Advanced Methods for Environment Analysis Based on Lidar Data

Theses of the *Ph.D.* Dissertation

Örkény Ádám H. Zováthi

Scientific advisor:
Csaba Benedek, D.Sc.

Roska Tamás Doctoral School of Sciences and Technology
Pázmány Péter Catholic University

Budapest, 2023

1 Introduction and aim

The utilization of new sensor technologies for three-dimensional (3D) data acquisition is a key step for deeply understanding and widely exploiting spatial information in our environment. Widespread sensor platforms are mapping systems based on Lidar (Light Detection and Ranging) technology, as they provide accurate 3D measurement flows with high acquisition speed [6]. However, due to a trade-off between the available laser scanners' temporal and spatial resolution, the provided point clouds show significantly different quality and density characteristics [7], limiting the general usability of standard point cloud processing techniques or methods developed for specific sensors with their own domain specific functional requirements.

On the one hand, autonomous vehicles and mobile event surveillance missions (such as traffic analysis or crowd monitoring) demand real-time 3D data acquisition and processing techniques operating onboard on mobile platforms. For dynamic environment perception and recognition tasks such as advanced scene analysis and understanding, repetitive, typically rotating multi-beam (RMB) Lidar sensors (e.g., Ouster OS1 or Velodyne Puck models) [1] are commonly utilized devices. RMB Lidars can produce real-time point cloud streams (300 thousand-2 million points/s), however, their measurements have low spatial density, and their field-of-view (FoV) coverage is constant through the whole scanning process: Their vertical resolution is fixed by the number of the laser beams (16-128), while their horizontal resolution depends on the sensor's rotation frequency (5-20 Hz). Alternatively to RMB Lidars, recent non-repetitive circular scanning (NRCS) Lidar sensors are also capable of providing measurements for real-time scene analysis, at a significantly lower cost compared to the RMB technology. Unlike RMB Lidars, NRCS Lidars (e.g., the Livox AVIA sensor) are able to densely map large areas from a given scanning position due to their special scanning technology which follows non-repetitive, e.g., rosetta patterns. The main challenge is here to efficiently balance between the spatial and the temporal resolution of the recorded

range data using a suitable integration window [2, 8].
On the other hand, city management applications such as urban development and planning, public place surveillance, and road maintenance need very detailed and accurate 3D spatial maps from the environment, which are obtained by offline scanning technologies and stored and maintained in new generation Geo-Information Systems (GIS). Recent Terrestrial (TLS) and Mobile Laser Scanning (MLS) platforms equipped with time synchronized Lidar sensors and navigation units are common choices for such applications, as they provide dense, accurate and feature rich point clouds precisely registered to a geo-referenced global coordinate system.
This thesis deals with two main tasks related to advanced environment analysis using Lidar sensors. First, I investigate how the perception capabilities of the state-of-the-art real-time RMB Lidars can be extended by prior location information, adopting offline and semantically evaluated 3D point cloud maps captured by an up-to-date MLS system. Second, I investigate how the perception of the latest real-time NRCS Lidars can be improved without any external information, exploiting the spatial and temporal characteristics of the sensor measurements by deep learning techniques.

