71. From simulation to field: a ground truth-free approach for 3D orchard monitoring

H. Murcia Moreno[1,2<*] and S. Lacroix[1]
[1]*LAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France; * hfmurciamo@laas.fr*
[2]*Facultad de Ingeniería, Universidad de Ibagué, Ibagué, Colombia*

Abstract

Lack of annotated data as well as model transfer challenges limit accurate structural analysis of 3D orchards in field conditions by LiDAR and machine learning techniques. This study explores a scalable framework that trains deep learning models on synthetic labelled data. It then applies them to unlabelled real point clouds acquired with an unmanned ground vehicle in an apple orchard. Integrating classification, skeletonisation, and contrastive learning the framework segments trees, generates structural maps, and detects anomalies without ground-truth annotations. These results suggest potential contributions to flexible orchard monitoring by supporting variability analysis based on descriptors derived from the structural representations of trees.

Keywords: contrastive learning, data-driven tree characterization, LiDAR point clouds

Introduction

Precision agriculture increasingly uses advanced technologies to enhance crop management and phenotyping. Light detection and ranging (LiDAR) has emerged as a key tool for acquiring 3D data, providing precision in tree mapping and structural analysis for agricultural environments (Rivera *et al*., 2023). Various computational approaches have been explored to enhance phenotypic analysis in orchard environments from point clouds. Nevertheless, automating segmentation and tree characterization remains challenging, especially in large or heterogeneous vegetation areas. This study identifies two primary obstacles: firstly, real-world databases are limited in number and often present format incompatibilities, require substantial preprocessing for complete reconstruction or focus on specific tree viewpoints. Consequently, these limitations affect the scalability and generalizability of the models. Secondly, LiDAR-derived data analyses face a dependency on large annotated datasets, the production of which is labour-intensive and restricts their practical application in agriculture (Jin *et al*., 2021). This dependency underlines the interest in developing autonomous data tagging methods to overcome scalability limitations in field phenotyping (Xu and Li, 2022).
Recent advances in tag learning, including active, semi-supervised, weakly supervised, self-supervised learning, and unsupervised clustering, seek to minimise reliance on manual annotations and enhance scalability (Li *et al*., 2023). However, robust plant phenotyping requires diverse and continuously updated datasets to enhance model generalization in variable environments. While active and semi-supervised learning reduce dependence on labelled data, they face limitations in heterogeneous conditions. Similarly, sparsely supervised methods using coarse labels struggle with fine-grained tasks like individual tree characterization. Addressing these challenges involves exploring approaches that can effectively extract meaningful information even in the absence of labelled or annotated data. These include identifying ground, understory vegetation (such as weeds and grass), and trees. Likewise, segmentation plays a vital role in orchard analysis. When accurately performed, it enables the extraction of biophysical variables like tree height, trunk diameter, and crown attributes using dedicated geometric algorithms. However, in the absence of annotated data, it is essential to explore methods that derive abstract, data-driven representations to support spatial analyses, group comparisons, and variability mapping. Deep learning models enable the extraction

of latent features that, while not directly interpretable, capture intrinsic patterns, supporting analyses such as clustering, anomaly detection, and spatial variability (Pang *et al.*, 2021).
To overcome the difficulty of generating diverse real-world databases, this study explores the use of synthetic data to replicate varied and complex orchard scenarios, making model training easier and transferable to real data without retraining. Specifically, this work utilizes datasets generated with the discrete anisotropic radiative transfer (DART) software, which accurately simulates LiDAR acquisitions under realistic conditions. The resulting simulations provide detailed radiative transfer modelling, enabling the generation of realistic LiDAR point clouds that replicate tree structures, canopy distributions, and laser vegetation responses. This study investigates the following questions: (i) Can real-world orchard data be classified into key categories – such as ground, weeds, and trees – using models trained exclusively on synthetic data? and (ii) Can contrastive learning applied to isolated tree point clouds provide additional insights, enabling the identification of non-tree objects and the distinction between different tree typologies? To address these questions, we propose a LiDAR-based framework that integrates different learning strategies to process raw orchard point clouds. The following sections provide a detailed description of the framework, the materials and methods used for its implementation and evaluation, and the key findings of the study.

