https://doi.org/10.2352/EI.2023.35.17.3DIA-105

DL-Based Floor Plan Generation from Noisy Point Clouds

Xin Liu, Egor Bondarev, Peter H.N. de With
Eindhoven University of Technology, SPS-VCA group of Electr. Eng.; Eindhoven, The Netherlands

Abstract

Remote inspection of unknown and hostile environments can be performed by military/police personnel via the deployment of sensors and SLAM-based 3D reconstruction techniques. However, the generated point clouds cannot be transmitted to coordinators in real time, because of their large volume sizes. A common data-reduction solution is to convert 3D point cloud models into 2D floor plans. In this paper, we propose an end-to-end network for automated floor plan generation from noisy point clouds to estimate the main building structures (doors, windows and walls). First, the noisy 3D point cloud is column-filtered to remove irrelevant or noisy points. Second, we project the remaining points onto a grid map. Finally, an end-to-end neural network is trained to generate an accurate line-based floor plan from the grid map. Experimental results reveal that the proposed method generates floor plans that accurately represent the main structures of a building. On average, the estimated floor plans reach a 0.66 F_1 score for the building-layout evaluation, which outperforms the state-of-the-art methods. Furthermore, using floor plans reduces the model size by thousands of times on average, which enables real-time communication about the building structure.

Introduction

Floor-plan generation is an essential topic in many fields, especially in remote inspections of unknown and hostile environments. For an interior inspection, defense/police personnel often need to enter unknown hostile buildings without any map, while being coordinated by a remotely located commander via radio communication only. To enhance the commander's global situational awareness, on-body sensors and the SLAM system [2, 9, 3, 4] are used to generate a 3D building model in real time. However, the radio channel bandwidth (0.1-1.0 Mbps) is insufficient to transmit the reconstructed point clouds (45-100 MB) to the commander. A common solution is to convert the point cloud into a 2D floor plan.

Current research [10, 6, 8, 12, 13] performs floor plan generation from noise-free 3D point clouds that are obtained by active sensors (LiDAR and ToF cameras). However, in our project, the choice is constrained to passive sensors (monocular and stereo cameras) only to avoid inspecting personnel being exposed. Since passive sensors are not able to compute depth data accurately, the resulting stereo-based point clouds are very noisy. The depth inaccuracy is caused by the small baseline of a stereo camera or poor indoor illumination. Therefore, an advanced method is required for robust floor plan generation from noisy point clouds.

In this paper, we introduce a DL-based floor plan generation method that generates accurate 2D floor plans for stereo-based point clouds. Stereo images are first sent to a SLAM baseline which generates raw point clouds. Afterwards, we apply a column filter to pre-process the noisy point. Then an orthogonally projected grid map is supplied to a trained network for floor plan generation. Our contributions are threefold. First, the proposed method based on the trained network creates floor plans that clearly and accurately depict the main structures of a building (walls, doors and windows). Second, we create a dataset containing a labeled grid map that describes various building structures. Third, the generated floor plans reduce the data size by thousands of times on average which enables 3D data transmission via radio communication.

