GENERATION OF GROUND TRUTH DATASETS FOR THE ANALYSIS OF 3D POINT CLOUDS IN URBAN SCENES ACQUIRED VIA DIFFERENT SENSORS

Y. Xu[1,*], Z. Sun[1], R. Boerner[1], T. Koch[2], L. Hoegner[1], U. Stilla[1]

[1] Photogrammetry and Remote Sensing, Technische Universität München, 80333 Munich, Germany
[2] Remote Sensing Technology, Technische Universität München, 80333 Munich, Germany
-(yusheng.xu, zhenghao.sun, richard.boerner, tobias.koch, ludwig.hoegner, stilla)@tum.de

Commission III, WG III/6

KEY WORDS: Different sensors, Point clouds, Multi-resolution voxel structure, 3D space labeling

ABSTRACT:

In this work, we report a novel way of generating ground truth dataset for analyzing point cloud from different sensors and the validation of algorithms. Instead of directly labeling large amount of 3D points requiring time consuming manual work, a multi-resolution 3D voxel grid for the testing site is generated. Then, with the help of a set of basic labeled points from the reference dataset, we can generate a 3D labeled space of the entire testing site with different resolutions. Specifically, an octree-based voxel structure is applied to voxelize the annotated reference point cloud, by which all the points are organized by 3D grids of multi-resolutions. When automatically annotating the new testing point clouds, a voting based approach is adopted to the labeled points within multiple resolution voxels, in order to assign a semantic label to the 3D space represented by the voxel. Lastly, robust line- and plane-based fast registration methods are developed for aligning point clouds obtained via various sensors. Benefiting from the labeled 3D spatial information, we can easily create new annotated 3D point clouds of different sensors of the same scene directly by considering the corresponding labels of 3D space the points located, which would be convenient for the validation and evaluation of algorithms related to point cloud interpretation and semantic segmentation.

1. INTRODUCTION

In the past decade, the automatic 3D scene analysis using point clouds has attracted increasing attentions in research fields of photogrammetry (Vosselman and Maas, 2010), remote sensing (Lefsky et al., 1999), computer vision (Buch et al., 2011), and robotics (Rusu et al., 2009). For the dense and accurate 3D scene analysis and interpretation, especially in the context of urban areas, plenty of algorithms and approaches have been developed for a wide variety of applications such as semantic interpretation (Weinmann et al., 2015, Landrieu and Simonovsky, 2017, Vosselman et al., 2017), segmentation (Rabbani et al., 2006, Vo et al., 2015), registration (Aiger et al., 2008, Yang and Zang, 2014, Ge and Wunderlich, 2016, Theiler et al., 2014), object recognition (Schnabel et al., 2007, Yao et al., 2011, Niemeyer et al., 2014, Aldoma et al., 2012, Yu et al., 2016). For any proposed algorithms and methods, satisfying experiments and convincing evaluations are always non-trivial and crucial steps to validate the feasibility and performance of the proposed method. To conduct such experiments and evaluations, the benchmark dataset or the ground truth are normally required. In fact, lots of efforts have been paid to the generation of the benchmark dataset in the community of point clouds processing, for example, the ISPRS Benchmark Test on Urban Object Detection and Reconstruction, containing challenging aerial laser scanning point clouds for 3D object reconstruction (Rottensteiner et al., 2014), semantic 3D benchmark for classification having large-scale terrestrial point clouds of various urban, suburban, and rural scene (Hackel et al., 2017).

However, for the majority of the benchmark that published, they only focus on data obtained from one type of sensor, for example photogrammetric or LiDAR point clouds. However, for test-

ing the generality, in many cases we need to test the proposed method on datasets acquired by more than one kind of sensors at the same scene, for example, the LiDAR points and TomoSAR points. Thus, for the testing scene, if the ground truth can be used for automatically annotating new datasets acquired from other sensors, it will contribute a lot to the validation of the generality of the proposed methods, and provide an possibility of fusing the datasets of multiple sensors. To this end, we present a novel strategy for generating annotated dataset for new point clouds from different sensors based on the labeled 3D grid structure. Instead of directly labeling the 3D points with time consuming manual work, multi-resolution 3D voxel grids for the testing site are generated. On the basis of a set of basic labeled reference points (i.e., point clouds with low density and easier to be stored and manually annotated), we can generate a 3D labeled space of the entire testing site with different resolutions. To be specific, the voxel structure is applied to voxelize the annotated reference point cloud (see Fig. 1b), by which all the points are organized by a 3D grid. Then, a voting based strategy is adopted to the voxels of multiple resolution levels, in order to assign a semantic label to the 3D space represented by the voxel. Lastly, robust line- and plane-based fast registration methods are developed for aligning point clouds obtained via various sensors. Benefiting from the labeled 3D spatial information, we can easily create new annotated 3D point clouds of different sensors of the same scene directly by considering the corresponding labels of 3D space the points located, which would be convenient for the validation and evaluation of algorithms related to point cloud interpretation and semantic segmentation. In Fig. 1c, we illustrate an example of the reference MLS point cloud we used.

