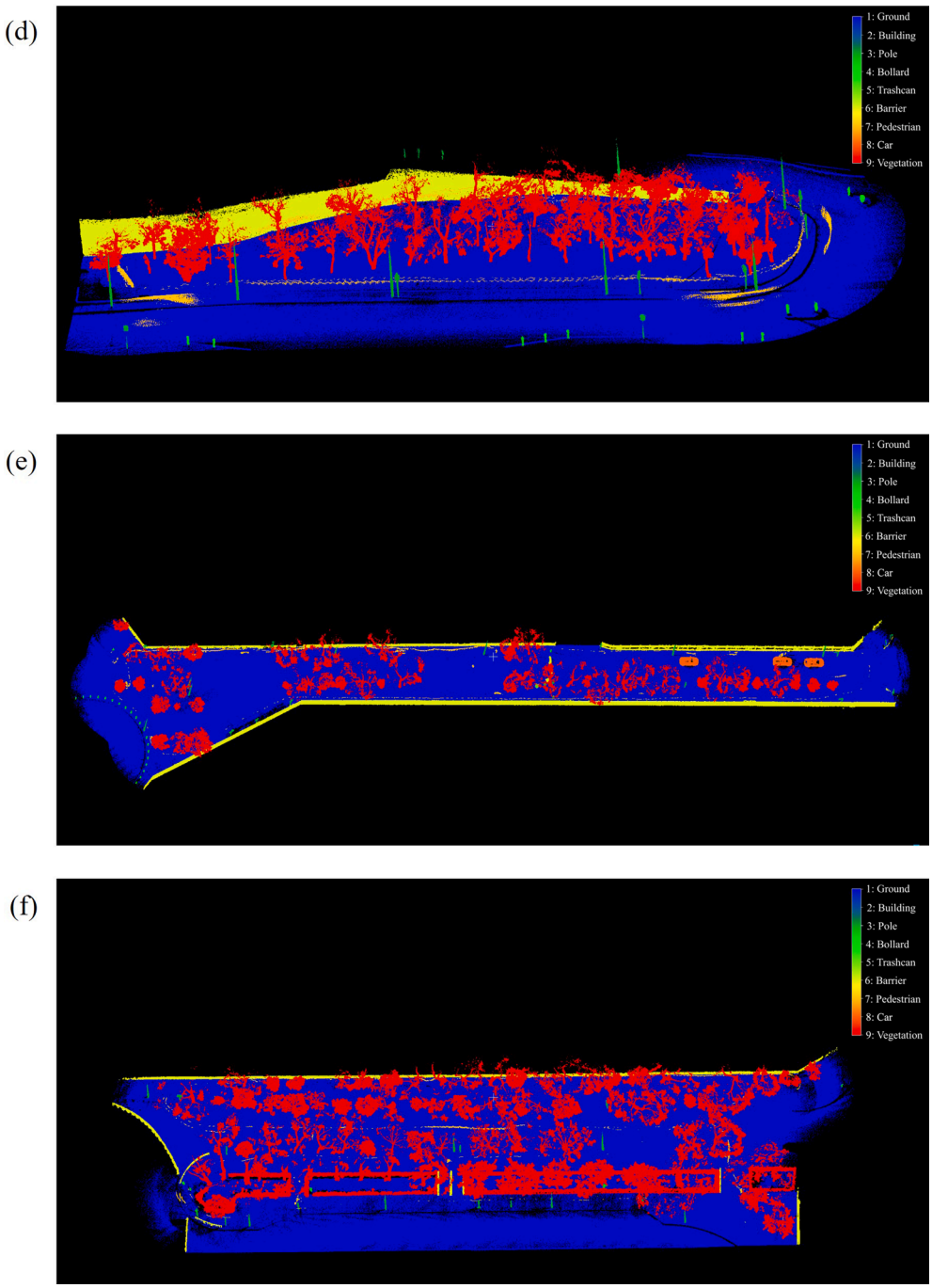


Fig. 6. (*continued*).

successfully detected trees at HKUST Site 3 and Mei Yuen Street Site 5 were selected for calibration which involves assuming the outer surface of the tree trunk to be located at a particular percentile of the distribution of the distance between the points and the tree trunk centre in a given sector. Thus, a trial-and-error approach was adopted to determine the value of the percentile which gives the least RMSE. Table 4 provides a summary of the RMSEs and rRMSEs associated with the estimation of DBH for trees located at HKUST Site 3 and Mei Yuen Street Site 5. The table reveals that the 35^{th} percentile yields the least RMSE value of 2.02 cm and rRMSE value of 10.01 %. The proposed method was also applied to obtain the RMSEs and rRMSEs of DBH estimation for trees located at other three sites. The results are presented in Table 5. The RMSE and the rRMSE for these three sites are 1.90 cm and 10.69 % respectively. Therefore, the proposed method has demonstrated feasibility in accurately estimating the DBH with an overall RMSE of 1.93 cm and rRMSE of 10.50 %. Table 5 also shows the RMSE and rRMSE obtained from the application of the ellipse least squares fitting method for estimating DBH. Furthermore, Fig. 10a illustrates a comparison between the field measurements and LiDAR estimation for trees in all 5 study areas, while Fig. 10b showcases the RMSE across various DBH value ranges in these study areas.

3.5. Tree image extraction

Finally, in Stage V, the results of individual tree segmentation in Stage III were utilized to determine the position of the tree in the panoramic image captured by the 360° camera for tree image extraction. An example of tree image extraction is illustrated in Fig. 11, which

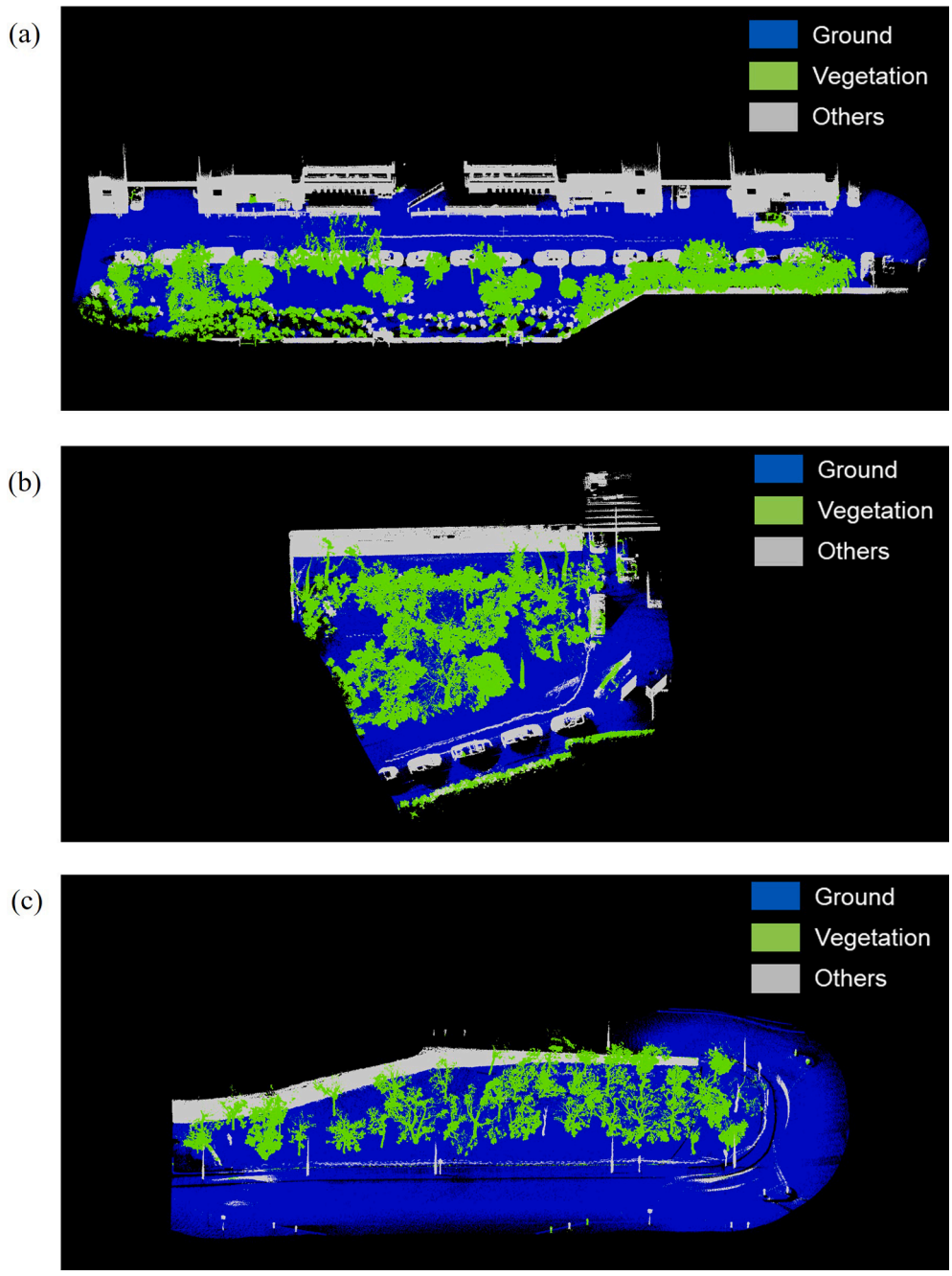


Fig. 7. The results of semantic segmentation for (a) Site 2, (b) Site 3, and (c) Site 4 within the HKUST campus, as well as for (d) Site 5 and (e) Site 6 at Mei Yuen Street were obtained using RandLA-Net and subsequently refined by SalsaNext. Note that ground and vegetation points are highlighted in blue and green, respectively, while all other points are shown in grey.