2 New Scientific Results

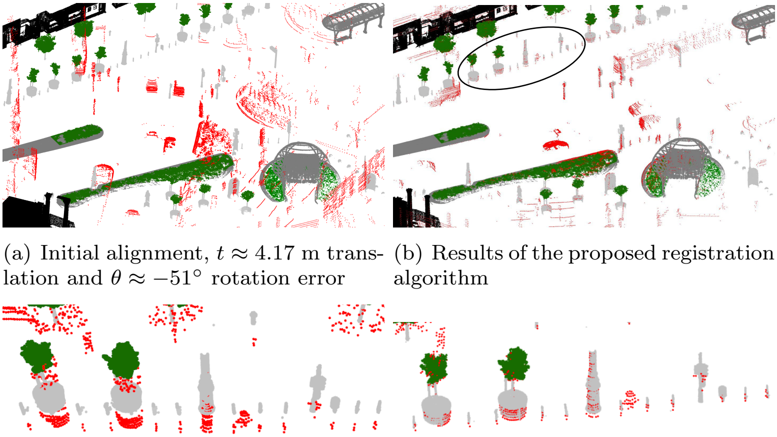
1. Thesis: I have proposed a new method for change detection, which comprises 3D point cloud registration and point-level change segmentation through fusing Lidar point clouds with significantly different density characteristics. I have constructed a new urban dataset by a state-of-the-art rotating multi-beam (RMB) Lidar scanner (with a point density of around 50-500 points/m^2) and an up-to-date Mobile Laser Scanning (MLS) system (more than 5000 points/m^2). Using this new dataset, I have quantitatively demonstrated the advantage of the proposed algorithm against various state-of-the-art reference techniques.

Published in [1][3][4][5]

This thesis deals with three consecutive subtasks: First, a cross-source point cloud registration is performed to precisely align the sparse RMB and dense MLS data, and a utilization of this algorithm is introduced for tracking the pose of the capturing vehicle in real time, even in partially occluded and dynamic environment. Second, a cross-source change detection algorithm is proposed for finding every point that changed since the map creation. Third, a new method is introduced by utilizing the output of the change detection technique for improved dynamic object detection.

> *1.1. I have proposed a novel approach for the registration of sparse RMB Lidar and dense MLS point clouds with a relatively poor initial alignment, which consists of a coarse pre-alignment step through detecting and matching landmark object candidates using geometry-based feature points from the whole scene, and a point-level refinement step that calculates the accurate transformation matrix based on the matched objects' local point cloud segments for reducing the computational need. I have demonstrated the advantage of the proposed algorithm against various state-of-the-art reference techniques in urban scenes, by comparing translation and rotation errors calculated by the decomposition of each transformation matrix and the manually labelled ground truth (GT), and by measuring point distances between the registered point clouds. I have shown an efficient utilization of the introduced algorithm for pose tracking, by estimating the planar pose of the vehicle from the object-matching result and fusing it in a constant velocity model-based Kalman filter.*

Several point cloud registration algorithms exist in the literature that perform correspondences between features of handcrafted [9, 10, 11] or learning-based [12, 13] keypoints, segments [14, 15] or points [16, 17, 18]. In the addressed cross-source application, the main challenge is that the RMB Lidar frames are too sparse for extracting meaningful 3D keypoints (which work between dense MLS point clouds), while the MLS point clouds are in several regions 100-1000 times denser than the corresponding RMB measurement



(a) Initial alignment, $t \approx 4.17$ m translation and $\theta \approx -51°$ rotation error

(b) Results of the proposed registration algorithm

(c) Results of the coarse alignment in case of an uneven road surface

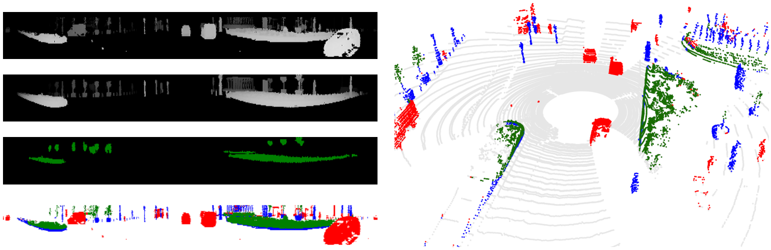
(d) Results after the point level refinement

Figure 1: Results of the proposed point cloud registration algorithm. Subfigures (c) and (d) refer to the same area circled by black in Subfigure (b). Color codes: RMB points are shown with red, the segmented MLS regions are marked by various colors depending on their semantic classes.