---

*Corresponding author

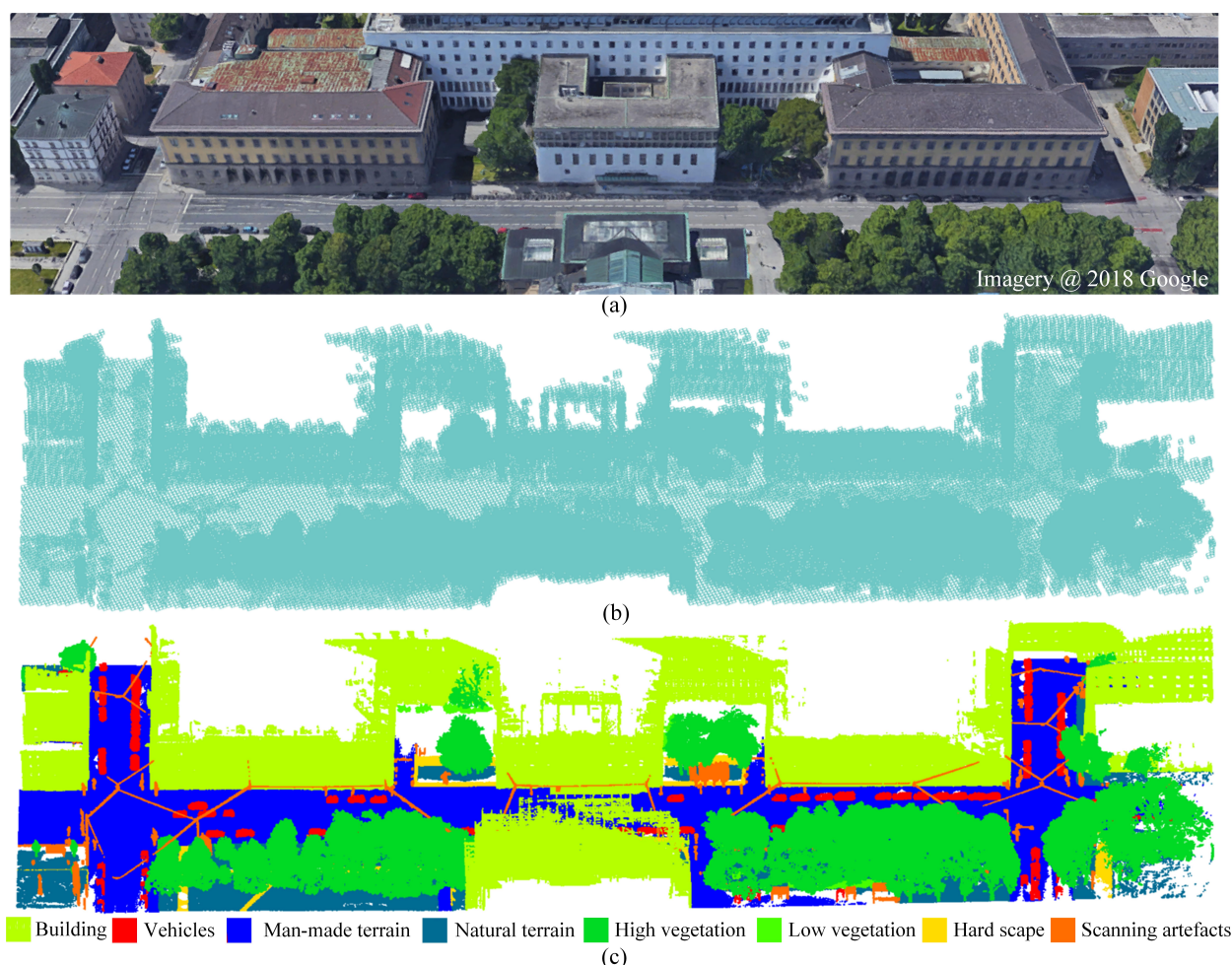| ■ Building | ■ Vehicles | ■ Man-made terrain | ■ Natural terrain | ■ High vegetation | ■ Low vegetation | ■ Hard scape | ■ Scanning artefacts |

(c)

Figure 1. (a) Real scene of the TUM main entrance from Google Maps, 2018. (b) Voxelized 3D space. (c) Labeled reference MLS point cloud.