showcases an image captured from Site 4. As mentioned in Section 2.6, the limited range of vertical field of view of the LiDAR sensor might lead to incompleteness of capturing the entire tree as shown in Fig. 11a. To overcome this limitation, this study proposed overlaying the panoramic images with the spherical projection of a 3D point cloud map, as depicted in Fig. 11b. Based on the results of individual tree segmentation, the 3D point cloud map was categorized into distinct trees which were labelled in different colors and rectangular bounding boxes were subsequently generated for every tree for image extraction. Two examples are shown in Fig. 11c & 11d. These extracted images held potential for further urban forestry applications such as tree species classification and tree health monitoring. It should be highlighted that Fig. 11a illustrates a total of 35 trees within the respective point cloud frame. A considerable number of these trees were positioned at significant distances from the sensor, with the furthest being up to 53 m away. Additionally, the minimum distance between recorded two adjacent trees was 1.63 m. These specific trees appeared relatively small in the captured image and were partially concealed by adjacent trees. Consequently, the extraction of images of these trees may not offer practical value for the intended objective. Therefore, Fig. 11b illustrates the detection of only four trees.

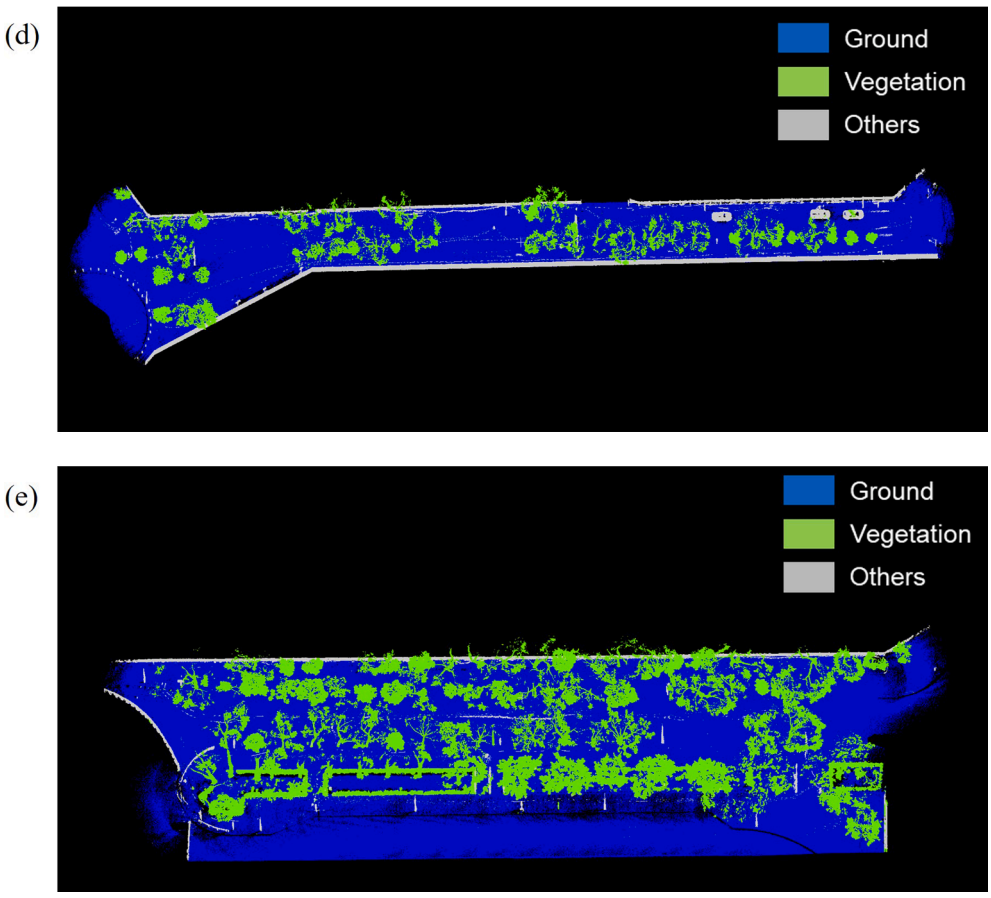


Fig. 7. (*continued*).

Table 2
Intersection over Union (IoU) of the semantic segmentation results.

Testing dataset	Ground	Vegetation	Others	mIoU
HKUST Site 2	0.97	0.92	0.89	0.93
HKUST Site 3	0.97	0.93	0.87	0.92
HKUST Site 4	0.99	0.98	0.95	0.97
Mei Yuen Street Site 5	0.98	0.97	0.94	0.97
Mei Yuen Street Site 6	0.95	0.93	0.78	0.89
Overall	0.97	0.94	0.90	0.93

4. Discussion

4.1. Performance of the proposed method for vegetation segmentation

In the studies conducted by Chen et al. (2021b), Luo et al. (2021), Lei et al. (2022), Wang et al. (2023b), Xia et al. (2023), and our own research, deep learning models are employed to extract the vegetation point cloud before performing instance tree segmentation. Thus, this step is crucial for accurate segmenting individual trees in the urban environments. Unlike previous studies, this study integrated the outputs of RandLA-Net and SalsaNext for vegetation point cloud extraction, rather than exclusively relying on the prediction of pointwise neural network. This dual-model approach leverages the strengths of each model, potentially leading to more robust and accurate vegetation segmentation. The effectiveness of this approach was demonstrated by the successful rectification of a significant proportion of moving object points misclassified by RandLA-Net as shown in Section 3.2. This precise identification allowed for the effective removal of moving objects from the vegetation point cloud, thus enhancing the segmentation of individual trees. Based on the data provided in Table 6, relying exclusively on RandLA-Net for semantic segmentation results in a total of 111 individual tree point clouds containing moving objects. However, by implementing the proposed approach that combines the results of RandLA-Net and SalsaNext, the removal of moving objects was successfully achieved in 52 individual tree point clouds (47 %), and at least 90 % of the moving objects were eliminated from 91 individual tree point clouds (82 %). As mentioned in Section 2.3, the appearance of moving objects within the 3D point cloud map can vary depending on their velocity and trajectory, potentially presenting as long trails or taking on different structures, shapes, and point distributions. Therefore, RandLA-Net or other pointwise neural network, which relies on relative spatial information of neighboring points for semantic segmentation, has a higher likelihood of misclassifying the moving object point clouds as various objects, including vegetation. Consequently, these misclassified moving object point clouds can aggregate within the cluster of tree trunks through HDBSCAN, as illustrated in Fig. 12. This aggregation directly impacts the accurate implementation of automatic DBH measurement when utilizing the resulting individual tree point cloud. In contrast, SalsaNext, which analyzes spatiotemporal information from the concatenated range-residual images, is able to effectively identify and distinguish moving objects from the background. The incorporation of SalsaNext to refine the segmentation outputs generated by RandLA-Net leads to improved accuracy in individual tree point cloud segmentation, which in turn facilitates automated tree geometry measurement. This advancement holds significant importance in the continuous progress of automated urban tree inventory systems. Note that the proposed method has yielded a marginal increase of 1.1 % in precision and 0.6 % in the F1 score for the detection of trees. This slight improvement can be attributed to the fact that a majority of the incorrectly classified moving object point clouds were in close proximity to the tree point clouds. As a result, during the process of segmenting

