



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Master's Thesis of Landscape Architecture

Mapping Carbon Stock of Individual
Street Trees Using LIDAR-Camera
Fusion-based Mobile Mapping
System

라이다-카메라 퓨전 기반 모바일 맵핑 시스템을
활용한 개별 가로수의 탄소 저장량 맵핑

August 2023

Graduate School of
Seoul National University

Department of Landscape Architecture and Rural
Systems Engineering, Landscape Architecture Major

Yang, Tackang

Mapping Carbon Stock of Individual Street Trees Using LIDAR-Camera Fusion-based Mobile Mapping System

지도 교수 류 영 렬

이 논문을 생태조경학 석사 학위논문으로 제출함
2023 년 8 월

서울대학교 대학원

생태조경지역시스템공학부 생태조경학전공

양 태 강

양태강의 생태조경학 석사 학위논문을 인준함
2023 년 8 월

위 원 장 _____ 이 동 근 _____ (인)

부위원장 _____ 류 영 렬 _____ (인)

위 원 _____ 송 영 근 _____ (인)

Abstract

Urban street trees constitute a considerable proportion of urban trees, yet their individual carbon stocks remain under-explored. Here, we map the carbon stock of individual street trees using LiDAR-camera fusion-based Mobile Mapping System (MMS) which enables extensive urban coverage, high spatial sampling, and concurrent acquisition of species and structural parameters. We implement a two-step approach to detect individual street tree, initially applying a U-Net to images for semantic segmentation, and subsequently applying a Random Forest classifier on point clouds, informed by the image segmentation results. To measure the carbon stock of street trees, we employed another Yolo-v3 to classify tree species from images and calculated Diameter at Breast Height (DBH) and height (H) from point clouds. Through experiment, we detected 35,247 street trees from scanned streets in the Suwon, Republic of Korea, which had the carbon stock of 2.16 ± 0.03 GgC. City-wide evaluations showed the average recall, precision, and F1-score of the proposed street tree extraction method were 78.89, 85.65, and 81.79, respectively. In addition, estimated DBH and H revealed slight overestimation by an average of 4.37 cm (15.13%) and 0.86 m (8.57%) with RMSEs of 8.17 cm (28.27%) and 2.18 m (21.82%), respectively. Our work contributes a practical framework for estimating individual street trees' carbon stocks using the LiDAR-camera fusion-based MMS, paving the way towards more accurate urban carbon management and progress in urban carbon management.

Keyword : Urban street trees, Carbon stocks, LiDAR-camera fusion, Mobile Mapping system

Student Number : 2021-27046

Table of Contents

Chapter 1. Introduction	1
Chapter 2. Materials and Method.....	5
2.1. Study site and data acquisition.....	5
2.2. Sensor Alignment.....	7
2.3. Preprocessing	9
2.4. Individual street tree extraction.....	11
2.5. Carbon stock estimation	17
Chapter 3. Results.....	21
Chapter 4. Discussion	27
Chapter 5. Conclusion	34
Bibliography	37
Abstract in Korean.....	40

List of Tables

[Table 1]	19
[Table 2]	20
[Table 3]	22

List of Figures

[Figure 1]	4
[Figure 2]	6
[Figure 3]	10
[Figure 4]	12
[Figure 5]	13
[Figure 6]	15
[Figure 7]	22
[Figure 8]	24
[Figure 9]	26

[Figure 10]	27
[Figure 11]	28
[Figure 12]	29
[Figure 13]	30
[Figure 14]	31
[Figure 15]	32
[Figure 16]	33

Chapter 1. Introduction

The rise in anthropogenic carbon emissions in urban regions, coupled with rapid global urbanization, has highlighted the need to comprehend the carbon stored in urban trees (Nowak et al., 2013; Seto et al., 2014). Street trees, which constitute a considerable proportion of urban trees, grow in isolation, unlike other urban trees which are typically clustered together. This distinct spatial distribution necessitates individual-level investigation of street trees, a task that generally accompany with labor-intensive field survey (McPherson, 1998; Nowak et al., 2013; Nowak & Crane, 2002; Timilsina et al., 2014). Advances in remote sensing technologies offer promising alternatives to traditional methods, but each remote sensing technology comes with its own set of challenges and limitations. Given these developments, the research questions arise: How can we reliably measure the carbon stock of individual street trees without resorting to labor-intensive field survey? Is that method practical to apply at a city-scale?

Accurate quantification of the carbon stock in street trees typically requires both structural parameters such as diameter at breast height (DBH) or height (H), and species information at an individual level and on a city-wide scale. Light Detection and Ranging (LiDAR) is a promising tool to measure trees' structural parameters. However, it has inherent limitations depending on the platform to which it is attached. Spaceborne LiDAR offers extensive coverage but lacks the spatial sampling density needed for individual-level analysis of street trees (Dubayah et al., 2020). In contrast, airborne LiDAR provides better point density but is generally associated with high costs. Terrestrial LiDAR offers sufficiently dense spatial sampling to identify the individual tree with species information, but it lacks mobility (Zou et al., 2017). Thus, the motivation behind this paper stems from the pursuit of a balanced method that can provide a dense enough point cloud to capture individual street trees while also offering a broad scope of coverage encompassing extensive urban areas.

Recently, the Mobile Mapping System (MMS) has emerged as a powerful means for surveying street environments (Pu et al., 2011; W. Xiao et al., 2015). MMS captures two- or three-dimensional geometric information of surrounding environments using LiDAR and/or camera mounted on a vehicle (Puentes et al., 2013). When MMS employs LiDAR as a mapping sensor for surveying street trees, it accurately captures structural parameters but falls short in providing species information (Safaie et al., 2021; Zhao et al., 2018). A few studies classified tree species using LiDAR-based MMS and recorded lower performances on classifying species of trees with similar geometric shapes than trees with distinct geometries (Chen et al., 2019; Guan et al., 2015). Classifying street tree species is more challenging when frequent management activities alter tree shapes, obscuring intra- and inter-species differences of geometric shapes. With camera-based MMS, deep learning (DL) has been employed in the classification of street tree species, but estimating structural parameters from images is less reliable than LiDAR-based MMS (Choi et al., 2022). These recent studies inspire applying LiDAR-camera fusion technique, which compensates each sensor's shortcomings, to enhance the accuracy of carbon stock estimation. Though DL-based methods have significantly improved the performance of the LiDAR-camera fusion technique (Cui et al., 2022; Fu et al., 2018; L. Xiao et al., 2018), there remains a scarcity of practical studies applying this fusion approach to map the carbon stock of street trees at a city scale.

The key challenge in application of DL-based LiDAR-camera fusion technique for map the carbon stock of street trees lies in generating a point cloud training dataset. Detecting individual street trees in the point cloud is essential, as their structural parameters are vital for carbon stock estimation. Conventional methods for extracting individual trees from point clouds remain indispensable (Ning et al., 2019; Zhong et al., 2017), but recent DL-based methods have exhibited remarkable performance in segmenting individual trees from point clouds (Jiang et al., 2023; Luo et al., 2021). The practical use of DL models necessitates the generation of a training

dataset tailored to the target site, leading to an increase in the volume of required training data across diverse urban scenarios on a city scale. Given that annotating point clouds is considerably more demanding than annotating images (Xie et al., 2020), finding a method that alleviates these difficulties while maintaining the performance of DL model is imperative.

The goal of this study is generating a carbon stock map of individual street trees at a city-wide scale using LiDAR-camera fusion-based MMS. The overview of our result is visualized in Figure 1. To achieve this goal, we detected individual street trees and estimated the carbon stock of each tree. In individual street tree detection, we applied DL exclusively to images to minimize the effort required to construct point cloud training data while maintaining the performance of DL. To test the robustness of proposed method, we used the dataset collected at a city-wide scale in Suwon, Korea.. The scientific questions we address include: (1) How many street trees are there in Suwon? (2) How much carbon are they storing?