2. METHODOLOGY

Conceptually, the implementation of our proposed method consists of three core steps: the voxelization of the reference dataset and the extraction of primitives, the primitive-based registration, and the voxel-based 3D labeling. In the first step, the reference data set is manually labeled and voxelized into the 3D grid structure with cubics. Simultaneously, geometric primitives (e.g., lines or planes) are extracted from the reference dataset and the testing point cloud, respectively. Combining the labeled reference dataset and the voxel structure, the 3D space covering the testing dataset can be labeled. In the second step, the primitive-based registration is conducted between the reference dataset and the testing point cloud, in order to align them into a same coordinate frame. In the last step, the multi-scale 3D labeled voxel space is applied to the aligned testing point clouds, labels of the points in the testing point cloud are assigned via the multi-scale voting strategy. The processing workflow is sketched in Fig. 2, with the key steps of involved methods and sample results illustrated. The detailed explanation of each step will be introduced in the following sections.

2.1 Annotation of the reference dataset

The annotation of the reference dataset is a crucial step in our work. Normally, the acquired measurements (e.g., point cloud) have a large amount of points, which is a challenge for the manual

annotation. However, in our method, it is not necessary to label all the points of the reference dataset. Instead, only those points representing or covering the 3D space should be labeled. Thus, in fact, there is no limitation for the type of reference dataset (e.g., point clouds, mesh, or BIM model). In our case, the original reference dataset (i.e., MLS point cloud) is considerably downsampled, in order to facilitate the manual labeling.

2.2 Voxelization and geometric primitive extraction

2.2.1 Octree based voxelization For labeling the 3D space covering the reference dataset, the reference dataset is voxelized with multiple resolutions. Here, we utilize the octree-based voxelization to decompose the entire point cloud with 3D cubic grids. As discussed in many of our former studies (Boerner et al., 2017, Xu et al., 2017c), the main reason of using the octree structure is that the nodes of an octree structure have explicit linking relations, which facilitate the traversal for searching the adjacent ones (Vo et al., 2015, Xu et al., 2017b). Besides, benefiting from the structure of the octree, it is also quite easier for us to generate the multi-resolution voxel spaces. In Fig. 3, we provide an illustration about the generated 3D labeled voxel spaces of different resolutions.

2.2.2 Primitive extraction To label the testing point cloud with the given labeled 3D space, it is needed to align the coordinate frame of the testing point cloud to that of the labeled 3D
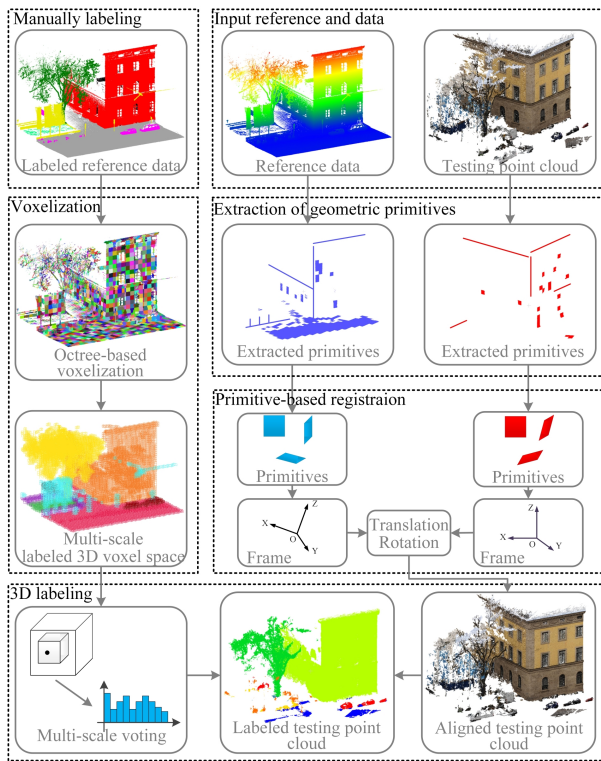
Figure 2. Workflow of generating new labeled point cloud by using labeled 3D space.
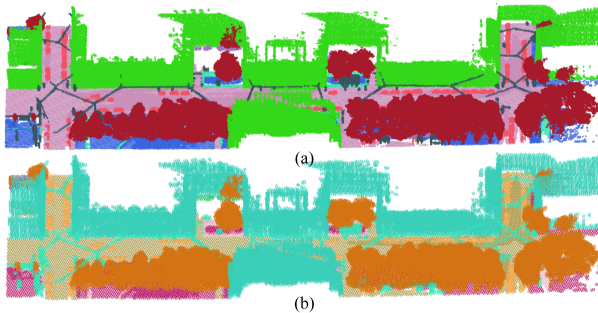