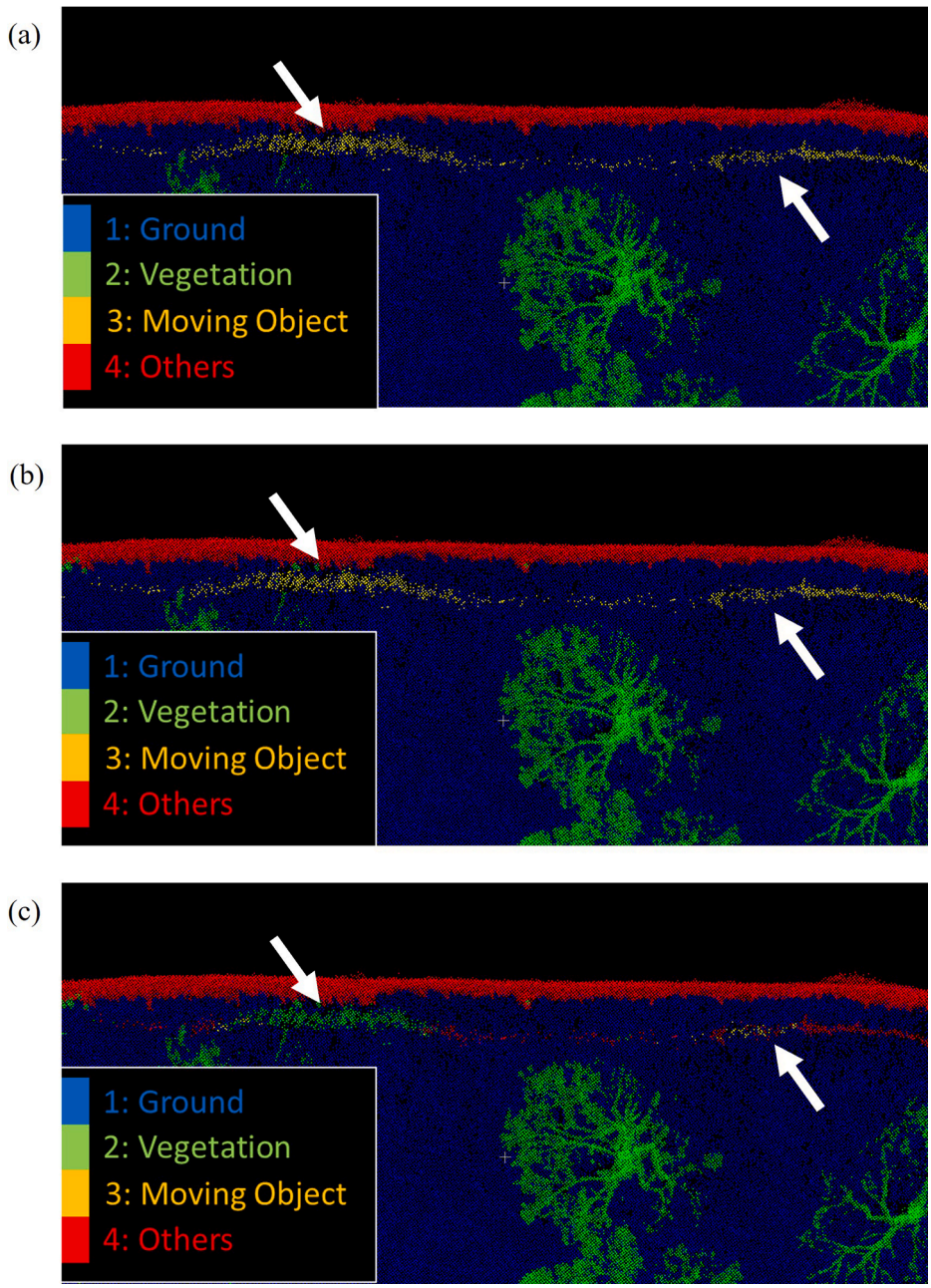


Fig. 8. Comparison of the semantic segmentation results: (a) the ground truth, achieved by (b) RandLA-Net exclusively and (c) RandLA-Net with SalsaNext. Note that the white arrows indicate the moving object point cloud.

individual tree trunks in Stage III, HDBSCAN merged these misclassified moving object point clouds with the individual tree trunks, as shown in Fig. 12. Consequently, the improvement in these evaluation metrics was not statistically significant. Although this enhancement does not have a substantial impact on the accuracy of tree detection in terms of the evaluation metrics, it does facilitate the extraction of non-biased individual tree point clouds. This, in turn, assists in the automation of DBH estimation. Indeed, it is crucial to acknowledge that the results may vary when conducting tests in environments with higher population densities. Therefore, additional research is necessary to investigate the effects on the outcomes of semantic segmentation and tree detection in such scenarios. It is crucial to underscore that within the proposed framework, the utilization of RandLA-Net and SalsaNext is not exclusive. Other deep learning models, such as PointNLM (Chen et al., 2021b) and enhanced versions of RandLA-Net (Lei et al., 2022; Xia et al., 2023), could serve as viable substitutes. These models have demonstrated notable improvements in overall accuracy for vegetation segmentation, with gains ranging from 0.17 % to 10.61 % compared to RandLA-Net. As the field of deep learning continues to evolve and new breakthroughs emerge, new models could also be integrated into the framework in the future.

4.2. *Performance of the proposed method for DBH estimation*

In accordance with Section 1, several factors can affect the accuracy of DBH estimation. These factors include the non-uniform shape of the tree trunk at breast height (Bauwens et al., 2016), the impact of point cloud density (Balenović et al., 2021), LiDAR measurement errors (Forsman et al., 2018), and bark roughness (Zeybek and Vatandaşlar, 2021; Tsuchiya et al., 2023). Upon analysis of the collected dataset, it becomes evident that irregularities in tree trunk shape are not uncommon. Consequently, it is more appropriate to consider the actual

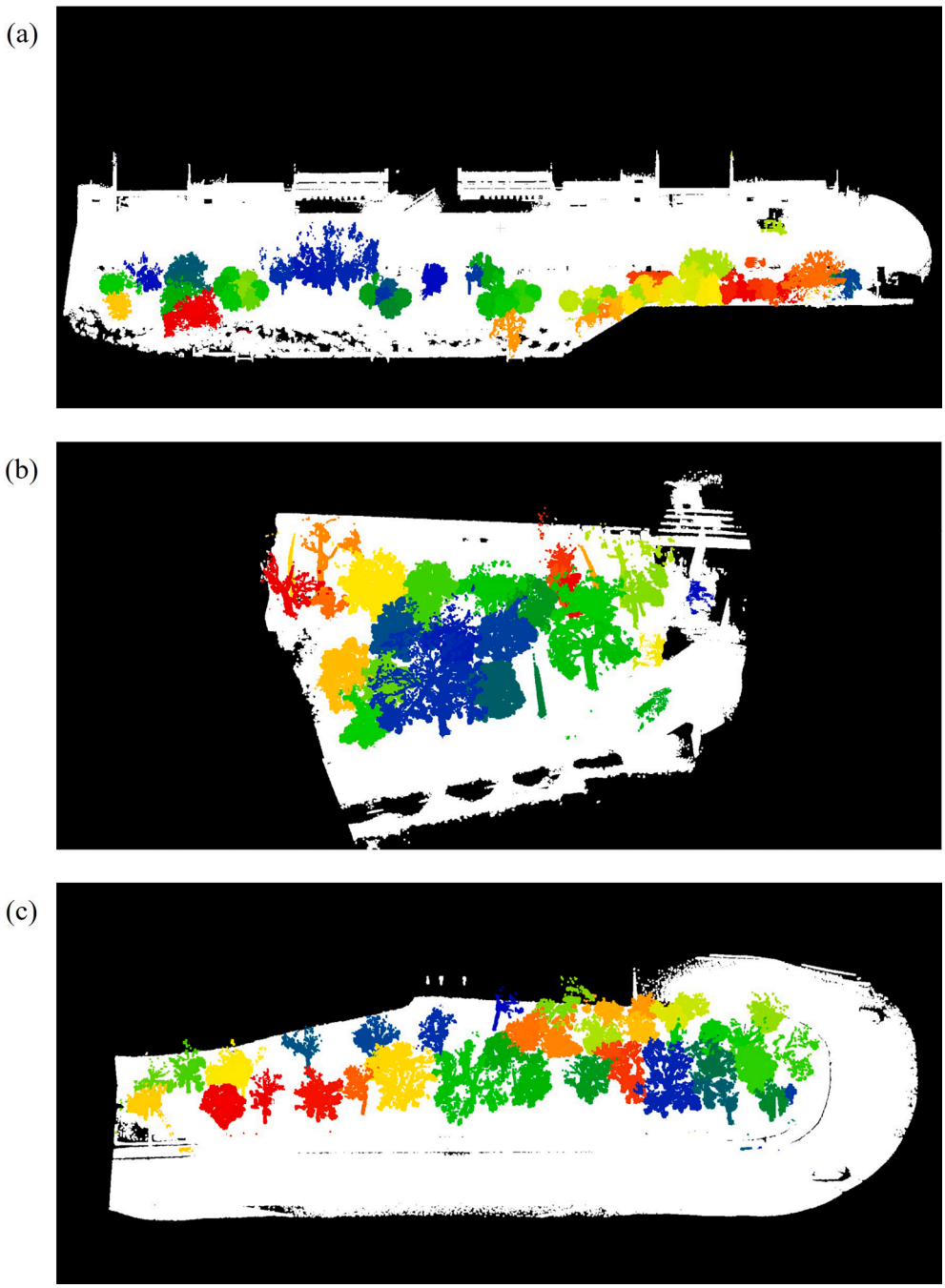


Fig. 9. The results of individual tree segmentation of (a) Site 2, (b) Site 3, and (c) Site 4 within the HKUST campus, as well as for (d) Site 5 and (e) Site 6 at Mei Yuen Street in Sai Kung. Note that non-tree point clouds are depicted in white colour, while each individual tree is presented by a distinct colour.

perimeter of the tree trunk cross section rather than relying on predetermined shape assumptions used in least squares fitting or RANSAC fitting methods. The data presented in Table 5 supports the superiority of the proposed method over the ellipse least squares fitting method, as indicated by lower rRMSE values ranging from 3.26 % to 12.66 %, demonstrating more accurate DBH estimation. Note that this proposed approach yields results consistent with previous studies (Huang et al., 2011; Bauwens et al., 2016; Koreň et al., 2020; Balenović et al., 2021; Proudman et al., 2021; Zeybek and Vatandaşlar, 2021; Tsuchiya et al., 2023), which reported RMSE values ranging from 0.56 cm to 7.00 cm.