Framework

The proposed framework integrates simulated and real LiDAR data for tree segmentation and analysis in an orchard with free-standing trees, where individual canopies do not overlap. It follows a multi-stage process, starting with data generation, followed by 3D scene interpretation and segmentation of individual trees (Figure 1). This leads to a contrastive learning phase for anomaly detection and tree clustering based on proximity. The framework concludes with the application of trained models to real-world data to validate their performance.

Synthetic data generation
The 3D orchard reconstruction models are simulated using the Discrete Anisotropic Radiative Transfer (DART) software (Gastellu-Etchegorry *et al.*, 2004). In these simulations, a virtual 3D LiDAR system is configured to replicate the characteristics of the real device, including its resolution, field of view, sampling rate, wavelength, and height above the ground. The process starts by moving the mobile laser scanner (MLS) in an environment with a set of simulated trees, categorised in different groups based on a predefined user-defined taxonomy. Each simulated point in the orchard is represented as $p_i^{\text{syn}} = (x_i, s_i)$, where where $x_i = (x_i, y_i, z_i) \in R^3$ denotes the spatial coordinates of the point in the world reference frame, and $s_i \in S$ represents the semantic class of the point. The simulated dataset is represented as a matrix $P_{\text{syn}} \in R^{nxd}$, where n is the total number of simulated points and is the dimensionality of each data point (including spatial coordinates and class label). If a point p_i^{syn} belongs to the category S = tree, an additional value $g_i \in G_{\text{syn}}$ is assigned, where $G_{\text{syn}} = \{0,1...,12\}$ represents specific tree group attributes. To enhance the learning process, data augmentation is performed via voxelisation with a fixed voxel size v_s, resulting in a tenfold increase in the number of trees. Voxel centres serve as seed points, and for each centre, the n_{NN} k-nearest neighbours are selected from the original dataset.

Orchard scale
Supervised classification is conducted using synthetic data subsets for training. The RandLA-Net model (Hu *et al.*, 2021), a state-of-the-art neural architecture designed for efficient semantic segmentation of large-scale point clouds, is employed. Once the classification model is established, the orchard-scale model is applied to real data, producing three spatial maps: ground, weeds, and trees. The data labelled as trees are extracted and a voxelisation step is applied to reduce the computational cost, followed by tree individualisation. Each potential tree is isolated using an algorithm based on

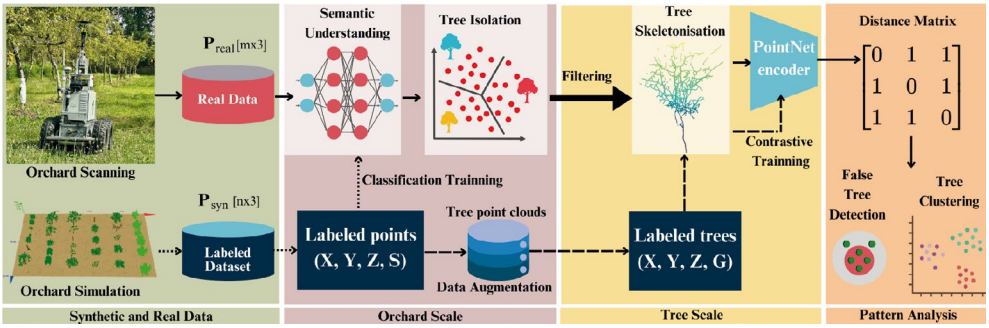

Figure 1. Multi-stage LiDAR framework for individual tree analysis with pre-trained models.

hierarchical density-based spatial clustering of applications with noise (HDBSCAN) (McInnes *et al.*, 2017). Sparse clusters are discarded, resulting in N_{trees} clusters. Then original point data within the voxels are retrieved by finding the points occupied for each voxel.

