Improving AdTree reconstruction from degraded point clouds using preprocessing

Hana Firer[1], **Gorazd Gorup**[1], **Uroš Šmajdek**[1], **Matija Marolt**[1], **Žiga Lesar**[1,2], **Ciril Bohak**[1]

[1]*Univerza v Ljubljani, Fakulteta za računalništvo in informatiko*
[2]*Univerza v Novi Gorici*
E-mail: hf99173@student.uni-lj.si, ziga.lesar@ung.si,
{gorazd.gorup, uros.smajdek, matija.marolt, ciril.bohak}@fri.uni-lj.si

Abstract

This paper evaluates the robustness of the AdTree algorithm for tree reconstruction from degraded point clouds. Six degradation scenarios were tested on a synthetic dataset of four tree models: density reduction, Gaussian noise, random outliers, occlusion, synthetic leaf clusters, and combined degradation. Reconstruction quality was measured using Chamfer Distance, Hausdorff Distance, precision, recall, and F1-score. The results show that AdTree is relatively stable under density reduction and moderate Gaussian noise, but highly sensitive to isolated outliers. With only 1 % random outliers, the average F1-score drops from 0.968 to 0.214. SOR almost restores the original reconstruction quality. For synthetic leaf noise, DBSCAN performs best in isolated cases, while the SOR+ROR pipeline is more stable under combined degradations. The results indicate that simple preprocessing can significantly improve AdTree reconstruction quality, especially when the input point cloud contains isolated noise.

1 Introduction

Tree reconstruction from point clouds is important in forestry, environmental monitoring, and autonomous systems. Modern LiDAR scanning and photogrammetry can capture dense three-dimensional point clouds, but reconstructing tree geometry remains difficult because trees contain thin branches, irregular topology, and partially hidden structures [1].

AdTree [2] is a reconstruction method that builds a tree skeleton from a point cloud and fits cylindrical primitives to individual branches. The method works well on clean point clouds, but its performance can decrease when the input contains common acquisition artifacts such as outliers, foliage, occlusion, or uneven point density.

This paper investigates how different types of point cloud degradation affect AdTree reconstruction quality. The goal is not to replace AdTree with a new reconstruction algorithm, but to evaluate whether simple preprocessing methods can make it more robust. The main contributions of this work are:

- an experimental comparison of six degradation types,
- an analysis of the most harmful failure cases, and
- an evaluation of preprocessing methods for improving reconstruction quality before applying AdTree.

2 Related work

Wang *et al.* [1] provide a comprehensive overview of methods for reconstructing 3D urban object models from light detection and ranging (LiDAR) point clouds and highlight robustness to noise as one of the major open problems. While reconstruction via methods such as Gaussian reconstruction tend to reduce this problem [3], they require additional modalities that may be unavailable in some cases. AdTree [2] is based on building minimal spanning tree (MST) from the point cloud, simplifying the skeleton and adapting cylindrical primitives to individual branches. The method achieves high accuracy on pure LiDAR clouds of real trees, but the authors also point out its sensitivity in the presence of foliage and solitary trees. Since foliage can introduce significant reconstruction errors, separating wood from foliage prior to reconstruction is an important preprocessing step. Ferrara *et al.* [4] demonstrated that DBSCAN can effectively identify leaf clusters and separate them from woody structures by exploiting the density difference between compact leaf clusters and sparser branch structures. Lowe and Pinskier [5] propose an alternative approach using topology optimization, which may be more robust to noise, but is computationally more demanding.

This work focuses on AdTree and examines to what extent its sensitivity can be reduced by preprocessing the input point cloud.

3 Method

A synthetic dataset consisting of four tree models with different branching complexity was created for the evaluation. For each reference mesh, a point cloud containing 100 000 sampled points was generated. Six degradation scenarios were simulated to represent common imperfections found in real point cloud acquisitions: point density reduction, Gaussian noise, outliers, occlusion, synthetic leaf noise, and combined degradation. The experimental setup is summarized in Table 1.

3.1 Degradation protocol

Each degradation type was evaluated at four to five severity levels in order to simulate progressively more challenging acquisition conditions.