It is important to note that the point cloud density is significantly influenced by the chosen surveying route. In the context of DBH estimation using the least squares fitting method, non-uniform point cloud density can have a significant impact on accuracy. The standard least squares fitting approach assumes equal precision and reliability for all data points, which may not reflect the reality of lower point concentrations in certain trunk sections. This uneven distribution can distort the fitting process and introduce errors in DBH estimation. In contrast, the proposed method in this study addresses this issue by dividing the trunk into multiple sectors and assuming the tree trunk surface corresponds to a specific percentile of the point distribution within these sectors. This eliminates potential influences on the estimation process due to uneven point density.

Additionally, a calibration process was employed to address LiDAR

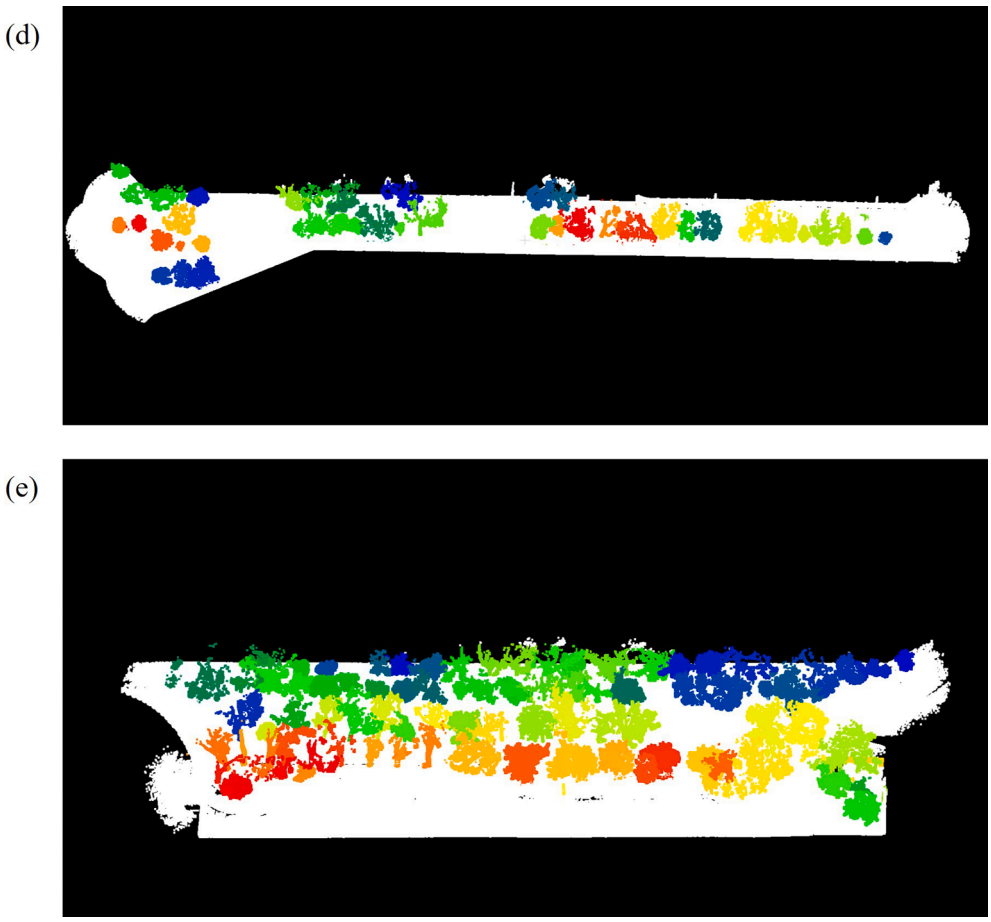


Fig. 9. (*continued*).

Table 3
Results of individual tree detection.

	TP	FP	FN	Precision	Recall	F1 score
HKUST Site 2	61	11	10	0.85	0.86	0.85
HKUST Site 3	22	7	0	0.76	1.00	0.86
HKUST Site 4	39	2	2	0.95	0.95	0.95
Mei Yuen Street Site 5	42	3	1	0.93	0.98	0.95
Mei Yuen Street Site 6	84	12	1	0.90	0.99	0.94
Total	248	35	14	0.88	0.95	0.91

measurement errors and potential variations between the trunk's inner and outer bark. This involves assuming a specific percentile of the distance distribution between points and the trunk center as the outer surface of the trunk. Fig. 10b illustrates the RMSE of DBH estimation for various ranges of DBH values. It is evident that, in the majority of cases, the proposed method in this study outperforms the least squares fitting method, as indicated by lower RMSE values. Furthermore, Balenović et al. (2021) reviewed that high beam divergence results in relatively low point density in small trees with a DBH of less than 10 cm, leading to a higher-level error of DBH estimation. As illustrated in Fig. 10b, the RMSE for trees with a DBH ranging from 6 to 10 cm demonstrates similar error levels to those observed in trees with a DBH greater than 10 cm when using the proposed method. This evidence substantiates that the proposed calibration-inclusive method contributes to error minimization. It is important to acknowledge that the RMSE exhibits an upward trend at both ends of the curve in Fig. 10b, due to the scarcity of calibration samples in those particular ranges. Hence, additional investigation is necessary to validate the proposed method specifically for trees falling within these particular ranges of DBH values.

Table 4
Summary of the RMSE and rRMSE of DBH estimation for trees at HKUST Site 3 and Mei Yuen Street Site 5. The analysis assumes that the outer surface of the tree trunk is positioned at different percentiles of the distribution of distances between points and the trunk center. Additionally, the table includes the average DBH value of the trees at both sites.

Study area	Average DBH (ground truth)	RMSE / rRMSE				
Percentile		30^{th}	35^{th}	40^{th}	45^{th}	50^{th}
HKUST Site 3	22.40 cm	1.97 cm / 8.80 %	1.88 cm / 8.41 %	1.91 cm / 8.52 %	2.02 cm / 9.03 %	2.22 cm / 9.82 %
Mei Yuen Street Site 5	19.43 cm	2.32 cm / 11.94 %	2.12 cm / 10.90 %	2.11 cm / 10.85 %	2.25 cm / 11.60 %	2.49 cm / 12.80 %
Overall	20.22 cm	2.19 cm / 10.82 %	2.02 cm / 10.01 %	2.03 cm / 10.03 %	2.16 cm / 10.69 %	2.38 cm / 11.75 %

4.3. Potential application of the extracted DBHs and tree images

This study primarily focuses on the extraction of individual tree point clouds for DBH estimation. Although it is possible to estimate other tree geometry parameters like tree height and tree crown width (Cabo et al., 2018; Fan et al., 2020; Tsuchiya et al., 2023), these parameters were not estimated in this study due to certain limitations: (1) accurate tree height estimation is compromised by the presence of a lower canopy

Table 5
Comparative analysis of RMSE and rRMSE of DBH estimation using the proposed method and the ellipse least squares fitting method. Note that the table includes the average DBH value of the trees at the study areas.

Study area	Average DBH (ground truth)	RMSE / rRMSE (proposed method)	RMSE / rRMSE (least squares method)
HKUST Site 2	13.39 cm	1.45 cm / 10.79 %	3.14 cm / 23.45 %
HKUST Site 3	22.40 cm	1.88 cm / 8.41 %	2.61 cm / 11.67 %
HKUST Site 4	17.45 cm	1.90 cm / 10.88 %	2.51 cm / 14.41 %
Mei Yuen Street Site 5	19.43 cm	2.12 cm / 10.90 %	3.32 cm / 17.08 %
Mei Yuen Street Site 6	21.08 cm	2.17 cm / 10.29 %	3.69 cm / 17.51 %
Overall (Site 2, 4 & 6)	17.77 cm	1.90 cm / 10.69 %	3.29 cm / 18.54 %
Overall (All)	18.40 cm	1.93 cm / 10.50 %	3.24 cm / 17.61 %

