GROUND FILTERING OF CO-REGISTERED MOBILE AND STATIONARY LASER SCANS BY USING SUPERPOINTS IN RANSAC PLANES

Dominik Stütz[1], Jianping Li[2,3], Jianzhu Huai[2], and Dimitri Bulatov[1]

[1] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Ettlingen, Germany
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Mapping, and Remote Sensing, Wuhan University, China
[3] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

KEY WORDS: Ground segmentation, mobile laser scanning, stationary laser scanning, superpoint, point cloud registration

ABSTRACT:

Ground filtering is an important tool for many applications. The high variability of landscapes makes it necessary to perform its computation with 3D points as the only input, that is, with as few as possible algorithm parameters and without any training data. In the case of terrestrial laser scans, an additional challenge comes from a highly inhomogeneous point density. The SiRP algorithm on ground filtering, relying on intermediate validating superpixels as ground or non-ground, was previously developed for airborne point clouds. We consider a dataset acquired by a mobile mapping system mounted on a car, which was extended by additional stationary laser scans. The registration algorithm is based on hierarchical merging. Afterwards, SiRP was applied to both uni-modal and multi-modal point clouds. Impressive qualitative and quantitative classification results with around 97 % on overall accuracy were obtained and discussed.

1. INTRODUCTION

Thanks to the increasing employment of mobile and stationary laser scanning and advanced techniques for registration of scans captured at different locations, there is also broader access to high density outdoor point clouds, which, in turn, reflects in many applications of surveillance, registration, and geometric modeling. For many of these applications, filtering out vegetation and clutter points is an essential step. Consider registration: One knows that temporarily dependent objects, such as trees and vehicles, are challenging for registration. Retrieving the ground is sensible if one wants to perform a stratified registration, since once the ground has been aligned, only the issues of rotation around the xy plane and planar translation remain to be solved. Consider clustering: subdivision into two categories is not only useful for the computation of DTMs (digital terrain models) but also allows facilitating hierarchical clustering of the non-manifold (tree, bush, vehicles) and manifold (buildings, terrain) points using a few relatively simple features, such as relative height and normal vector.

Recently, an approach called SiRP (Superpoints in RANSAC Planes), by (Bulatov et al., 2020), allowing subdivision of airborne point clouds into locally planar and non-planar regions, was proposed. The advantage of SiRP is that it can handle both purely 3D structures of data and scarcity, or even absence, of training examples. In (Bulatov et al., 2021), this method was applied to both an actively-sensed airborne point cloud and a result of a photogrammetric reconstruction from a UAV-borne sequence of high-resolution images. The UAV flight around the Gubbio wall (Italy) at a moderate altitude ensured a sufficient overlap of images. This means that, though the point density varied strongly from dataset to dataset, it was relatively constant within the dataset, with a few exceptions in the area of overhanging structures. In the case of stationary laser scans, the inhomogeneous point density is systematic, with orders of magnitude density difference between points close to the scan position and those only a few meters further away.

The task of this paper is to find out how well SiRP can perform ground filtering for laser point clouds collected with a stationary laser scanner (SLS) and a mobile laser scanner (MLS) in urban areas. The scans from the MLS are geo-referenced. The stationary scans are registered to the mobile scans using a multi-view registration algorithm based on hierarchical merging. We describe this algorithm and the successive (ground) point filtering algorithm in Section 3; however, we start with the description of the relevant previous works (Section 2). The methodology is applied to the multi-modal point clouds as described in Section 4 while the main conclusions and ideas for future works are presented in Section 5.

2. PREVIOUS WORKS

2.1 Point clouds registration

Point cloud registration has been receiving continuing research attention. A recent survey is given in (Dong et al., 2020). Here we briefly summarize the research trends. The existing point cloud registration methods could be categorized into coarse registration and fine registration. The coarse registration tries to find the initial transformation between two point clouds without any prior information. In contrast, the fine registration refines the initial transformation information using local matches.

Coarse registration methods can be grouped into local ones, based on local features, and global ones, based on correlation or convolution. The local methods start with feature extraction and then match features by nearest neighbor search, and then solve for the relative pose with a robust least squares method. Common feature extraction methods include fast point feature histograms (FPFHs) (Rusu et al., 2009), USIP (Li and Lee, 2019), D3Feat (Bai et al., 2020), etc. The global methods directly transform the entire point clouds and derive the relative pose, including those based on correlation and those based on deep learning. Examples of the correlation approaches are (Tsin and Kanade, 2004) and (Bernreiter et al., 2021). The end-to-end deep-learning approaches for registration include the DeepVCP

(Lu et al., 2019), (Deng et al., 2019), and OverlapNet (Chen et al., 2021).

For fine registration methods, the nearest corresponding primitives searching and parameter estimation are two keys for the accurate transformation (Rusinkiewicz and Levoy, 2001). In the corresponding primitives searching step, correspondences are selected between two point clouds according to the nearest distances or nearest descriptors (Honda et al., 2022). In the parameter estimation step, the transformation between the two point clouds is estimated using the selected correspondences. The above two steps are performed iteratively until the geometric error converges and results in accurate transformation (Yang and Li, 2022). An epitome of fine registration methods is the iterative closest point (ICP) algorithm proposed in (Besl and McKay, 1992).