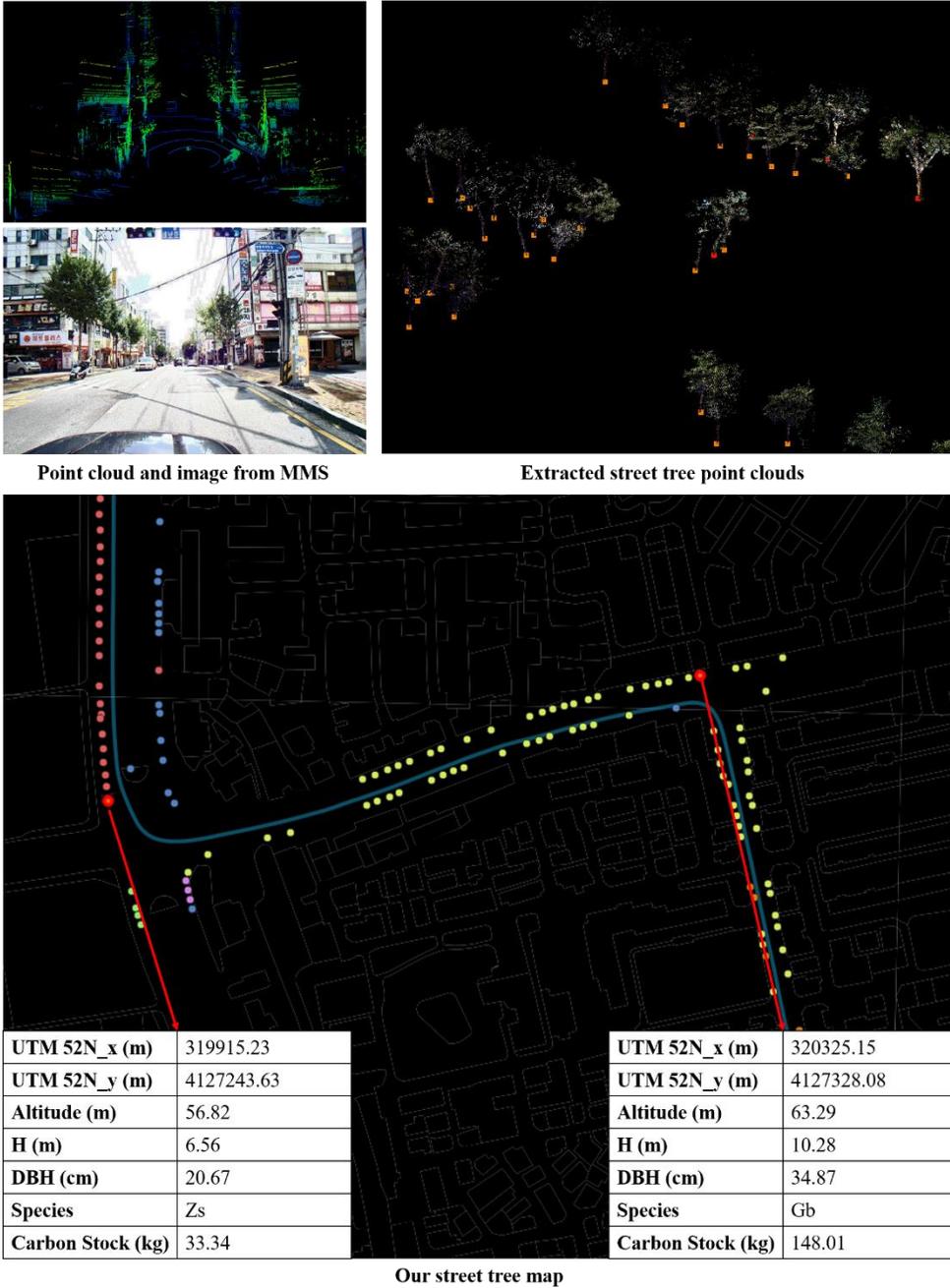


Figure 1. We processed data on individual street trees from point cloud and image datasets, collected using a MMS. Our final results are depicted in a comprehensive map that delineates the location, species, H, DBH, and carbon stock of each street tree. The point clouds of street trees are visualized in RGB color with red dots marking the determined location of each tree. Additionally, the distinct colors used in our street tree map correspond to different

species

Chapter 2. Materials and Method

2.1. Study site and data acquisition

The study area encompassed the majority of vehicle-accessible roads where street trees are planted in Suwon city, Republic of Korea. The area of Suwon city is 121 km² and the center of city is located in 37° 16 ' 50" N, 127° 00 ' 27" E). In this study, street trees are defined as the trees planted on the pedestrian strips adjacent to the road and the median strips. The majority of street trees in Suwon city comprise 12 species: *Acer buergerianum* (*Ab*), *Acer palmatum* (*Ap*), and *Aesculus turbinata* (*At*), *Chionanthus retusus* (*Cr*), *Ginkgo biloba* (*Gb*), *Metasequoia glyptostroboides* (*Mg*), *Pinus densiflora* (*Pd*), *Platanus occidentalis* (*Po*), *Prunus yedoensis* (*Py*), *Quercus palustris* (*Qp*), *Styphnolobium japonicum* (*Sj*), *Zelkova serrata* (*Zs*).

To test the robustness and practicality of the proposed street tree detection method in various urban scenarios, we collected data from across the entire streets of Suwon city. The trajectories of collected datasets are depicted in Figure 2 (a). Data collection was conducted from August 2021 to October 2021, during daytime hours from 9 am to 5 pm. Then, we annotated 35,395 street trees from the collected datasets to validate the result of street tree detection. Annotated trees are randomly sampled from the entire datasets. There were overlapped area in the datasets and some trees are detected multiple times among datasets.

To test the accuracy of estimated structural parameters, we sampled 287 street trees from randomly chosen streets. We conducted a tape measurement at 1.3 m of the trunk for DBH measurement and employed a terrestrial LiDAR (VZ-400i from RIEGL Laser Measurement Systems GmbH, Horn, Lower Austria, Austria) to measure the H. These data was collected from May 2022

to April 2023 during the day time.

We used an off-the-shelf product (Mobiltech, Seoul, Republic of Korea) to collect point cloud and image data as shown in Figure 2(b). The mapping sensors comprised a LiDAR sensor (VLP-32C from Velodyne, San Jose, California, USA), an RGB camera (FLIR Blackfly from Teledyne FLIR, Wilsonville, Oregon, USA), a longwave infrared (LWIR) camera (FLIR A65 from Teledyne FLIR, Wilsonville, Oregon, USA) and the positioning sensors (APX-15 UAV from Trimble Applanix, Richmond Hill, Ontario, Canada) that included a Global Navigation Satellite System (GNSS) receiver and an inertial measurement unit (IMU). It is worth noting that cameras were oriented in the front direction of the vehicle. The LiDAR sensor scans 360° in 0.1 seconds, operating at a frequency of 10 Hz. The RGB camera operated at frequency of 10 Hz. The vehicle's six degrees-of-freedom (DOF) was obtained from the trajectory, processed using the commercial software (POSPac-UAV 8.4 from Trimble Applanix, Richmond Hill, Ontario, Canada) with data from the positioning sensors.

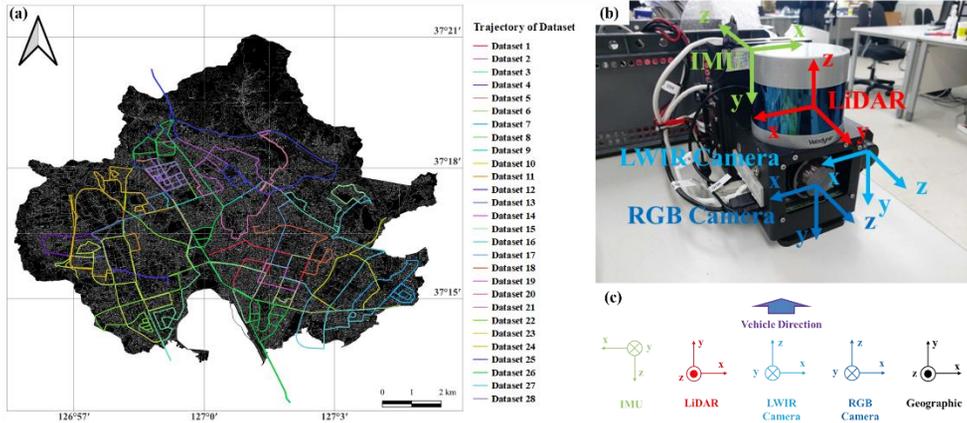


Figure 2. Overview of materials related to the data acquisition (a) Overview of data collection trajectory. We drove about 324 km to collect datasets reflecting various urban scenarios. (b) The product of MMS used in this study. (c) Three-dimensional local coordinates of each sensor

2.2. Sensor Alignment

We matched the spatiotemporal coordinate of LiDAR and camera to transfer information from images to the point cloud. Since MMS records the exact data collection time from different sensors, matching temporal coordinates of these sensors is achievable. First, we matched the temporal coordinates of LiDAR and camera, then the spatial coordinates. Temporal matching must be precedent to spatial matching because each sensor’s spatial coordinate changes overtime as the vehicle moves. The local coordinate of each sensor is described in Figure 2 (c).

To match the temporal coordinate of LiDAR and camera, we transformed the LiDAR’s local coordinate to that at the time when images are collected. For instance, transforming the local coordinate of the MMS point cloud collected at time frame t_n to the LiDAR’s local coordinate at t_{n-1} is required when the point cloud is collected at t_n and the image is collected at t_{n-1} . This transformation necessitates the relative position between the LiDAR sensor origin at t_n and t_{n-1} . The relative position can be obtained from transforming the both point cloud collected at different time to the geographic coordinate, which exploits the six DOF information of vehicle position measured by positioning sensors. Since the vehicle position is recorded with respect to the local coordinates of the IMU, the point cloud is first transformed to the local coordinate of the IMU using Equation (1). Inversely, Equation (2) shows transformation from IMU local coordinate to LIDAR coordinate using the same rotation and translation matrix. Equation (3) demonstrates transforming the point cloud in IMU local coordinate at t_n to the geographic coordinate and transforming it again to IMU local coordinate at t_{n-1} .

$$P_I = \begin{pmatrix} R_{LI} & T_{LI} \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot P_L \quad (1)$$

$$P_L = \begin{pmatrix} R_{LI}^T & -R_{LI}^T \cdot T_{LI} \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot P_I \quad (2)$$

$$P_{It_{n-1}} = \begin{pmatrix} R_{IGt_{n-1}}^T & -R_{IGt_{n-1}}^T \cdot T_{IGt_{n-1}} \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} R_{IGt_n} & T_{IGt_n} \\ 0 & 0 & 0 & 1 \end{pmatrix} \cdot P_{It_n} \quad (3)$$

Here, P, R, and T are point cloud of 4 x 1 matrix with the dummy value of 1 at the last row, 3 x 3 rotation matrix, and 3 x 1 translation matrix, respectively. For subscripts, L, I, C, and G denote LiDAR, IMU, camera local coordinate, and geographic coordinate, respectively, and t is used when a specific time frame is relevant. In P, subscripts refer the coordinate where the point cloud is located in. In R and T, two sensor subscripts are used together to indicate the transformation from one coordinate to another. For example, R_{IGt_n} is rotation matrix from IMU local coordinate to geographic coordinate at t_n . Time notation is not included in Equation (1) and (2), as the transformation matrix between sensors' local coordinates remains constant regardless of time. In summary, transforming the point cloud's local coordinate at time t_n to t_{n-1} can be achieved in four steps: P_{Lt_n} to P_{It_n} , P_{It_n} to P_G , P_G to $P_{It_{n-1}}$, and $P_{It_{n-1}}$ to $P_{Lt_{n-1}}$.