Figure 3. (a) and (b) Labeled 3D space with resolutions of 0.5 m and 1.0 m.

space. However, since they are normally different types of dataset and obtained via various sensors. Thus, traditional point-based registration algorithms may work not so well. Considering that in the same scene, the basic geometric structures of the buildings are always consistent (Boerner et al., 2018), we developed the strategy of using geometric primitives to align these two coordinate frames. Here, the primitives we used include lines (Koch et al., 2016) and planes (Xu et al., 2017a)

The ways of extracting 3D lines have been widely reported. For example, the 3D lines can be reconstructed from a set of images using the method of (Hofer et al., 2017) or (Jain et al., 2010). While for extracting 3D lines from LiDAR point clouds, methods are reported in (Lin et al., 2017) or (Hackel et al., 2016), aim at extracting boundaries and contours of the point cloud. As for the extraction of planes, there are also plenty of work like (Xu et al., 2017a, Nguyen et al., 2017, Dong et al., 2018). The benefits of using 3D lines and planes lie on less heavy computational cost than points, because such primitives always represent important structures of the environment for finding correspondences (esp.
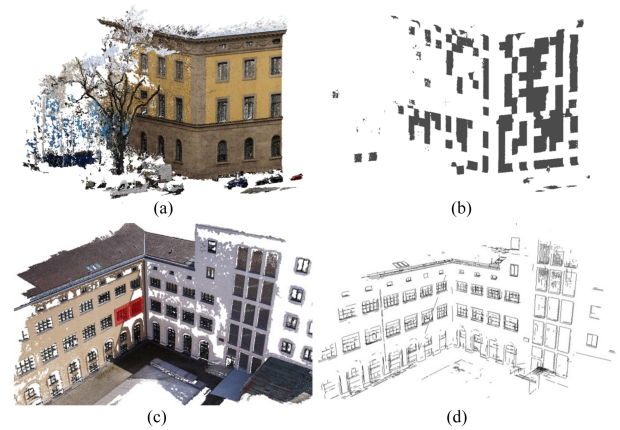

Figure 4. (a) Point cloud and (b) extracted planar primitives.(c) Point cloud and (d) corresponding linear primitives.

in urban scenes, which are rich of straight edges), which can reduce the number of candidates. Besides, primitives have more dimensions for limiting the degree of freedom which can make the estimation of transformation more robust. Especially for the case of urban scenes, buildings have flat facades, the registration is easier in 2D if corresponding planar structures can be found. In Fig. 4, an illustration of the extracted lines and planes from the corresponding photogrammetric point clouds are given.

2.3 Primitive-based registration

The registration of the reference and test point clouds can be expressed by a simple 3D similarity transformation $\mathbf{T} = (\mathbf{T}_r, \mathbf{T}_t, s)$, with $\mathbf{T}_t$, $\mathbf{T}_r$ and s defining the parameters of the transformation as a 3D translation vector, a 3×3 rotation matrix and a scale. In the following, we will briefly introduce the line- and plane-based registration methods.

2.3.1 Line-based registration The 3D line-based registration follows a modification of a methodology for aligning individual indoor and outdoor 3D building models (Koch et al., 2016). Unlike aligning non-overlapping image-based 3D reconstructions from different views, the 3D models to be aligned are captured from different sensors but represent identical structures from similar viewpoints. After extracting 3D line segments from LiDAR points $L_L = \{l_L^1, ..., l_L^n\}$ and images $L_I = \{l_I^1, ..., l_I^m\}$ as explained in Section 2.2.2, the transformation $\mathbf{T}$ aligning L_L to L_I can be estimated by minimizing perpendicular distances between k corresponding 3D line segments of both line sets

$$\mathbf{T} = \arg\min_T \sum_{i=1}^{k} d\left(l_I^i, \pi\left(l_L^i, \hat{\mathbf{T}}\right)\right), \qquad (1)$$

where $\pi\left(l, \hat{\mathbf{T}}\right)$ projects a line segment l with $\hat{\mathbf{T}}$, and $d(l_I, l_L)$ computes the length of the perpendicular of two 3D line segments extended to infinity.

In order to find line correspondences, planar characteristics of urban environments are exploited by first extracting a set of 3D planes in both line sets in a RANSAC-based scheme. These 3D planes mainly contain 3D lines representing façade boundaries and openings like window and door frames, which are suitable geometrical matching structures. For each detected 3D plane, corresponding inlier 3D lines are projected to these plane hypotheses for generating 2D binary images. The matching step is

finally conducted in 2D, by exhaustively comparing all 3D plane hypotheses using chamfer matching on all 2D binary images. Corresponding 3D planes share identical geometrical structures and therefore result in a low chamfer score. The identification of corresponding 3D plane hypotheses solves the 3D rotation, while the 3D translation is given by the result of the binary matching. This prior transformation can be applied to L_L and corresponding 3D lines in L_I can be detected in the 3D space by a simple nearest neighbor search. These 3D line matches are finally used to refine the transformation $\mathbf{T}$ using Equation 1.

