ICRA2026 Late Breaking Results Poster presented at
2026 IEEE International Conference on Robotics and Automation (ICRA 2026)
June 1-5, 2026. Vienna, Austria

Semantic 3D Skeleton Extraction for Precision Agricultural Robotics : preliminary result

Dayeon Yang[1,2] and Chanyoung Ju[2*]

Abstract— **A multi-modal dataset was constructed in a real orchard environment under leaf-off conditions using an RGB-D camera and LiDAR, enabling clear observation of branch and trunk structures. The complementary geometric information from both sensors allows for more precise 3D structural reconstruction. Dense point clouds obtained from the RGB-D camera are fused with LiDAR point clouds via ICP registration, followed by ground removal and DBSCAN clustering to segment individual trees. AdTree is then applied to each segmented tree to extract the 3D skeletal structure and generate Ground Truth. The constructed GT explicitly represents the hierarchical branch structure of each tree, and additional data collection under leaf-on conditions is planned to enable quantitative evaluation of skeleton extraction performance across varying foliage conditions. Furthermore, the constructed dataset will be utilized for training and evaluation of a Flow Matching-based generative model for tree skeletonization. Flow Matching enables stable skeleton reconstruction even from noisy and heavily occluded point clouds in real orchard environments, and the dataset is expected to facilitate quantitative analysis of performance differences between leaf-off and leaf-on conditions.**

I. INTRODUCTION

Real-world orchard environments present complex structures where recovering branch topology from sensor data is challenging due to dense foliage and severe self-occlusion [1]. Tree topology, describing the connectivity of trunks and branches, is essential for agricultural robotics tasks such as pruning, growth monitoring, and harvesting [2].

These challenges are further amplified by species-dependent differences in tree architecture. Pear trees, which are the focus of this work, exhibit more complex branching patterns and denser canopies than apple trees [3], resulting in more severe self-occlusion in point clouds. Conventional geometric or regression-based skeletonization methods struggle under such sparse and partial observations [4], [5], limiting reliable reconstruction of branch connectivity.

To address this, we construct a multi-modal dataset in a real pear orchard using RGB-D cameras and LiDAR under leaf-off conditions. Ground truth skeletons are generated via ICP-based fusion, DBSCAN clustering, and ADTree-based extraction [7], enabling evaluation under varying occlusion levels.

Recent work has explored generative models such as diffusion for tree skeletonization [6]. Motivated by this, we adopt a Flow Matching-based generative framework to enable robust reconstruction from heavily occluded point clouds.

II. METHOD

A. Skeleton Extraction

We propose a generative framework for extracting tree skeleton structures from 3D point clouds using Conditional Flow Matching [8]. Given an input point cloud, the model learns a continuous transformation that maps randomly initialized nodes to target skeleton positions via a learned velocity field. During inference, the nodes are progressively updated through an ODE solver, and a Minimum Spanning Tree (MST) is constructed to enforce a valid and connected tree topology.

To define a continuous transformation between the initial node distribution and the target skeleton nodes, we adopt a linear interpolation path:

$$x_t = (1-t)x_0 + tx_1 \tag{1}$$

where $x_0 \sim p_0(x)$ denotes a randomly sampled initial node, $x_1 \sim q(x)$ is a target skeleton node, and $t \in [0,1]$ is the interpolation time step.

The model is trained using Conditional Flow Matching (CFM), which aligns the predicted velocity field with a target vector field along the interpolation path:

$$\mathcal{L}_{CFM}(\theta) = \mathbb{E}_{t,x_0,x_1}\left[\|v_\theta(x_t,t) - u_t(x_t \mid x_1)\|^2\right] \tag{2}$$

where $v_\theta(x_t,t)$ is the velocity field predicted by the network, and $u_t(x_t \mid x_1)$ represents the target flow directing samples toward the skeleton nodes.

At inference time, skeleton nodes are generated by solving the following ordinary differential equation:

$$\frac{dx}{dt} = v_\theta(x,t) \tag{3}$$

which iteratively updates node positions from x_0 toward the final skeleton configuration.

To construct the final skeleton graph from the predicted nodes, we apply a Minimum Spanning Tree (MST) algorithm:

$$T^* = \arg\min_{T} \sum_{(i,j)\in T} d(p_i,p_j) \tag{4}$$

where $d(p_i,p_j)$ denotes the Euclidean distance between nodes p_i and p_j. This step enforces a connected and cycle-free structure, ensuring that the resulting graph represents a valid tree topology.

---

*Corresponding author

[1] Department of Mechanical and Robotics Engineering, Gwangju Institute of Science and Technology, 123, Cheomdangwagi-ro, Buk-gu, Gwangju, Republic of Korea `ydy4890@kitech.re.kr`

[2] Purpose Built Mobility Group, Korea Institute of Industrial Technology, Cheomdangwagi-ro 208beon-gil 6, Buk-gu, Gwangju, 61012, Republic of Korea `cyju@kitech.re.kr`

