Vo-SmoG: A Versatile, Smooth Segment-based Ground Filter for Point Clouds via Multi-scale Voxelization

Erzhuo Che¹, Andrew Senogleś, Michael J. Olsen⁴

¹ Oregon State University, Corvallis, Oregon, 97331, United States – (erzhuo.che, senoglea, michael.olsen)@oregonstate.edu

KEY WORDS: Lidar, Laser Scanning, SfM, UAS, Point Cloud, Ground Filtering, Voxelization, DEM.

ABSTRACT:
Point clouds acquired by light detection and ranging (lidar) and photogrammetry technology (e.g., structure from motion/multi-view stereo-SfM/MVS) are widely used for various applications such as topographic mapping due to their high resolution and accuracy. To generate a digital elevation model (DEM) or extract other features in the data, the ground points and non-ground points usually need to be separated first. This process, called ground filtering, can be tedious and time consuming as it requires substantial manual effort for high quality results. Although many have developed automated ground filtering algorithms, very few have the capability to process data acquired from different scenes and systems. In this paper, we propose a versatile ground filter based on multi-scale voxelization and smooth segments, named Vo-SmoG. The proposed method introduces a novel voxelization approach, followed by isolated voxel filtering, lowest point filtering, local smooth filtering, and ground clustering. The result of the Vo-SmoG ground filtering is a classified point cloud. The effectiveness and efficiency of our method are demonstrated qualitatively and quantitatively. The quantitative evaluation consists of both point-wise and grid-wise comparisons. The recall, precision, and F1-score are over 97% in terms of classification while the root mean squared error (RMSE) of the DEM is within 0.1 m, which is on par with the reported vertical accuracy of the tested data. We further demonstrate the versatility of the Vo-SmoG via large-scale, real-world datasets collected from different environments with mobile laser scanning, airborne laser scanning, terrestrial laser scanning, uncrewed aircraft system (UAS)-SfM, and UAS-lidar.

1. INTRODUCTION

Lidar (light detection and ranging) and structure from motion/multi-view stereo (SfM/MVS) photogrammetry technology have revolutionized terrain mapping and offer many benefits over other techniques such as radar and conventional photogrammetry including resolution, accuracy, and canopy penetration. Entities ranging from local to international in scale have invested heavily to update digital elevation models (DEMs) using these technologies given the wide array of applications supported by these high-quality, versatile data (e.g., Sugarbaker et al., 2014). Both lidar and SfM/MVS data can be acquired from terrestrial, Uncrewed Aircraft Systems (UAS), mobile, or airborne platforms depending on the desired accuracy, resolution, and area to be captured (Olsen and Gillins, 2015). The resulting point cloud from lidar requires additional processing to extract ground points from points representing other objects or noise in the data.

1.1 Related Work
A wide range of techniques have been developed to identify ground points within the point cloud since commercial lidar systems became available. Sithole & Vosselman (2004) evaluated several ground filters for ALS data in the early days and categorized those based on their concepts into slope-based, block-minimum, surface-based, and clustering/segmentation. Later, Meng et al. (2010) classified the ground filters for ALS data into segment/cluster, morphology, directional scanning, contour, TIN (triangular irregular network), and interpolation, and completed a comprehensive review. Then Chen et al. (2017) further reviewed and summarized the more recent approaches using a similar classification. Besides the aforementioned categories and characteristics, deep learning has become more and more popular in point cloud processing including ground filtering (e.g., Rizaldy et al., 2018, Jin et al., 2020). Following the previous work covered in the review papers, Ni et al. (2018) improved the progressive TIN densification (PTD) filtering algorithms by adding a clustering step and adopting iterative graph cuts. Similarly, with PTD as the final refinement of the terrain model, Cai et al. (2019) implemented cloth simulation filtering (CSF) to acquire the seed points to generate the initial terrain model. Instead of using a TIN to organize the point cloud, some other methods structure the data into 2D or 3D grid (e.g., voxels). Additionally, Wang et al. (2017) developed a voxel-based ground filtering approach where the constraint of slope and local elevation variation is embedded in the voxelization settings. Kumar et al. (2018) analysed the local elevation variation in a set of neighbour searching results at each point to identify the ground points iteratively. To assist the ground filtering and terrain modelling to better follow the topography, splines are also commonly used to model the terrain from the seed ground points (e.g., Sánchez et al., 2019, Liu et al., 2020, Abdeldayem, 2020, Chen et al., 2021).