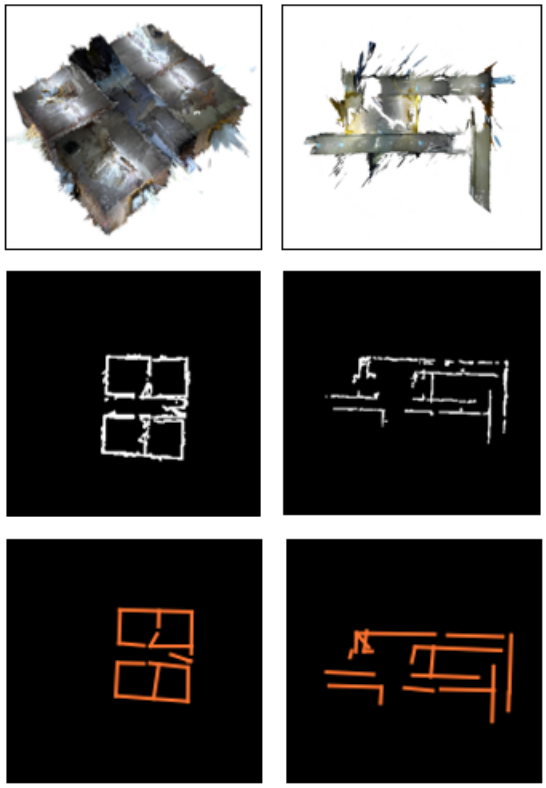

Figure 1. *Floor plans generated by the proposed deep learning network. Top row: Point cloud models. Middle row: grid maps. Bottom row: Generated floor plans.*

Related work

In 3D research, floor plan is a good replacement for the point cloud model, since it can represent the main structure of a building, but it requires a much smaller data size. Most researchers

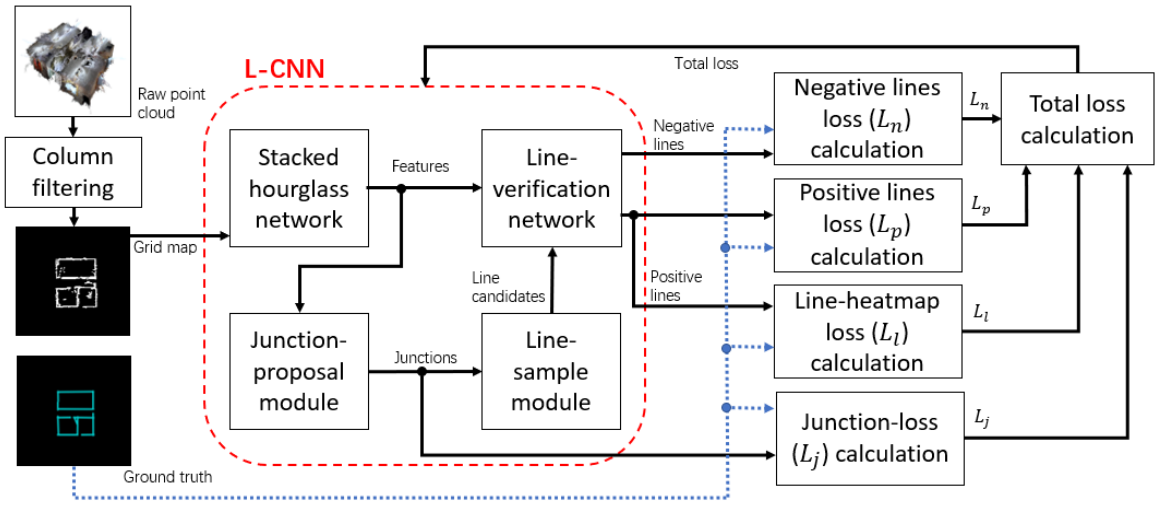

Figure 2. *Overview of the proposed DL-based floor plan generation. A modified L-CNN network is adopted as a baseline. A stacked hourglass network is utilized for primary feature extraction. Then, junction and line candidates are generated by the junction-proposal module and the line-sample module, respectively. Afterwards, a line-verification network classifies line candidates into positive and negative lines. Finally, we modify the function of the total loss calculation to improve the positive line sensitivity of the network.*

utilize lines and planes as primary blocks to construct floor plans. A common way in the literature is to apply the RANSAC algorithm [14] to fit plane candidates in a point cloud [16, 15, 13]. However, this computation is expensive and the algorithm is influenced by outliers. Research in [7] draws horizontal and vertical line candidates for each wall pixel. Then lines are filtered by counting the overlapping area between the line candidates and the projected wall points. This method provides clean and accurate floor plans, but it is limited by the Manhattan world assumption using vertical and horizontal lines only. Luperto [11] detects line features in a 2D grid map by combining the Canny edge detection and the directional Hough transform [5]. Cai [6] estimates the boundary of a building by calculating the confidence score that one unit region is part of the external boundaries. However, for all of the aforementioned methods, a noise-free point cloud is required as the input.