The above papers mainly focus on registration of point clouds captured by the same type of LiDAR. Authors (Zang et al., 2021) proposed a patch-based approach consisting of a cascaded neural network for pose estimation and a global refinement to register SLS points to TLS points, but it could take up to 3 minutes for one pair of registration in our tests (30 million points). As such, this study used an established hierarchical approach (Dong et al., 2018) for co-registering point clouds captured by a MLS and a SLS.

2.2 Point filtering

There are numerous methods for 3D point cloud filtering. As has been mentioned, for example, in the survey article of (Chen et al., 2017), it is very difficult to separate a complex terrain relief from a variety of non-ground features using a set of limited parameters since ground DTM generation methods are usually applied to large-scale sites. Therefore, methods for ground filtering have been continuously developed.

The pioneering work of (Kraus and Pfeifer, 1998) for extracting ground surface and thus a DTM from airborne laser scanning data using iterative robust plane fitting has been followed by algorithms based on slope-based filtering (Vosselman, 2000, Sithole and Vosselman, 2001), edge-based filtering (Brovelli et al., 2002), segment-based filtering (Sithole and Vosselman, 2004), progressive morphological filtering (Zhang et al., 2003) and hierarchical filtering (Mongus and Žalik, 2012), as well as based contour analysis (Elmqvist et al., 2001).

The more recent articles constitute probabilistic pipelines (Hui et al., 2019). The method of (Mousa et al., 2021) allows dealing with incomplete point clouds by adjusting the filter shape. The increasing availability of training data made it possible to apply very robust deep-learning pipelines, such as (Hu and Yuan, 2016, Gevaert et al., 2018, Rizaldy et al., 2018, Jin et al., 2020).

Still, as the name "manifold filtering" reveals, we are searching for a quite basic set of surfaces in the point clouds, which makes the application of deep-learning-based methods an overkill. With respect to the manifold extraction from point clouds, we are inspired by the RANSAC-based procedure of (Schnabel et al., 2007), who have applied RANSAC to unorganized point clouds in order to fit selected manifolds of first and second degree (planes, cylinders, cones, etc.), whereby the points were given a normal vectors. Other authors use Hough transform (Rabbani and Van Den Heuvel, 2005). Also here, deep learning is unstoppable, finding its way into application-based research in recognizing advanced shapes in 3D (Deprelle et al., 2019).

3. METHODOLOGY

This section presents our method to register the point clouds from two modalities, the MLS, and the SLS, and our method to segment ground from the registered points.

3.1 Point cloud registration

We have point clouds collected from two types of devices, the mobile mapping system (MMS) including an MLS mounted on a car, and an SLS setup at several stops for capturing stationary data.

The MMS utilizes GNSS-IMU locations to directly geo-reference laser scanning data for mapping an area. The car moves around while the MLS collects profiling laser data, color images, as well as GNSS and tactical-grade IMU data. The Waypoint Inertial Explorer GNSS-IMU post-processing software package is used to compute the car's trajectory at 100 Hz by combining the commercial RTK correction data to the GNSS and IMU data collected by the on-board GNSS-IMU unit. The computed trajectory is used to geo-reference the raw laser data considering the sensor calibration parameters to get the aggregated 3D point clouds. These points are colored by the corresponding pixels from the image data, considering the sensor calibration information.

The MLS data has missed some parts of road surfaces and building facades due to occlusions. The SLS collection is done to complement the area mapping. From the geo-referenced MLS point clouds, we visually identified several areas with sparse laser point coverage, e.g., fractured road surfaces. Then, a SLS was set up at these area centers to gather more points.

To register the points from the SLS data to the MLS points, we adopt a hierarchical registration approach (Dong et al., 2018) which consists of a descriptor-based linear least squares method for relative transformation initialization and the subsequent ICP method for refinement. This pipeline is designed for efficiency and robustness in view of the existing geometric and learning-based registration approaches.

In the descriptor-based solver, we first detect keypoints with the method in (Mian et al., 2010). Then, for these keypoints, binary shape context descriptors are extracted as in (Dong et al., 2018). Keypoints between two point clouds are matched by the nearest neighbor search, and mutually the best matches are accepted. To deal with outlier matches, we iteratively check the difference of distances between two pairs of matches, and build groups of matches with similar inner distances. With the largest group of consistent matches, we compute the relative transform between the two point clouds using the algorithm of (Umeyama, 1991). The resultant relative transform is further refined by the generalized ICP method (Segal et al., 2009). With the refined relative transform, we map the SLS points to the MLS coordinate frame. The merged point clouds are processed by the ground filtering method described next.

3.2 Ground filtering of co-registered points

As illustrated in Figure 1, the algorithm SiRP could be divided into five main steps. First, the point cloud $\{p\}$ is approximated with so called superpoints s. Second, for each superpoint a RANSAC plane r_s is calculated, based on its nearest point cloud neighbors $G(s, p)$. Third, the superpoints are classified concerning their distance to the RANSAC plane $d(s, r_s)$. To

refine this result, the isolated ground superpoints are removed in the fourth step by clustering. Finally, the result of the superpoint classification is transferred to the original point cloud $p(s)$. While the formal description of the algorithm can be found (Bulatov et al., 2020), we give a more detailed description of these steps below.
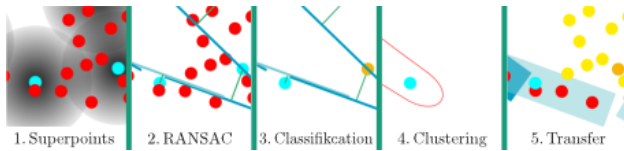

Figure 1. SiRP functionality

Superpoints can be thought of as flexible voxels. In a first step, the coordinates are discretized, which is analogous to the voxels. The grid size is an important hyperparameter. Afterward, all discretized points with the same coordinate are assigned to one superpoint. Its geometric coordinates are determined by the mean of the coordinates of points in the voxel.