The point cloud from UAS-SIM shares a lot of similarities with ALS data except that ALS can penetrate vegetation to capture the ground and record multiple returns for a laser pulse. To process UAS-SIM point clouds, Yilmaz et al. (2018) evaluated the performance of a variety of ground filtering algorithms on UAS-SIM point clouds. In contrast to ALS and UAS-SIM that capture the site from a top view with relatively consistent point density, terrestrial laser scanning (TLS) usually captures objects with much more detail from the side but with varying density. Thus, it is not straightforward to simply apply the ground filters developed for ALS to TLS dataset. In our prior work (Che & Olsen, 2017), we reviewed several TLS-specific ground filtering
approaches and proposed a ground filter by analysing density anomalies in each scanline. Additionally, Roberts et al. (2019) tested some of the ALS ground filtering approaches on some TLS datasets. Although Mobile laser scanning (MLS) has a similar scanning geometry with TLS, recognizing ground is a much easier task with MLS given it is operated in areas with relatively less vegetation and flatter slopes. Che et al. (2019) performed a comprehensive review in object recognition in MLS point clouds. Recently, for MLS point clouds with more complex road side environments including steep slopes, Che et al. (2021) proposed a segment-based ground filtering method relying on reconstructing the scan pattern grid and Yadav et al. (2021) proposed a hybrid approach based on PTD.

1.2 Objectives

The general challenges in ground filtering for point cloud data can be summarized as follows:

1. Most of these algorithms were only tested on some smaller datasets (typically no more than several millions points) such that the scalability is unclear considering the fact that the computation time often increases exponentially with larger datasets.
2. Some of the parameters are not intuitive for users to provide, and it can be challenging to tune these parameters for a complex scene without extensive knowledge and experience with certain technologies or concepts that are involved in the workflow.
3. For the machine learning and deep learning approaches, collecting training datasets is very time-consuming, and the same model may not work as effectively for a different dataset with differences in terrain. Further, it is also unclear how large the training datasets needs to be for the learning process.
4. Very few approaches are demonstrated to be able to effectively handle datasets collected from various platforms.

To overcome these challenges, we propose a novel ground filtering algorithm for point cloud data that utilizes a multi-scale voxelization segmentation approach to organize point clouds to operate efficiently on very large point cloud datasets (hundreds of millions of points). This versatile algorithm is applicable to point clouds acquired from various methods and systems (e.g., ALS, TLS, MLS, UAS-lidar, UAS-SIM). A wide variety of real-world point clouds from multiple sources were tested to demonstrate this versatility, and the effectiveness of the proposed method was evaluated both qualitatively and quantitatively.

2. METHODOLOGY

With an unorganized point cloud data and intuitive parameters as input, our method is able to label or extract the ground points automatically. The proposed ground filtering methodology consists of five steps: best-fit plane rotation (if applicable), isolated voxel filtering, lowest point filtering, local smooth filtering, and ground clustering. In this workflow, a proposed multi-scale voxelization technique is introduced to sample and organize the data.

2.1 Voxelization-based Sampling

Voxelization is a common approach to organize and process point cloud data. Ordinarily in voxelization, the point cloud is sampled to voxels and the voxels occupied by at least a certain number of points are treated as a single point or valid voxel. These processes can substantially reduce the complexity of the point cloud data while enriching the point cloud with connectivity information. However, because each voxel is often presented by its centre coordinates, the voxelized point cloud can be very sensitive to the voxel dimensions, often resulting in loss of or change to geometric information. To overcome this challenge, the proposed approach instead uses the point closest to the centre of its corresponding voxel to represent the voxel. Sampling the point cloud in such way preserves the geometry better because the coordinates of the original point cloud are used. Meanwhile, the voxel indices still provide the structure and connectivity needed for efficient processing. The original point cloud can always retrieved via the voxel structure and mapped the processing result. As a result, the proposed ground filtering is able to apply such voxelization process multiple times throughout the workflow with different cell size to analyse the data across different spatial scales.