Recently, deep learning techniques have become capable to solve the floor plan generation problem [10] [8] [12]. Most researchers train specialized end-to-end networks to convert point cloud models into floor plans. However, these deep learning methods are trained with clean point clouds generated by active-sensor SLAM systems. In our case, incorrect depth information from a stereo sensor can lead to a very noisy 3D point cloud. Unfortunately, the above-mentioned methods show poor performance on point clouds obtained by passive stereo or monocular sensors, because of the noisy characteristics. L-CNN is an end-to-end network that takes an RGB image as the input and shows good performance on wireframe parsing [1]. However, the current L-CNN provides insufficient results, since it is not designed for floor plan generation. Therefore, we adopt L-CNN as a baseline and re-train it with modifications. The next section introduces a robust floor plan generation from noisy point clouds.

Method

First, we present an overview of the proposed system. Figure 2 depicts the proposed DL-based generation method. The raw point cloud is originally obtained by a stereo-based RTAB-Map system. In the proposed method, first, a column-filtering component is applied to pre-process the noisy point cloud. The pre-processed point cloud is then orthogonally projected onto a black-and-white 2D grid map with a grid size of 10×10 cm. In this grid map, we assign a white pixel to a grid cell, in case at least one point appears in this grid. The obtained grid map is the input for the floor plan generation network.

As shown in Fig. 2 (middle), the L-CNN network contains four components. An RGB image is first fed into a stacked hourglass network for the primary feature extraction. Then a junction-proposal module estimates the junction map that indicates the location of candidate junctions (end points of line segments). Every two junctions are able to constitute a line candidate, but the amount of positive lines (existing lines) and negative lines (not existing lines) is highly unbalanced. Imbalanced data causes the trained network to be biased towards the majority class only. Therefore, L-CNN utilizes a line-sample module to solve the above-mentioned issue. Finally, a line-verification network takes the estimated line candidates and primary feature maps as inputs to implement the positive and negative line classification. As shown at the left of Fig. 2, the input grid map in our case differs from the RGB image. A grid map contains less information compared to an RGB image. Therefore, we modify the network and re-train it using a grid-map dataset.

Besides the changed input data, we have found that the default loss function of the L-CNN network is not suited for floor plan generation. The loss of the junction map receives the highest weight in the default loss function, while other losses share the same weight. In our scenario, the impact of missing line segments is more important than the extraction of noisy lines. It is possi-