The last averaging step is the core difference regarding a voxel grid. The averaging helps us achieve a better approximation to the point cloud that is not fixed to a grid. A comparison of both methods could be seen in Figure 2.
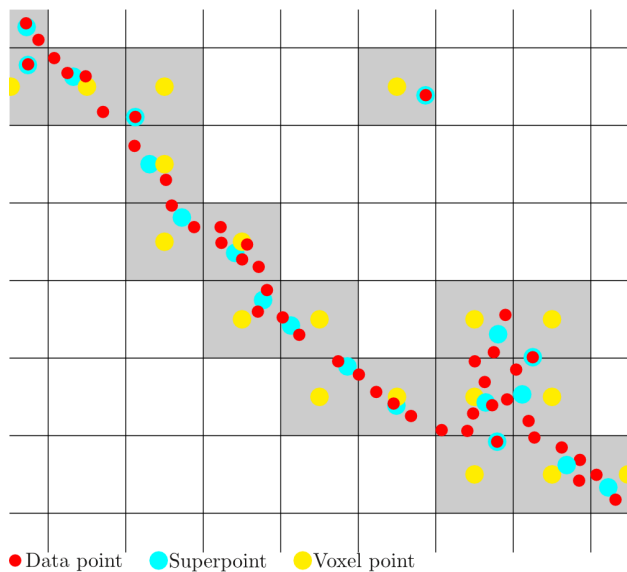

Figure 2. Point cloud approximation using voxels or superpoints

For each superpoint s, a RANSAC plane r is calculated. RANSAC is an iterative method to fit a geometry according to a point set, without influence of outliers. Therefore, we determine the nearest point cloud neighbors, also called support set, of the superpoint G(s) by a radius search (RNN). By this, we use the full data resolution and have few computation steps. With these, the RANSAC plane s is fitted as

$$r_s = \mathrm{RANSAC}(G(s, p))$$

For RANSAC, we calculate 100 planes and compute their parameters and inlier sets in parallel. 100 planes are a good tradeoff between computation time and accuracy. This number is found by tests. As a consequence, the computation time does *not* increase significantly with an increase or decrease of the superpixels density: higher density means many superpixels, but, at the same time, smaller support sets.

Based on the distance of the RANSAC plane, calculated by the nearest point cloud neighbors $G(s, p)$, according to the position of the superpoint $d(s, r_s)$, these are classified as manifold (= ground) or non-manifold. The distance threshold only depends on the superpoint resolution.

To remove superpoint outliers (small manifold-like segments, e.g. roofs) the ground superpoints are clustered by a further, greater distance value. All clusters containing too few number of superpoints are assigned as non-ground and removed.

Finally, the classification result of the superpoints is transferred to the original point cloud (Figure 3). The point criteria to classify as ground is to be within a specific number of ground superpoint RANSAC planes. Again, the tolerance threshold is automatically derived from of the superpoint resolution.
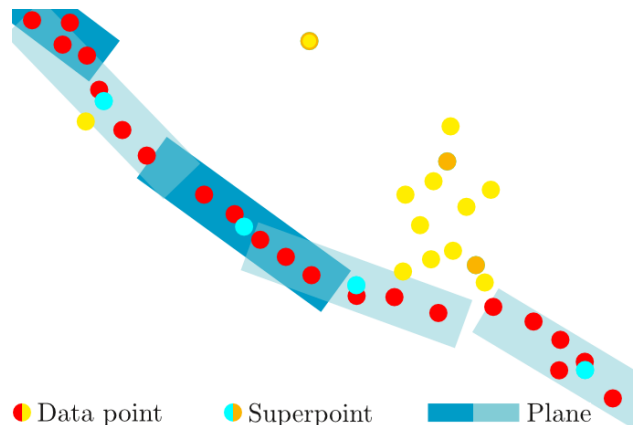

Figure 3. Transfer of superpoint classification to point cloud

4. RESULTS

4.1 Data description

The mobile data was acquired by using the Alpha3D MMS [1] in the Wuhan University campus, Hubei, China, at about 3 pm Nov 18 2021 (see Figure 4). The Alpha3d MMS consists of a single RIEGL VUX-1 laser scanner of a 360 ° field of view (FOV) and a 475 m maximum range, a GNSS/INS unit of a tactical-grade IMU of gyro's in-run bias stability 0.25 °/hr, and a spherical camera rig of six synchronized cameras each acquiring 5 MP images at 10 Hz. The point clouds captured by the laser scanner have a relative accuracy of 5 mm in the 100 m range. The MMS churns out up to 1.8 million points/s. All sensors were calibrated with the proprietary software shipped with the MMS before the data collection.

The complementary stationary data were collected using a Leica RTC360 SLS with a horizontal FOV of 360° and a vertical FOV of 300°. The measurement accuracy was 2.9 mm at the 20 m distance. The scans were captured in the "medium density" mode, with a point spacing of 6 mm at 10 m both horizontally and vertically.