segments which misleads the general segment level matching processes. Following a different approach, as the first contribution, instead of aligning the original point clouds, I have separated and matched landmark objects in the RMB Lidar frames and from the MLS map. Here, as a remaining challenge, many falsely detected object candidates can present (e.g., traffic participants, partially occluded objects), which may result in a possibly large ratio of outlier matches. To handle their effect, I have applied the voting schema of the generalized Hough transform. As the second contribution, I have used the coarse alignment step to initialize the point-level Iterative Closest Point algorithm [19], which I have executed only for point cloud segments corresponding to the previously aligned object pairs. In comparison to six different point cloud registration methods [13, 20, 16, 21, 22, 23], the median value of point-level distances is decreased by 1–2 orders of magnitude by the proposed approach. To overcome the problem of heavily occluded scenarios without a sufficient number of matchable object pairs, I have extracted the planar (3DoF) pose (planar position and yaw orientation) of the capturing vehicle from the result of the object matching process and integrated the estimated pose parameters by a constant velocity (CV) model-based Kalman filter. Starting from a poor GPS-based positioning with 5-10 meters error, the proposed pose tracking approach is able to reduce the location error of the vehicle by one order of magnitude and to keep the yaw angle error around 1° during its whole trajectory without considerable drift, while running in real time (20-25 Hz).

1.2. I have proposed a new Markov Random Field-based approach (RangeMRF) for multi-class change extraction and classification (dynamic, seasonal, or no change) between registered RMB and MLS point clouds using 2D range image representations. I have demonstrated the advantage of the proposed algorithm against various state-of-the-art reference methods by qualitative and quantitative evaluations.

Three-dimensional change detection is a highly discussed topic in the literature [24], however, existing point [25, 26], segment [27] or voxel [28] based methods cannot handle well when the characteristics of the two comparable point sets are significantly different. In the addressed scenario, they show notable trade-off between false positive (e.g., noise, vegetation changes) and false negative hits (i.e., loss of details). As the first advantage, the proposed RangeMRF method detects changes between 2D range images derived from the point clouds. Using a compact range image representation, the proposed method is notably quick, meanwhile, it can robustly handle the significantly different density characteristics of the two point sets by containing only relevant parts of the dense MLS data. Second, I have applied a Markov Random Field model, which is highly robust though noisy measurement data. Third, I have distinguished three classes in the segmentation model: seasonal changes in veg-



(a) Range domain (b) 3D domain

Figure 2: Change segmentation results in the range image domain and in the 3D space on real measurement data. The first two rows of Subfigure (a) are depth images where brighter pixels denote closer distance, and black pixels contain no measurements. The third row displays semantic labels of the MLS data, where vegetation is marked by green. The last row shows the segmentation output of the RMB data. Subfigure (b) displays the segmentation of the same area in the 3D space. The pixels/points for static background are displayed by blue, for dynamic change by red, and for seasonal change by green.

etation regions, foreground changes caused by moving objects or changed/re-located static street furniture elements, and unchanged background regions. By handling the vegetation areas with a specific sensitivity, the proposed method can eliminate several false hits in vegetation areas, while it is able to sharply recognize even small foreground changes (i.e., pedestrians standing near stations or facades) between the input point clouds. In comparison to four reference techniques [27, 24, 28, 26], the proposed method outperforms them either in F1-scores (by around 10-25%) or in computational complexity, running 10–1000 times faster.

> *1.3. I have proposed a new method to utilize the introduced change detection approach for improving the performance of Lidar-only dynamic object detection algorithms. I have demonstrated in high-traffic road sections that the proposed approach*

can efficiently balance the precision and recall values with significant overall improvement for both vehicles and pedestrians, using a state-of-the-art object detection method.