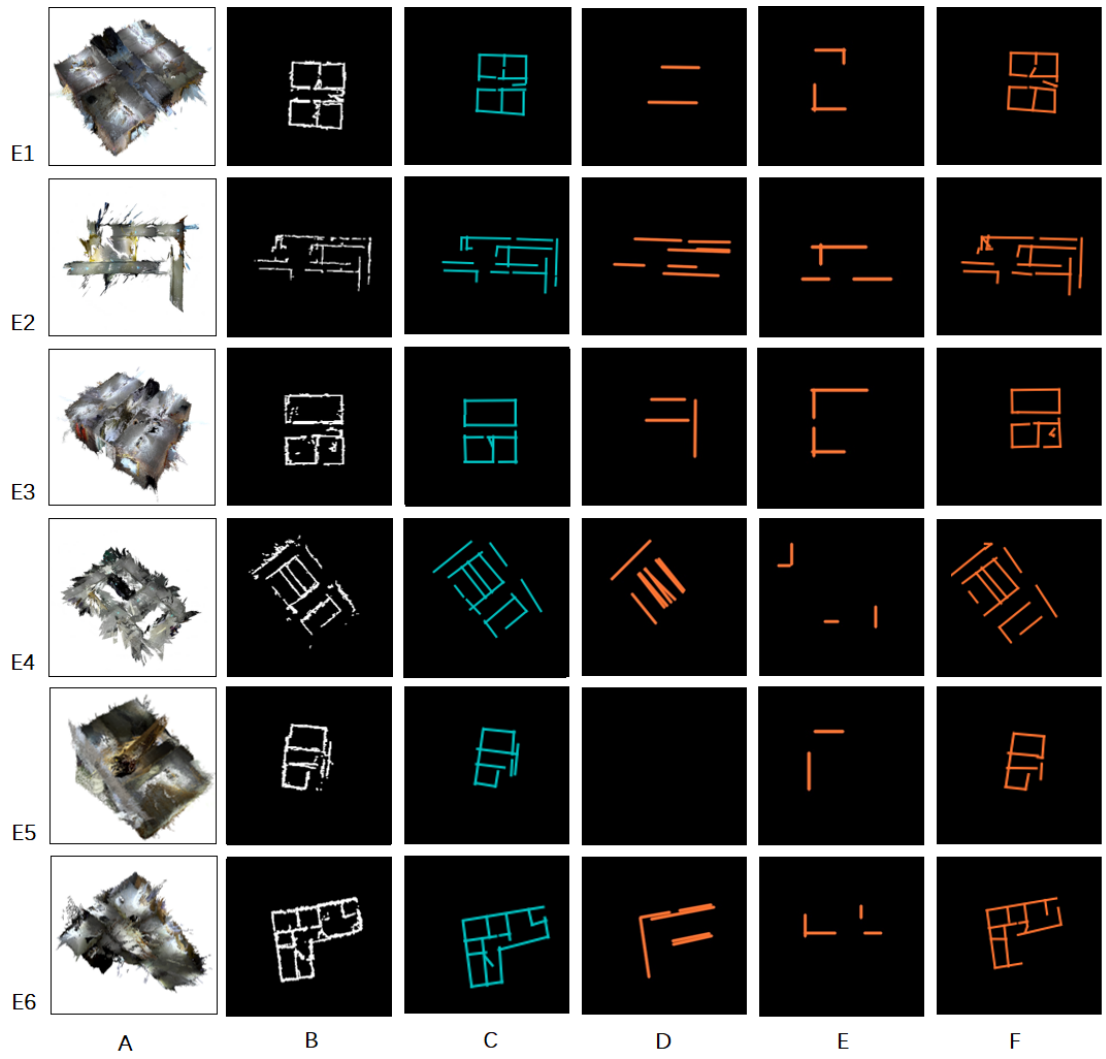

Figure 3. *Six examples of qualitative evaluation and comparison. Column A: Raw stereo-based point clouds. Column B: Black-white grid maps after column filtering. Column C: Manually labeled ground-truth floor plan. Column D: floor plans estimated by the Hough method. Column E: floor plans estimated by the Cai method. Column F: floor plans estimated by the proposed method.*

ble to remove noisy lines by applying post-processing techniques. However, missing lines lead to an incomplete global floor plan, which can influence the decision of the commander. Therefore, it is not fair to assign the same weight for the losses of positive and negative lines. In order to improve the network sensitivity to positive lines, we modify the loss function as follows. Because of the higher importance of the positive lines, we assign a higher weight factor W_H to the use of positive lines. Moreover, the higher loss weight also applies to the junction map because they are essential for the construction of the building. The new loss function is defined as the sum of several loss components, where the loss components for junction L_j and positive lines L_p are assigned higher weights, so that:

$$\mathscr{L}_{\text{Total}} = W_H(L_j + L_p) + W_L(L_n + L_l). \quad (1)$$

Parameters L_n and L_l represent the loss components for negative lines and line map, respectively. We assign a higher weight W_H=8 to parameters that are significant for the floor plan extraction. The lower weight W_L=1 is applied for less important parameters. In order for the network to generate lines in noisy grid maps, we retrain the network using our labelled grid-map dataset. This dataset is needed because passive-sensor data for floor plan generation is not publicly available. The details of this dataset are explained in the next section.