Once the data is co–registered by the method in Section 3.1 (see Figure 4.2), two different point cloud fragments about $50 \times 40 \times 30$ m are chosen. These test segments were selected from several views of the Wuhan University campus along one main road, as shown by two white polygons in Figure 4. We

---

[1] https://chcnav.com/product-detail/alpha3d

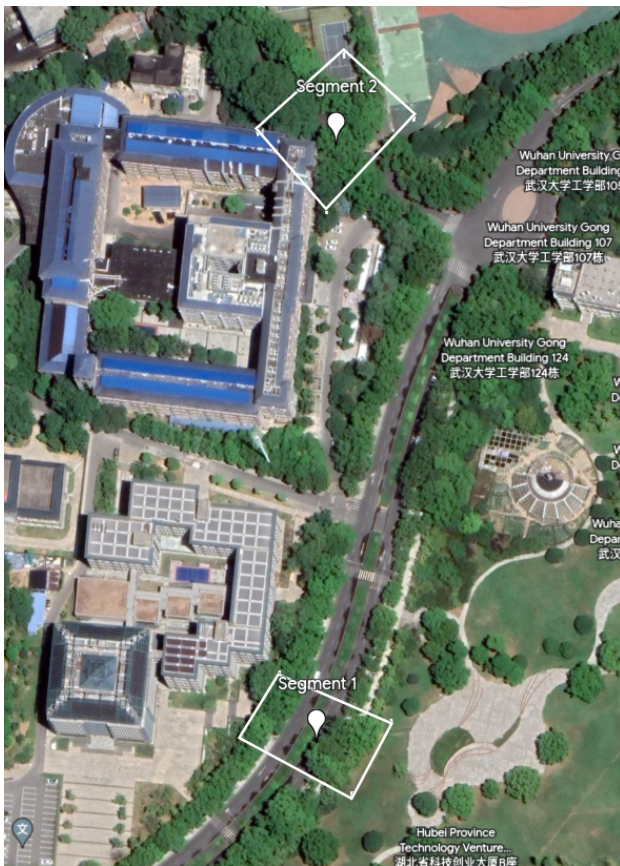(a) Segment 1   (b) Segment 2

Figure 5. Segments in RGB

Figure 4. The two test areas demarcated by white rectangles on Google Earth.

try to cover the variety of the whole point cloud in these two segments. One segment was in a scene of mainly road surface, trees, bushes, with a rectangular shape of an area 1668.0 m^2. In the other segment, the point clouds fall on road surface, trees, ironwire meshes, and buildings, with a rectangular shape of an area 2071.6 m^2. We refer to Figure 5 for illustration of the corresponding point clouds.

In both test areas, points were manually labeled with the point labeler [2] in order to evaluate the classification performance. The point labeler was originally created for the SemanticKITTI benchmark (Behley et al., 2019) to label point clouds. We adapted its interface to load our point cloud chunks. The tool supports up to 26 classes, but we collapsed the classes down to five (structures, vegetation, ground, moving agents, and others) since our primary goal is about ground and vegetation filtering. The structure class include buildings, lamps, traffic signs, wire fence, short posts, trash bins, guardrails. The vegetation class include trees, trunks, bushes, but excludes grassland. The agent class include pedestrians, cars, trucks, bicycles, motorcycles and riders. The ground class include road, sidewalk, parking lots, grassland, lane markings.

The reference labels were created by labeling points manually within the point labeler tool. At least one other person inspected the labeling and suggests corrections, per which the labels were retouched.

4.2 Registration

The SLS data were registered to the MLS data by the registration method in Section 3.1. For the test site 1, the point
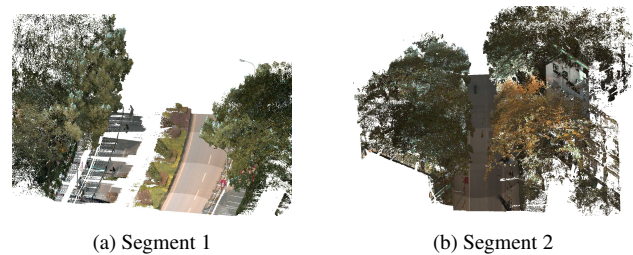
---

[2] `https://github.com/jbehley/point_labeler`

clouds before and after the registration are shown in Figure 6 which shows that the SLS and MLS point clouds were precisely aligned. By setting a distance threshold on inliers for registration, the moving objects and growing plants have negligible adverse effects on the registration as the lampposts and road curbs from the two data sources are well aligned.

(a) Segment 1
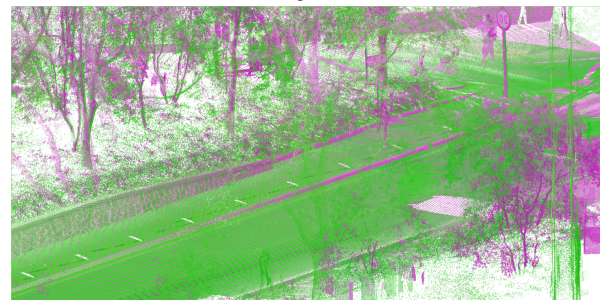
(b) Segment 2

Figure 6. Blended views of the registered green MLS data and pink SLS data; left before and right after the registration.