Table 1: Experimental setup.

Component	Setting
Dataset	4 synthetic trees
Input cloud	100k points/tree
Evaluation	100k samples/mesh
Method	AdTree
Metrics	CD, HD, precision, recall, F1-score
Threshold	$\tau = 0.10$ m
Outliers	1 %, 5 %, 10 %, 20 %
Leaf noise	50, 200, 500, 1000 clusters
Occlusion	90°, 135°, 180°
Density	50 %, 25 %, 10 %, 5 % retained
Gaussian noise	1, 2, 5, 10 cm

- **Density reduction** was performed by uniformly random sampling, retaining 50 %, 25 %, 10 %, and 5 % of the original points.
- **Gaussian noise** was added independently to each point with standard deviations of 1, 2, 5, and 10 cm.
- **Random outliers** were generated inside a volume twice the size of the original bounding box at ratios of 1 %, 5 %, 10 %, and 20 %.
- **Occlusion** was simulated by removing one side of the point cloud using viewing angles of 90°, 135°, and 180°.
- **Synthetic leaf noise** was generated by placing spherical point clusters around the tree crown, using between 50 and 1000 clusters.
- **Combined degradation:** Finally, two combined degradation scenarios were generated by simultaneously applying multiple degradation types.
  - *Severe level*: 10 % outliers + 500 leaf clusters + 50 % density retained,
  - *Mild level*: 5 % outliers + 200 leaf clusters + 75 % density retained.

3.2 Degradation evaluation

Reconstruction quality was evaluated by sampling 100 000 points from the predicted and reference meshes. Then, Chamfer distance (CD), Hausdorff distance (HD), precision, recall, and the F1-score were computed using nearest neighbor search (k-d tree):

$$\mathrm{CD} = \frac{1}{|P|}\sum_{p\in P} d(p,Q) + \frac{1}{|Q|}\sum_{q\in Q} d(q,P), \tag{1}$$

$$\mathrm{HD} = \max\left(\max_{p\in P} d(p,Q),\ \max_{q\in Q} d(q,P)\right), \tag{2}$$

$$F_1 = \frac{2\cdot \mathrm{prec}\cdot \mathrm{recall}}{\mathrm{prec}+\mathrm{recall}}, \tag{3}$$

where $\tau = 0.10$ m is the distance threshold for precision and recall, P is the set of predicted points, and Q is the set of reference points.

The Chamfer distance measures the average bidirectional distance between the predicted and reference point sets, providing an overall measure of geometric accuracy. The Hausdorff distance captures the largest deviation between the two point sets, making it sensitive to local reconstruction errors and outliers. The F1-score combines precision and recall into a single metric, indicating the proportion of points that are correctly reconstructed within the specified distance threshold.

3.3 Preprocessing improvements

To improve reconstruction quality, four preprocessing methods were evaluated. Statistical outlier removal (SOR) was applied to remove isolated outliers, while radius outlier removal (ROR), principal component analysis (PCA)-based filtering, and DBSCAN clustering were compared for reducing synthetic leaf noise. In addition, combined preprocessing pipelines were tested for point clouds containing multiple degradation types.

These methods were selected because they address different types of noise. SOR and ROR remove isolated points based on local neighborhoods, PCA exploits local geometric structure, whereas DBSCAN groups nearby points into clusters based on point density.

DBSCAN groups points into density-connected clusters based on a neighborhood radius (ε) and a minimum number of neighbors ($n_{\min}$). In the synthetic dataset, the branch structure usually forms one large connected component, whereas the synthetic leaf clusters are often separated from it. Therefore, the largest connected component was retained and smaller components were removed as leaf noise. This approach assumes that the distance between branches and leaf clusters is larger than the selected neighborhood radius [6]. The evaluated configuration was $\varepsilon = 0.05$ m, $n_{\min} = 5$.

4 Results

The following experiments evaluate degradation impact and the effectiveness of each preprocessing approach.

4.1 Impact of degradations

Figure 1 summarizes the impact of different degradation types on AdTree reconstruction quality. The results show that AdTree is robust to uniform density reduction and moderate Gaussian noise, but highly sensitive to isolated outliers and combined degradation.