2.3.2 Plane-based registration The 3D plane-based registration is a modification of the methodology for aligning 3D building scenes in urban areas (Xu et al., 2017a). The matching between corresponding plane sets is conducted by finding a triple of planes, forming a 3D corner in the urban scene, which is used as a constraint to define a coordinate frame determining six degrees of freedom. Normal vectors of three planes forming the corner are firstly calculated. Afterwards, the coordinate frame defined by this corner is estimated by a cross product of normal vectors. Based on a pair of corners from different scans, we can define the transformation to align one scan to another, which consists of the estimation of rotations and the determination of translations.

As stated in (Xu et al., 2017a), in theory, if we can find the corresponding corners standing for the corresponding plane sets, the correct transformation can be calculated. For this purpose, we assume that if the majority of the planes in one scan can be matched to another scan, the transformation parameters are estimated optimally. Similar to the line-based registration, a RANSAC-based strategy is adopted, sampling a triple of planes from two scans concurrently in each random selection round. These planes should be non-parallel and their corresponding coordinate frame will be estimated. In RANSAC process, we apply the estimated transformation parameters to transform all the selected planes in the target scan, and count the number of matched planes. The triple of planes having the largest number of matched planes are identified as the optimal corresponding plane sets, and used for estimating the final transformation parameters. For rotations, the angle differences α, β, and γ of the rotation matrix $\mathbf{T}_r$ are calculated by comparing the vectors of axis in corresponding coordinate frames F_s and F_t. Then, is obtained as follows:

$$\mathbf{T}_r = F_s{}^{-1} * F_t \qquad (2)$$

For the translation, the intersecting points $\vec{P}$ of triples of planar patches are used as the reference:

$$\vec{P} = (|\vec{N_x} \quad \vec{N_y} \quad \vec{N_z}|)^{-1}[(\vec{X_x} \cdot \vec{N_x})(\vec{N_y} \times \vec{N_z}) + \\ (\vec{X_y} \cdot \vec{N_y})(\vec{N_z} \times \vec{N_x}) + (\vec{X_z} \cdot \vec{N_z})(\vec{N_x} \times \vec{N_y})] \qquad (3)$$

The translation matrix $\mathbf{T}_t$ is calculated as follows:

$$\mathbf{T}_t = \vec{P_t} - \vec{P_s} \qquad (4)$$

By applying $\mathbf{T}_r$ and $\mathbf{T}_t$ to coordinate frame of the testing dataset can be aligned to the reference dataset.

2.4 Voxel-based 3D labeling

To avoid directly labeling 3D points of testing point clouds, the annotated 3D space with labels is used. More specifically, the 3D voxel-grids with different resolutions are generated with merely a subset of the point cloud in order to further label the testing dataset. Here, a simple voting approach is applied to label the

voxels in the grid with only basic labeled points. Multiple resolution can be achieved by means of changing parameter settings the division levels in the process of octree based voxelization.

For giving labels to the points of the aligned testing point clouds with the labeled 3D space, the position of the testing point is considered. To be specific, if the point is located in a voxel with a given label k, then this point will be tagged with label k. For the labeled 3D space with n multiple resolutions, each testing point will finally get maximum n different tags. These n tags will conduct a voting process, the tag having the highest voting score will be assigned to the point. Here, the use of multiple resolution voxel space is designed to deal with the ambiguity caused by the boundaries and connections between different structural components. In theory, the smaller the voxel, more accurate the assignment of testing points will be. Howerver, this requires a detailed manual labels, which increase the workload definitely. The use of multi-resolution strategy can make a good balance between the preservation of details and the accuracy of the tags giving to testing points.