Tree scale

Statistical outlier removal (SOR) filtering is applied to reduce noise efficiently. The Laplacian-based contraction method simplifies the 3D tree point cloud into a skeleton, preserving its main structure while reducing data size. This process extracts key components like the trunk and main branches, enhancing tree architecture analysis and computational efficiency (Meyer *et al.*, 2023). The contrastive learning stage follows a supervised approach to learn 128-dimensional feature descriptors $f \in R^{128}$ from simulated point cloud data $p_i^{\text{real}} = (x_i, s_i)$. This model builds upon the foundational PointNet architecture (Qi *et al.*, 2017), but introduces key enhancements to process both spatial and additional point descriptors. It leverages permutation invariance, local and global feature aggregation, and a spatial transformer network (STN) to robustly align the spatial (XYZ) coordinates. Furthermore, additional descriptors are concatenated after the STN transformation, and a sequence of convolutional and fully connected layers with increased capacity and regularisation via dropout is used to extract robust and discriminative global features.

Pattern analysis

Starting from the pre-trained PointNet encoder, a distance matrix $D \in R^{N_{trees} \times N_{trees}}$ is computed to represent the pairwise Euclidean distances between point clouds corresponding to individual trees. The diagonal elements of D are set to zero to ensure self-similarity does not contribute to the analysis. For each tree, the mean distance to all others is computed, yielding a vector of average distances $\mu \in R^{N_{trees} \times 1}$. The reconstructed point cloud is subdivided into segments of n_{NN} points, like the simulation point clouds, and is fed into the pre-trained model for the classification stage, to produce labelled spatial coordinate data.

Materials and methods

Real data collection

Field data were collected using an Ouster OS0-128 LiDAR (2048×x128 beams at 10 Hz) mounted at 1.65 m above ground on an unmanned ground vehicle (UGV). In August 2024, the remotely operated UGV moved through three 50 m rows of an apple orchard in Toulouse, France. Each row had about 12 trees, and the UGV travelled at approx. 1 m/s. The orchard consisted of established trees planted in a spaced-row orchard layout, planted in an approx. 3.5 m×5.0 m pattern, with an average height of around 3.2 m. Data fusion from GNSS + local RTK base station and an inertial

measurement unit (IMU) were used for position estimation. LiDAR scans and navigation data were recorded in robot operating system (ROS) format. An extended Kalman filter fused the navigation data, providing an estimate of the robot's pose T_t for each scan. This pose was represented as a transformation matrix, which defined the relationship between the sensor reference frame F_{Lidar} and the absolute reference frame F_{World}. Iterative closest point and pose graph optimization refined the transformations for precise alignment. Post-processing involved down-sampling and orchard area delimitation. The final output was a 3D reconstruction matrix $P_{real} \in R^{mxd}$, where m is the number of points number of points collected from the real-world LiDAR scans and d is the dimensionality of each data point, similar to the simulated dataset. Real data $P^{real}[m \times 3]$ shown in Figure 2a comprised approx. 28 million points over 728 m².

Framework setup and configuration

The simulation was configured with 22 trees subsequently increased to 22 with data augmentation, which were categorised in 13 groups. The generated point cloud comprising 22.2 million points over 700 m², was voxelised with a of 0.25 m to enhance learning data. For each voxel centre, 30 000 nearest points were selected, resulting in 18,936 training subsets, and 6,313 each for validation and testing. The RandLA-NET model, structured in five layers with random sampling and a local feature aggregation, preserved geometric details while optimizing computational efficiency. Subsampling rates of 4, 4, 4, 4 and 2 were applied, progressively increasing output dimensions to 16, 64, 128, 256 and 512. The training was executed on a GPU-enabled machine, with a batch size of 8 and 30 000 n_{NN} points per batch for 50 epochs. The Adam optimizer was used with an initial learning rate of 0.001, reduced using a cosine annealing scheduler. In the contrastive learning stage, synthetic tree skeletons were randomly subsampled to standardize input dimensions, ensuring GPU compatibility and consistency. The classifier trained on synthetic data utilised a reserved 20% subset for a priori classification evaluation, with a total testing time of 3.35 h. To identify potential trees, a voxelisation of 5 cm was empirically chosen to balance resolution and computational cost, followed by HDBSCAN clustering with experimentally selected parameters (minimum cluster size: 200 voxels; neighbourhood: 200 voxels) to optimize structure preservation and noise reduction. Sparse clusters were discarded and the original point data contained in each voxel were recovered. SOR filter was fine-tuned through multiple trials with a standard deviation multiplier of 2 and 10 nearest neighbours. Subsequently, in the training stage of contrastive learning, a triplet loss function was employed to maximize inter-class separation while minimizing intra-class distances, fostering the formation of compact, distinct clusters (Wu *et al*., 2017). Triplets (A,P,N) were selected based on tree groups , ensuring that the anchor and positive samples belong to the same group, while the negative sample comes from a different group. Furthermore, Weighted losses are assigned in proportion to the number of positive and negative cases to balance the learning process. Finally, an Isolation Forest-based approach established a threshold to classify trees as regular or outliers (Liu *et al*., 2008). After removing outliers, agglomerative clustering grouped the data, with the optimal cluster count determined using graph Laplacian eigenvalues and the elbow method.