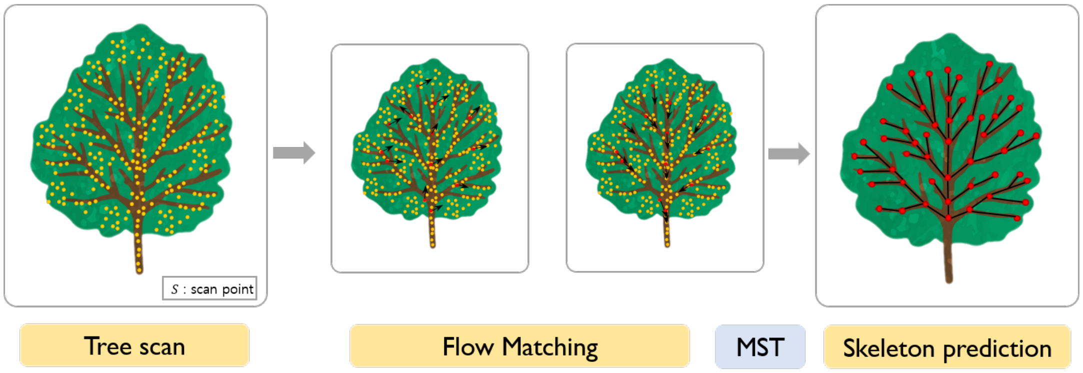

Fig. 1. Overview of the diffusion-based tree skeletonization pipeline. Given an input 3D point cloud scan of a tree, the diffusion process generates skeleton node candidates under noise perturbations. A minimum spanning tree (MST) is then constructed to enforce a valid tree topology, producing the final skeleton prediction.

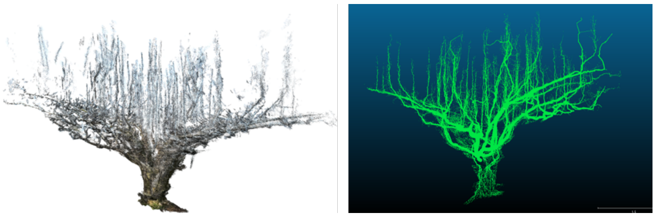

Fig. 2. A real tree point cloud captured in an orchard environment together with its corresponding ground truth skeleton generated using the ADTree model, illustrating the structural representation used for training.

B. *Real Tree Data and Ground Truth*

We construct a dataset consisting of real tree point clouds captured in orchard environments using LiDAR sensors. The acquired point clouds exhibit realistic challenges such as noise, sparsity, and severe occlusions caused by dense foliage, as shown in Fig. 2.

To obtain ground truth skeleton structures, we utilize ADTree-based modeling, which generates structured representations of tree topology. Each real tree point cloud is paired with its corresponding skeleton generated from the ADTree framework, providing supervision for training and evaluation.

This dataset enables the model to learn robust skeleton extraction under real-world conditions while preserving the underlying topological structure of trees.

III. CONCLUSION

This paper presented a generative framework for tree skeleton extraction from 3D point clouds using Conditional Flow Matching. The proposed method learns a continuous transformation that robustly maps noisy and incomplete point clouds to structured skeleton representations. By integrating flow-based node evolution with MST-based topology construction, the approach ensures geometric accuracy and valid tree structure.

Experimental results on real orchard data demonstrate that the proposed method effectively handles noise, sparsity, and occlusions, enabling reliable skeleton reconstruction under challenging conditions. The resulting skeletons provide meaningful structural information that can support various agricultural robotics applications, including pruning, growth analysis, and yield estimation.

ACKNOWLEDGMENT

This work was supported by the Korea Institute of Planning and Evaluation for Technology in Food, Agriculture and Forestry (IPET) through the High Value-added Food Technology Development Program, funded by the Ministry of Agriculture, Food and Rural Affairs (MAFRA) (No. 1545027880), and by the Korea Institute of Industrial Technology (KITECH) internal project (JK260002, Development of All-Diamond CMOS Technology based on a Smart Lab).

References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications.

REFERENCES

[1] J. L. Cardenas, C. J. Ogayar, F. R. Feito, and J. M. Jurado, ÒModeling of the 3D tree skeleton using real-world data: A survey,Ó IEEE Trans. Vis. Comput. Graph., vol. 29, no. 12, pp. 4920–4935, 2022.
[2] L. He and J. Schupp, ÒSensing and automation in pruning of apple trees: A review,Ó Agronomy, vol. 8, no. 10, p. 211, 2018.
[3] J. Li, H. Sun, G. Wu, H. Xu, S. Tao, W. Guo, K. Qi, H. Yin, S. Zhang, and S. Ninomiya, ÒStructural parameter determination and pruning pattern analysis of pear tree shoots for dormant pruning,Ó Plant Phenomics, p. 100136, 2025.
[4] Y. Livny, F. Yan, M. Olson, B. Chen, H. Zhang, and J. El-Sana, ÒAutomatic reconstruction of tree skeletal structures from point clouds,Ó in ACM SIGGRAPH Asia, 2010, pp. 1–8.
[5] J. Cao, A. Tagliasacchi, M. Olson, H. Zhang, and Z. Su, ÒPoint cloud skeletons via Laplacian based contraction,Ó in Shape Modeling International Conf., 2010, pp. 187–197.
[6] E. A. Marks, L. Nunes, F. Magistri, M. Sodano, R. Marcuzzi, L. Zimmermann, J. Behley, and C. Stachniss, ÒTree skeletonization from 3D point clouds by denoising diffusion,Ó in Proc. IEEE/CVF Int. Conf. Comput. Vis., 2025, pp. 27607–27617.
[7] S. Du, R. Lindenbergh, H. Ledoux, J. Stoter, and L. Nan, ÒAdTree: Accurate, detailed, and automatic modelling of laser-scanned trees,Ó Remote Sens., vol. 11, no. 18, p. 2074, 2019.
[8] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, ÒFlow matching for generative modeling,Ó arXiv preprint arXiv:2210.02747, 2022.