3. EXPERIMENTAL RESULTS

The testing site is the Arcisstrasse along the main entrance of Technische Universität München (TUM) city campus, covering approximately an area of 29000 m^2. In Fig. 1a, we provide an overview of the entire scene. To some extent, this scene is a representative scenario of the urban area, including rich information of buildings, vehicles, vegetations, ground surfaces, et al. Here, the basic point cloud is original acquired by Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) (Gehrung et al., 2017). The utilized point clouds are acquired by two Velodyne HDL-64E mounted with an angle of $35°$ on the front roof of the van. Fig. 5 provides sketch about how the two scanners are mounted (Gehrung et al., 2017). The original raw point clouds are also preprocessed by a statistical outlier removal for down-sampling and noise suppressing. The raw point cloud is downsampled from 0.5 billion points to 20 million points, in order to facilitate the manually labeling work. For testing the per-
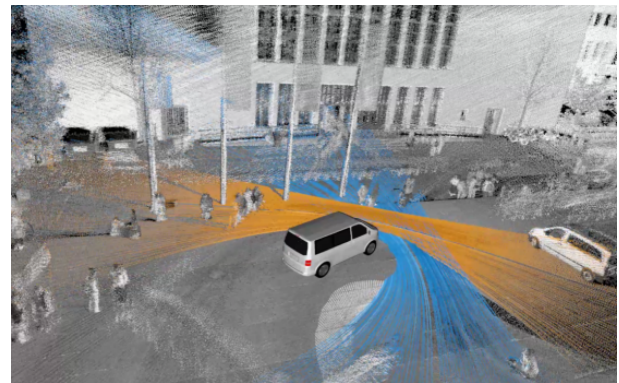

Figure 5. Two oblique mounted laser scanners of the MLS system. Figure courtesy of (Gehrrng et al., 2017).

formance of automatically labeling new point clouds acquired by other sensors, we also generate photogrammetric point clouds using hand held camera with 500 images via stereo matching based on the method depicted in (Xu et al., 2018). This photogrammetric point cloud will be annotated by the use of our 3D labeled space. In Fig. 6, we provide the labeling result of the photogrammetric point cloud. Comparing with the RGB textured original point cloud, it is clear that the majority part of this point clouds

(a)

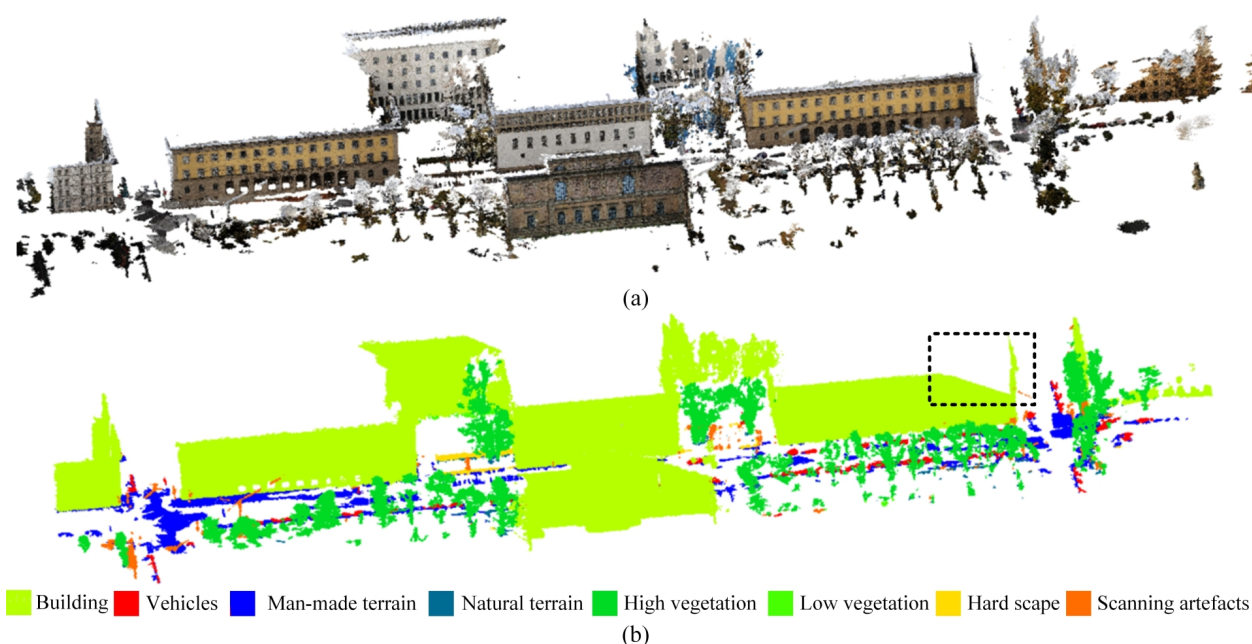| ⬛ Building | 🟥 Vehicles | 🟦 Man-made terrain | 🟦 Natural terrain | 🟩 High vegetation | 🟩 Low vegetation | 🟨 Hard scape | 🟧 Scanning artefacts |

(b)

Figure 6. (a) Original testing point cloud and (b) Labeled testing point cloud.