Results and discussion

As summarised in Table 1, the assessment of the classification model prior to the transfer demonstrated high performance in semantic understanding, achieving high accuracy (over 99%) across the evaluated orchard elements. Specifically, it attained an overall accuracy (OA) of 99.99%, a mean intersection over union (mIoU) of 99.93%, and a mean accuracy (mAcc) of 99.97%. These results underscore the capability of the proposed classification framework to safely distinguish among the considered classes.

Field data showed differences from simulations, such as larger bushes and poles, which generate false positives or outliers. Despite these variations, the color-coded results in Figure 2b visually indicate

Table 1. Confusion matrix for classification at the orchard scales with simulated data

True/predicted class	Ground	Tree	Grass/weed
Ground	99.999%	0.032%	0.022%
Tree	0.267%	99.976%	0.020%
Grass/weed	0.085%	0.053%	99.945%

effective classification, distinguishing key elements like ground, vegetation, and tree structures. In this way, three spatial maps of the orchard are generated: ground model, weed or grass, and potential trees. Continuing the framework, the points for which a tree class prediction was obtained were used to feed the tree isolation stage. As illustrated in Figure 2c, 49 clusters with between 38,000 and 55,000 points were detected.

The contrastive learning model was evaluated on the labelled synthetic data using the adjusted Rand index (ARI) and normalized mutual information (NMI), achieving scores of 0.90 and 0.95, respectively. Next, the model was applied to real tree data. The distance matrix (Figure 3a) generated from the skeletons of each potential tree, reveals clustering patterns. The determined threshold separates four false positives from 45 regular trees (Figure 3b), corresponding to dead trees, poles, and a large wood weed, as shown in Figure 3c.

Figure 4 presents the results of the point cloud clustering of individual skeletons, represented by the dendrogram derived from the distance matrix and the 2D x- and z-axis projections. This visualisation shows the hierarchies of similarity between trees, and their grouping according to distances, and the eight identified classes.

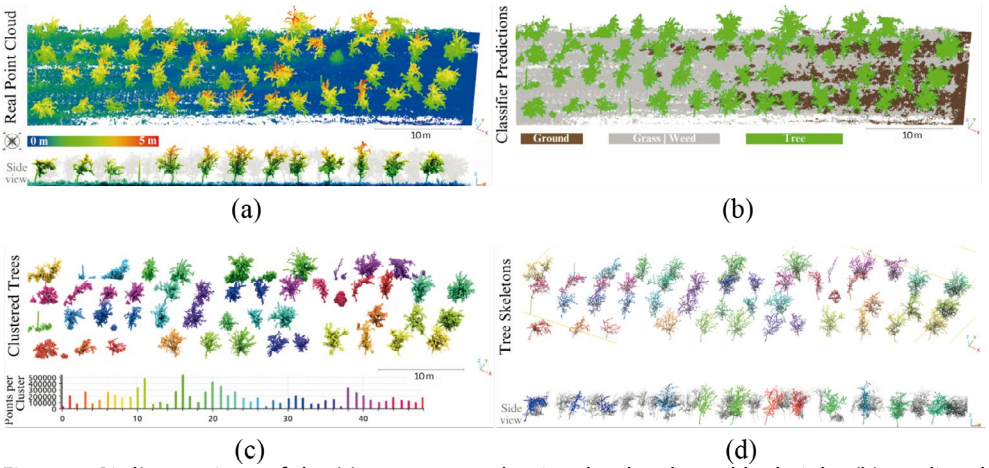

Figure 2. Bird's-eye views of the (a) reconstructed point cloud coloured by height; (b) predicted classes at orchard scale (brown represents the ground, grey indicates grass or weeds, and green denotes trees); (c) 49 isolated potential trees; and (d) derived skeletons.