Experiments

We have created a stereo-based grid map dataset for training and evaluation. The data generation procedure is explained in Subsection A. For evaluation, the proposed network is compared to two state-of-the-art methods: Hough [5] and Cai [6]. Subsection B and Subsection C describe the qualitative and quantitative results, respectively. Furthermore, the model-size reduction is discussed in Subsection D.

A. Dataset generation

We have scanned 16 buildings with various structures, including houses and university buildings using the stereo ZED Mini camera (Stereolabs Inc., San Francisco, USA). Table 1 lists the properties of the 16 different buildings inside the dataset. As shown in Table 1, the first eight buildings are of the house type. We scan Buildings B1 and B4 twice with the internal lights on and off. In other house-type buildings, we had to scan without internal lights, due to the absense of electricity. We have noticed that there is a drift in the generated point cloud model due to the poor illumination conditions. Therefore, we have adjusted several parameters of the SLAM baseline, to obtain point cloud models with lower drift. All university-type buildings (B9 to B16) are scanned with internal lights on. These buildings have no windows, and the interior of the building is too dark for the visual SLAM algorithm when the lights are turned off. Furthermore, when generating the filtered grid map, we apply grid sizes of 5 cm and 10 cm except for Building B16, since the building size of B16 is too large. As a result, we have 47 grid maps for 16 buildings in total. All scanned data is sent to the SLAM baseline, in order to generate a filtered black-white grid map.

The ground truth of the floor plan is manually labeled. Furthermore, data augmentation techniques are applied to enlarge the dataset and improve the network generalization capability. We have applied vertical and horizontal flipping, rotation every 45 degrees, blurring, adjusting the aspect ratio, and scaling in the x-axis and y-axis. Afterwards, we have split the dataset into training (90%), validation (10%) and test set (10%). We have deliberately placed all the grid maps of Building B9 and B13 into the test set to ensure that there are unseen building structures in the test set. Moreover, all blurred grid maps in the test set are discarded, since the augmentation by blurring is applied purely to improve the network generalization capability during training. The resulting dataset contains 20,072 training and 1,994 testing grid-map images.

B. Qualitative evaluation

Figure 3 shows examples of the qualitative comparison for six buildings reconstructed with the stereo SLAM baseline. We compare the performance of the proposed method with two state-of-the-art methods. Cai assumes in his work [6] that the main structures of the buildings are perpendicular or parallel to the x-axis or y-axis. To make a fair comparison, we rotate the input grid map, as a pre-processing step for the method in [6]. In Fig. 3, the orange lines in Columns D, E, and F present the estimated floor plan. It can be observed that the Hough transform performs worst. It fails to detect lines in example E5. With a pre-processed input, it can be noticed that the method from [6] can only detect a few lines. By comparing Columns C and F, we observe that the results of the proposed method almost coincide with the ground truth. However, we have also found a few limitations. First, a nonuniform thickness of structures results in a missing line which can be seen in Example E6. Second, Example E3 shows that the network cannot distinguish noise caused by objects (e.g., furniture, people).

C. Quantitative evaluation

Table 2 depicts the quantitative comparison using three metrics adopted from [8], [12] and [10].

Buildings	Lights condition	grid size (cm)	Baseline parameter
B1	off,on	5,10	✗
B2	off	5,10	✓
B3	off	5,10	✓
B4	off,on	5,10	✗
B5	off	5,10	✓
B6	off	5,10	✓
B7	off	5,10	✓
B8	off	5,10	✓
B9	on	5,10	✗
B10	on	5,10	✗
B11	on	5,10	✗
B12	on	5,10	✗
B13	on	5,10	✗
B14	on	5,10	✗
B15	on	5,10	✗
B16	on	10	✗

Table 1. The generation strategies for the grid map of Buildings B1-B16 in our dataset. Buildings B1-B8 are house-type constructions, while the rest are university-type buildings.