However, the median strip on the road is measured from the opposite side of the street, we have aligned it manually in figure 7 to compare the dive in of the laser pulse. We could clearly see that the SLS has a superior depth penetration, and it can measure many of the ground points below the vegetation. The depth penetration of the SLS is likely because that the SLS is closer to one side of the bushes and its scanning time is much longer than the MLS which moves along the road quickly at a median speed 5.55 m/s. We will examine in the next section the performance of SiRP in this challenging regions.

The accuracy of the registration is measured by the root-mean-square error (RMSE) of the inlier matches from the point-to-point ICP step. The inlier matches are those point pairs of a distance less than 0.1 m. The inlier RMSE is 0.161 m and 0.109 m for test site 1 and test site 2, respectively.

The seasonal change and gardening cause the vegetation points from the MLS and the SLS to not match precisely. This discrepancy actually has little bearing on the registration because
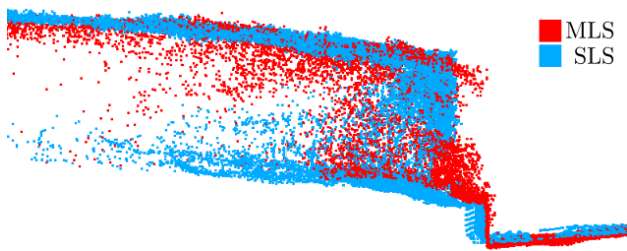
Figure 7. Comparison of planted median strip measured by MLS in red and SLS in blue.

the ICP in the refinement step uses closest point matches for registration and point matches of distances larger than a threshold (0.1 m in tests) are disregarded.

4.3 Classification

Two experiments were performed with the SiRP algorithm. The superpoint resolution and the minimum cluster size are set to 0.1 and 5000, respectively. First, we evaluate it using the MLS only. Second, we apply SiRP on the merged point cloud from mobile and stationary laser scans (MSLS), to prove the benefit of point cloud fusion. We start with the quantitative evaluation, which reflects both point-wise classification accuracy and also that of a homogenized reference point cloud, with equal point density. Then, we discuss qualitative results. Remarks on computation time conclude this section.

First, we look at the confusion matrices in Tables 1 to 4. The first two tables show the overall results for both test segments 1 and 2. The overall accuracy is roughly the same for MLS and MSLS point clouds. To highlight the improvement of additional scans and make classification output independent on the point density, we show the confusion matrices for the homogenized point cloud of the more challenging segment 1. These results are shown in the last two tables.

ground	yes	no	$\sum$	%
yes	2542398	120977	2663375	42.84
no	82156	3471851	3554007	57.16
$\sum$	2624554	3592828	6217382	
%	42.21	57.79		96.73

Table 1. Confusion matrix of MLS classification result for both segment 1 and 2; the rows depict the prediction, the columns the ground truth. Gray and green colors in the background depict correctly identified vegetation and ground points, respectively. Magenta and yellow denote misclassifications.

ground	yes	no	$\sum$	%
yes	2600518	134243	2734761	43.99
no	24036	3458585	3482621	56.01
$\sum$	2624554	3592828	6217382	
%	42.21	57.79		97.45

Table 2. Confusion matrix of MSLS classification result for both segment 1 and 2, see structure of Table 1 for further explanations. Since we use the same reference point cloud, the aggregated last row of the confusion matrix is the same as in Table 1.

Both overall results are around 97 %, which is very satisfying. The enhancement by adding more scans is marginal by means of the confusion matrices (Table 1 vs. 2). As we will see in our

qualitative evaluation, the reason for the small numerical improvement is the unbalanced point resolution. On the opposite road side from the MLS in segment 1, the resolution is much lower. If we balance the resolution for evaluation and compare the result only for the more challenging segment 1, the numerical improvement is clearer (Tables 3 and 4). The overall accuracy raises by 1.3 %. Concerning the recognition rate of correct labelled road points, it raises from 72% by mobile scanner to 94% by combining mobile and stationary laser scans.

ground	yes	no	$\sum$	%
yes	39921	10542	50463	7.68
no	15853	590510	606363	92.32
$\sum$	55774	601052	656826	
%	8.49	91.51		95.98

Table 3. Segment 1, Confusion matrix of homogenized MLS classification result. See structure of Table 1 for further explanations.

ground	yes	no	$\sum$	%
yes	52634	14826	67460	10.27
no	3140	586226	589366	89.73
$\sum$	55774	601052	656826	
%	8.49	91.51		97.26

Table 4. Segment 1, Confusion matrix of homogenized MSLS classification result; see structure of Table 1 for further explanations.

All tests for Tables 1 to 4 were carried out with the superpoint resolution 0.1 m, which turned out to be a quite good choice for this dataset and has the same order of magnitude as for the photogrammetric dataset in (Bulatov et al., 2021). Experiments with larger resolutions, up to about 1 m, perform increasingly better on recognizing correctly more street points, especially with sparse point cloud density in areas away from the scan. However, more vegetation points are spuriously being assigned to this ground class. This tendency happens to a larger extent for MLS than for MSLS. Even if the superpoint resolution is not optimally selected, the overall accuracy remains around 90 %.

To compare our result and rank its performance, we also run tests with CANUPO, the built-in classifier of CloudCompare. We needed some time to find a good set-up. Finally, we set the scales to [0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28] and use 1000 max core points. However, it was only possible to run the test with the MLS point cloud, because of the much lower number of points. To save labeling time and get the best case prediction, we use the labelled point cloud parts for training and classifying. The overall accuracy is 94 % which is worse than our result.