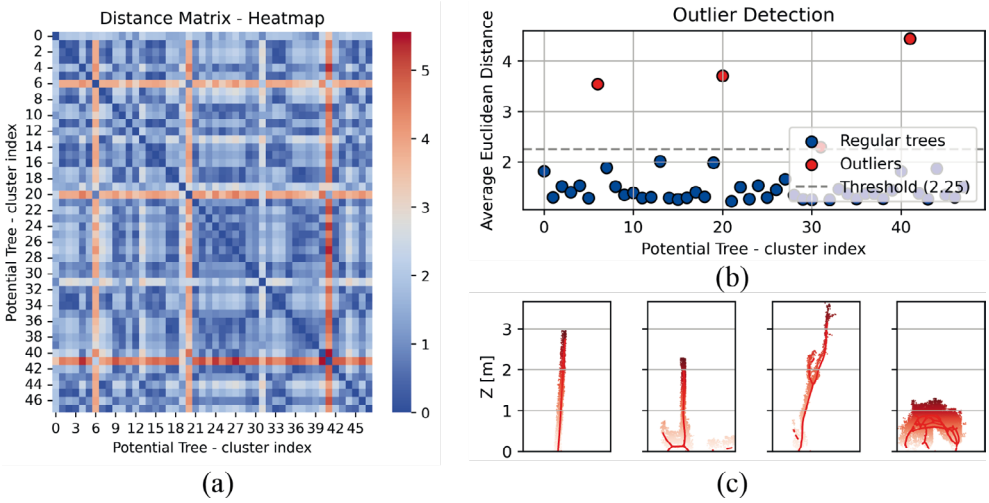

Figure 3. Analysis of tree clusters based on: (a) Similarity matrix showing pairwise distances between isolated potential trees; (b) distance-based threshold detection; and (c) false trees detected.

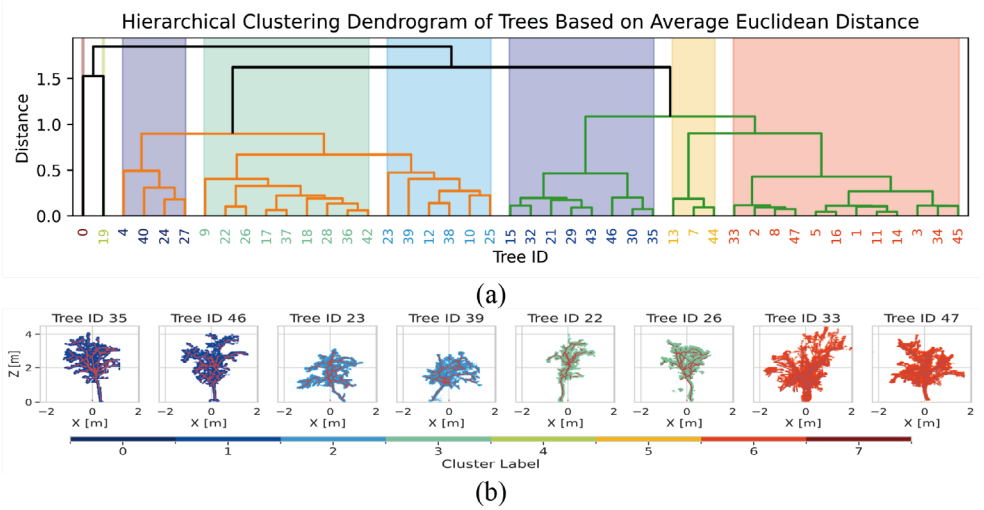

Figure 4. Clustering results of trees: (a) The dendrogram illustrates groupings and assignments; and (b) examples of point clouds and skeletons colour-coded by cluster.