has been correctly labeled. However, there are still some part wrongly labeled or missed. For example, the area marked by the black dash box is missing in the labeled results. The original photogrammetric point cloud includes around 33 million points, but after the labeling only 30 millions point corrected labeled. The reason of this error is due to the error caused by the registration process, which will cause a bias of the position of the 3D voxel, so that the testing points cannot fall into correct voxels or locate outside the voxel space. In addition, our generated 3D labeled can also be used to generate annotated MLS point clouds with ultra high density. In Fig. 7, we display an illustration of the annotated original MLS point clouds. In this annotated dataset, the points of each scan have been labeled individually. This work is collaborated with Fraunhofer-IOSB, and currently this dataset is already published. In this case, the voxelized 3D space has fixed resolution and the resolution of the voxel is 0.2m. Therefore, at the connections between the ground surface and facades, there are still some obvious "zig-zag" edges, which can be further improved by the use of multi-resolution strategy.
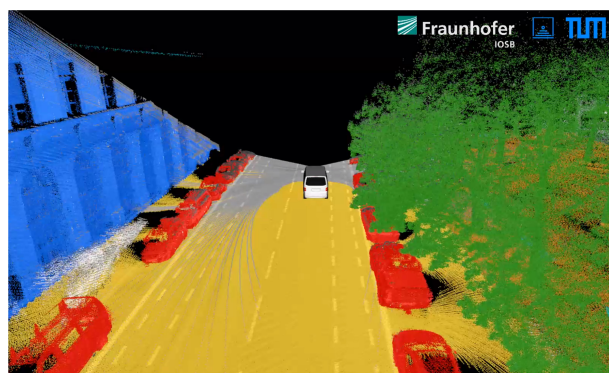

Figure 7. Labeled each scan of the MLS dataset.

4. CONCLUSION

In this work, we proposed a novel strategy of automatically generating ground truth dataset for analyzing point cloud from different sensors and validation of algorithms. Without directly labeling the 3D points, a multi-resolution 3D voxel grid for the testing site is generated, with 3D the entire 3D space labeled based on a simple annotated basic point cloud. Benefiting from the labeled 3D spatial information, we can easily create new annotated 3D point clouds of different sensors of the same scene directly considering the corresponding labels of 3D space where points located, which can provide benchmark datasets of various sensors for the validation and evaluation of algorithms related to point cloud interpretation. Moreover, our proposed voting based approach for assigning labels to the 3D space can also be used for the purpose of fusing datasets of multiple sensors, which is also a promising research topic of the community of remote sensing(Schmitt and Zhu, 2016).

REFERENCES

Aiger, D., Mitra, N. J. and Cohen-Or, D., 2008. 4-points congruent sets for robust pairwise surface registration. *ACM Transactions on Graphics (TOG)* 27(3), pp. 85.

Aldoma, A., Marton, Z.-C., Tombari, F., Wohlkinger, W., Potthast, C., Zeisl, B., Rusu, R. B., Gedikli, S. and Vincze, M., 2012. Tutorial: Point cloud library: Three-dimensional object recognition and 6 dof pose estimation. *IEEE Robotics & Automation Magazine* 19(3), pp. 80–91.

Boerner, R., Hoegner, L. and Stilla, U., 2017. Voxel based segmentation of large airborne topobathymetric lidar data. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLII-1/W1, pp. 107–114.

Boerner, R., Xu, Y., Hoegner, L. and Stilla, U., 2018. Registrierung von flugzeug getragenen kameraaufnahmen und uav-aufnahmen zur anreicherung von 3d daten. In: *38. Wissenschaftlich-Technische Jahrestagung der DGPF*, pp. 76–85.

Buch, N., Velastin, S. A. and Orwell, J., 2011. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems* 12(3), pp. 920–939.

Dong, Z., Yang, B., Hu, P. and Scherer, S., 2018. An efficient global energy optimization approach for robust 3d plane segmentation of point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing* 137, pp. 112–133.

Ge, X. and Wunderlich, T., 2016. Surface-based matching of 3d point clouds with variable coordinates in source and target system. *ISPRS Journal of Photogrammetry and Remote Sensing* 111, pp. 1–12.

Gehrung, J., Hebel, M., Arens, M. and Stilla, U., 2017. An approach to extract moving objects from mls data using a volumetric background representation. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Hackel, T., Savinov, N., Ladicky, L., Wegner, J. D., Schindler, K. and Pollefeys, M., 2017. SEMANTIC3D.NET: A new large-scale point cloud classification benchmark. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. IV-1-W1, pp. 91–98.

Hackel, T., Wegner, J. D. and Schindler, K., 2016. Contour detection in unstructured 3d point clouds. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1610–1618.