Real-time dynamic object detection in 3D sparse point clouds is a hot topic in autonomous driving with several geometric [29] and deep learning [30, 31, 32, 33, 34] based algorithms in the literature. However, there are a number of limitations of these approaches: False positive hits may appear in point cloud regions containing static scene objects with similar appearance and context parameters to the focused dynamic scene objects, while the point cloud blobs of several dynamic objects can be heavily merged or occluded by static street furniture elements, yielding many unrecognized traffic participants. I have proposed a new approach that utilizes dense MLS maps in order to decrease in parallel both the false negative and false positive hits of object detection algorithms. The proposed approach includes a map-based object validation, the introduced cross-source change extraction, and an object-level change analysis step between registered RMB and MLS map data. As a basis of comparison, I have chosen the *PointPillars* [34] state-of-the-art object detection method, with which the proposed method achieved an improvement of 5.96% in precision, 9.21% in recall and 7.93% in F1-score metrics on our own dataset.

2. Thesis: I have proposed a novel depth completion method from sparse consecutive measurements of a non-repetitive circular scanning (NRCS) Lidar using a deep learning model (ST-DepthNet). I have constructed a new urban dataset that comprises various simulated and real-world NRCS Lidar data samples. Using this new dataset, I have qualitatively and quantitatively demonstrated the superiority of the proposed method against a densified depth map obtained from the raw sensor stream, and against two independent state-of-the-art Lidar-only depth completion algorithms.

Published in [2]

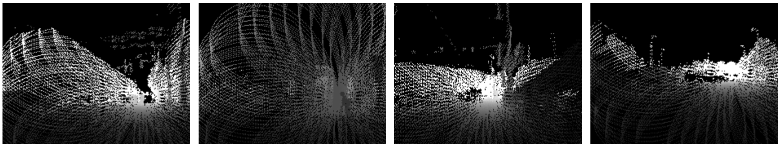
This thesis deals with efficient data-driven completion of NRCS Lidar data. Due to their non-repetitive scanning technology, NRCS Lidars are able to map different areas of their field of view (FoV) from a given scanning position in consecutive times [8]. The main challenge of analysing their point cloud streams is to efficiently balance between the spatial and the temporal resolution of the recorded range data using a suitable integration window. Allowing a larger integration time (e.g., 1 s) yields high spatial measurement resolution with various artifacts, such as blurred shapes of the observed vehicles, pedestrians or buildings, which phenomena complicate dynamic event analysis, while a narrow time window (e.g., 200 ms) yields spatially more precise but notably sparse measurements with a significant loss of spatial details. To overcome this spatio-temporal trade-off of the NRCS Lidar-based perception, I have proposed a novel deep learning-based approach for the densification of sparse NRCS Lidar depth data streams while keeping their temporal resolution and spatial accuracy high.

2.1. I have proposed a new training framework for the densification of 3D data streams provided by NRCS Lidars, by mapping their consecutive point cloud measurements to sparse 2D depth images, each collected within 200 ms to enable 40% field-of-view coverage. I have constructed a new synthetic dataset which contains depth images acquired by simulating the behaviour of a NRCS Lidar, and high-quality dense depth images for each sparse sample exploiting the complete spatial information of the virtual world. I have extended the dataset with sparse real samples using the same depth image representation and I have made them both publicly available. The proposed framework enables to train and to test depth completion algorithms on synthetic scenarios, and to validate their reliability in real-world data as well.

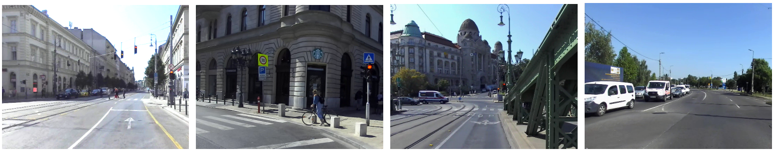
For the 2D representation, I have converted the captured sparse point clouds of the Livox AVIA NRCS Lidar sensor from the Cartesian (x, y, z) to the spherical (distance, azimuth, elevation) polar coordinate system, then the horizontal and vertical FoVs were quantized onto a 400×400 pixel lattice using an integration time window of 200 ms for collecting the consecutive time frames. By this approach, around 60% of the range image pixels receive undefined range values, however, they are not notably effected by blurring and this representation permits the use of two-dimensional convolutional neural networks to fill in the missing structural information in the image domain. As higher integration time induces blurred silhouettes due to the independent movements of dynamic objects of the scene including the ego robot or vehicle, it is challenging to provide dense, spatially precise GT depth information for real data. Therefore, besides the real measurements, I have constructed a synthetic range image dataset from a realistic virtual world using the CARLA simulator [35], where the behaviour of the Livox AVIA NRCS Lidar sensor was implemented. The virtual world allows to extract dense, spatially precise depth information, used as GT for the Lidar's sparse depth sample data.