Since the results of the quantitative evaluation were quite accurate for all configurations, we present in Figures 8 and 9 the qualitative results for both test areas. The improved quality of the MSLS result could be clearly seen. Especially road points are better classified. In the homogenized case, vegetation has a strong overbalance against the road. However, areas of different resolution are equal weighted. This has the consequence, that the improvement by MSLS gets clearer.

In Figure 8 parts of street which are mislabelled as vegetation are mostly distant from the sensor platform and have less density and rough edges. But even parts are recognized as street, which is due to the superpoint generalization possibility. If the

superpoint grid size is increased, e.g., to 1 m, this part of road is classified better, at the cost of more vegetation being labelled as ground elsewhere. On the other side of the test segment, which is not shown in Figure 8, a footway exists and has a connection to the correctly labelled street, it is also completely marked correctly. Furthermore, it can be seen, that beside the improvement in the area of the road, the classification of objects on the road like bollards or guardrails is improved too. The Misclassification in vegetation remains, even though it gets better.

In the bottom half of Figure 9, many ground points are misclassified because of the low point density. In contrast, in the upper half, the MSLS result has more false vegetation points. These points are below bushes and can not clearly assign to vegetation or road. SiRP decides more differentiated if ground points below vegetation exist, because it seizes the opportunity to set the estimated plane to the ground and remove vegetation.
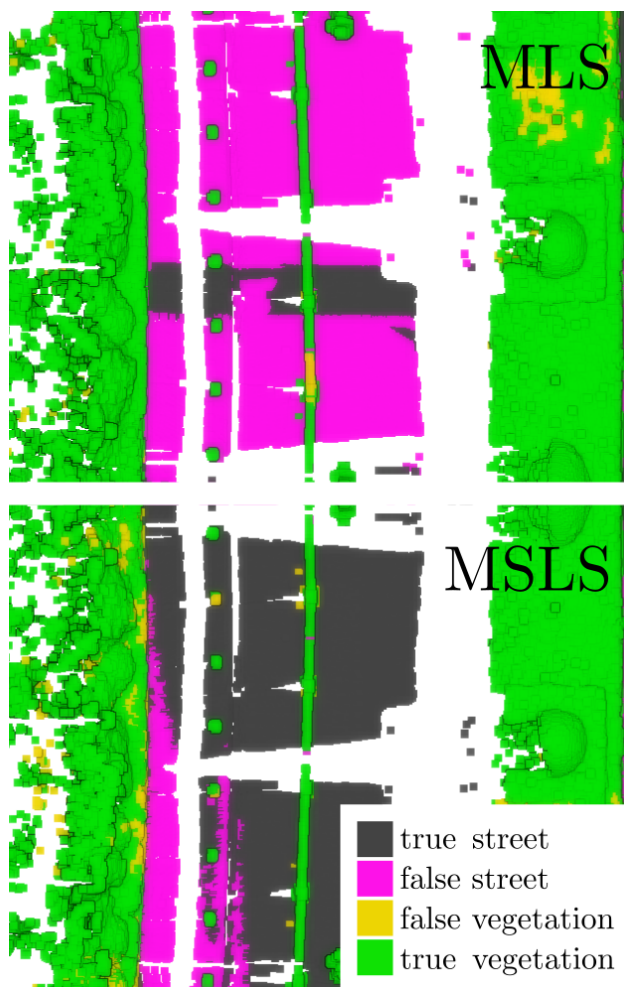

Figure 8. A bird-eye view of ground and low objects in segment 1. Both halves show the same point cloud part mirrored along the white middle axis. The colors are the same as in the confusion matrices in the previous chapter.

To summarize, challenging areas, such as ground below vegetation, are better classified in MSLS. If we have a sufficient density, SiRP could tell apart noisy vegetation and flat ground.

We show in Figure 10 the classification results of the whole point cloud described in Section 4.1. Overall, they look very plausible. There are some outliers visible, but they are not all
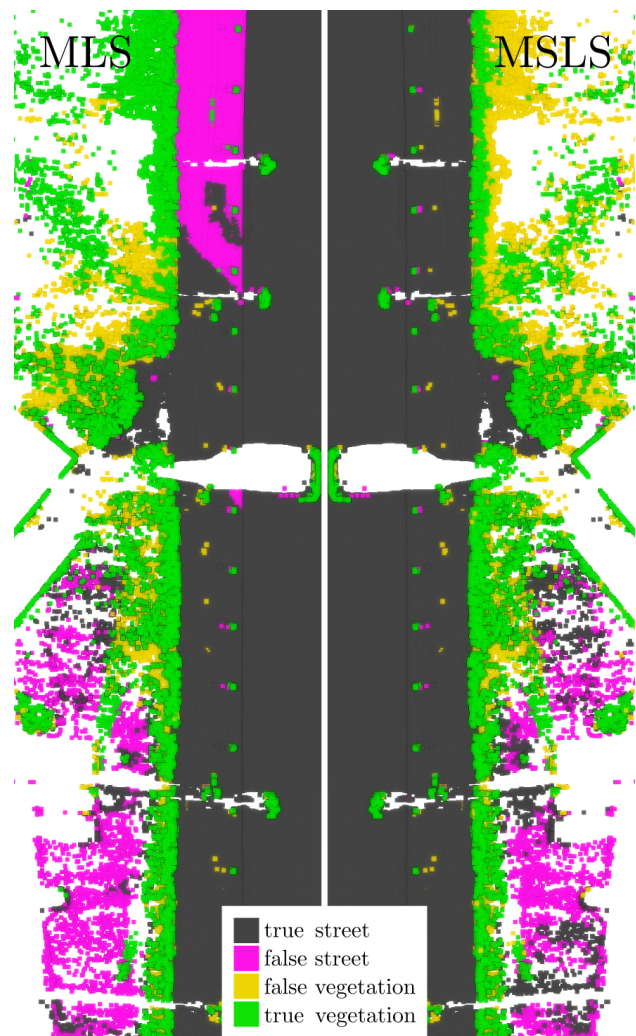