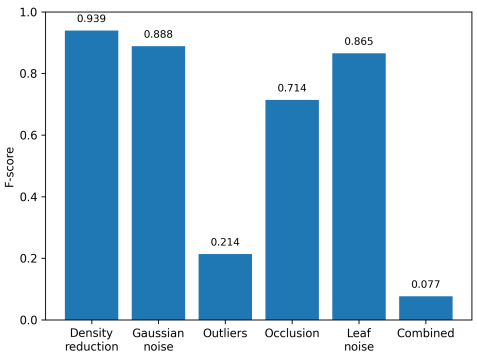


Figure 1: Average F1-score for the most severe degradation level.

With density reduction, the average F1-score remains high even when only 5 % of the original points are retained. This suggests that the MST-based skeleton can

still preserve the global branch structure when the remaining points are uniformly distributed. Gaussian noise has a slightly stronger effect, but the reconstruction remains stable up to $\sigma = 5\,\text{cm}$, with a larger drop appearing at $\sigma = 10\,\text{cm}$.

As shown in Figure 1, the most critical degradation is the presence of isolated outliers. Since the MST connects all input points, random outliers can be included in the skeleton and produce phantom branches. In the 1 % outlier case, the F1-score drops from 0.968 to 0.214, while CD increases from 0.060 to 1.233. Occlusion mainly affects recall, because hidden or poorly supported branches are omitted rather than reconstructed incorrectly. Leaf clusters have a moderate impact and mainly reduce precision.

The combination of all degradations leads to an almost complete breakdown of the reconstruction (F1-score below 0.08). The results in this scenario are not entirely monotonic—the mild level is worse in some metrics than the severe level—confirming that under heavy noise, the final success of the MST depends not only on the amount of interference but also on its random spatial position relative to the actual tree structure.

4.2 Outlier removal with SOR

Table 2 summarizes the reconstruction quality before and after SOR preprocessing for all four synthetic trees.

Table 2: Effect of SOR preprocessing on point clouds with 1 % isolated outliers.

Tree	Before			After		
	F1-score	CD	HD	F1-score	CD	HD
synthetic_01	0.438	1.032	4.308	0.939	0.070	1.284
synthetic_02	0.057	1.724	9.227	0.989	0.054	0.398
synthetic_03	0.216	1.020	5.015	0.991	0.051	1.120
synthetic_04	0.114	1.155	4.819	0.987	0.043	0.242
Average	**0.206**	**1.233**	**5.842**	**0.977**	**0.055**	**0.761**

As shown in Table 2 and Figure 2, SOR almost completely eliminates the effect of 1 % isolated outliers. The average F1-score increases from 0.206 to 0.977, while CD decreases from 1.233 to 0.055. This reduces the occurrence of phantom branches and brings the reconstruction close to the baseline quality. This improvement is also evident in the representative example shown in Figure 2. Since a more aggressive SOR configuration ($k = 20$, $\lambda = 1{,}0$) slightly degraded some reconstructions, the conservative setting ($k = 20$, $\lambda = 2{,}0$) was used in later experiments.

Table 3: F1-score per tree for 1000 synthetic leaf clusters after preprocessing.

Tree	Degr.	ROR	PCA	DBSCAN
synthetic_01	0.800	0.807	0.770	**0.864**
synthetic_02	0.891	0.889	0.940	**0.952**
synthetic_03	0.917	0.921	**0.957**	0.946
synthetic_04	0.853	0.854	0.953	**0.955**
Average	0.865	0.867	0.905	**0.929**

4.3 Comparison of methods for removing leaves

Table 3 shows reconstruction quality at the most severe leaf noise level (1000 clusters), where differences between methods are most pronounced. ROR provides only limited improvement, since synthetic leaf clusters are locally dense and are therefore not always detected as outliers. PCA-based filtering improves reconstruction quality mainly for Tree 3, while DBSCAN performs best overall. It achieves the highest F1-score in three out of four trees and the best average (0.929 vs. 0.865 without preprocessing). This result is partly caused by the synthetic setup, where leaf clusters are often spatially separated from the branch structure.