> *2.2. I have proposed a new depth completion deep neural network called ST-DepthNet, which extends the classical U-Net architecture with a spatio-temporal downscaling branch for utilizing five consecutive sparse measurements captured by NRCS Lidars and produces spatially precise high-density depth data in real time. I have demonstrated the advantage of the proposed algorithm against the state of the art in both synthetic and real-world scenarios.*

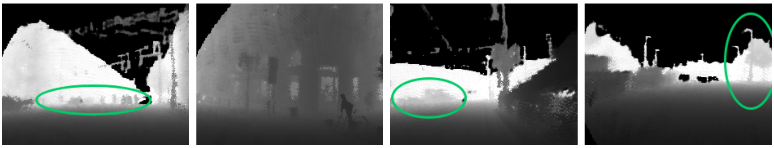
First, I have exploited that using the applied Livox AVIA sensor, a time interval of 1 s contains enough dense range information from the scene with almost full FoV coverage. Thus, I have taken five consecutive sparse depth images – each one recorded in 200 ms – as the network's input to have enough information about the complete FoV. To accurately restore the single output image, I have adopted an image-to-image U-Net [36] architecture and I have extended the downscaling part of the U-Net network by utilizing Conv2DLSTM layers [37] to exploit temporal connections between the features derived from the input image sequence. Second, the upscaling branch of the proposed network remained purely two-dimensional and skip



(a) Sparse input data captured in a 200 ms time window



(b) RGB image for visual reference only



(c) Output of the proposed ST-DepthNet deep network

Figure 3: Depth completion results on real measurements from Budapest, Hungary. Accurately predicted fine object structures by *ST-DepthNet* are highlighted with green ellipses.

connections at each level were performed by recurrent pooling utilizing the last output of a Conv2DLSTM layer which represents features of the last 200 ms measurement. Third, I have directly connected the last input image to the output to exploit that the last and most up-to-date input image contains spatially precise points, and therefore, the network only has to learn the missing regions of the range image. I have trained the model on the introduced synthetic dataset and I have quantitatively shown that the proposed approach outperforms two state-of-the-art reference methods [38, 39] on synthetic data, reducing their root-mean-squared error (RMSE) by more than 1 meter in the range domain, and achieving around a half meter less error in the 3D domain by the normalized Chamfer distance and median distance. For real NRCS measurement data, a survey from 20 computer vision-related experts demonstrated that the proposed method performs significantly better than the reference techniques.

3 Application of the Results

All the developed algorithms can be used in advanced perception platforms of mobile robots and intelligent vehicles equipped with RMB or NRCS Lidar sensors. The first thesis can be applied in urban environment where detailed 3D point cloud maps are available about the city for assisting vehicles that are equipped with a RMB Lidar and a GPS receiver. The proposed methods can contribute to map-based real-time scene understanding like accurate self-localization and pose tracking, change-based scene segmentation and improved dynamic object detection. The second thesis can be applied for robot or vehicle platforms that are equipped with a NRCS Lidar sensor to produce accurate and dense depth maps from the environment while keeping high temporal resolution, which can be a crucial middle step for more advanced scene understanding or mapping. As part of my Cooperative Doctoral Program, many of the proposed algorithms directly contributed to R&D projects conducted with the participation of the Institute for Computer Science and Control (SZTAKI) and the Pázmány Péter Catholic University (PPCU), and we also submitted two patent applications related to the methods.