Figure 9. A bird-eye view of ground and low objects in segment 2.

misclassifications because SiRP tries to assign points lying far away, too, which also satisfies the ground conditions.

The far beam power of the three additional static laser scans is astonishing. Even not measured areas are improved. This can be seen by the blue areas, which are additionally classified as ground after adding these scans.

Computing time is about 300 µs per point, whereby all experiments were performed with an Intel® Xeon® Gold 6154 CPU, using multiprocessing. For the fused point cloud from Figure 10, which contains 122.5 million points, the total computing time was 11.5 hours. For comparison, the performance of CANUPO is about 37 µs/point on the same device. However, the test classified only 10 million points. The whole point cloud was not computable. Additionally, time for labelling has to be spent, ranging from 30 minutes to up to an hour.

5. CONCLUSION

We presented a modular workflow consisting of co-registering multi-modal 3D point clouds using hierarchical merging and ground filtering using the SiRP method from (Bulatov et al., 2020). The ground filtering method, originally developed for airborne point clouds, has several key parameters. For the terrestrial dataset, most of them were kept unchanged. The su-

perpoint grid size, the key parameter, could be chosen in an intuitive way, which is also an encouraging finding. Its range, in which pretty high values for overall accuracy could be obtained, is sufficiently broad. The decay in overall accuracy is, however, faster if only 3D points from MLS are used. This validates the benefit of additional point clouds from the TLS.

SiRP is known to be rotationally invariant and does not filter points with respect to the surface normal and color content. This partly explains a few remaining misclassifications, such as cropped bushes. While we present the results of SiRP without any kind of post-processing, we note that for applications of autonomous driving, these vegetable remnants are less critical than moving cars and pedestrians.

There are several immediately connected topics worth exploring in future work. Validation of the registration approach in a diverse set of large scenes, possibly with aerial point clouds fused in, would be similarly interesting in understanding the effect of sensor noise and registration errors on the classification results. With respect to classification, we look to explore deep-learning approaches for ground and vegetation filtering, especially, in the context of sparse training data.

REFERENCES

Bai, X., Luo, Z., Zhou, L., Fu, H., Quan, L., Tai, C.-L., 2020. D3Feat: Joint learning of dense detection and description of 3D local features. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Virtual, 6359–6367.

Behley, J., Garbade, M., Milioto, A., Quenzel, J., Behnke, S., Stachniss, C., Gall, J., 2019. SemanticKITTI: A dataset for semantic scene understanding of LiDAR sequences. *Proc. of the IEEE/CVF International Conf. on Computer Vision (ICCV)*.

Bernreiter, L., Ott, L., Nieto, J., Siegwart, R., Cadena, C., 2021. PHASER: A robust and correspondence-free global point cloud registration. *IEEE Robotics and Automation Letters*, 6(2), 855–862. Conference Name: IEEE Robotics and Automation Letters.

Besl, P., McKay, N. D., 1992. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 239–256.

Brovelli, M. A., Cannata, M., Longoni, U., 2002. Managing and processing LiDAR data within GRASS. *GRASS Users Conf.*, 29.

Bulatov, D., Stütz, D., Hacker, J., Weinmann, M., 2021. Classification of airborne 3D point clouds regarding separation of vegetation in complex environments. *Applied Optics*, 60(22), F6–F20.

Bulatov, D., Stütz, D., Lucks, L., Weinmann, M., 2020. Superpoints in RANSAC planes: A new approach for ground surface extraction exemplified on point classification and context-aware reconstruction. *Proc. Int. Conf. on Compute Graphics Theory and Applications*, 25–37.

Chen, X., Läbe, T., Milioto, A., Röhling, T., Behley, J., Stachniss, C., 2021. OverlapNet: A Siamese network for computing LiDAR scan similarity with applications to loop closing and localization. *Autonomous Robots*, 46(1), 61–81. https://link.springer.com/10.1007/s10514-021-09999-0.

Figure 10. Point cloud ground points filtered by SiRP. The corresponding aerial image is shown in Figure 4.

Chen, Z., Gao, B., Devereux, B., 2017. State-of-the-art: DTM generation using airborne LiDAR data. *MDPI Sensors*, 17(1), 150.

Deng, H., Birdal, T., Ilic, S., 2019. 3D local features for direct pairwise registration. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Long Beach, CA, USA, 3239–3248.

Deprelle, T., Groueix, T., Fisher, M., Kim, V., Russell, B., Aubry, M., 2019. Learning elementary structures for 3D shape generation and matching. *Advances in Neural Information Processing Systems*, 32.

Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J., Stilla, U., 2020. Registration of large-scale terrestrial laser scanner point clouds: A review

and benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 163, 327–342.