Hofer, M., Maurer, M. and Bischof, H., 2017. Efficient 3d scene abstraction using line segments. *Computer Vision and Image Understanding* 157, pp. 167–178.

Jain, A., Kurz, C., Thormaehlen, T. and Seidel, H., 2010. Exploiting global connectivity constraints for reconstruction of 3d line segments from images. In: *International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Koch, T., Körner, M. and Fraundorfer, F., 2016. Automatic alignment of indoor and outdoor building models using 3d line segments. *Proceedings of Computer Vision and Pattern Recognition 2016* pp. 10–18.

Landrieu, L. and Simonovsky, M., 2017. Large-scale point cloud semantic segmentation with superpoint graphs. *arXiv preprint arXiv:1711.09869*.

Lefsky, M. A., Cohen, W., Acker, S., Parker, G. G., Spies, T. and Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests. *Remote sensing of environment* 70(3), pp. 339–361.

Lin, Y., Wang, C., Chen, B., Zai, D. and Li, J., 2017. Facet segmentation-based line segment extraction for large-scale point clouds. *IEEE Transactions on Geoscience and Remote Sensing* 55(9), pp. 4839–4854.

Nguyen, H. L., Belton, D. and Helmholz, P., 2017. A comparative study of automatic plane fitting registration for mls sparse point clouds with different plane segmentation methods. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*.

Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing* 87, pp. 152–165.

Rabbani, T., Van Den Heuvel, F. and Vosselmann, G., 2006. Segmentation of point clouds using smoothness constraint. *International archives of photogrammetry, remote sensing and spatial information sciences* 36(5), pp. 248–253.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U. and Jung, J., 2014. Results of the isprs benchmark on urban object detection and 3d building reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing* 93, pp. 256–271.

Rusu, R. B., Blodow, N. and Beetz, M., 2009. Fast point feature histograms (fpfh) for 3d registration. In: *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, IEEE, pp. 3212–3217.

Schmitt, M. and Zhu, X. X., 2016. Data fusion and remote sensing: An ever-growing relationship. *IEEE Geoscience and Remote Sensing Magazine* 4(4), pp. 6–23.

Schnabel, R., Wahl, R. and Klein, R., 2007. Efficient ransac for point-cloud shape detection. In: *Computer Graphics Forum*, Vol. 26number 2, Wiley Online Library, pp. 214–226.

Theiler, P. W., Wegner, J. D. and Schindler, K., 2014. Keypoint-based 4-points congruent sets–automated marker-less registration of laser scans. *ISPRS Journal of Photogrammetry and Remote Sensing* 96, pp. 149–163.

Vo, A.-V., Truong-Hong, L., Laefer, D. F. and Bertolotto, M., 2015. Octree-based region growing for point cloud segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing* 104, pp. 88–100.

Vosselman, G. and Maas, H.-G., 2010. *Airborne and terrestrial laser scanning*. CRC Press.

Vosselman, G., Coenen, M. and Rottensteiner, F., 2017. Contextual segment-based classification of airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing* 128, pp. 354–371.

Weinmann, M., Jutzi, B., Hinz, S. and Mallet, C., 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 286–304.

Xu, Y., Boerner, R., Yao, W., Hoegner, L. and Stilla, U., 2017a. Automated coarse registration of point clouds in 3d urban scenes using voxel based plane constraint. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 4, pp. 185.

Xu, Y., Hoegner, L., Tuttas, S. and Stilla, U., 2017b. Voxel- and graph-based point cloud segmentation of 3d scenes using perceptual grouping laws. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1/W1, pp. 43–50.

Xu, Y., Tuttas, S., Hoegner, L. and Stilla, U., 2017c. Geometric primitive extraction from point clouds of construction sites using vgs. *IEEE Geoscience and Remote Sensing Letters* 14(3), pp. 424–428.

Xu, Y., Tuttas, S., Hoegner, L. and Stilla, U., 2018. Voxel-based segmentation of 3d point clouds from construction sites using a probabilistic connectivity model. *Pattern Recognition Letters* 102, pp. 67–74.

Yang, B. and Zang, Y., 2014. Automated registration of dense terrestrial laser-scanning point clouds using curves. *ISPRS Journal of Photogrammetry and Remote Sensing* 95, pp. 109–121.

Yao, W., Hinz, S. and Stilla, U., 2011. Extraction and motion estimation of vehicles in single-pass airborne lidar data towards urban traffic analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* 66(3), pp. 260–271.

Yu, Y., Li, J., Wen, C., Guan, H., Luo, H. and Wang, C., 2016. Bag-of-visual-phrases and hierarchical deep models for traffic sign detection and recognition in mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 113, pp. 106–123.