4 Datasets and Implementation Details

For joint utilization of both RMB Lidar measurements and MLS data, I created a new Benchmark called *SZTAKIBudapest* that contains RMB point cloud streams captured by a Velodyne HDL 64E 64-beam RMB Lidar sensor from different downtown areas of Budapest, where high-density, geo-referred point cloud maps are also recorded by a Riegl VMX450 MLS scanner. The Benchmark contains three different road scenarios, each one covering a path of

around 300 meters with segmented reference MLS data [40]. For quantitative evaluation, Ground Truth (GT) annotations verified by operators are available for registration, pose tracking, change and object detection.
For training and evaluation of the proposed method in the second thesis, I constructed a synthetic range image dataset called *LivoxCARLA* from a realistic virtual world using the CARLA simulator [35], where I simulated the behaviour of the Livox AVIA NRCS Lidar sensor. The *LivoxCARLA* dataset consists of 11726 randomly sampled input-output range image pairs, from which 10000 were used for training, 500 as validation and 1226 for testing. Besides the *LivoxCARLA* dataset, I also collected real measurement sequences from Budapest. In these experiments, the Livox AVIA sensor was mounted on the front-top of our test vehicle on a driving path of total 5.5 kilometers in both speedways and in the city center.
The main platform for point cloud handling and processing was implemented in C++ with the OpenCV and PCL libraries, while the neural network models were implemented and trained in Python with the Pytorch or Keras frameworks.

5 Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Csaba Benedek for his continuous support, motivation and patience during my Ph.D. study and work. Special thanks to Prof. Tamás Szirányi, head of the Machine Perception Research Laboratory for supporting my research at the Institute for Computer Science and Control (SZTAKI). I thank my colleagues for their works and advices: Csaba Benedek, the leader of our research group, Zsolt Jankó, the company advisor of my KDP research, furthermore Balázs Nagy, Lóránt Kovács, Balázs Pálffy, József Kövendi, Yahya Ibrahim and László Tizedes. Pázmány Péter Catholic University (PPCU) is also gratefully acknowledged, thanks to Prof. Gábor Szederkényi for the opportunity to study there. Special thanks to those whom I may have not mentioned by name but who supported me directly or indirectly in accomplishing my Ph.D. research.

I thank the reviewers of my dissertation for their work and valuable comments.
For financial support, thanks to the National Research, Development and Innovation Fund, under the ÚNKP-20-3 and ÚNKP 21-3 New National Excellence Program and under the KDP-2020 Cooperative Doctoral Program (KDP-977852). My research at SZTAKI was also supported within the frameworks of the Autonomous Systems National Laboratory and the Artificial Intelligence National Laboratory programs and the TKP2021-NVA-01, OTKA K-120233, and K-143274 projects.
Last but not least, I would like to thank my family for supporting me throughout these Ph.D. years and my life in general. My mother who raised me with her unconditional love that I will never be able to pay back even if I live for a thousand years, my father who always motivated me to dream big and taught me that hard work always pays off, my brothers, Bendi, Kende, Csani and Domi and my beautiful wife Kata who motivated and supported me even in the hardest times.
To my grandparents for their neverending blessings and to my godmother for her continuous support.

6 Publications

6.1 The Author's Journal Publications

[1] **Ö. Zováthi**, B. Nagy, and C. Benedek, "Point cloud registration and change detection in urban environment using an onboard lidar sensor and MLS reference data," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 110, p. 102767, 2022. IF = 7.5, Scimago Q1/D1.

[2] **Ö. Zováthi**, B. Pálffy, Z. Jankó, and C. Benedek, "ST-DepthNet: A spatio-temporal deep network for depth completion using a single non-repetitive circular scanning lidar," *IEEE Robot. and Autom. Lett.*, vol. 8, no. 6, pp. 3270–3277, 2023. IF = 5.2*, Scimago Q1/D1.