2.2 Best-fit Plane Rotation

In some cases, there can be a dominant orientation of the terrain, which poses challenges to perform analysis plotting the point cloud onto the horizontal (x-y) plane or based on the assumption that the terrain tends to have a lower slope. For example, in cases where the area of interest is a steep rock slope or coastal cliff, a global best-fit plane can be estimated and used to re-project the point cloud such that the analysis can be adapted to the dominant slope (Olsen et al. 2020). Hence, when desired, the propose approach initially rotates the point cloud to perform the analysis in the best-fit plane using Principal Component Analysis (PCA). Thus, the coordinates $x'$, $y'$, and $z'$ refer to the coordinates after the rotation (or original coordinates if no rotation) hereafter. After the filtering is complete, the point cloud will be projected back to its original coordinate system.

2.3 Isolated Voxel Filtering

Point cloud data can contain various forms of noise. In addition to the random noise caused by the ranging error, some artefacts from moving objects in the scene can contradict to the assumptions made in the point cloud processing and analysis. Taking ground filtering as an example, many approaches initially consider the points with lower elevations in a local area as the starting or seed points to identify ground. However, as shown in Figure 1, noise points with low elevation, if not removed, can result in significant error, if not complete failure of the ground filtering process. Applying one or multiple thresholds to define the range of elevation can cope with such noise in an area without a substantial change in elevation. However, for the point cloud covering a larger area, it can be challenging to select these thresholds without extensive prior knowledge about the site. Additionally, the noise points can also occur at a similar elevation with the ground (e.g., the distant noise in Figure 1). Another approach is to identify isolated points with no close neighbours; nonetheless, such an approach cannot tackle the noise points in small clusters.

To solve these issues, in the proposed method, the point cloud is first sampled into voxels. Then, all of the points in the isolated voxels (without neighbour voxels that contain any points) are labelled as noise points. This approach is able to cope with both the noise of isolated points and small clusters. The voxel size depends on the noise level and point density. In
the presented tests, it is typically set over 10 times of $V_{\text{ground}}$.

![Image](https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-59-2021)

Figure 1. Example of noise points in a mobile lidar point cloud.

### 2.4 Lowest Point Filtering

After removing noise from the point cloud, we apply a lowest point filter by organizing the remaining data into voxels with a given voxel size $V_{\text{ground}}$. The voxel size set here should be given based on the general ground point density in the dataset. Because the voxelization result in this step will be also utilized in the following analysis, the parameter $V_{\text{ground}}$ may need to be adjusted accordingly. With the point cloud being sampled via the proposed voxelization, the lowest voxel point for each 2D voxel index on the $x$-$y$ plane is preserved (Figure 2).

![Image](https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-59-2021)

Figure 2. Example of the proposed lowest point filtering.

### 2.5 Local Smooth Filtering

Although the previous step of lowest filtering removes a substantial portion of the non-ground points from objects such as powerlines, poles, and so on, some non-ground objects are not fully removed from the point cloud (e.g., vehicles, thick vegetation, etc.), especially when the ground is not adequately captured. As a result, we propose a local smooth filter to analyse the remaining points to identify ground points.

The highly efficient normal variation analysis (Norvana) segmentation algorithm was proposed in our prior work, Che & Olsen (2018) and Che & Olsen (2019), to segment and classify terrestrial or mobile point clouds into smooth and rough surfaces. However, these approaches needed to be modified to address challenges when applying them to extract the smooth points from the filtered point cloud. First, both approaches require the point cloud data to be structured into a 2D scan pattern grid. Nonetheless, the input point cloud may not be organized and the scan pattern information is not necessarily stored in the file format. Although the scan pattern can be reconstructed in some cases (Che & Olsen, 2019), the voxelization and the filtering processes will create discontinuities in the scan pattern. Second, Norvana is sensitive to rough surface and noise in the data because it considers the maximum normal gradient locally as the criterion to determine whether a point is on a smooth surface. For a natural terrain, it is challenging to use this as the only condition to detect ground points. Thus, our approach still takes advantage of the concept of Norvana but modifies it for ground filtering to overcome the aforementioned challenges.