4.4 Pipeline for combined degradations

For combined degradations, several preprocessing pipelines were evaluated: SOR, SOR+ROR, SOR+PCA, and SOR+DBSCAN. The goal was to determine whether methods that perform well on individual degradations also remain effective when isolated outliers, leaf clusters, and density thinning are present at the same time. Table 4 summarizes the quantitative comparison of all evaluated pipelines, while Figure 3 shows a representative reconstruction example.

Table 4: Comparison of preprocessing pipelines for combined degradations.

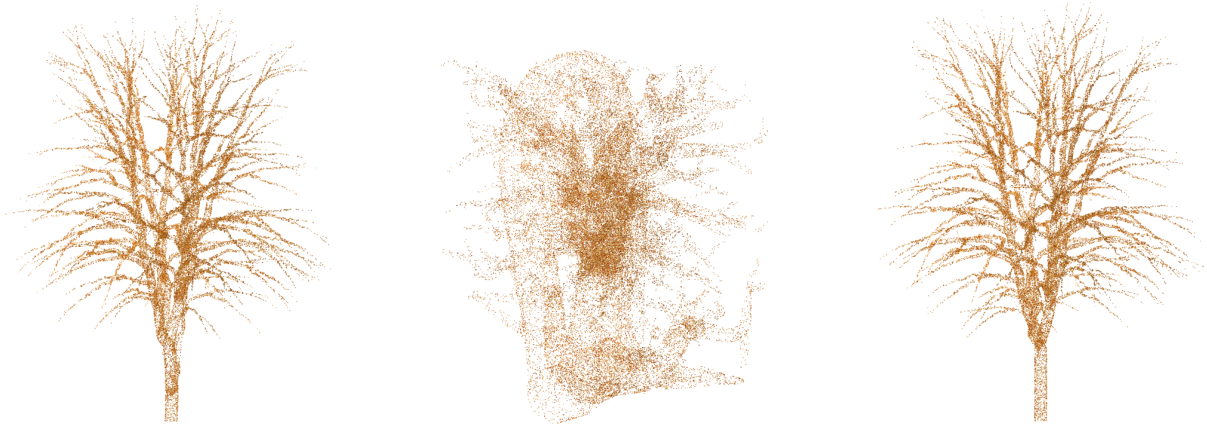
Pipeline	Mild			Severe		
	F1	CD	HD	F	CD	HD
No filter	0.051	3.355	8.932	0.077	2.436	7.512
SOR	0.796	0.274	3.376	0.234	3.269	7.009
SOR+ROR	**0.804**	0.276	**3.351**	**0.516**	**0.879**	**4.427**
SOR+PCA	0.803	**0.270**	3.368	0.379	2.128	5.880
SOR+DBSCAN	0.399	1.543	5.021	0.044	3.765	7.020

As shown in Table 4, SOR, SOR+ROR, and SOR+PCA achieve very similar results under mild combined degradation. The qualitative improvement obtained with the SOR+ROR pipeline can also be observed in Figure 3, where most of the original branch topology is successfully recovered.

For severe combined degradations, the differences between pipelines become much larger. SOR alone is no longer sufficient, while SOR+ROR remains the most stable option, achieving the highest F1-score and the lowest CD. In contrast, SOR+DBSCAN performs poorly, which suggests that DBSCAN is less reliable when leaf clusters are combined with strong outlier noise and density reduction.