Once the temporal alignment between LiDAR and camera is completed, we conducted the spatial alignment and projected point cloud to the images to transfer information from images to point clouds. We transformed the LiDAR's spatial coordinate to the camera's one using the Equation (1) with R_{LC} and T_{LC} , which are rotation matrix and translation matrix calculated from extrinsic parameters between LiDAR and camera. In this study, the extrinsic parameters between different sensors are provided by the MMS manufacturer. After the transformation, the projection of point clouds

in the camera's local coordinate onto the image plane is feasible by using the camera's geometry and intrinsic parameters. Equation (4) represents the projection of point clouds in the camera's local coordinate to the normalized image plane. Equation (5) is calculation of radial distance, which is used in Equation (6) to correct the distortion of image. Equation (7) projects from the normalized image plane to the pixel plane. Point cloud retrieves two-dimensional information of the pixels where the points are projected. At this stage, color and temperature information are extracted from the RGB image and LWIR image, respectively, to the point cloud.

$$\begin{pmatrix} X_{Nu} \\ Y_{Nv} \end{pmatrix} = \begin{pmatrix} X_c/Z_c \\ Y_c/Z_c \end{pmatrix} \quad (4)$$

$$r_u^2 = X_{Nu}^2 + Y_{Nv}^2 \quad (5)$$

$$\begin{pmatrix} X_{Du} \\ Y_{Dv} \end{pmatrix} = (1 + k_1 r_u^2 + k_2 r_u^4) \begin{pmatrix} X_{Nu} \\ Y_{Nv} \end{pmatrix} + \begin{pmatrix} 2p_1 X_{Nu} Y_{Nv} + p_2 (r_u^2 + 2X_{Nu}^2) \\ p_1 (r_u^2 + 2Y_{Nv}^2) + 2p_2 X_{Nu} Y_{Nv} \end{pmatrix} \quad (6)$$

$$\begin{pmatrix} X_u \\ X_v \\ 1 \end{pmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} X_{Du} \\ Y_{Dv} \\ 1 \end{pmatrix} \quad (7)$$

Here, f_x , f_y , c_x , and c_y are camera's focal length and optical center. k_1 , k_2 , p_1 , and p_2 are radial and tangential distortion coefficients. X_{Nu} and Y_{Nv} are pixel coordinates on the normalized image plane, X_{Du} and Y_{Dv} are distortion-corrected pixel coordinate on the normalized image plane, and X_u and Y_v are pixel coordinate on the pixel plane.

2.3. Preprocessing

We removed the ground points from the point clouds to reduce data volume, which consequently decreases the computational

resources needed for processing point clouds in subsequent stages of the method. During the preprocessing phase, point clouds are divided into horizontal tiles before ground removal takes place. We assigned each tile a height value which is calculated by subtracting the minimum z -value from the maximum z -value of points within each tile. Tiles made entirely of ground points, referred to as ground tiles, have low height values and can therefore be discarded using a threshold value. This threshold is set at 0.25 m, a value derived considering the curb height at the study site.

To remove ground points in non-ground tiles, which we call carpet-like ground points, we calculated the average of ground tiles' height values and filtered out the points in non-ground tiles which have z -values within the threshold of this average height. This threshold is also set at 0.25 m, split into an upper and lower bound of 0.125 m each. This split accounts for the height variance of ground tiles due to the differential height of pedestrian strips and roads. The tile size is set at 0.5 m x 0.5 m, which is empirically decided with respect to the point cloud resolution. The smaller tile size results in better ground removal, but increase in processing time. The entire ground removal process is depicted in Figure 3.

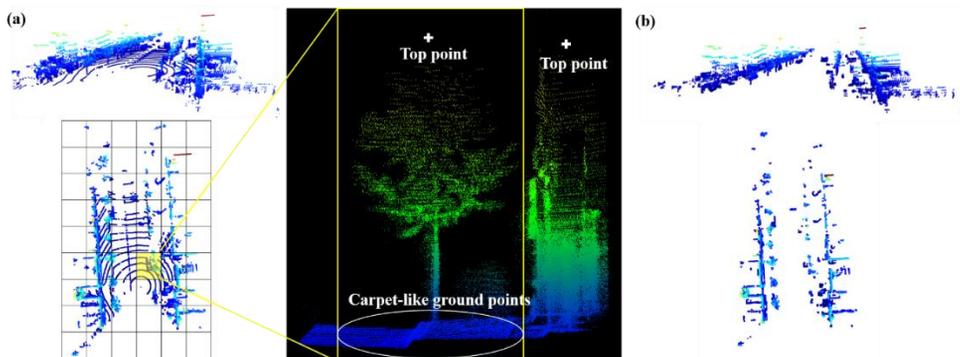


Figure 3. Ground removal process from point clouds. (a) Division of

point clouds into tiles and elimination of ground tiles. The subsequent step removes carpet-like ground points from non-ground tiles. (b) Visualization of point clouds after the completion of the ground removal process.

We applied the ground removal process to the single frame of point cloud to mitigate the effect of dynamic topography. The proposed ground removal method is susceptible to the topography with varying elevation since it uses the average height value of ground tiles. For example, the average height value of ground tiles increases when a street has a slope. This results in the removal of non-ground points during the removal step of the carpet-like ground points. Thus, the smaller scanned area is included in point cloud, there is the less possibility of dynamic topography being involved in the point cloud in that urban topography is generally flat. We selected the single frame of point cloud (a 360 ° scan in 0.1 seconds) as the processing unit for ground removal regarding that all the subsequent processes are based on single frame.

In addition, we created the digital surface model (DSM) and digital elevation model (DEM) to calculate the H of individual street trees. When two dimensional tiles are generated during ground removal process, we use the maximum z-values of points within the tiles to create DSM and the minimum z-values of to create DEM. We employ DSM and DEM to overcome the constraint in H calculation, which we discuss in detail in Section 2.5.

2.4. Individual street tree detection

We propose a two-step approach to detect individual street trees that includes: (1) semantic segmentation on images using DL,

and (2) instance segmentation on point clouds using clustering and cluster refinement. We developed a filter using pseudo-plane and applied machine learning (ML) classifier to refine the clusters. The overview of framework appears in Figure 4. The motivation behind this bifurcation is to decrease the need for extensive point cloud training data for DL models and prioritize information from images which is denser than the point cloud. Images typically produce a higher resolution representation of trees, whereas point clouds generated by MMS often have a sparser sampling density.

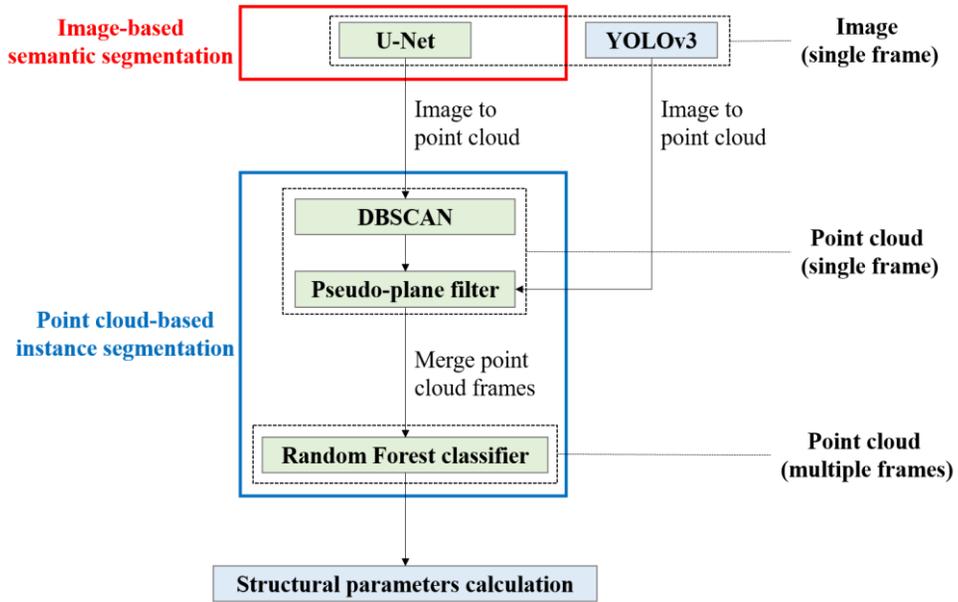


Figure 4. The flowchart of the proposed framework. Green boxes represent the flow of individual street tree detection, and blue boxes represent species detection and structural parameters calculation

To achieve semantic segmentation of street trees, we employ the U-Net convolutional neural network (CNN), known for its proficiency in segmenting biomedical images (Weng & Zhu, 2021). This CNN performs pixel-level binary classification, discerning tree

and non-tree areas in images. Subsequently, point clouds are projected onto these images according to the methodology outlined in Section 2.2, yielding the semantic segmentation result depicted in Figure 5. Extracted point clouds corresponding to the segmented pixels are identified as potential street trees. These candidates may include occluded or background objects when lifting two-dimensional information to a three-dimensional space (Gong et al., 2020). We aim to increase the likelihood of street tree detection by applying the aforementioned task to a single point cloud frame. Given the 10 Hz frequency of LiDAR and image collection, the same street tree is captured multiple times across frames. Therefore, processing single frames allows the U-Net model to detect the same tree multiple times. Despite the trade-off of decreased point cloud density, this issue can be resolved by merging frames later in the process. Images are unaffected by single frame processing due to their constant resolution.

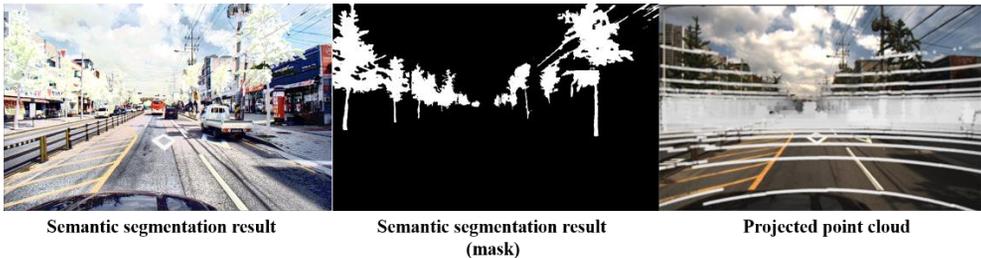


Figure 5. The result of street tree semantic segmentation on collected image and projected point cloud. Point cloud is visualized with the distance from the LiDAR.

Next, we used The Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to group the points belonging to the same object. DBSCAN is a density-based clustering algorithm

capable of identifying clusters of varied shapes and sizes based on the density of points (Ester et al., 1996). DBSCAN parameters, including an epsilon of 1.9 and a minimum point count of 10, should be empirically decided given that each LiDAR point cloud has a different point density. Due to the imperfectness of DBSCAN, the resulting clusters are categorized into single street tree cluster, multiple street tree cluster, or non-tree cluster, and each cluster requiring subsequent refinement procedures.