Effective decision-making in precision agriculture relies on the accurate identification of orchard elements: terrain, weeds, and trees; as well as tree-specific traits, yet data-driven models face challenges in data generation and annotation. While most studies rely on supervised methods, few explore tree characterisation using label-free approaches, underscoring to opportunities for further investigation. This study presents a multi-scale framework leveraging synthetic data and machine learning to overcome these limitations. Results show that models trained on synthetic data can be applied to field conditions, but their success depends on how closely synthetic and real data match. Also, the results could improve with specific targets or partial annotations. Tree isolation

assumes sufficient spacing to avoid branch overlap, though denser formations remain challenging, as explored in forest LiDAR research (Wielgosz *et al*., 2024). Contrastive learning enables adaptive tree characterisation by generating features that support outlier detection, leading to the identification of false positives. However, it assumes that most samples are trees, resulting in latent representations to highlight group similarities and differences, which, while useful for clustering and anomaly detection, are more difficult to interpret and validate in an explanatory sense.

Conclusions

This study proposes a ground truth-free approach to proximal sensing with LiDAR, reducing dependence on labour-intensive field labelling and annotations. The framework trains models on synthetic data for application to real-world scenarios, integrating supervised and contrastive learning with clustering methods to achieve precise tree segmentation, accurate identification of misclassified trees, and individual pattern analysis. Although the approach supports broad applicability, its generalisation to denser systems, such as hedgerow orchards, remains untested. Morphology-based clustering on synthetic data and outlier detection on real data suggests the contrastive learning method can extract representative features from tree skeletons, though validation with annotated datasets and expert review is needed to confirm their relevance.

Acknowledgements

This research was supported by LAAS-CNRS, Toulouse, France, in collaboration with the Universidad de Ibagué, Colombia. Field experiments were carried out at the Lycée Agricole de Toulouse Auzeville apple orchard. The authors gratefully acknowledge financial support from "Fundación para el Futuro de Colombia" Colfuturo.

References

Gastellu-Etchegorry, J.P., Martin, E., & Gascon, F. (2004). DART: a 3D model for simulating satellite images and studying surface radiation budget. International Journal of Remote Sensing, 25(1), 73–96.

Hu, Q., Yang, B., Xie, L., Rosa, S., Guo, Y., Wang, Z., Trigoni, N., & Markham, A.. (2021). Learning semantic segmentation of large-scale point clouds with random sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), 8338–8354.

Jin, S., Sun, X., Wu, F., Su, Y., Li, Y., Song, S., Xu, K., Ma, Q., Baret, F., Jiang, D., Ding, Y., & Guo, Q. (2021). Lidar sheds new light on plant phenomics for plant breeding and management: Recent advances and future prospects. ISPRS Journal of Photogrammetry and Remote Sensing, 171, 202–223.

Li, J., Chen, D., Qi, X., Li, Z., Huang, Y., Morris, D., & Tan, X. (2023). Label-efficient learning in agriculture: a comprehensive review. Computers and Electronics in Agriculture, 215, 108412.

Liu, F.T., Ting, K.M., & Zhou, Z.H. (2008). Isolation forest. In Eighth IEEE International Conference on data Mining. IEEE, 2008, pp. 413–422.

McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: hierarchical density based clustering. J. Open Source Software, 2(11), 205.

Meyer, L., Gilson, A., Scholz, O., & Stamminger, M. (2023). CherryPicker: semantic skeletonization and topological reconstruction of cherry trees. IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6244–6253.

Pang, G., Shen, C., Cao, L., & Hengel, A.V.D. (2021). Deep learning for anomaly detection: a review. ACM Computing Surveys (CSUR), 54(2), 1–38.

Qi, C.R., Su, H., Mo, K., & Guibas, L.J. (2017). Pointnet: deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 652–660.

Rivera, G., Porras, R., Florencia, R., & Sánchez-Solís, J.P. (2023). LiDAR applications in precision agriculture for cultivating crops: a review of recent advances. Computers and Electronics in Agriculture, 207, 107737.
Wielgosz, M., Puliti, S., Xiang, B., Schindler, K., & Astrup, R. (2024). Automated forest inventory: Analysis of high-density airborne LiDAR point clouds with 3D deep learning. *Remote Sensing of Environment, 305*, 114078.
Wu, C.Y., Manmatha, R., Smola, A.J., & Krahenbuhl, P. (2017). Sampling matters in deep embedding learning. IEEE International Conference on Computer Vision, pp. 2840–2848.
Xu, R., & Li, C. (2022). A review of high-throughput field phenotyping systems: focusing on ground robots. Plant Phenomics, 9760269.