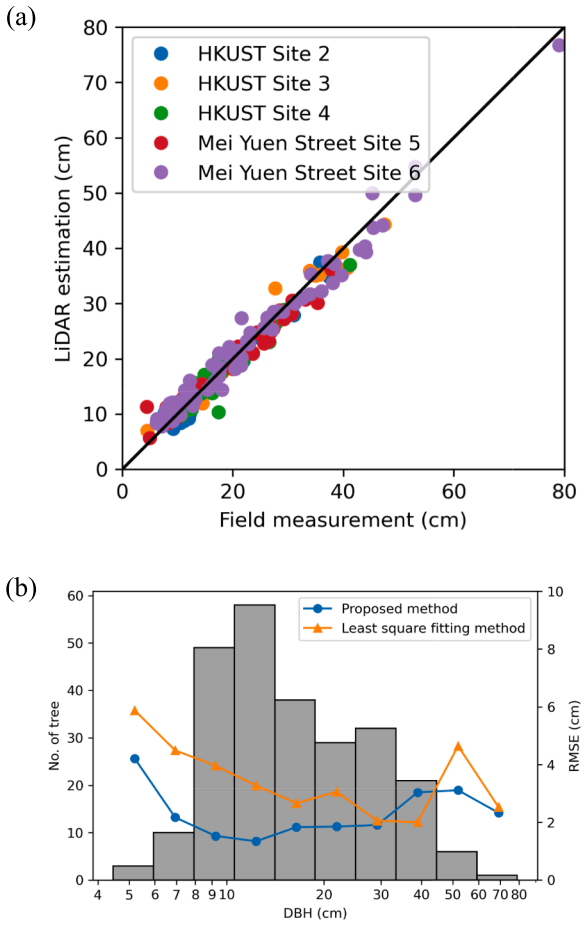


Fig. 10. Accuracy of DBH values estimated using the proposed method in this study: (a) a comparison with field measurements using a tape measure; (b) the RMSE of DBH estimation by the proposed method and the ellipse least squares fitting, and the number of trees in different DBH ranges across all five study areas.

obstructing the collection of point cloud data from the upper canopy (Chiappini et al., 2022); (2) segmentation of tree crowns proves challenging due to crown touching and interlocking (Li et al., 2021). Consequently, the study concentrates solely on extracting the DBH of the trees. Notwithstanding these constraints, it is crucial to acknowledge that the utilization of allometric equations solely reliant on DBH remains viable for estimating above ground biomass (Yoon et al., 2013). This has substantial implications for urban planning and the formulation of efficacious climate change mitigation strategies (Pataki et al., 2021). Despite the limited research conducted on developing allometric modeling for Hong Kong's urban trees, it is possible to utilize models derived from neighboring cities or countries and refine them by calibrating with fallen trees resulting from typhoons or urban development. With the proposed framework, the DBH could be frequently updated to monitor the carbon sequestration of urban trees in Hong Kong. Moreover, in the context of urban forestry practices in Hong Kong, the physical dimensions of trees are primarily recorded for those classified as potentially hazardous. However, this framework allows for the documentation of all trees' physical dimensions in a specific area, including low-risk ones, improving the tree inventory's comprehensiveness.

In addition to the pivotal role of tree carbon sequestration, the monitoring and assessment of tree health are of utmost significance. To address this, the current study employed a 360° camera to capture a diverse array of panoramic images, encompassing trees from various angles throughout the site. This methodology enables the classification of tree species, the estimation of growing stock volume (Liu et al., 2019; Choi et al., 2022), and facilitates a thorough visual assessment of the trees. By integrating the results of visual tree assessment with long-term monitoring data obtained through the utilization of Internet of Things (IoT) technology, a reliable tree database could be established. This will enable arborists to track the tree history, enhance diagnosis accuracy, and make informed decisions for efficient tree management (Chau et al., 2023). Furthermore, these extracted images hold valuable potential for developing deep learning models that are specifically designed to identify and detect diseases in trees (Anagnostis et al., 2021; Singh et al., 2021; Ye et al., 2024). The recent developments in large-scale multimodal models, as demonstrated by Liu et al. (2024), demonstrate their capacity to facilitate complex reasoning, utilize extensive knowledge of the world, and offer explanatory responses. These advancements present a promising prospect for future research to investigate the viability of conducting thorough analysis of tree health using the images collected as a preliminary screening method for hazardous trees. This approach could be integrated into tree risk assessment practices in Hong Kong, initially identifying potentially hazardous trees and subsequently performing detailed health assessments on the individual trees of concern. The integration of artificial intelligence into tree risk assessments holds significant potential in providing valuable assistance and insights to arborists. This methodology has the potential to effectively mitigate the risk of human errors, subsequently enhancing the accuracy and reliability of evaluations. As a result, it could potentially decrease the occurrence of fatal accidents stemming from tree failure.

4.4. Operational challenges and limitations of the proposed mobile mapping system

Several operational issues necessitate improvements to the framework's practicality. The current post-processing data analysis could be enhanced by real-time processing, requiring algorithm optimization and hardware upgrades, thereby improving immediate decision-making, efficiency, and resource management. The weight of the StructXray system, which is approximately 8–10 kg, could potentially lead to operator fatigue during prolonged periods of data collection. This, in turn, may have an impact on the quality of the data gathered. This could be resolved by mounting the system on a robotic dog, reducing operator strain and ensuring consistent survey speed. Carrying StructXray may also limit access to complex environments and cause occlusion in urban areas, leading to incomplete tree data. This can be mitigated by careful survey route planning (Bauwens et al., 2016), using sensors with a larger vertical field of view, and integrating aerial-based sensor data (Chen et al., 2024). Addressing these challenges is crucial for the framework's practical implementation in various urban settings.

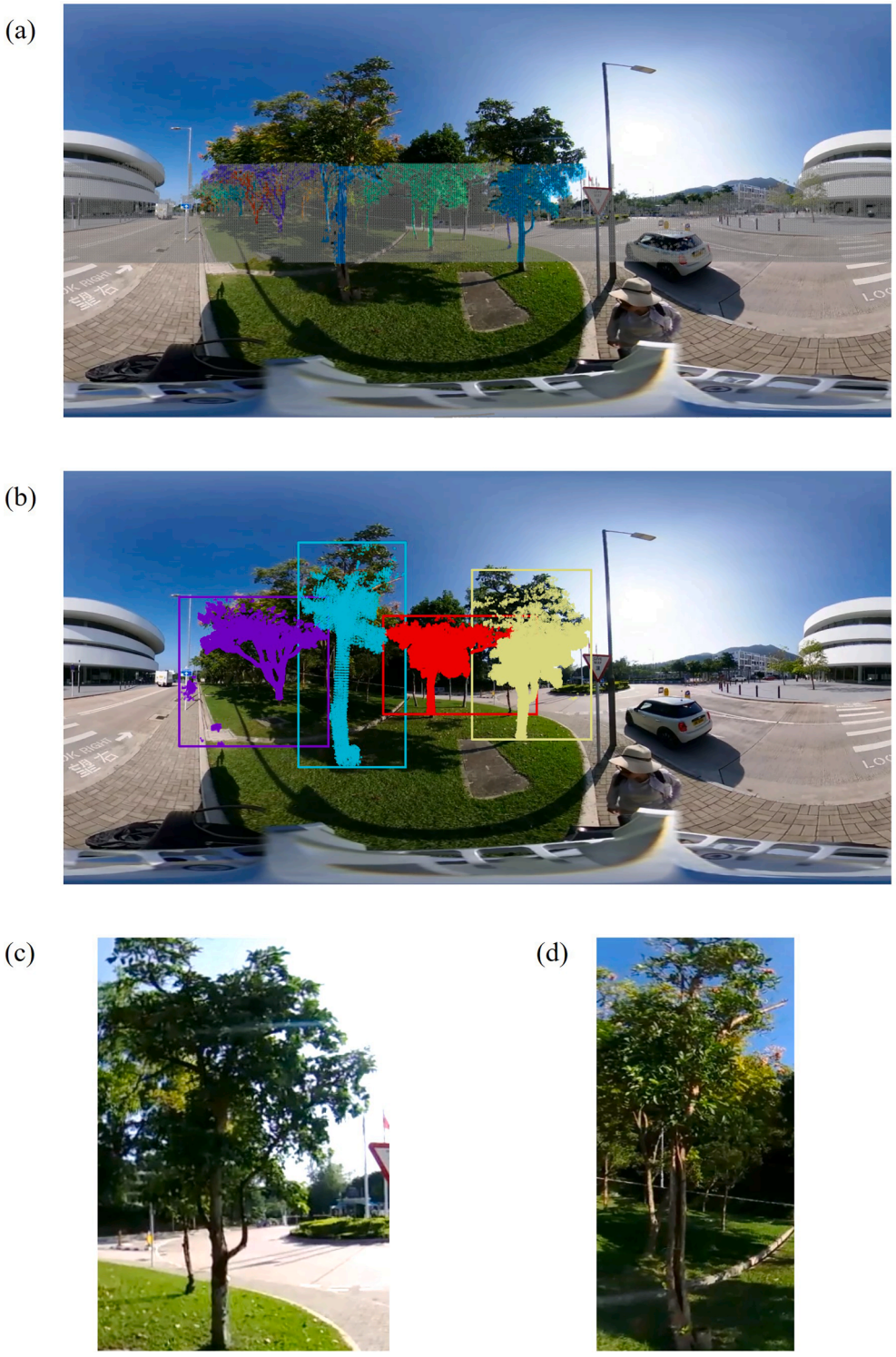


Fig. 11. An example of tree image extraction involves overlaying (a) the spherical projection of a point cloud frame and (b) the 3D point cloud map onto a panoramic image; and (c) & (d) extracting tree images from each bounding box. It is worth noting that the point cloud frame depicted in (a) shows a total of 35 visible trees. However, it is important to acknowledge that 31 of these trees are represented by fewer than 500 points. This suggests that these particular trees are either situated at a significant distance from the sensor's position or obstructed by other trees. Therefore, extracting images of these trees may not be appropriate for urban forestry applications. As a result, only four trees are featured in (b). Note that the point clouds and bounding boxes of each tree are depicted in unique colors in (a) & (b). In contrast, the point clouds of non-tree objects, as well as those measured at distances greater than 10 m from the sensor, which are excluded from the analysis as detailed in Section 2.2, are shown solely in grey color in (a).