6.2 The Author's International Conference Publications

[3] **Ö. Zováthi**, B. Nagy, and C. Benedek, "Exploitation of dense MLS city maps for 3D object detection," in *Int. Conf. Image Anal. Recognit.*, vol. 12131 of *Lecture Notes in Computer Science*, pp. 393–403, Springer, 2020.

[4] **Ö. Zováthi**, B. Pálffy, and C. Benedek, "Real-time vehicle localization and pose tracking in high-resolution 3D maps," in *European Signal Process. Conf.*, pp. 1786–1790, 2022.

[5] **Ö. Zováthi**, L. Kovács, B. Nagy, and C. Benedek, "Multi-object detection in urban scenes utilizing 3D background maps and tracking," in *Int. Conf. Control, Artif. Intell., Robot. Optim.*, pp. 231–236, 2019.

6.3 Selected Publications Connected to the Dissertation

[6] C. Benedek, A. Majdik, B. Nagy, Z. Rózsa, and T. Szirányi, "Positioning and perception in LIDAR point clouds," *Digit. Signal Process.*, vol. 119, p. 103193, 2021.

[7] J. Behley, V. Steinhage, and A. B. Cremers, "Performance of histogram descriptors for the classification of 3D laser range data in urban environments," in *IEEE Int. Conf. Robot. Autom.*, (St. Paul, MN, USA), pp. 4391–4398, May 2012.

[8] L. Kovács, M. Kégl, and C. Benedek, "Real-time foreground segmentation for surveillance applications in NRCS lidar sequences," *ISPRS Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLIII-B1-2022, pp. 45–51, 05 2022.

[9] R. Hänsch, T. Weber, and O. Hellwich, "Comparison of 3D interest point detectors and descriptors for point cloud fusion," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. II-3, pp. 57–64, 09 2014.

[10] I. Sipiran and B. Bustos, "Harris 3D: A robust extension of the Harris operator for interest point detection on 3D meshes," *The Vis. Comput.*, vol. 27, pp. 963–976, 11 2011.

[11] Y. Zhong, "Intrinsic shape signatures: A shape descriptor for 3D object recognition," in *Int. Conf. Comput. Vis. Workshops*, (Kyoto, Japan), pp. 689–696, 11 2009.

[12] J. Li and G. H. Lee, "USIP: Unsupervised stable interest point detection from 3D point clouds," in *IEEE/CVF Int. Conf. Comput. Vis.*, (Seoul, Korea), pp. 361–370, 2019.

[13] F. Lu, G. Chen, Y. Liu, Z. Qu, and A. Knoll, "RSKDD-Net: Random sample-based keypoint detector and descriptor," in *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 21297–21308, 2020.

[14] B. Douillard, A. Quadros, P. Morton, J. P. Underwood, M. D. Deuge, S. Hugosson, M. Hallström, and T. Bailey, "Scan segments matching for pairwise 3D alignment," in *IEEE Int. Conf. Robot. Autom.*, (St. Paul, MN, USA), pp. 3033–3040, May 2012.

[15] R. Dubé, A. Cramariuc, D. Dugas, J. Nieto, R. Siegwart, and C. Cadena, "SegMap: 3D segment mapping using data-driven descriptors," in *Robot.: Sci. Syst.*, 06 2018.

[16] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.*, vol. 13, p. 119–152, Oct. 1994.

[17] A. Gressin, C. Mallet, and N. David, "Improving 3D LIDAR Point Cloud Registration Using Optimal Neighborhood Knowledge," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. I-3, pp. 111–116, 08 2012.

[18] A. Gressin, B. Cannelle, C. Mallet, and J.-P. Papelard, "Trajectory-Based Registration of 3D LIDAR Point Clouds Acquired with a Mobile Mapping System," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. I-3, pp. 117–122, 08 2012.