Metric 1. Precision/recall of end points: A point is considered true positive if the distance to the nearest point in the ground truth is lower than 10 pixels. Similarly, false positive means that there are no ground-truth points near the target point within 10 pixels. Conversely, false negative counts contain the amount of ground-truth points without estimated points within 10 pixels.

Metric 2. Precision/recall of lines: A line is considered to be true positive if the largest distance between the two start points or the two end points of two line segments is smaller than 10 pixels compared to the nearest ground-truth line.

Metric 3. F_1 score of global layout: To compute this metric, we first generate a global layout of line maps by the rasterizing line technique [17]. Parameter TP_l presents the amount of overlapping pixels between the ground truth and the estimated line map. Likewise, FP_l and FN_l count the number of the residual pixels of lines in the estimated and the ground-truth layout (line) map, respectively. The F_1 score is defined by:

$$F_1 = \frac{2 \cdot TP_l}{(FP_l + FN_l + 2 \cdot TP_l)}. \tag{2}$$

However, single-pixel accuracy is difficult to obtain with Metric 3, because noisy indoor structures are on average 6 pixels thick in the grid map. Therefore, similar to research in [8] [10] [7], a relaxation strategy is applied to the estimated layout map by marking 5 pixels around each line.

Table 2 depicts the quantitative results. It can be noticed that our results reach the highest score under each metric. For Metric 1, the proposed DL-based method reaches 0.98 precision and 0.95 recall scores. It proves that the trained network is able to accurately estimate the location of end points. However, the residual noisy lines and missing lines decrease the score of Metric 2 and 3.

	Metric 1		Metric 2		Metric 3
Methods	Precision	Recall	Precision	Recall	F_1 score
Hough	0.71	0.26	0.10	0.02	0.13
Method [7]	0.67	0.25	0.17	0.05	0.02
DL-based	**0.98**	**0.95**	**0.45**	**0.45**	**0.66**

Table 2. Quantitative evaluation results of several floor plan creation methods. We have selected three metrics and calculate the average scores for the test dataset. Metric 1 evaluates the junction (end points) positions. Metric 2 focuses on the evaluation of each connected junction pair (estimated line object). Metric 3 compares the heat map of lines between the estimated building layout and the ground truth. The highest scores are in bold.

D. Size reduction of the model

The average size of our raw point cloud model is 76.7 MB, while the average size of a generated floor plan is only 1.1 KB. The proposed method is able to reduce the model size by several thousands times on average.

Conclusions

In this paper, we introduce a DL-based floor plan generation method that enables the conversion of very noisy point clouds into a 2D floor plan, where the main building structures become clearly visible. Moreover, the resulting models are small in data size and can be transmitted via radio communication. The main scientific contribution is in the combination of a vertical column-filtering technique with the improved L-CNN network, where the new loss function is introduced to account for the high amount of point cloud noise. Experimental results show that the proposed method outperforms the state-of-art method and reaches 0.66 F_1 score for the building-layout evaluation. In the future, we consider involving semantic information to avoid noisy lines that generate from object-type point clouds. Moreover, prepossessing is required for grid maps that have structures with nonuniform thickness. This prepossessing should be able to solve the current problem of missing lines.

References

[1] Y. Zhou, H. Qi, and Y. Ma, "End-to-end wireframe parsing," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 962–971.

[2] M. Labb'e, "Rtab-map as an open-source lidar and visual slam library for large-scale and long-term online operation, 2018.

[3] Z. Teed and J. Deng, "Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras," Advances in Neural Information Processing Systems, vol. 34, 2021.

[4] Newcombe R.A., Izadi S., Hilliges O., Molyneaux D., Kim D., Davison A.J., Kohi P., Shotton J.H., Steve, and Fitzgibbon A. Kinectfusion: Real-time dense surface mapping and tracking. In 2011 10th IEEE Int. Symp. Mixed and Augmented Reality, pages 127–136, 2011.