Once the lowest point filtering has been applied to the point cloud, we can organize the point cloud in a 2D grid to reduce the computing and searching complexity to achieve higher efficiency. The neighbourhood of a point is defined as all the points in its adjacent neighbour grids with a maximum of 8 neighbour points for a given point. The normal vector for each point is first estimated via PCA with its neighbouring points in the grid where the $z'$ component of normal is forced to point upward ($+z'$). If the normal at a point cannot be estimated because of insufficient neighbours, the point will be labelled as noise point and removed. Then, similar to Norvana, a local triangular mesh is generated around the point under analysis. Before computing the normal of each facet, we smooth the local triangular mesh by adjusting the vertices in the direction of the normal at the point under analysis with a given maximum adjustment, $T_{\Delta \text{Norm}}$, such that the proposed analysis is less sensitive to rough surfaces. Note that the actual coordinates of the point cloud are not modified during this process—just a temporary copy for the analysis. After the adjustment, the normal of each facet is computed and the maximum normal gradient over the point under analysis is compared with the user-given threshold, $T_{\Delta \text{Norm}}$. If the local normal gradient exceeds $T_{\Delta \text{Norm}}$, the point under analysis is labelled as non-ground. As a result of adjusting the local triangular mesh, the proposed approach is able to cope with surfaces with a moderate curvature or sudden change in height while still successfully removing the objects such as vegetation (Figure 3).
2.6 Ground Clustering

After all the aforementioned steps, there are still some points lying on objects such as vegetation and building roofs (Figure 3) because the prior filters have predominately only considered local features when classifying ground candidates. We apply region growing, a common approach for segmentation and modelling of point cloud data, to cluster the ground candidate points and further analyse the segments to refine the ground filtering result. To ensure the effectiveness and robustness of the region growing process, there are three key factors including seed point selection, neighbour searching approach, and growing criteria, that need to be determined to improve the ground filtering results while being generally consistent with the previous steps in terms of their assumptions.

In the proposed approach, we organize the ground candidate points via voxelization with a given voxel size of $V_{S_{seed}}$. The ground candidate with the lowest $z'$ value in each voxel will be selected as the seed points with the assumption that it is more likely to be a ground point, which is consistent with the proposed lowest point filtering. Then, starting with the seed points, a 2D neighbour searching is performed with a given radius of $R_{search}$ to close some of the gaps (e.g., occlusions) in the point cloud. The growing criteria for each neighbour point are consistent with the proposed smooth filtering where the normal difference between the seed point and this neighbour point should be no larger than $\Delta \text{Norm}_{\text{seed}}$. If a neighbour point satisfies both criteria, it will be labelled as seed point for the next iteration.

Finally, we rotate the ground points to their original coordinate system, if needed, and project them back to the original point cloud to obtain the final ground filtering result.

### Table 1

<table>
<thead>
<tr>
<th></th>
<th>MLS</th>
<th>ALS</th>
<th>TLS</th>
<th>UAS-lidar</th>
<th>UAS-SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point count</td>
<td>19.4 M</td>
<td>18.2 M</td>
<td>131.2 M</td>
<td>380.4 M</td>
<td>119.4 M</td>
</tr>
<tr>
<td>Ground points</td>
<td>14.5 M</td>
<td>7.7 M</td>
<td>43.8 M</td>
<td>54.8 M</td>
<td>28.9 M</td>
</tr>
<tr>
<td>Grid cell count</td>
<td>0.61 M</td>
<td>3.9 M</td>
<td>4.5 M</td>
<td>5.8 M</td>
<td>6.5 M</td>
</tr>
<tr>
<td>Efficiency (points per second)</td>
<td>0.46 M</td>
<td>0.12 M</td>
<td>0.47 M</td>
<td>0.50 M</td>
<td>0.26 M</td>
</tr>
<tr>
<td>Efficiency (cells per second)</td>
<td>8.5 K</td>
<td>25.2 K</td>
<td>16.0 K</td>
<td>7.6 K</td>
<td>14.0 K</td>
</tr>
</tbody>
</table>