To decide whether the cluster contains multiple street trees, we developed a pseudo-plane filter which identifies a trunk of each tree. First, each cluster is sectioned into 0.3 m x 0.3 m horizontal tiles, as mentioned in ground removal in Section 2.3, and we assigned each tile the value of minimum z-value of points in the tile. Only the points having z-values less than 2 m is used to calculate the tile value to avoid bottom of crown engaging to the identification of trunk. Then, we made the pseudo-plane using the values of tiles and rasterized the pseudo-plane. The purpose of generating the pseudo-plane lies in smoothing the gradient of tile values. A 3 x 3 window search is conducted to identify pixels with minimum value which are centrally located within the window – these pixels are designated as trunk tiles. Direct application of window search to horizontal tiles without making the pseudo-plane often leads to the detection of false trunk tiles by extra local minima. . Furthermore, the tile size must be chosen large enough to avoid a single tree having multiple trunk tiles. Finally, the point with the minimum z-value in a trunk tile is deemed the center of a cluster. The clusters having multiple trunk tiles require the refinement step to break down clusters to contain only single street tree.

To separate multiple street tree clusters into single tree

clusters, we apply K-means clustering with the k-value equal to the number of identified trunk tiles. Since the centers have already been determined, the K-means clustering process can be conducted without iteration, by simply calculating the distance between points and centers, and grouping points closer to the same center. Clusters that do not include any trunk tiles are filtered out at this stage. The refined clusters consist of individual street tree clusters and non-tree clusters that include false trunk tiles. Overall process of refining the clusters is illustrated in Figure 6.

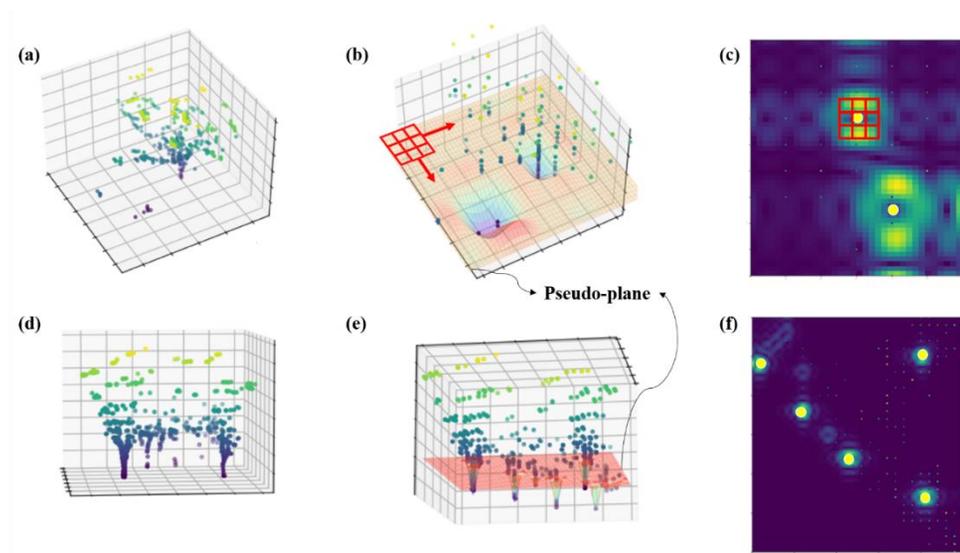


Figure 6. Process of dividing clusters with multiple objects into clusters with single object. (a) (b) (c) Example of processing cluster containing a single tree and noise. (d) (e) (f) Example of processing cluster containing five trees. (a) (d) Overview of given clusters. (b) (e) Overview of pseudo-plane. The plane is convex downward at the position where the trunk or noise are located.. (c) (f) Detected individual trees and noise within the clusters on rasterized pseudo-plane. Yellow circles represent the centers of detected objects. Example of window is depicted with red color in (b) and (c).

To increase point density for accurate calculation of structural parameters, we merged the point clouds across multiple frames. This is achieved by transforming the first frame of the point cloud into geographic coordinates (as explained in Section 2.2), and then sequentially transforming subsequent frames. Each time a frame is transformed into geographic coordinates; the center of each cluster is matched with the centers of neighboring clusters from previously transformed frames, considering a neighbor search threshold of 0.6 m in light of the typical street tree's DBH. During the merging stage, the threshold often results in the unintended merging of apart objects. To separate them again, the filtering step outlined in Figure 6 is reapplied. We also apply DBSCAN with a smaller epsilon parameter of 1 to remove noise. If multiple clusters are detected, we only retain the cluster with the most points, assuming that the rest are noise clusters. These noise clusters are filtered out at the later stage because the point cloud of a single frame does not always have sufficient point density to define the cluster as noise. At this stage, we assume that each cluster contains only a single object and that these clusters consist of either trees or non-tree objects.

Lastly, we apply a Random Forest (RF) classifier (Breiman, 2001) to the merged multiple point cloud frames to filter out non-tree clusters. The primary challenge with learning-based approaches for point clouds is the need for detailed annotations that require per-point labels. Our method circumvents this issue by generating training data for the RF classifier from the street tree point cloud candidates identified in the preceding steps. This greatly simplifies the training data creation process, requiring only the classification of non-tree clusters from the candidates, thus

eliminating the need for per-point labeling. The features used for the RF classifier include general statistical features like the median, mean, and standard deviation of RGB, thermal, and intensity values of clusters, linearity, planarity, and sphericity derived from the principal components of clusters, the angle between the first principal component and the z-axis, height and related statistical features, and the number of points. The selection of general statistical features avoids the need for traditional feature design processes

2.5. Carbon stock estimation

To estimate the carbon stock of each street tree non-destructively, we obtained species information and structural parameters of street trees and applied the allometric equations. To classify street tree species, we applied DL model to replace expert's role in tree species classification. For the estimation of H, we employed DSM and DEM to overcome the limitation caused by sensors' FOV. In addition, we developed the H-based allometric equation to set the boundary for estimated DBH which often reveals high uncertainty because only the half of trunk is scanned with MMS. Lastly, we estimated volume (V) or dry biomass of tree from DBH using allometric equations that are developed for street trees species in Korea. Detailed explanation of each step is followed in this section.

We utilized the You Only Look Once v3 (YOLOv3) system (Redmon & Farhadi, 2018), a CNN-based object detection system, to classify the species of street trees. The twelve street tree species are targeted (see Section 2.1). We applied the same model trained from the previous study (Kwon et al., 2023), which includes thirteen labels, one of which is assigned to species not included among the identified twelve. The species detection results are then projected

onto the point cloud, similar to the projection of semantic segmentation result described in Section 2.4. . Majority voting takes place for each cluster's species information when single frames merge into multiple frames. In this way, the credibility of species information increases because the same tree is classified multiple times throughout the frames.

To calculate tree H, we used the difference between DSM and DEM generated in Section 2.2. When LiDAR's vertical field-of-view (FOV) is restricted, the emitted laser can reach the tree top only if the tree is sufficiently distant from the vehicle. However, our method tends to detect trees when they are closer because of the pixel resolution of images, which often results in losing the treetop points. Hence, we calculate the height by subtracting the DEM from the DSM at the location of the street trees detected in Section 2.4, rather than calculating it within the extracted tree clusters.

To estimate DBH, we used a circle-fitting method and developed allometric equation between H and DBH to filter outliers from the fitted circles. Accurate DBH estimation is challenging as MMS typically scans only the single side of street trees, thus fitted circles often significantly deviates from the truth value. One strategy to increase the accuracy of DBH estimation is fitting circles to multiple transections of the trunk point cloud. The previous study showed quantifying a tree's DBH based on the average diameter of the multi-height diameters can increase the accuracy of DBH estimation (Liu et al., 2021). We generated transections at 2 cm intervals from 20 cm above the starting point of each tree trunk to 150 cm of each tree trunk. Furthermore, we set the 95 % confidence interval (CI) of estimated DBH using the developed allometric equation in Table 1, which is developed in this study using TLS

measured H and tape measured DBH. We developed general allometric equation because we could not sample the entire species targeted in this study. In case of all estimated diameters from transections fall outside the confidence interval, we derived the tree's DBH from H using allometric equation instead of circle-fitting.

Table 1. The allometric equation having DBH as a dependent variable and H as an independent variable. Generic allometric equation is developed due to the lack of sufficient samples for each species.

H		$\ln(DBH) = a + b \times \ln(H)$					
Range (m)	n	<i>a</i>	95% CI	<i>b</i>	95% CI	R^2	RMSE (%)
5.46 ~ 20.72	34	0.9086	[0.776, 1.0412]	1.0653	[1.010, 1.1203]	0.81	0.15 (4.42)

In the calculation of carbon stock, we used allometric equations that estimating V or dry biomass from DBH developed from street tree samples in Korea to decrease the uncertainty lies in the allometric equation itself. It is because allometric equations vary by location and climate, even within the same species. Therefore, we utilized allometric equations of seven street tree species in Korean cities, which are denoted in Table 2 (Kim & Lee, 2016; Park et al., 2018; Yoon et al., 2013). The estimated V was then multiplied by the wood basic density to calculate each tree's dry biomass. We used allometric equations estimating V because we could not find species-specific allometric equation estimating dry biomass developed for street trees in Korea. For the remaining species which we could not find any allometric equations related to the V or biomass, we calculated the dry biomass directly using the generic allometric

equation developed for street trees (Yoon et al., 2013).

Table 2. Species-specific allometric equations of street trees having V as a dependent variable and DBH as an independent variable. In addition, wood basic density of each species is demonstrated.