[5] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. Communications of the ACM, 15(1):11–15, 1972.

[6] Ruifan Cai, Honglin Li, Jun Xie, and Xiaogang Jin. Accurate floor plan reconstruction using geometric priors. Computers & Graphics, 102:360–369, 2022.

[7] Uuganbayar Gankhuyag and Ji-Hyeong Han.Automatic 2d floor plan cad generation from 3d point clouds. Applied Sciences, 10(8):2817, 2020.

[8] Jiacheng Chen, Chen Liu, Jiaye Wu, and Yasutaka Furukawa. Floorsp: Inverse cad for floor plans by sequential room-wise shortest path. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 2661–2670, 2019.

[9]] Raul Mur-Artal and Juan D. Tardos. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras. IEEE Transactions on Robotics, 33(5):1255–1262, 2017.

[10] Techasarntikul N., Tsuchida K., and Mashita T. Room layout estimation using a machine learning technique. In 2021 Int. Conf. Electrical, Comp. and Energy Technol. (ICECET), pages 1–6. IEEE, 2021.

[11] Matteo Luperto and Francesco Amigoni. Extracting structure of buildings using layout reconstruction. In International Conference on Intelligent Autonomous Systems, pages 652–667. Springer, 2018.

[12] Chen Liu, Jiaye Wu, and Yasutaka Furukawa. Floornet: A unified framework for floor plan reconstruction from 3d scans. In Proceedings of the European conference on computer vision (ECCV), pages 201–217, 2018.

[13] Hao Fang, Florent Lafarge, Cihui Pan, and Hui Huang. floor plan generation from 3d point clouds: A space partitioning approach. ISPRS Journal of Photogrammetry and Remote Sensing, 175:44–55, 2021.

[14] Ruwen Schnabel, Roland Wahl, and Reinhard Klein. Efficient ransac for point-cloud shape detection. Comput. Graph. Forum, 26:214–226, 06 2007.

[15] Sebastian Ochmann, Richard Vock, and Reinhard Klein. Automatic reconstruction of fully volumetric 3d building models from oriented point clouds. ISPRS journal of photogrammetry and remote sensing, 151:251–262,2019.

[16]] Sebastian Ochmann, Richard Vock, Raoul Wessel, Martin Tamke, and Reinhard Klein. Automatic generation of structural building descriptions from 3d point cloud scans. In 2014 International Conference on Computer Graphics Theory and Applications (GRAPP), pages 1–8. IEEE, 2014.

[17] Marloes LP van Lierop, Cornelius WAM van Overveld, and Huub MM van de Wetering. Line rasterization algorithms that satisfy the subset line property. Computer vision, graphics, and image processing, 41(2):210–228, 1988.

Author Biography

Xin Liu received her MSc degree in Electrical Engineering in 2019 from the Eindhoven University of Technology. She is currently a PhD candidate at the same university. Her research interests include computer vision and simultaneous localization and mapping.

Egor Bondarev obtained his PhD degree in the Computer Science Department at TU/e, in research on performance predictions of real-time component-based systems on multiprocessor architectures. He is an Assistant Professor at the Video Coding and Architectures group, TU/e, focusing on sensor fusion, smart surveillance and 3D reconstruction. He has written and co-authored over 50 publications on real-time computer

vision and image/3D processing algorithms. He is involved in large international surveillance projects like APPS and PS-CRIMSON.

Peter H.N. de With is Full Professor of the Video Coding and Architectures group in the Department of Electrical Engineering at Eindhoven University of Technology. He worked at various companies and was active as senior system architect, VP video technology, and business consultant. He is an IEEE Fellow and member of the Royal Holland Society of Sciences, has (co-)authored over 600 papers on video coding, analysis, architectures, and 3D processing and has received multiple papers awards. He has been a program committee member of several IEEE conferences and holds some 30 patents.