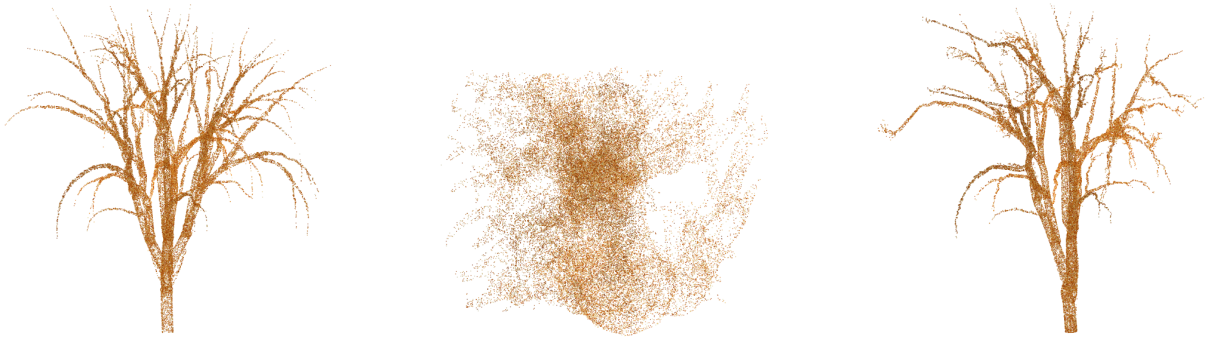
5 Conclusion

This paper evaluated the robustness of AdTree reconstruction under six common point cloud degradation scenarios and investigated whether simple preprocessing can improve reconstruction quality. The experiments show that AdTree is particularly sensitive to isolated outliers, whereas density reduction and moderate Gaussian noise have a considerably smaller impact. Appropriate preprocessing substantially improves robustness. SOR proved



(a) Baseline (ground truth) (b) 1 % outliers (c) After SOR

Figure 2: Representative reconstruction results for Tree 3. 1 % isolated outliers cause phantom branches that collapse the skeleton (b), but SOR preprocessing almost fully restores the baseline quality (c).



(a) Baseline (ground truth) (b) Mild combined degradation (c) After SOR + ROR

Figure 3: Representative reconstruction results for Tree 2. Mild combined degradation almost completely destroys the branch structure (b), whereas the SOR+ROR pipeline largely restores the original topology (c).

most effective for isolated outliers, DBSCAN performed best for synthetic leaf noise, but only if leaf clusters are spatially separated from the branch structure, and the SOR+ROR pipeline provided the best overall performance under combined degradations.

The main limitation of this study is the use of synthetic point clouds, which do not fully represent the complexity of real LiDAR acquisitions. Real-world datasets may contain connected foliage, background objects, varying point density, and missing branch structures that were not considered in the present experiments.

Future work should evaluate the proposed preprocessing pipeline on real LiDAR datasets and investigate more advanced approaches for separating branches from foliage, such as skeleton-guided filtering or learning-based semantic segmentation.

6 Acknowledgment

The research was conducted within the project Geospatial Information Technologies for Resilient and Sustainable Society (GeoAI) (GC-0006), funded by the Slovenian Research and Innovation Agency (ARIS).

References

[1] R. Wang et al. “LiDAR Point Clouds to 3-D Urban Models: A Review”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.2 (2018), pp. 606–627. DOI: 10.1109/JSTARS.2017.2781138.

[2] S. Du et al. “AdTree: Accurate, Detailed, and Automatic Modelling of Laser-Scanned Trees”. In: *Remote Sensing* 11.18 (2019), p. 2074. DOI: 10.3390/rs11182074.

[3] E. Mažgon et al. “Wood log reconstruction from Gaussian representations”. In: *Zbornik štiriintridesete mednarodne Elektrotehniške in računalniške konference ERK 2025 = Proceedings of the 34th International Electrotechnical and Computer Science Conference ERK 2025 : Portorož, Slovenija, 25. - 26. september 2025*. 2025, pp. 378–381.

[4] R. Ferrara et al. “An automated approach for wood-leaf separation from terrestrial LiDAR point clouds using the density based clustering algorithm DBSCAN”. In: *Agricultural and Forest Meteorology* 262 (2018), pp. 434–444. DOI: 10.1016/j.agrformet.2018.04.008.

[5] T. Lowe and J. Pinskier. “Tree Reconstruction Using Topology Optimisation”. In: *Remote Sensing* 15.1 (2023), p. 172. DOI: 10.3390/rs15010172.

[6] M. Ester et al. “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise”. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*. AAAI Press, 1996, pp. 226–231. DOI: 10.5555/3001460.3001507.