Table 1: Summary of the testing datasets and test results.
Regarding the parameter settings, we did not conduct extensive fine-tuning, but rather provided a reasonable estimate and kept them as consistent as possible across the datasets to show the robustness of the proposed approach (Table 2). The ALS dataset is the sample data provided in the swissSURFACE3D database (Swisstopo, 2021). These data are used in the quantitative evaluation. All the other point clouds are used to demonstrate the versatility of Vo-SmoG. Note that the result for MLS data has been shown in the methodology section to illustrate each step of the algorithm.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLS</th>
<th>ALS</th>
<th>TLS</th>
<th>UAS-lidar</th>
<th>UAS-SIM</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{Ground}}$</td>
<td>0.15 m</td>
<td>0.5 m</td>
<td>0.05 m</td>
<td>0.5 m</td>
<td>0.3 m</td>
</tr>
<tr>
<td>$T_{\text{False}}$</td>
<td>30 m</td>
<td>100 m</td>
<td>50 m</td>
<td>50 m</td>
<td>50 m</td>
</tr>
<tr>
<td>$R_{\text{Start}}$</td>
<td>0.05 m</td>
<td>0.1 m</td>
<td>0.02 m</td>
<td>0.2 m</td>
<td>0.05 m</td>
</tr>
<tr>
<td>$S_{\text{Tuning}}$</td>
<td>0.3 m</td>
<td>1.5 m</td>
<td>0.1 m</td>
<td>5 m</td>
<td>5 m</td>
</tr>
<tr>
<td>$T_{\text{Size}}$</td>
<td>100 voxels</td>
<td>10 voxels</td>
<td>100 voxels</td>
<td>10 voxels</td>
<td>100 voxels</td>
</tr>
</tbody>
</table>

Table 2 Summary of the parameters used in the tests.

### 3.2 Quantitative Evaluation

The swissSURFACE3D sample dataset was chosen for the quantitative evaluation because of its pre-existing point classification (used as ground truth) generated by automated filters followed by rigorous manual clean-up. The dataset also includes a wide variety of objects (e.g., trees, buildings, powerlines, etc.) and topography (e.g., steep slopes, highways, bridges) throughout the scene (Figure 5). To assess the quantitative accuracy of Vo-SmoG at identifying ground points, both point-wise and grid-wise comparisons were performed.

For this assessment, the points in this dataset were reclassified using Vo-SmoG in order to create two ground models of the same point cloud, one ground model containing the ground points classified by Vo-SmoG (referred to as the Vo-SmoG points) and one model containing the ground points as classified in the pre-existing dataset (referred to as the reference points). Notice that we consider the bridges are part of the ground classification although they are manually assigned in a separate class in the reference data for the hydro enforcement process. All non-ground points were then removed from the data to create two distinct ground terrain models.

The first comparison consisted of a direct point-wise comparison between the classification of the Vo-SmoG points and the reference points (Figure 6). The Precision, Recall, and F1-Score for this comparison are all greater than 97% (Table 3), which shows that the proposed Vo-SmoG method is capable of yielding accurate and robust results. These results are more promising considering that a substantial number of the false positive points are constrained to 3 areas (A, B, and C of Figure 6). Areas A and B represent roofs of buildings that are connected to the surrounding ground leading to the misclassification where there is a ramp on the side of the building in area A while the building in area B is smoothly connected to terrain behind it. Area C represents a water body that is separately classified as such in the reference dataset. In practice, further classification methods could be combined with Vo-SmoG to successfully classify these points as a water body; however, this work is outside of the scope of this paper. Nevertheless, in all three of these cases, these false positives could easily be manually cleaned up by a user with much less effort than is typically required with conventional ground filters. These simple cleanup efforts would then result in even better accuracy results than represented in the above analysis.

Since many applications of ALS rely on derivative products such as DEMs, the second assessment compares the resulting DEMs from each of the ground classifications (Vo-SmoG and reference). This demonstrates how biases introduced in the ground modelling may propagate into derivative products. To ensure a consistent comparison, DEMs were generated for both datasets using a cell size of 0.5m. The mean elevation of all ground points lying within that grid cell was used without any further refinement or interpolation. A cell size of 0.5m was chosen to be consistent with the resolution of the publicly available digital surface model (DSM), which is derived from the same point cloud.