Species	DBH range (cm)	$V = a \times DBH^b$				rRMSE	Reference	Wood basic density (kg/m ³)	Reference
		a	b	R ²					
Ab	12.8 ~ 41.0	0.0000709	2.511	0.97	8.50	(Yoon et al., 2013)	620	ICARAF database	
Cr	5.4 ~ 29.9	0.000108	2.199	0.93	2.52	(Park et al., 2018)	705*	(Chave et al., 2009)	
Gb	10.5 ~ 34.5	0.0000453	2.656	0.99	2.70	(Yoon et al., 2013)	523	(Kang et al., 2011)	
Mg	12.0 ~ 61.4	0.000527	2.029	0.87	4.76	(Kim & Lee, 2016)	284	ICARAF database	
Po	17.1 ~ 55.0	0.000207	2.158	0.67	35.70	(Park et al., 2018)	448	ICARAF database	
Py	12.3 ~ 66.4	0.000664	1.819	0.97	8.30	(Yoon et al., 2013)	582*	(Chave et al., 2009)	

	48.					2013)	2009)
	2						
	11.					(Yoon	(Kang
Zs	8 ~	0.0000	3.0	0.9	5.80	et al., 751	et al.,
	38.	078	84	6		2013)	2011)
	4						

* Average wood basic density of the same genus substitutes the species specific value

Chapter 3. Result

We performed a quantitative analysis to assess the accuracy of street tree detection. There are three types of street tree detection results: true positive (TP) represents the detected tree located within 2 m boundary of truth tree location; false negative (FN) represents the truth tree which does not have detected tree within 2 m boundary; false positive (FP) represents the detected tree located outside of 2 m boundary of truth tree location. We utilized three evaluation metrics which are recall, precision, and F1-score as shown in Equation (8).

$$\left\{ \begin{array}{l} recall = \frac{TP}{TP + FN} \\ precision = \frac{TP}{TP + FP} \\ F1-score = 2 \times \frac{precision \times recall}{precision + recall} \end{array} \right. \quad (8)$$

Recall and precision represent completeness and correctness of street tree detection result, and F1-score is the harmonic mean of recall and precision.

The results of quantitative analysis are shown in Table 3. The averages for recall, precision, and F1-score were 78.89%, 85.65%, and 81.79% respectively. The highest recorded recall, precision, and F1-score were 93.41%, 93.26%, and 90.24% respectively. As depicted in Figure 7, precision demonstrated better consistency and performance than recall. Although our street tree detection method showed proficiency in avoiding FPs, recall exhibited more variability, indicating fluctuations in the model's ability to correctly identify all actual positives, i.e., all street trees present.

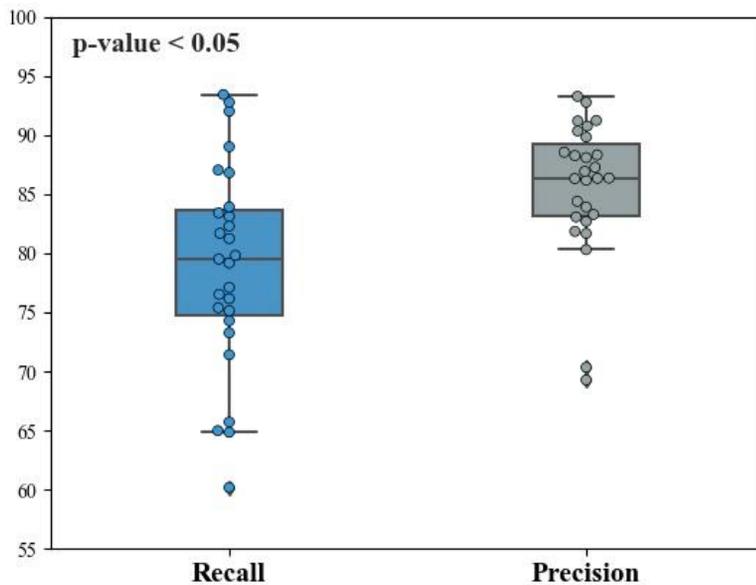


Figure 7. Statistical comparison of recall and precision. The p-value of recall and precision are 0.04.

Table 3. Evaluation of proposed individual street tree detection method.

Datasets	TP	FN	FP	Recall	Precision	F1
----------	----	----	----	--------	-----------	----

						score
Dataset 1	1570	573	168	73.26	90.33	80.91
Dataset 2	922	72	188	92.76	83.06	87.64
Dataset 3	949	67	213	93.41	81.67	87.14
Dataset 4	943	82	122	92.00	88.54	90.24
Dataset 5	2009	665	372	75.13	84.38	79.49
Dataset 6	850	163	115	83.91	88.08	85.95
Dataset 7	1362	203	216	87.03	86.31	86.67
Dataset 8	682	153	66	81.68	91.18	86.17
Dataset 9	941	238	151	79.81	86.17	82.87
Dataset 10	1128	335	115	77.10	90.75	83.37
Dataset 11	401	61	29	86.80	93.26	89.91
Dataset 12	692	226	54	75.38	92.76	83.17
Dataset 13	1082	250	144	81.23	88.25	84.60
Dataset 14	950	329	233	74.28	80.30	77.17
Dataset 15	297	155	57	65.71	83.90	73.70
Dataset 16	731	229	153	76.15	82.69	79.28
Dataset 17	644	349	94	64.85	87.26	74.41
Dataset 18	558	111	84	83.41	86.92	85.13
Dataset 19	310	63	49	83.11	86.35	84.70
Dataset 20	512	205	227	71.41	69.28	70.33
Dataset 21	256	66	29	79.50	89.82	84.35
Dataset 22	2254	692	357	76.51	86.33	81.12
Dataset 23	2425	638	321	79.17	88.31	83.49
Dataset 24	1640	353	364	82.29	81.84	82.06
Dataset 25	1055	698	212	60.18	83.27	69.87

Dataset 26	987	532	95	64.98	91.22	75.89
Dataset 27	1546	191	652	89.00	70.34	78.58
			Min	60.18	69.28	69.87
			Max	93.41	93.26	90.24
			Average	78.89	85.65	81.79

Figure 8 shows the accuracy of estimated DBH and H. Estimated DBH and H revealed slight overestimation by an average of 4.37 cm (15.13%) and 0.86 m (8.57%), respectively. This deviation is highlighted by the RMSE values, with DBH and H exhibiting RMSEs of 8.17 cm (28.27%) and 2.18 m (21.82%), respectively.

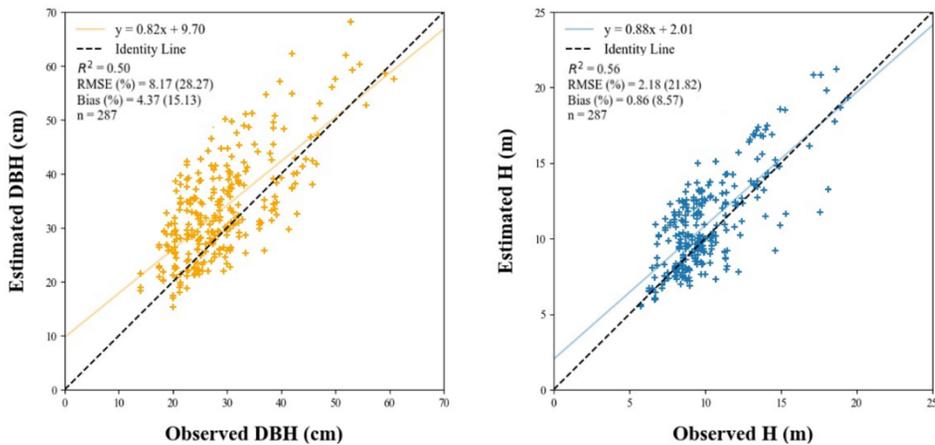


Figure 8. Comparison of estimated structural parameters to observed structural parameters.

We detected 35,247 street trees after omitting the street trees detected multiple times throughout the dataset. Gb (20.95 %), Py (21.63 %), and Zs (22.46 %) occupied more than the half

population of street trees in the study site. Figure 9 shows the distribution of H, DBH, and carbon stock of detected street trees. The median value of H of the entire detected street trees was 8.68 m, with a first quantile of 6.85 and a third quantile of 11.02 m. The median value of DBH of the entire detected street trees was 25.02 cm, with a first quantile of 19.40 cm and a third quantile of 32.17 cm. The median value of carbon stock of the entire detected street trees was 60.90 kg, with a first quantile of 35.48 kg and a third quantile of 113.39 kg.

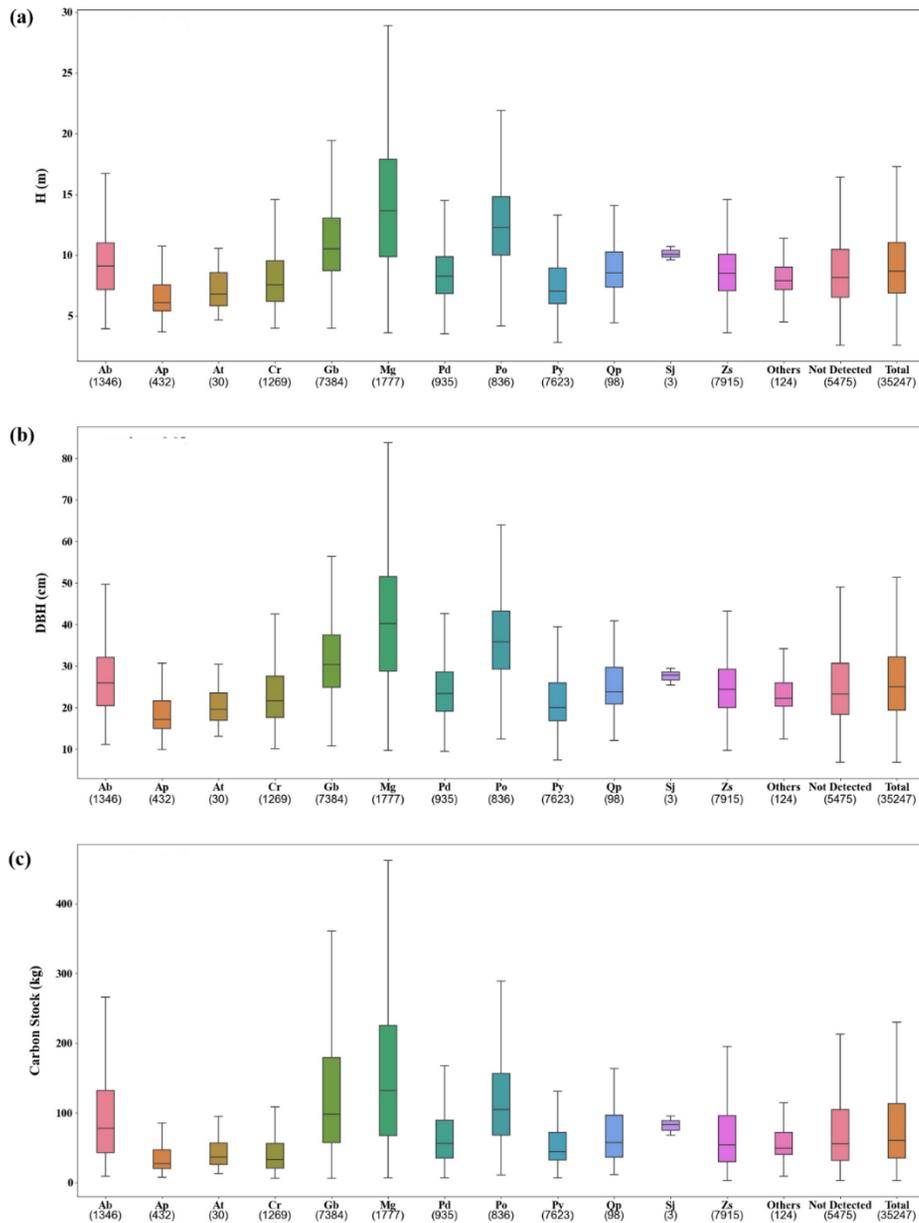


Figure 9. Distribution of structural parameters and carbon stock of detected street trees per species. (a) Distribution of H per species. (b) Distribution of DBH per species. (c) Distribution of carbon stock per species. The numbers below species names are the number of detected street trees.