Table 6
Count of trees with varying percentages of moving objects removed by SalsaNext.

	Number of trees						
Percentage of moving objects removed by SalsaNext in individual tree point clouds	100 %	>90 %	>80 %	>70 %	>60 %	>50 %	≥0 %
Site 2	12	16	18	19	19	20	20
Site 3	4	5	6	6	6	7	8
Site 4	8	13	13	13	14	14	16
Site 5	2	19	19	20	20	20	21
Site 6	26	38	43	44	44	44	46
Total	52	91	99	102	103	105	111

Fig. 12. HDBSCAN erroneously clusters misclassified moving object point clouds near tree trunks into a unified tree point cloud, thereby presenting a formidable challenge for the automated measurement of DBH. Note that the misclassified moving object point cloud is indicated by the white arrow.

5. Conclusions

The findings of this study underscore the transformative potential of integrating advanced sensing technologies and artificial intelligence into urban forestry management in Hong Kong. By developing a robust framework for estimating DBH and extracting tree images, this research has demonstrated improvements in the accuracy and efficiency of urban tree inventories. The innovative use of StructXray for data acquisition, coupled with deep learning and machine learning techniques for vegetation point cloud extraction and tree segmentation, has demonstrated promising results, highlighting the reliability of the proposed framework. Additionally, the study effectively addresses the challenge of identifying moving objects within 3D point cloud maps using the SalsaNext model, which enhances unbiased individual tree segmentation, thereby facilitating the automatic extraction of geometric parameters. The introduction of a chord length-based method for DBH estimation offers significant improvements in reducing estimation errors compared to traditional methods. These advancements not only contribute to more precise urban tree monitoring but also pave the way for future research to explore additional applications of automated systems in urban forestry. The insights gained from this study suggest promising directions for further enhancing urban tree management practices, ultimately contributing to more sustainable and resilient urban environments.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

CRediT authorship contribution statement

Wai Yi Chau: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation. **Jun Kang Chow:** Methodology, Formal analysis, Data curation. **Tun Jian Tan:** Methodology, Formal analysis, Data curation. **Jimmy WU:** Methodology, Formal analysis. **Mei Ling Leung:** Methodology, Data curation. **Pin Siang Tan:** Methodology, Conceptualization. **Siu Wai Chiu:** Writing – review & editing, Supervision, Conceptualization. **Billy Chi Hang Hau:** Writing – review & editing, Supervision, Conceptualization. **Hok Chuen Cheng:** Writing – review & editing, Supervision, Conceptualization. **Yu-Hsing Wang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: [Yu-Hsing Wang reports financial support was provided by the Innovation Technology Fund, Midstream Research Programme for Universities. Yu-Hsing Wang reports financial support was provided by the Hong Kong Research Grants Council].

Data availability

Data will be made available on request.

Acknowledgement

This study was supported by the Innovation Technology Fund, Midstream Research Programme for Universities [Project No. MRP/003/21X], and the Hong Kong Research Grants Council [Project No. 16205021]. The authors are also grateful to the reviewers for their valuable comments.

Appendix

Table A1
Summary of training, validation and testing datasets used for training the RandLA-Net models.

Model	1		2	3	4
Training datasets	HKUST Sites 2, 3 & 4 Lille1_1 Lille1_2 Paris	HKUST Sites 2, 3 & 4 Lille1_1 Lille1_2 Paris	HKUST Sites 3 & 4 Lille1_1 Lille1_2 Paris	HKUST Sites 2 & 4 Lille1_1 Lille1_2 Paris	HKUST Sites 2 & 3 Lille1_1 Lille1_2 Paris
Validation datasets	HKUST Site 1 Lille2	HKUST Site 1 Lille2	HKUST Site 1 Lille2	HKUST Site 1 Lille2	HKUST Site 1 Lille2
Testing dataset	Mei Yuen Street Site 5	Mei Yuen Street Site 6	HKUST Site 2	HKUST Site 3	HKUST Site 4

Table A2
Summary of training, validation and testing datasets used for training the SalsaNext models.

Model	1		2	3	4
Training & Validation datasets	HKUSTSites 1, 2, 3 & 4	HKUSTSites 1, 2, 3 & 4	HKUSTSites 1, 3 & 4	HKUSTSites 1, 2 & 4	HKUSTSites 1, 2 & 3
Testing dataset	Mei Yuen Street Site 5	Mei Yuen Street Site 6	HKUST Site 2	HKUST Site 3	HKUST Site 4

Table A3
Comparison of the distances measured by the Gowin total station and estimated by reconstructed 3D point cloud map of Site 1.

Distance between			Distance measured by Gowin (m)	Distance estimated by LiDAR (m)	AbsoluteError (m)	Absolute Relative Error
point 1	&	point 2	22.004	21.910	0.094	0.428 %
point 2	&	point 3	16.771	16.837	0.066	0.394 %
point 3	&	point 4	19.014	19.116	0.102	0.532 %
point 4	&	point 5	16.835	16.731	0.104	0.619 %
point 5	&	point 6	14.533	14.575	0.042	0.286 %
point 6	&	point 7	19.101	19.049	0.052	0.270 %
point 7	&	point 8	14.447	14.507	0.060	0.414 %
point 8	&	point 9	16.904	16.805	0.099	0.588 %
point 9	&	point 10	10.555	10.486	0.068	0.650 %
point 10	&	point 1	62.022	62.074	0.052	0.084 %
				Mean	0.074	0.427 %
				Median	0.067	0.421 %
				Std.	0.024	0.179 %

Table A4
Comparison of the distances measured by the Gowin total station and estimated by reconstructed 3D point cloud map of Site 2.

Distance between			Distance measured by Gowin (m)	Distance estimated by LiDAR (m)	Absolute Error (m)	Absolute Relative Error
point 1	&	point 3	23.792	23.832	0.040	0.169 %
point 3	&	point 4	17.581	17.520	0.062	0.354 %
point 4	&	point 6	20.143	20.091	0.052	0.261 %
point 6	&	point 7	23.718	23.679	0.039	0.165 %
point 7	&	point 8	17.103	17.113	0.010	0.061 %
point 8	&	point 9	10.522	10.484	0.038	0.364 %
point 9	&	point 5	46.105	46.034	0.071	0.154 %
point 5	&	point 2	26.149	26.177	0.027	0.104 %
point 2	&	point 1	23.134	23.041	0.093	0.402 %
				Mean	0.048	0.226 %
				Median	0.040	0.169 %
				Std.	0.025	0.123 %

Table A5
Comparison of the distances measured by the Gowin total station and estimated by reconstructed 3D point cloud map of Site 3.

Distance between			Distance measured by Gowin (m)	Distance estimated by LiDAR (m)	Absolute Error (m)	Absolute Relative Error
point 1	&	point 2	33.487	33.581	0.094	0.280 %
point 2	&	point 3	19.379	19.337	0.042	0.215 %
point 3	&	point 4	11.328	11.257	0.071	0.631 %
point 4	&	point 5	20.771	20.783	0.012	0.058 %
point 5	&	point 1	15.007	14.864	0.143	0.959 %
				Mean	0.072	0.429 %
				Median	0.072	0.355 %
				Std.	0.045	0.325 %

Table A6
Comparison of the distances measured by the Gowin total station and estimated by reconstructed 3D point cloud map of Site 4.

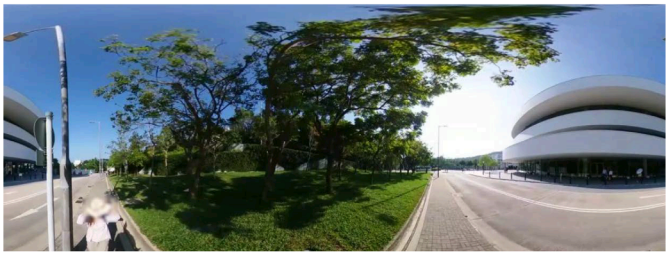
Distance between			Distance measured by Gowin (m)	Distance estimated by LiDAR (m)	Absolute Error (m)	Absolute Relative Error
point 1	&	point 2	16.204	16.219	0.007	0.045 %
point 2	&	point 3	34.402	15.809	0.013	0.038 %
point 3	&	point 4	28.521	25.108	0.041	0.145 %
point 4	&	point 5	34.337	28.405	0.181	0.530 %
point 5	&	point 1	30.281	30.319	0.186	0.612 %
				Mean	0.086	0.274 %
				Median	0.041	0.145 %
				Std.	0.090	0.276 %

(a)

(b)

(c)

(d)



(e)

(f)

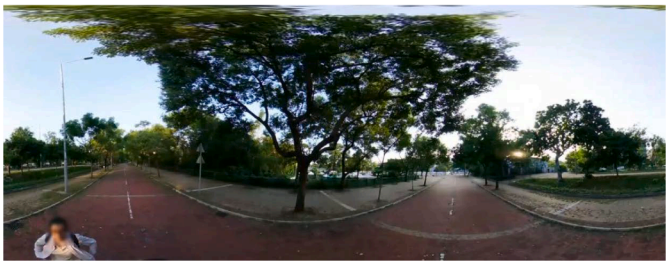


Fig. A1. Six datasets were collected from (a) Site 1; (b) Site 2; (c) Site 3; (d) Site 4 at the HKUST campus; (e) Site 5; and (f) Site 6 at Mei Yuen Street in Sai Kung, Hong Kong.