A grid-wise comparison was then performed between the two datasets (Figure 7). The precision, recall, and F1-Score are again are all greater than 97% (Table 3) and contain the same three false positive areas as described in the point-wise analysis above. Notice that the evaluation here only examines whether a cell contains at least one ground point or not. The mean, and RMSE difference between the two models is 0.00m and 0.10m respectively. It is worth noting that the reported vertical accuracy of this dataset is ± 0.1m, which is consistent with the RMSE.
3.3 Versatility Test

Vo-Smog was also tested on data from multiple sensors (e.g. TLS, MLS, UAS-SfM, and UAS lidar) and scenes (e.g. urban, rural, and forest). Vo-Smog was effective across the different sensors and scenes. Select examples are provided herein to demonstrate the versatility. For TLS data in an urban scene (Figure 8), Vo-Smog was effective in removing trees and buildings and provides a high-quality DEM of the grassy field, sidewalks, and roadways present in the scene. Note that Vo-Smog also effectively removed the clusters at the outer edges of the dataset that ordinarily would need to be manually cropped when applying conventional ground filters.

When applied to rugged terrain surveyed via UAS-lidar for a coastal landslide (Figure 9), Vo-Smog successfully removed the trees and tall grasses while still preserving the complex topographic features across the landslide that are important for monitoring efforts. Many conventional ground filters struggle with preserving the rugged terrain while removing the vegetation. It also removed noise from the lidar sensor appearing in the bottom right of the figure introduced by a water body. When applied to a similar scene with UAS-SfM (Figure 10), the proposed ground filter was effective in areas where the UAS-SfM data provides sufficient ground samples. The UAS-SfM has a marked increase in data gaps given that the ground points cannot be generated through the SfM process in dense forest canopy. Because UAS lidar is an active sensor, it is much more effective in penetrating this dense canopy to obtain ground points. Note that from the profile view in Figure 10, Vo-Smog shows its robustness to significant noise in the data. Most conventional ground filters would erroneously select many of the artificially low noise points as ground points.

Figure 6. Point-wise comparison between the Vo-SmoG points and the reference points.

Figure 7. The grid wise comparison between DEMs (0.5 m) generated by the Vo-SmoG and reference data.

Figure 8. Vo-SmoG result ($V_{\text{Ground}} = 0.05$ m) for the TLS point cloud.
4. CONCLUSION

In this paper, we propose Vo-SmoG, a novel smooth segment-based ground filtering method based on multi-scale voxelization for processing point cloud data. It was tested and evaluated both qualitatively and quantitatively in the experiment. The primary contributions can be summarized as follows:

1. We introduce a voxelization approach that can mitigate the loss of information comparing against traditional voxelization approaches. The point cloud can be organized in voxelization structures across multiple scales to cope with different types of noise and objects.

2. Vo-SmoG only requires a few intuitive parameters that can be easily provided by users.

3. Vo-SmoG is effective, efficient, and scalable, which was demonstrated by processing large point clouds including hundreds of millions of points within minutes. This high efficiency is achievable given the algorithm is structured effectively such that it can take advantage of parallel programming.

4. Vo-SmoG is very versatile and we demonstrate its ability to handle point clouds collected by several platforms and methods (e.g., TLS, MLS, ALS, UAS-lidar, UAS-SfM) from a variety of environments (e.g., urban, rural, forest, steep slope, coastal landslide, etc.) with different types and levels of noise.

In the future, we will focus on: 1) further classifying the point cloud based on the Vo-SmoG ground filtering results to consider other object types, and 2) extending the use of the proposed voxelization to more types of point cloud processing (e.g., segmentation, feature extraction).

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. CMMI-1351487 and Oregon State University under University Venture Development Fund (UVDF). The authors also appreciate the support of Leica Geosystems and David Evans and Associates who provided hardware and software used in this research. The data used in this work is provided by Oregon Department of Transportation under the project SPR-807. The UAS lidar data was collected by NHERI RAPID Facility. The authors also would like to thank the reviewers that have helped review and improve this paper.

REFERENCES


Che, E. and Olsen, M.J., 2018. Multi-scan segmentation of terrestrial laser scanning data based on normal variation...
This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.

https://doi.org/10.5194/isprs-annals-VIII-4-W2-2021-59-2021 | © Author(s) 2021. CC BY 4.0 License.