The carbon stock of detected street trees amounted to 2.16

± 0.03 GgC. Figure 10 shows the spatial distribution of carbon stock in street trees. Without the use of species-specific allometric equations, the carbon stock was estimated at 2.34 ± 0.03 GgC. The bias of DBH in Figure 8 is fixed before the calculation of carbon stock. To quantify this uncertainty, we assumed a normal distribution of errors from DBH estimation. We then conducted a Monte Carlo simulation with 10,000 iterations, using the 95% CI of the error distribution. The lower and upper bounds of the 95% CI of the simulation results were reported.

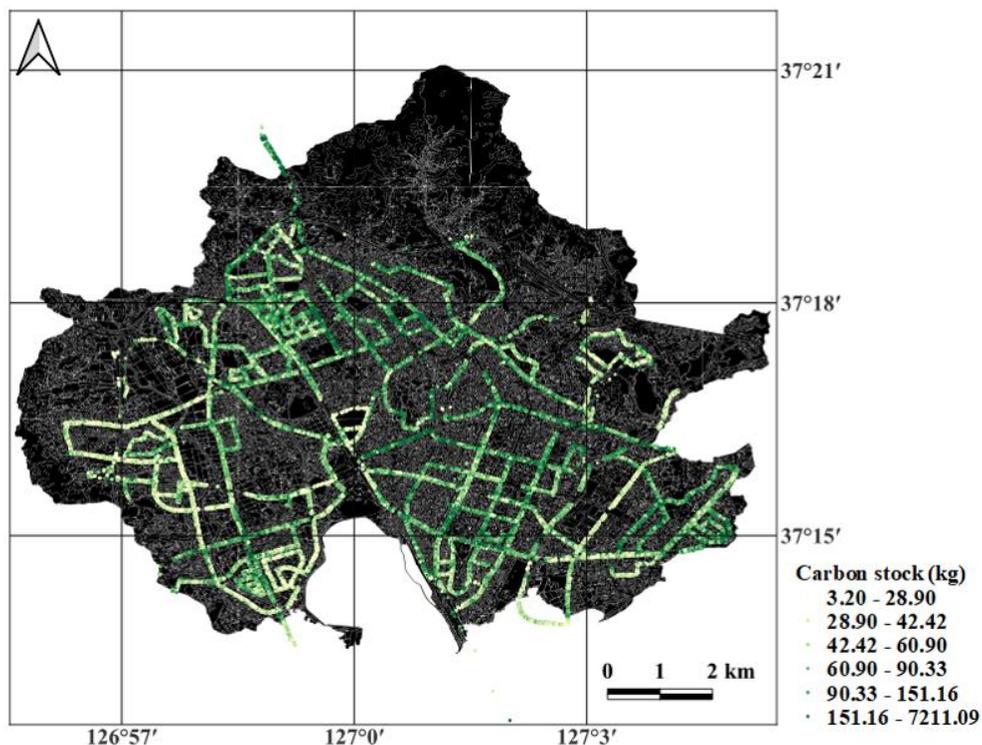


Figure 10. Carbon stock map of individual street trees in the study site.

Chapter 4. Discussion

To gain a deeper understanding of the factors causing this variability in recall, we undertook a qualitative analysis. We found

that low recall tends to occur when street trees are undetected in sequence, as illustrated in Figure 11. There is a low probability of sequential street trees being missed in the point cloud processing, as these processes incorporate a refinement step from the image's semantic segmentation result – a characteristic of our proposed method. Through a comprehensive qualitative analysis of image processing step, as depicted in Figure 12, we determined that the majority of FNs originate from the camera's vulnerability to varying environmental conditions (Cui et al., 2022).



Figure 11. Tree detection result from Dataset 25. Most of the missed street trees were in sequence rather than distributed randomly. Yellow boxes represent the sequences of missed trees. Red dots represent ground truths, and blue dots represent detected street trees.



Figure 12. (a) Example of trees obscured by a bus, thus the semantic segmentation result does not include these trees. (b) Example of a green background hindering the accurate segmentation of trees. (c) Example where the tree trunk is not visible because the area beneath the tree is dark.

Furthermore, we identified two factors that might decrease the performance of our proposed method during the qualitative analysis of the point cloud. Firstly, a considerable number of detected trees were counted as FPs because they did not fit the definition of a street tree as defined in Section 2.1, as shown in Figure 13. This suggests our method may extend its applicability to trees other than

street trees. Secondly, the MMS used in this study could not capture street trees located at a higher elevation than the vehicle. For instance, the trees in Figure 14 are located outside the LiDAR FOV and categorized as FNs, although they were detected in the semantic segmentation step.

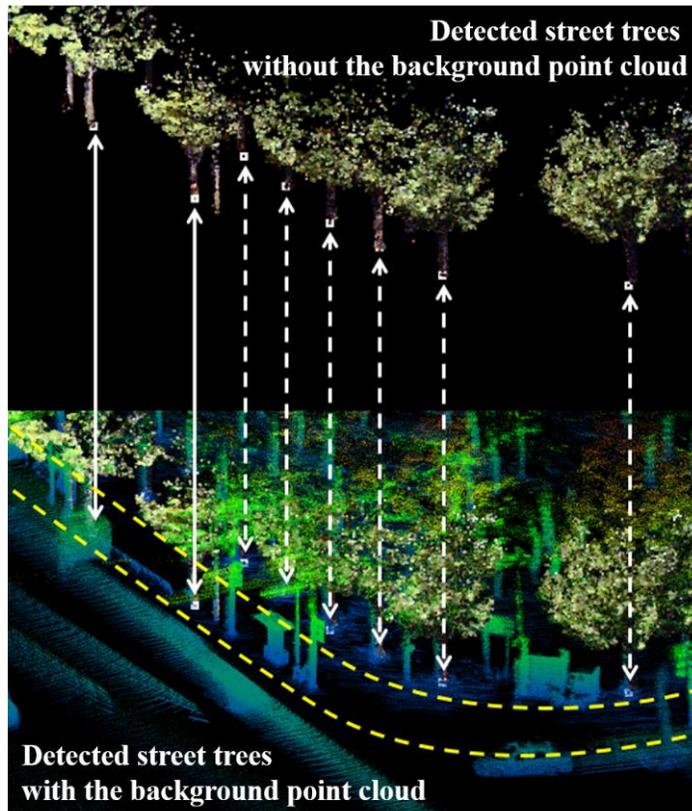


Figure 13. Example of detected trees that do not fit the definition of a street tree. Dashed line arrows denote trees that are not located on the pedestrian strip, while solid line arrows point to street trees located on the pedestrian strip. The yellow dashed line demarcates the border of the pedestrian strip. Segmented trees are visualized in RGB color.

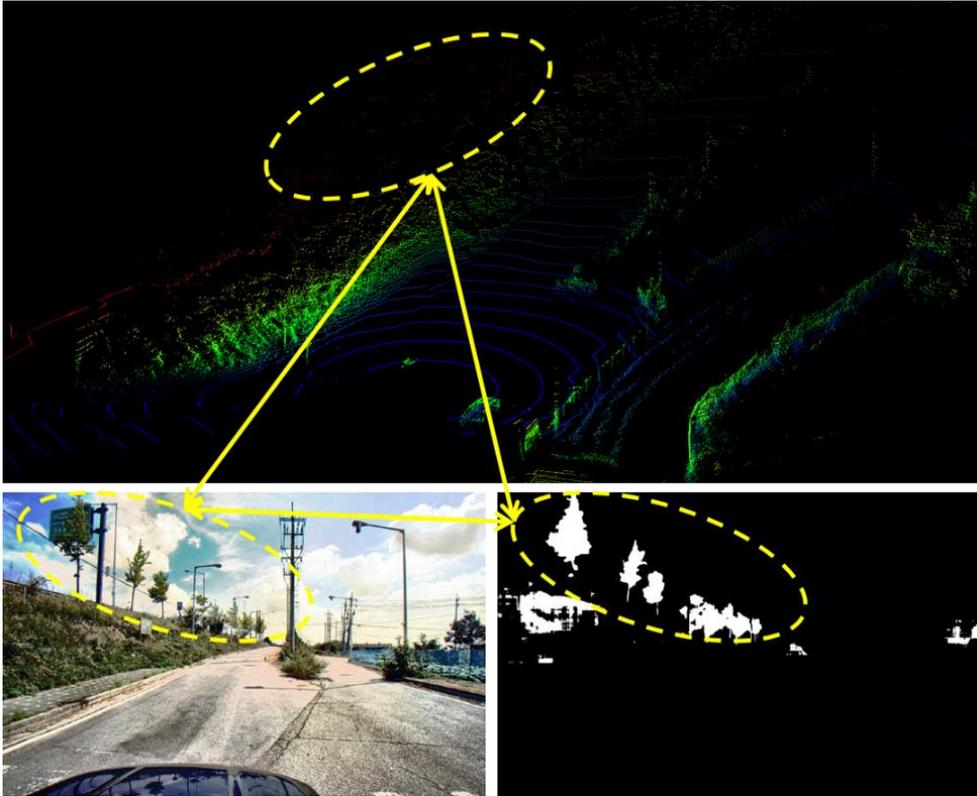
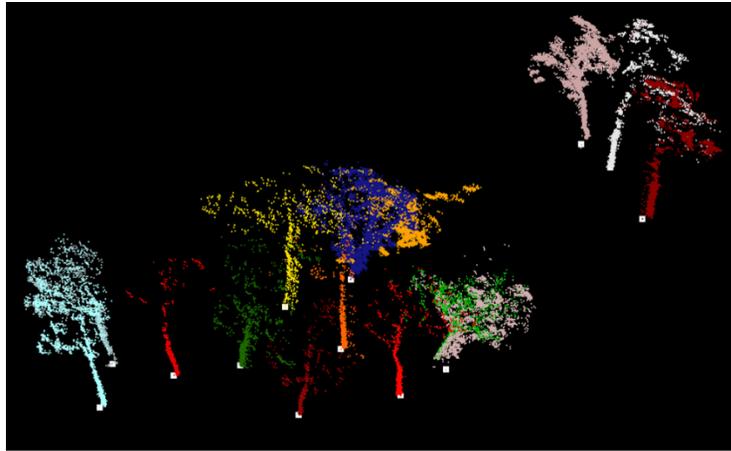