Fig. A2. The Gowin TKS-202 total station was selected to validate the 3D point cloud reconstruction using FAST-LIO2.

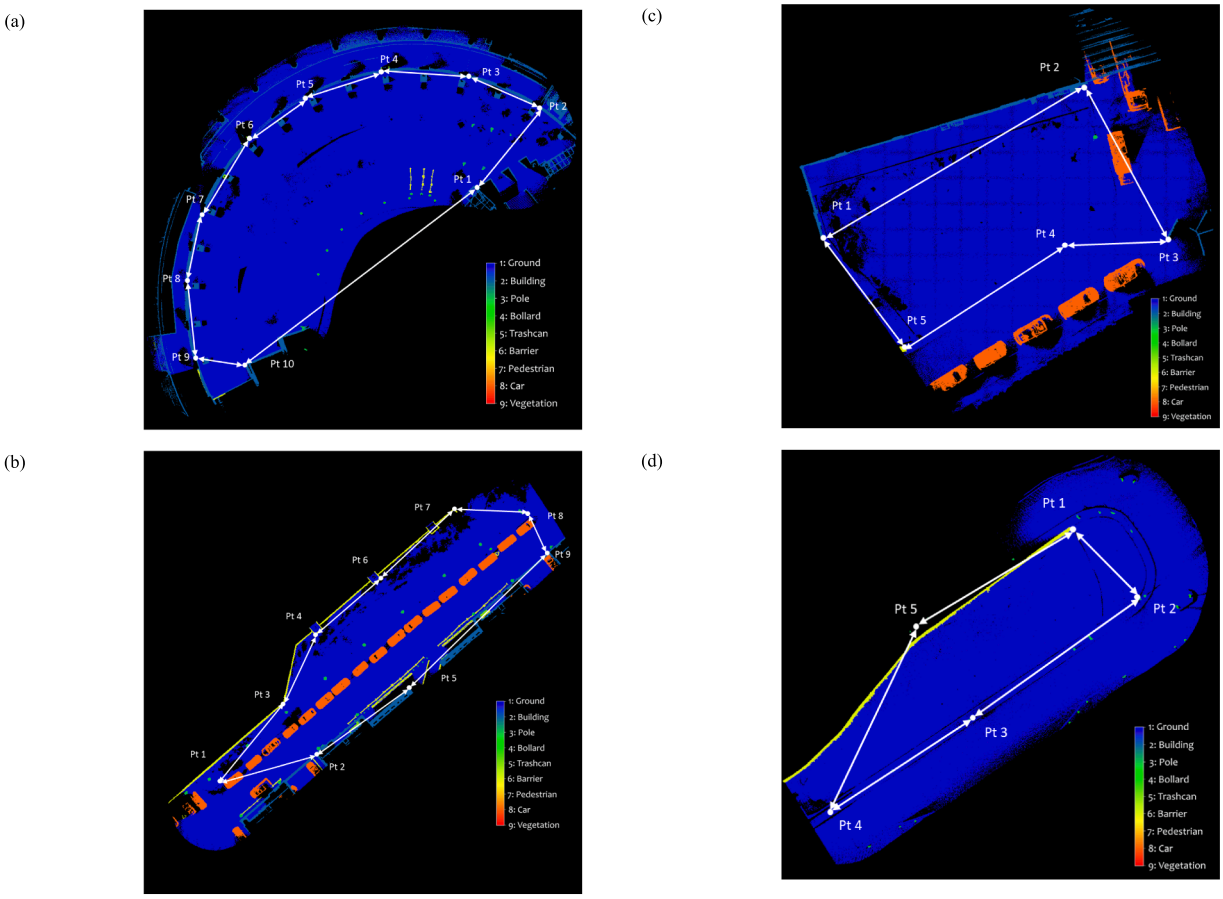


Fig. A3. The locations selected for SLAM validation in the reconstructed 3D point cloud maps of (a) Site 1; (b) Site 2; (c) Site 3; and (d) Site 4 within the HKUST campus. Note that vegetation point clouds are removed in the above figures for better visualization.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2024.109378.

References

Anagnostis, A., Tagarakis, A.C., Asiminari, G., Papageorgiou, E., Kateris, D., Moshou, D., Bochtis, D., 2021. A deep learning approach for anthracnose infected trees classification in walnut orchards. Comput. Electron. Agric. 182, 105998 https://doi.org/10.1016/j.compag.2021.105998.

Balenović, I., Liang, X., Jurjević, L., Hyyppä, J., Seletković, A., Kukko, A., 2021. Hand-held personal laser scanning–current status and perspectives for forest inventory application. Croatian Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering 42 (1), 165–183. https://doi.org/10.5552/crojfe.2021.858.

Bauwens, S., Bartholomeus, H., Calders, K., Lejeune, P., 2016. Forest inventory with terrestrial LiDAR: a comparison of static and hand-held mobile laser scanning. Forests 7 (6), 127. https://doi.org/10.3390/f7060127.

Cabo, C., Ordóñez, C., López-Sánchez, C.A., Armesto, J., 2018. Automatic dendrometry: tree detection, tree height and diameter estimation using terrestrial laser scanning. Int. J. Appl. Earth Obs. Geoinf. 69, 164–174. https://doi.org/10.1016/j.jag.2018.01.011.

Campello, R. J., Moulavi, D., Sander, J. (2013). Density-based clustering based on hierarchical density estimates. In *Advances in Knowledge Discovery and Data Mining: 17th Pacific-Asia Conference, PAKDD 2013, Gold Coast, Australia, April 14-17, 2013, Proceedings, Part II 17* (pp. 160-172). Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-37456-2_14.

Chau, W.Y., Wang, Y.H., Chiu, S.W., Tan, P.S., Leung, M.L., Lui, H.L., Wu, J., Lau, Y.M., 2023. AI-IoT integrated framework for tree tilt monitoring: a case study on tree failure in Hong Kong. Agric. For. Meteorol. 341, 109678 https://doi.org/10.1016/j.agrformet.2023.109678.

Chen, X., Li, S., Mersch, B., Wiesmann, L., Gall, J., Behley, J., Stachniss, C., 2021a. Moving object segmentation in 3D LiDAR data: a learning-based approach exploiting sequential data. IEEE Rob. Autom. Lett. 6 (4), 6529–6536. https://doi.org/10.1109/LRA.2021.3093567.

Chen, Y., Wang, S., Li, J., Ma, L., Wu, R., Luo, Z., Wang, C., 2019. Rapid urban roadside tree inventory using a mobile laser scanning system. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12 (9), 3690–3700. https://doi.org/10.1109/JSTARS.2019.2929546.

Chen, Y., Wu, R., Yang, C., Lin, Y., 2021b. Urban vegetation segmentation using terrestrial LiDAR point clouds based on point non-local means network. Int. J. Appl. Earth Obs. Geoinf. 105, 102580 https://doi.org/10.1016/j.jag.2021.102580.

Chen, J., Zhao, D., Zheng, Z., Xu, C., Pang, Y., Zeng, Y., 2024. A clustering-based automatic registration of UAV and terrestrial LiDAR forest point clouds. Comput. Electron. Agric. 217, 108648 https://doi.org/10.1016/j.compag.2024.108648.

Chiappini, S., Pierdicca, R., Malandra, F., Tonelli, E., Malinverni, E.S., Urbinati, C., Vitali, A., 2022. Comparing Mobile Laser Scanner and manual measurements for dendrometric variables estimation in a black pine (*Pinus nigra* Arn.) plantation. Comput. Electron. Agric. 198, 107069 https://doi.org/10.1016/j.compag.2022.107069.

Choi, K., Lim, W., Chang, B., Jeong, J., Kim, I., Park, C.R., Ko, D.W., 2022. An automatic approach for tree species detection and profile estimation of urban street trees using deep learning and Google street view images. ISPRS J. Photogramm. Remote Sens. 190, 165–180. https://doi.org/10.1016/j.isprsjprs.2022.06.004.

Chow, J.K., Liu, K.F., Tan, P.S., Su, Z., Wu, J., Li, Z., Wang, Y.H., 2021. Automated defect inspection of concrete structures. Autom. Constr. 132, 103959 https://doi.org/10.1016/j.autcon.2021.103959.

Cortinhal, T., Tzelepis, G., & Erdal Aksoy, E. (2020). Salsanext: Fast, uncertainty-aware semantic segmentation of lidar point clouds. In *Advances in Visual Computing: 15th International Symposium, ISVC 2020, San Diego, CA, USA, October 5–7, 2020, Proceedings, Part II 15 (pp. 207-222)*. Springer International Publishing. DOI: 10.1007/978-3-030-64559-5_16.

Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. Numer. Math. 1, 269–271.

Fan, W., Yang, B., Liang, F., Dong, Z., 2020. Using mobile laser scanning point clouds to extract urban roadside trees for ecological benefits estimation. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 43 (211–216), 2020. https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-211-2020.

Fernández-Sarría, A., Velázquez-Martí, B., Sajdak, M., Martinez, L., Estornell, J., 2013. Residual biomass calculation from individual tree architecture using terrestrial laser scanner and ground-level measurements. Comput. Electron. Agric. 93, 90–97. https://doi.org/10.1016/j.compag.2013.01.012.

Forsman, M., Börlin, N., Olofsson, K., Reese, H., Holmgren, J., 2018. Bias of cylinder diameter estimation from ground-based laser scanners with different beam widths: a simulation study. ISPRS J. Photogramm. Remote Sens. 135, 84–92. https://doi.org/10.1016/j.isprsjprs.2017.11.013.

Hacinecipoglu, A., Konukseven, E.I., Koku, A.B., 2020. Pose invariant people detection in point clouds for mobile robots. International Journal of Mechanical Engineering and Robotics Research 9 (5), 709–715. https://doi.org/10.18178/ijmerr.9.5.709-715.

Hough, P. V. (1962). *U.S. Patent No. 3,069,654*. Washington, DC: U.S. Patent and Trademark Office.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., Markham, A., 2020. Randla-net: efficient semantic segmentation of large-scale point clouds. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 11108–11117. https://doi.org/10.48550/arXiv.1911.11236.

Huang, H., Li, Z., Gong, P., Cheng, X., Clinton, N., Cao, C., Ni, W., Wang, L., 2011. Automated methods for measuring DBH and tree heights with a commercial scanning lidar. Photogramm. Eng. Remote Sens. 77 (3), 219–227. https://doi.org/10.14358/PERS.77.3.219.

Koreň, M., Hunčaga, M., Chudá, J., Mokroš, M., Surový, P., 2020. The influence of cross-section thickness on diameter at breast height estimation from point cloud. ISPRS Int. J. Geo Inf. 9 (9), 495. https://doi.org/10.3390/ijgi9090495.

Kwong, I.H.Y., 2022. Physical environment, species choice and spatio-temporal patterns of urban roadside trees in Hong Kong. Trees, Forests and People 10, 100358. https://doi.org/10.1016/j.tfp.2022.100358.

Lei, J., Li, H., Zhao, S., Wang, Y., Jiang, Y., Zhu, G., 2022. Automatic identification of street trees with improved RandLA-net and accurate calculation of shading area with density-based iterative α-shape. IEEE Access 10, 132384–132395. https://doi.org/10.1109/ACCESS.2022.3229901.

Li, J., Cheng, X., Wu, Z., Guo, W., 2021. An over-segmentation-based uphill clustering method for individual trees extraction in urban street areas from MLS data. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 2206–2221. https://doi.org/10.1109/JSTARS.2021.3051653.

Liu, H., Li, C., Wu, Q., Lee, Y.J., 2024. Visual instruction tuning. Adv. Neural Inf. Proces. Syst. 36 https://doi.org/10.48550/arXiv.2304.08485.

Liu, J., Wang, X., Wang, T., 2019. Classification of tree species and stock volume estimation in ground forest images using deep learning. Comput. Electron. Agric. 166, 105012 https://doi.org/10.1016/j.compag.2019.105012.

Luo, H., Khoshelham, K., Chen, C., He, H., 2021. Individual tree extraction from urban mobile laser scanning point clouds using deep pointwise direction embedding. ISPRS J. Photogramm. Remote Sens. 175, 326–339. https://doi.org/10.1016/j.isprsjprs.2021.03.002.

Nielsen, A.B., Östberg, J., Delshammar, T., 2014. Review of urban tree inventory methods used to collect data at single-tree level. Arboricult. Urban For. 40 (2), 96–111. https://doi.org/10.48044/jauf.2014.011.

Ning, X., Ma, Y., Hou, Y., Lv, Z., Jin, H., Wang, Z., Wang, Y., 2023. Trunk-constrained and tree structure analysis method for individual tree extraction from scanned outdoor scenes. Remote Sens. (Basel) 15 (6), 1567. https://doi.org/10.3390/rs15061567.

Pataki, D.E., Alberti, M., Cadenasso, M.L., Felson, A.J., McDonnell, M.J., Pincetl, S., Pouyat, R.V., Setälä, H.M., Whitlow, T.H., 2021. The benefits and limits of urban tree planting for environmental and human health. Front. Ecol. Evol. 9, 603757 https://doi.org/10.3389/fevo.2021.603757.

Proudman, A., Ramezani, M., & Fallon, M. (2021, August). Online estimation of diameter at breast height (DBH) of forest trees using a handheld LiDAR. In *2021 European Conference on Mobile Robots (ECMR)* (pp. 1-7). IEEE. DOI: 10.1109/ECMR50962.2021.9568814.

Roynard, X., Deschaud, J.E., Goulette, F., 2018. Paris-Lille-3D: A large and high-quality ground-truth urban point cloud dataset for automatic segmentation and classification. The International Journal of Robotics Research 37 (6), 545–557. https://doi.org/10.1177/0278364918767506.

Singh, P., Verma, A., Alex, J.S.R., 2021. Disease and pest infection detection in coconut tree through deep learning techniques. Comput. Electron. Agric. 182, 105986 https://doi.org/10.1016/j.compag.2021.105986.

Tsuchiya, B., Mochizuki, H., Hoshikawa, T., Suzuki, S., 2023. Error estimation of trunk diameter and tree height measured with a backpack LiDAR system in Japanese plantation forests. Landsc. Ecol. Eng. 19 (1), 169–177. https://doi.org/10.1007/s11355-022-00530-w.

Wang, W., Fan, Y., Li, Y., Li, X., Tang, S., 2023b. An individual tree segmentation method from mobile mapping point clouds based on improved 3-D morphological analysis. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 2777–2790. https://doi.org/10.1109/JSTARS.2023.3243283.

Wang, P., Tang, Y., Liao, Z., Yan, Y., Dai, L., Liu, S., Jiang, T., 2023a. Road-side individual tree segmentation from urban MLS point clouds using metric learning. Remote Sens. (Basel) 15 (8), 1992. https://doi.org/10.3390/rs15081992.

Weinmann, M., Weinmann, M., Mallet, C., Brédif, M., 2017. A classification-segmentation framework for the detection of individual trees in dense MMS point cloud data acquired in urban areas. Remote Sens. (Basel) 9 (3), 277. https://doi.org/10.3390/rs9030277.

Xia, K., Li, C., Yang, Y., Deng, S., Feng, H., 2023. Study on single-tree extraction method for complex RGB point cloud scenes. Remote Sens. (Basel) 15 (10), 2644. https://doi.org/10.3390/rs15102644.

Xiao, X., Zhao, Y., Zhang, F., Luo, B., Yu, L., Chen, B., Yang, C., 2023. BASeg: Boundary aware semantic segmentation for autonomous driving. Neural Netw. 157, 460–470. https://doi.org/10.1016/j.neunet.2022.10.034.

Xu, W., Cai, Y., He, D., Lin, J., Zhang, F., 2022. Fast-lio2: fast direct lidar-inertial odometry. IEEE Trans. Rob. 38 (4), 2053–2073. https://doi.org/10.1109/TRO.2022.3141876.

Yao, W., Fan, H. (2013, May). Automated detection of 3D individual trees along urban road corridors by mobile laser scanning systems. In *Proceedings of the International Symposium on Mobile Mapping Technology, Tainan, Taiwan* (Vol. 6).

Ye, X., Pan, J., Shao, F., Liu, G., Lin, J., Xu, D., Liu, J., 2024. Exploring the potential of visual tracking and counting for trees infected with pine wilt disease based on improved YOLOv5 and StrongSORT algorithm. Comput. Electron. Agric. 218, 108671 https://doi.org/10.1016/j.compag.2024.108671.

Yoon, T.K., Park, C.W., Lee, S.J., Ko, S., Kim, K.N., Son, Y., Lee, K.H., Oh, S., Lee, W.K., Son, Y., 2013. Allometric equations for estimating the aboveground volume of five common urban street tree species in Daegu. Korea. *Urban Forestry & Urban Greening* 12 (3), 344–349. https://doi.org/10.1016/j.ufug.2013.03.006.

Zeybek, M., Vatandaşlar, C., 2021. An automated approach for extracting forest inventory data from individual trees using a handheld mobile laser scanner. Croatian

Journal of Forest Engineering: Journal for Theory and Application of Forestry Engineering 42 (3), 515–528. https://doi.org/10.5552/crojfe.2021.1096.
Zhang, C., Zhou, Y., Qiu, F., 2015. Individual tree segmentation from LiDAR point clouds for urban forest inventory. Remote Sens. (Basel) 7 (6), 7892–7913. https://doi.org/10.3390/rs70607892.
Zhou, Q. Y., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. DOI: 10.48550/arXiv.1801.09847.