[19] Z. Zhang, "Iterative point matching for registration of free-form curves and surfaces," *Int. J. Comput. Vis.*, vol. 13, no. 2, pp. 119–152, 1994.

[20] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-squares fitting of two 3-D point sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 698–700, 1987.

[21] S. Choi, Q. Zhou, and V. Koltun, "Robust reconstruction of indoor scenes," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, (Boston, MA, USA), pp. 5556–5565, 2015.

[22] H. Yang, J. Shi, and L. Carlone, "Teaser: Fast and certifiable point cloud registration," *IEEE Trans. Robot.*, vol. 37, no. 2, pp. 314–333, 2021.

[23] X. Huang, G. Mei, and J. Zhang, "Feature-metric registration: A fast semi-supervised approach for robust point cloud registration without correspondences," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2020.

[24] W. Xiao, B. Vallet, M. Brédif, and N. Paparoditis, "Street environment change detection from mobile laser scanning point clouds," *ISPRS J. Photogramm. Remote Sens.*, vol. 107, pp. 38 – 49, 2015.

[25] D. Liu, D. Li, M. Wang, and Z. Wang, "3D change detection using adaptive thresholds based on local point cloud density," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 3, 2021.

[26] D. Girardeau-Montaut, M. Roux, R. Marc, and G. Thibault, "Change detection on point cloud data acquired with a ground laser scanner," *ISPRS Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 36, 01 2005.

[27] A. Schlichting and C. Brenner, "Vehicle localization by lidar point correlation improved by change detection," *ISPRS Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLI-B1, pp. 703–710, 06 2016.

[28] K. Liu, J. Boehm, and C. Alis, "Change detection of mobile lidar data using cloud computing," *ISPRS Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLI-B3, pp. 309–313, 06 2016.

[29] A. Börcs, B. Nagy, and C. Benedek, "Instant object detection in lidar point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, pp. 992–996, July 2017.

[30] K. Shin, Y. Kwon, and M. Tomizuka, "RoarNet: A robust 3D object detection based on region approximation refinement," *IEEE Intelligent Vehicles Symposium*, pp. 2510–2515, 2018.

[31] M. Simon, S. Milz, K. Amende, and H.-M. Groß, "Complex-YOLO: Real-time 3D object detection on point clouds," *ArXiv*, vol. abs/1803.06199, 2018.

[32] Y. Yan, Y. Mao, and B. Li, "SECOND: Sparsely embedded convolutional detection," *Sensors*, vol. 18, p. 3337, 10 2018.

[33] B. Yang, W. Luo, and R. Urtasun, "PIXOR: Real-time 3D object detection from point clouds," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 7652–7660, 2018.

[34] A. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "PointPillars: Fast Encoders for Object Detection from Point Clouds," in *IEEE Conf. Comput. Vis. Pattern Recognit.*, (Long Beach, CA, USA), pp. 12689–12697, 06 2019.

[35] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. Ann. Conf. Rob. Learn.*, pp. 1–16, 2017.

[36] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comp.-Ass. Interv.*, pp. 234–241, 2015.

[37] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-k. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Int. Conf. NIPS*, p. 802–810, 2015.

[38] J. Ku, A. Harakeh, and S. L. Waslander, "In defense of classical image processing: Fast depth completion on the CPU," *Conf. Comput. Rob. Vis.*, pp. 16–22, 2018.

[39] M. F. F. Khan, N. D. Troncoso Aldas, A. Kumar, S. Advani, and V. Narayanan, "Sparse to dense depth completion using a generative adversarial network with intelligent sampling strategies," in *Proc. ACM Int. Conf. Multimedia*, p. 5528–5536, 2021.

[40] B. Nagy and C. Benedek, "3D CNN-based semantic labeling approach for mobile laser scanning data," *IEEE Sens. J.*, vol. 19, pp. 10034–10045, Nov 2019.