Dong, Z., Yang, B., Liang, F., Huang, R., Scherer, S., 2018. Hierarchical registration of unordered TLS point clouds based on binary shape context descriptor. *ISPRS Journal of Photogrammetry and Remote Sensing*, 144, 61–79.

Elmqvist, M., Jungert, E., Lantz, F., Persson, A., Söderman, U., 2001. Terrain modelling and analysis using laser scanner data. *Int. Arch. of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3/W4), 219–226.

Gevaert, C., Persello, C., Nex, F., Vosselman, G., 2018. A deep learning approach to DTM extraction from imagery using rule-based training labels. *ISPRS Journal of Photogrammetry and Remote Sensing*, 142, 106–123.

Honda, K., Koide, K., Yokozuka, M., Oishi, S., Banno, A., 2022. Generalized LOAM: LiDAR Odometry Estimation With Trainable Local Geometric Features. *IEEE Robotics and Automation Letters*, 7(4), 12459–12466.

Hu, X., Yuan, Y., 2016. Deep-learning-based classification for DTM extraction from ALS point cloud. *MDPI Remote Sensing*, 8(9), 730.

Hui, Z., Li, D., Jin, S., Ziggah, Y. Y., Wang, L., Hu, Y., 2019. Automatic DTM extraction from airborne LiDAR based on expectation-maximization. *Optics & Laser Technology*, 112, 43–55.

Jin, S., Su, Y., Zhao, X., Hu, T., Guo, Q., 2020. A point-based fully convolutional neural network for airborne LiDAR ground point filtering in forested environments. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 3958–3974.

Kraus, K., Pfeifer, N., 1998. Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Rem. Sensing*, 53(4), 193–203.

Li, J., Lee, G. H., 2019. USIP: Unsupervised stable interest point detection from 3D point clouds. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 361–370.

Lu, W., Wan, G., Zhou, Y., Fu, X., Yuan, P., Song, S., 2019. DeepVCP: An end-to-end deep neural network for point cloud registration. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, 12–21.

Mian, A., Bennamoun, M., Owens, R., 2010. On the repeatability and quality of keypoints for local feature-based 3D object retrieval from cluttered scenes. *International Journal of Computer Vision*, 89(2), 348–361. https://doi.org/10.1007/s11263-009-0296-z.

Mongus, D., Žalik, B., 2012. Parameter-free ground filtering of LiDAR data for automatic DTM generation. *ISPRS J. of Photogrammetry and Remote Sensing*, 67, 1 - 12.

Mousa, Y. A., Bulatov, D., Abed, F. M., Helmholz, P., 2021. Dtm extraction and building detection in dsms having large holes. *Remote Sensing Technologies and Applications in Urban Environments VI*, 11864, SPIE, 77–85.

Rabbani, T., Van Den Heuvel, F., 2005. Efficient hough transform for automatic detection of cylinders in point clouds. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3, 60–65.

Rizaldy, A., Persello, C., Gevaert, C., Oude Elberink, S., 2018. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4(2).

Rusinkiewicz, S., Levoy, M., 2001. Efficient variants of the icp algorithm. *Proceedings third international conference on 3-D digital imaging and modeling*, IEEE, 145–152.

Rusu, R. B., Blodow, N., Beetz, M., 2009. Fast point feature histograms (FPFH) for 3D registration. *IEEE Int. Conf. on Robotics and Automation*, IEEE, 3212–3217.

Schnabel, R., Wahl, R., Klein, R., 2007. Efficient RANSAC for point-cloud shape detection. 26(2), 214–226.

Segal, A., Haehnel, D., Thrun, S., 2009. Generalized ICP. *Robotics: science and systems*, 2, Seattle, WA, 435. Issue: 4.

Sithole, G., Vosselman, G., 2001. Filtering of laser altimetry data using a slope adaptive filter. *Int. Arch. of Photogrammetry Remote Sensing and Spatial Information Sciences*, 34(3/W4), 203–210.

Sithole, G., Vosselman, G., 2004. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS J. of Photogrammetry and Remote Sensing*, 59(1), 85 - 101.

Tsin, Y., Kanade, T., 2004. A correlation-based approach to robust point set registration. T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, O. Nierstrasz, C. Pandu Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Y. Vardi, G. Weikum, T. Pajdla, J. Matas (eds), *Computer Vision - ECCV 2004*, 3023, Springer Berlin Heidelberg, Berlin, Heidelberg, 558–569. Series Title: Lecture Notes in Computer Science.

Umeyama, S., 1991. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4), 376–380. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Vosselman, G., 2000. Slope based filtering of laser altimetry data. *Int. Arch. of Photogrammetry, Rem. Sensing and Spacial Information sciences*, 33(B3/2; PART 3), 935–942.

Yang, B., Li, J., 2022. A hierarchical approach for refining point cloud quality of a low cost UAV LiDAR system in the urban environment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 183, 403–421.

Zang, Y., Meng, F., Lindenbergh, R., Truong-Hong, L., Li, B., 2021. Deep localization of static scans in mobile mapping point clouds. *Remote Sensing*, 13(2), 219.

Zhang, K., Chen, S.-C., Whitman, D., Shyu, M.-L., Yan, J., Zhang, C., 2003. A progressive morphological filter for removing nonground measurements from airborne LiDAR data. *IEEE Transactions on Geoscience and Remote Sensing*, 41(4), 872–882.