report

December 8, 2025

1 Leaf Counting and Area Estimation using 3D Point Cloud Data

TAU02E - Tomi Maijala (LUT), **Vinh Van** (TAU)

1.1 Introduction

The quantification of foliar elements within vegetation canopies—specifically the counting of leaves and the estimation of leaf area—stands as a central problem in contemporary remote sensing, forestry, and precision agriculture. Leaves function as the primary biological interface for gas exchange, driving the essential processes of photosynthesis, transpiration, and carbon sequestration. Consequently, the ability to derive precise, spatially explicit metrics of leaf distribution is not merely an academic exercise but a fundamental requirement for robust ecological modeling, crop yield forecasting, and the monitoring of forest health in the face of climate change.

Historically, the assessment of canopy structure relied heavily on direct, destructive sampling or passive optical methods. Destructive techniques, while accurate, are labor-intensive, site-specific, and impossible to scale; optical methods such as hemispherical photography or localized light sensors provide valuable estimates of Leaf Area Index (LAI) but are inherently limited by their two-dimensional nature and saturation issues in dense canopies. [1] The emergence of Light Detection and Ranging (LiDAR) technology has fundamentally altered this landscape. As an active remote sensing modality capable of penetrating canopy gaps and recording multiple returns, LiDAR offers a unique capacity to digitize the three-dimensional (3D) architecture of vegetation with millimeter-level precision.

However, the transition from raw LiDAR point clouds to meaningful biological integers—such as a “leaf count”—is fraught with complexity. It requires navigating the “semantic gap” between geometric coordinates and biological organs. This challenge necessitates a sophisticated interplay of data acquisition strategies, noise filtration, semantic segmentation (differentiating wood from leaf), and instance segmentation (distinguishing individual leaves).

1.1.1 Dataset

This report demonstrates a workflow for leaf counting and area estimation using 3D point cloud data acquired via terrestrial LiDAR scanning. For the Basic Course on Mathematical Modelling final project, we were given LiDAR data of a graphically simulated tree. The point cloud has been measured by LiDAR from [0 0 1.5]. The LiDAR is able to measure from the same direction only the object closest to it. The measured point can be a leaf, tree branch, tree trunk, or ground surface. There are no obstacles to the line connecting the measured point and the laser, the point [0 0 1.5]. The data is obtained from HELIOS++.

This dataset only contains the 3D coordinates (X, Y, Z) of the points in the point cloud. So, there is no ground truth information for semantic segmentation (differentiating wood from leaf), or instance segmentation (distinguishing individual leaves), or total leaf area or leaf count. Therefore, we will rely mainly on qualitative analysis and comparison with simulated data from HELIOS++ to assess the reliability and accuracy of our methods.

1.1.2 Challenges

Several challenges arise when attempting to accurately count leaves and estimate leaf area from 3D point cloud data:

1. **Data Quality**: The distances between points in the point cloud can vary significantly, leading to uneven point density. This is due to factors such as the diffrent sampling density between horizontal and vertical directions, distance from the LiDAR to the object (the further the object is, the sparser the points are). In general, the sparser the points are, the harder it is to accurately identify and segment individual leaves.
2. **Occlusion**: Leaves can be occluded by other leaves or branches, making it difficult to accurately count them. This is especially true in dense canopies where leaves overlap significantly. Additionally, **this dataset is obtained from a single scan position, which can lead to significant occlusion issues**.
3. **Leaf Size and Shape Variability**: Leaves can vary greatly in size and shape, making it challenging to develop a one-size-fits-all algorithm for leaf detection and area estimation.
4. **Leaf Overlap**: In dense canopies, leaves often overlap each other, leading to hard-to-seperate clusters of points that may represent multiple leaves. This can lead to undercounting of leaves and inaccurate leaf area estimates.

1.1.3 Assumptions

To address these challenges, we make several assumptions in our analysis. These assumptions also based on what we observe visually from the point cloud data:

1. **Point Density**: The point cloud density is adequate to capture the essential structure of the leaves. Its resolution is finer than the most leaves' size.
2. **Leaf Shape**: Leaves are mostly planar and can be approximated as flat surfaces for area estimation.
3. **Environmental Conditions**: The data was obtained in still air conditions, minimizing motion blur or distortion in the point cloud.

1.1.4 Approach Summary

Our method goes through several steps:

1. Keep only the tree half which faces the LiDAR scanner. (to mitigate occlusion issues)
2. Cluster the tree into smaller clusters.
3. Use different clustering proposing methods **(DBSCAN, RANSAC plane fitting)** to find the best clustering which gives the most leaf-like clusters for each smaller cluster. The leaf-like clusters are identified by **shape analysis**.

4. Compute leaf area for each identified leaf cluster and sum them up to get the total leaf area. The leaves areas are aproximated by **half the surface area of the convex hull of the leaf clusters**, because we assume the leaves are flat surfaces.

5. Only keep the leaves which have area below 0.15 m^2 to avoid suspiciously large clusters.

6. Visualize the final result and **show CT scans along different axes, for qualitative evaluation.** Clusters which look like leaves should be detected as leaves, and different leaves should be separated in the CT scans.

7. Using **simulated data from HELIOS++** for estimate the effect of occlusion.

1.2 Model: Explain your solution to the problem, also explain the process how you created this solution. Explain each equation or algorithm part. This is the most important part of your report, we are doing modelling after all.

1.3 Solutions: The solutions your model provides. Consider other possible solutions and different starting situations, how would your model work?

1.4 Methods: Compare different solutions with each other

1.5 Results: Explain the results your solution of choice provides. Discuss the robustness and/or weakness of your solution. Try to consider your model's performance as broadly as possible.

1.6 Summary: Summarise your results, the strengths and weaknesses of your model, and discuss what could be done further.

1.7 References