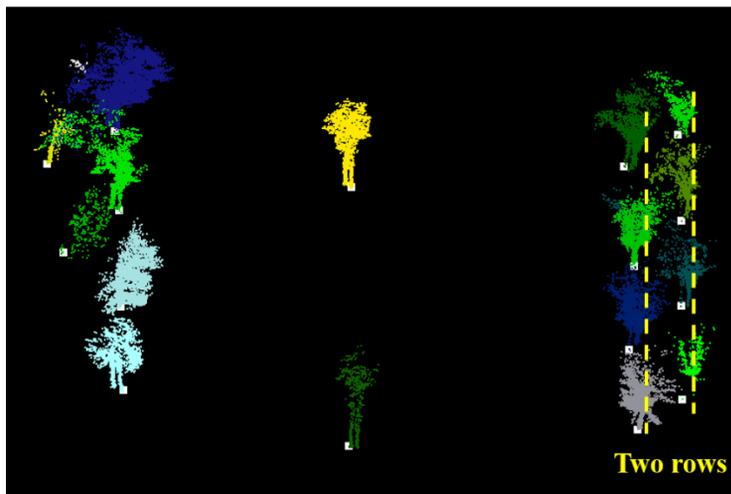
Figure 14. Example of FNs that are detected in the semantic segmentation step, but missed in subsequent steps. The yellow dashed-line circle and solid-line arrow refer to the corresponding region between the image and point cloud.

To evaluate the robustness of our method to various urban scenarios, we present examples of segmentation results for different street tree planting scenarios and challenging situations. Besides the regularly planted street trees in rows adjacent to the road, street tree planting scenarios in the study site can be categorized into three types: densely planted in planters, planted in multiple rows, and planted in median strips. As shown in Figure 15, our proposed method performed robustly in all planting scenarios. The most challenging situation to segment street trees in urban streets is the case when urban furniture near street trees or pole-like objects exist. As

shown in Figure 16, our method succeeded in segmenting the tree point cloud in challenging situations.

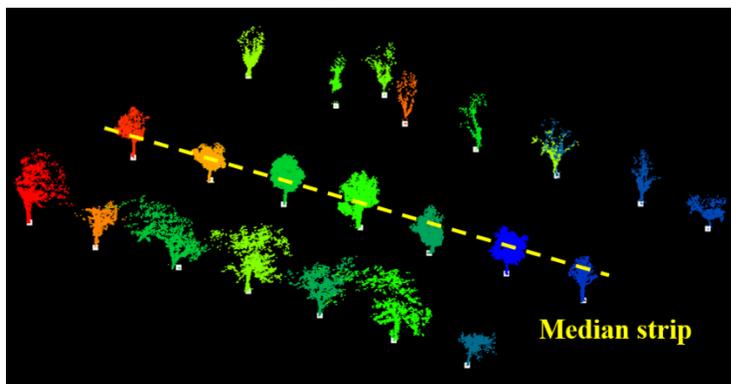


Street trees densely planted in the planters



Two rows

Street trees planted in two rows



Median strip

Street trees planted in the median strip

Figure 15. Examples of individual street tree detection results for each planting scenario, excluding the regular type. In our collected dataset, the maximum number of rows for the planting scenario of street trees planted in multiple rows was two. The colors of the point cloud represent each detected tree, and white squares indicate their locations.

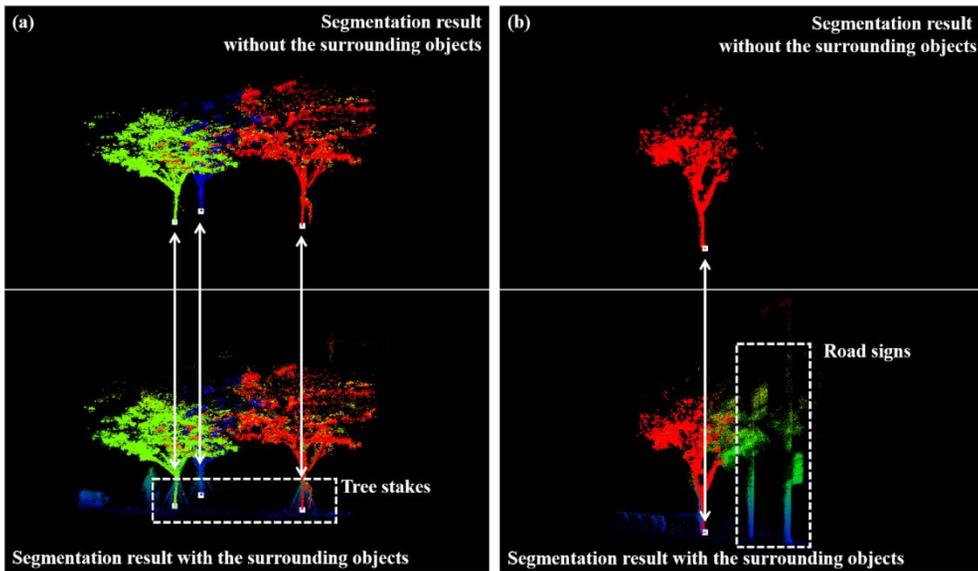


Figure 16. Examples of instance segmentation results. Top images represent the segmentation result without surrounding objects and bottom images represent the segmentation result with surrounding objects. (a) Instance segmentation result of street trees surrounded by tree stakes. (b) Instance segmentation result of street trees near road signs. Colors of point cloud represent each detected tree, and white squares indicate their locations.

The reliability of estimated carbon stock for each street tree is influenced by the accuracy of species classification and estimated DBH. The difference in carbon stock estimation, when using species-

specific allometric equations versus general allometric equations, was found to be 0.18 GgC at the city scale. This value provides a rough benchmark to determine the need for species classification, though the difference may vary depending on the species composition within the city. The uncertainty in DBH estimation primarily arises from the quality of segmentation and the characteristic of MMS scanning only half of the trunk. Our qualitative analysis demonstrates that our method accurately captures the location of street trees, even when the segmentation quality is imperfect. This contributes to the reduction in the uncertainty of DBH estimation by setting the boundary of estimated DBH from the H, which is calculated using the tree location.

The specifications of sensors implemented in MMS significantly influence the results. Regarding LiDAR FOV, its limited vertical range could potentially cause errors in H estimation and result in low recall. Meanwhile, the camera's FOV was limited to the front of the vehicle, so that only a partial scan of the point cloud is used within our proposed framework to synergize with the images. Modifying the camera to capture side and rear views can help mitigate the occlusion problem caused by moving objects. We anticipate that enabling the MMS to capture the entire surroundings of the vehicle could greatly enhance the performance of our method.

Chapter 5. Conclusion

Our study presents a novel framework for mapping the carbon stock of individual street trees using a LiDAR-camera fusion-based MMS. This shows its potential as a means to measure the carbon

stock of street trees without the need for labor-intensive field surveys. Our framework demonstrated commendable performance in detecting street trees and accurately estimating their DBH and H, which are critical parameters for reliable carbon stock estimation. Furthermore, our framework can be employed more practically in various urban environments without the need for parameter tuning, as most empirical parameters are designed to be sensor-specific rather than site-specific. We validated the reliability and practicality of our method by reporting the number and carbon stock of street trees through extensive experiments in Suwon, Republic of Korea.

Future research could focus on refining our current framework and measuring the carbon stock of all vegetation in urban areas. Improving the sensors in MMS, employing better fusion techniques, and implementing more sophisticated learning-based models could be considered to advance our current framework. Since the vehicle has limitations as it cannot access the entirety of urban vegetation, developing means to create tree inventories in those areas at an individual level, without labor-intensive field surveys, is essential for thorough coverage of urban vegetation. Such insights hold promising implications for urban planning and climate change mitigation strategies, as more accurate and detailed carbon stock information can guide the placement and care of urban trees for maximum carbon sequestration.

The importance of accurate carbon stock estimations is becoming more apparent, and our study sets a significant precedent in urban carbon management. We are optimistic that our work will inspire further research, leading to more effective, reliable, and sophisticated systems for estimating the carbon stock of urban trees, thus enabling more sustainable and resilient cities.

Overall, this research marks a significant step towards creating a sustainable urban future. By leveraging cutting-edge technologies like LiDAR-camera fusion and a learning-based approach, we can gain a deeper understanding of our urban environments and make more informed decisions about their management. This ultimately contributes to the global fight against climate change.

Acknowledgement

This paper was supported by Technology Development Project for Creation and Management of Ecosystem based Carbon Sinks (202300218237) through KEITI, Ministry of Environment.

Bibliography

- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324/METRICS>
- Chave, J., Coomes, D., Jansen, S., Lewis, S. L., Swenson, N. G., & Zanne, A. E. (2009). Towards a worldwide wood economics spectrum. *Ecology Letters*, 12(4), 351–366. <https://doi.org/10.1111/J.1461-0248.2009.01285.X>
- 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>
- 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 Journal of Photogrammetry and Remote Sensing*, 190, 165–180. <https://doi.org/10.1016/J.ISPRSJPRS.2022.06.004>
- Cui, Y., Chen, R., Chu, W., Chen, L., Tian, D., Li, Y., & Cao, D. (2022). Deep Learning for Image and Point Cloud Fusion in Autonomous Driving: A Review. *IEEE Transactions on Intelligent Transportation Systems*, 23(2), 722–739. <https://doi.org/10.1109/TITS.2020.3023541>
- Dubayah, R., Blair, J. B., Goetz, S., Fatoyinbo, L., Hansen, M., Healey, S., Hofton, M., Hurtt, G., Kellner, J., Luthcke, S., Armston, J., Tang, H., Duncanson, L., Hancock, S., Jantz, P., Marselis, S., Patterson, P. L., Qi, W., & Silva, C. (2020). The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography. *Science of Remote Sensing*, 1, 100002. <https://doi.org/10.1016/J.SRS.2020.100002>
- Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. www.aaai.org
- Fu, C., Hu, P., Dong, C., Mertz, C., & Dolan, J. M. (2018). Camera-Based Semantic Enhanced Vehicle Segmentation for Planar LIDAR. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018–November*, 3805–3810. <https://doi.org/10.1109/ITSC.2018.8569413>
- Gong, Z., Lin, H., Zhang, D., Luo, Z., Zelek, J., Chen, Y., Nurunnabi, A., Wang, C., & Li, J. (2020). A Frustum-based probabilistic framework for 3D object detection by fusion of LiDAR and camera data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 159, 90–100. <https://doi.org/10.1016/J.ISPRSJPRS.2019.10.015>
- Guan, H., Yu, Y., Ji, Z., Li, J., & Zhang, Q. (2015). Deep learning-based tree classification using mobile LiDAR data. *Remote Sensing Letters*, 6(11), 864–873. <https://doi.org/10.1080/2150704X.2015.1088668>
- Jiang, T., Wang, Y., Liu, S., Zhang, Q., Zhao, L., & Sun, J. (2023). Instance recognition of street trees from urban point clouds using a three-stage neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 199, 305–334. <https://doi.org/10.1016/J.ISPRSJPRS.2023.04.010>
- Kang, W., Lee, Y. H., Kang, C. W., Chung, W. Y., Xu, H. L., & Matsumura, J. (2011). Using the inverse method to estimate the solar absorptivity and emissivity of

- wood exposed to the outdoor environment. *Journal of the Faculty of Agriculture, Kyushu University*, 56(1), 139–148. <https://doi.org/10.5109/19542>
- Kim, H. J., & Lee, S. H. (2016). Developing the volume models for 5 major species of street trees in Gwangju metropolitan city of Korea. *Urban Forestry & Urban Greening*, 18, 53–58. <https://doi.org/10.1016/J.UFUG.2016.05.004>
- Kwon, R., Ryu, Y., Yang, T., Zhong, Z., Im, J. (2023). Unpublished manuscript.
- 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 Journal of Photogrammetry and Remote Sensing*, 175, 326–339. <https://doi.org/10.1016/J.ISPRSJPRS.2021.03.002>
- Mcperson, E. G. (1998). ATMOSPHERIC CARBON DIOXIDE REDUCTION BY SACRAMENTO'S URBAN FOREST. *Journal of Arboriculture*, 24(4).
- Ning, X., Tian, G., & Wang, Y. (2019). Top-Down Approach to the Automatic Extraction of Individual Trees from Scanned Scene Point Cloud Data. *Advances in Electrical and Computer Engineering*, 19(3). <https://doi.org/10.4316/AECE.2019.03002>
- Nowak, D. J., & Crane, D. E. (2002). Carbon storage and sequestration by urban trees in the USA. *Environmental Pollution*, 116(3), 381–389. [https://doi.org/10.1016/S0269-7491\(01\)00214-7](https://doi.org/10.1016/S0269-7491(01)00214-7)
- Nowak, D. J., Greenfield, E. J., Hoehn, R. E., & Lapoint, E. (2013). Carbon storage and sequestration by trees in urban and community areas of the United States. *Environmental Pollution*, 178, 229–236. <https://doi.org/10.1016/J.ENVPOL.2013.03.019>
- Park, J. H., Baek, S. G., Kwon, M. Y., Je, S. M., & Woo, S. Y. (2018). Volumetric equation development and carbon storage estimation of urban forest in Daejeon, Korea. *Forest science and technology*, 14(2), 97–104. <https://doi.org/10.1080/21580103.2018.1452799>
- Puente, I., González-Jorge, H., Martínez-Sánchez, J., & Arias, P. (2013). Review of mobile mapping and surveying technologies. *Measurement*, 46(7), 2127–2145. <https://doi.org/10.1016/J.MEASUREMENT.2013.03.006>
- Pu, S., Rutzinger, M., Vosselman, G., & Oude Elberink, S. (2011). Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), S28–S39. <https://doi.org/10.1016/J.ISPRSJPRS.2011.08.006>
- Redmon, J., & Farhadi, A. (2018). *YOLOv3: An Incremental Improvement*. <https://arxiv.org/abs/1804.02767v1>
- Safaie, A. H., Rastiveis, H., Shams, A., Sarasua, W. A., & Li, J. (2021). Automated street tree inventory using mobile LiDAR point clouds based on Hough transform and active contours. *ISPRS Journal of Photogrammetry and Remote Sensing*, 174, 19–34. <https://doi.org/10.1016/J.ISPRSJPRS.2021.01.026>
- Timilsina, N., Staudhammer, C. L., Escobedo, F. J., & Lawrence, A. (2014). Tree biomass, wood waste yield, and carbon storage changes in an urban forest. *Landscape and Urban Planning*, 127, 18–27. <https://doi.org/10.1016/J.LANDURBPLAN.2014.04.003>
- Weng, W., & Zhu, X. (2021). U-Net: Convolutional Networks for Biomedical Image Segmentation. *IEEE Access*, 9, 16591–16603. <https://doi.org/10.1109/ACCESS.2021.3053408>

- Xiao, L., Wang, R., Dai, B., Fang, Y., Liu, D., & Wu, T. (2018). Hybrid conditional random field based camera-LIDAR fusion for road detection. *Information Sciences*, *432*, 543–558. <https://doi.org/10.1016/J.INS.2017.04.048>
- Xiao, W., Vallet, B., Brédif, M., & Paparoditis, N. (2015). Street environment change detection from mobile laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, *107*, 38–49. <https://doi.org/10.1016/J.ISPRSJPRS.2015.04.011>
- Xie, Y., Tian, J., & Zhu, X. X. (2020). Linking Points with Labels in 3D: A Review of Point Cloud Semantic Segmentation. *IEEE Geoscience and Remote Sensing Magazine*, *8*(4), 38–59. <https://doi.org/10.1109/MGRS.2019.2937630>
- 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>
- Zhao, Y., Hu, Q., Li, H., Wang, S., & Ai, M. (2018). Evaluating Carbon Sequestration and PM_{2.5} Removal of Urban Street Trees Using Mobile Laser Scanning Data. *Remote Sensing 2018*, Vol. 10, Page 1759, *10*(11), 1759. <https://doi.org/10.3390/RS10111759>
- Zhong, L., Cheng, L., Xu, H., Wu, Y., Chen, Y., & Li, M. (2017). Segmentation of Individual Trees from TLS and MLS Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(2), 774–787. <https://doi.org/10.1109/JSTARS.2016.2565519>
- Zou, X., Cheng, M., Wang, C., Xia, Y., & Li, J. (2017). Tree Classification in Complex Forest Point Clouds Based on Deep Learning. *IEEE Geoscience and Remote Sensing Letters*, *14*(12), 2360–2364. <https://doi.org/10.1109/LGRS.2017.2764938>

Abstract

도시의 가로수들은 도시 나무의 상당 부분을 차지하고 있음에도 불구하고, 그들의 개별 탄소 저장량에 대한 연구는 아직 미흡하다. 본 연구에서는 광범위한 도시 지역을 커버하고, 고정밀 공간 샘플링을 가능케 하며, 종과 구조적 파라미터들의 동시 취득을 가능케 하는 LiDAR-카메라 퓨전 기반의 모바일 매핑 시스템(MMS)을 사용하여 개별 가로수들의 탄소 저장량을 매핑하였다. 우리는 개별 가로수를 탐지하기 위한 두 단계 접근법을 구현하였는데, 이는 이미지에 U-Net을 적용하여 의미론적인 분할을 수행하고, 이미지의 분할 결과에 근거하여 포인트 클라우드에 랜덤 포레스트 분류기를 적용하는 것이다. 가로수의 탄소 저장량을 측정하기 위해, 우리는 이미지로부터 나무 종을 분류하기 위해 YOLOv3를 사용하였고, 포인트 클라우드로부터 흉고직경(DBH)과 수고(H)를 계산하였다. 실험을 통해, 우리는 한국 수원시에 스캔된 거리에서 35,247그루의 가로수를 식별하였고, 이들의 탄소 저장량이 2.16 ± 0.03 GgC 이었다. 도시 규모의 검증을 통해 제안된 개별 가로수 탐지법의 평균 재현율, 정밀도, F1 점수는 각각 78.89, 85.65, 81.79로 나타났다. 또한, 예측된 DBH와 H는 각각 평균 4.37 cm (15.13%)과 0.86 m (8.57%)으로 약간 과대평가되는 경향을 보였으며, 이에 대한 RMSE는 각각 8.17 cm (28.27%)과 2.18 m (21.82%)였다. 이 연구는 LiDAR-카메라 퓨전 기반의 MMS를 사용하여 개별 가로수의 탄소 저장량을 추정하기 위한 실용적인 프레임워크를 제공하며, 더욱 정확한 도시 탄소 관리와 도시 탄소 관리의 발전을 위한 길을 여는데 기여한다.