Three-dimensional Scene Understanding in Mobile Laser Scanning data

Theses of the *Ph.D.* Dissertation

Balázs Nagy
computer engineer

Scientific adviser:
Csaba Benedek, Ph.D

Faculty of Information Technology and Bionics
Pázmány Péter Catholic University

Budapest, 2020

1 Introduction and aim

In recent decades, huge progress has been made in the field of sensors for environment perception and mapping, which greatly influences the remarkable scientific progress in the field of object detection and classification, and scene segmentation and understanding.

The breakthrough of deep learning methods significantly improved the quality of computer vision and scene understanding applications since they are able to provide more reliable results with high accuracy. However, we are still far from human-level performance. To train these models, a huge amount of precisely labeled data is required, furthermore in challenging situations such as *heavy traffic scenarios in dense urban areas*, *complex unknown regions as off-road scenes*, and *bad weather and illumination conditions* they still often underperform. Considering multiple sensor data can improve the accuracy of the perception and relying on more sensors can lead to more robust algorithms which are able to work independently from the environmental effects such as *heavy rain* and *illumination changes.*

Using detailed background information such as *High Definition (HD) maps* and *Geographic Information Systems (GIS)* to achieve more accurate scene understanding and more reliable localization is also an important research direction. Data fusion between onboard sensor data and offline detailed map databases is used in several industrial projects such as *Waymo and Uber self-driving vehicles.* By the wide-spreading of smart cities with detailed map information, autonomous vehicles can utilize the static background information for navigation, localization, and contextual based scene analysis by registering their real-time captured onboard sensors data to the corresponding part of the detailed map.

Though HD maps facilitate the localization and environment understanding problems, the presence of dynamic objects such as *vehicles* and *pedestrians* and the continuously changing different traffic scenarios require real-time environment perception. While optical cameras provide high resolution, feature-rich data with color information, real-time Lidar sensors such as Velodyne HDL-64 are able to obtain accurate 3D geometric information (up to 100 meter). Lidar sensors are less influenced by the lighting and weather conditions. So accurate sensor fusion i.e. calibration between the camera and Lidar sensors for more accurate and reliable perception is essential. This thesis proposes novel approaches for the following three research problems: The first task is automatic segmentation i.e. labeling of dense point clouds obtained by a mobile mapping system in an urban environment. The second problem introduces a novel solution for a robust, real-time registration between different types of point clouds and it proposes a method to solve the localization problem of self-driving vehicles by aligning the sparse real-time captured point cloud data to a segmented dense point cloud map. The last investigated task is an automatic camera and Lidar calibration task which can be performed on-the-fly.

2 New Scientific Results

1. Thesis: I have proposed a new scene segmentation model which is able to semantically label dense 3D point clouds collected in an urban environment. Using the segmented dense point clouds as detailed background maps I have proposed a Lidar-based self-localization approach for self-driving vehicles. I have quantitatively compared and evaluated the proposed approaches against state-of-the-art methods in publicly available databases.

Published in [2][9][10][11]

The proposed scene segmentation model aims to semantically label each point of a dense point cloud obtained by a mobile mapping system such as Riegl VMX-450 into different categories. Since the collected point clouds are geo-referenced i.e. the sensor registers all incoming scans into the same coordinate system, dynamic objects moving concurrently with the scanning platform appear as a long-drawn, noisy region called *phantom objects* in this thesis. Further-

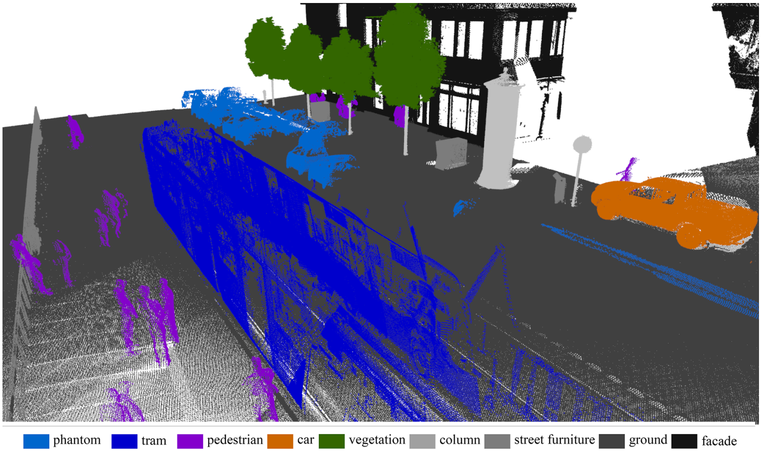


Figure 1: Labeling result of the proposed scene segmentation model [Thesis 1].

more, during the scanning process, the mapping system may scan the same regions several times yielding an inhomogeneous point density with the mentioned noisy phantom region. Though the mapping system assigns color information to the recorded 3D points, because of several occlusion and illumination artifacts this color information is often unreliable.

The proposed model is able to robustly segment the dense point cloud scenes collected in an urban environment and according to our experiments, the model can be adapted to segment point clouds with different data characteristics. To demonstrate the utility of the segmented dense map, I have proposed a real-time localization algorithm, which is able to register the Lidar data of a self-driving vehicle to the dense map.

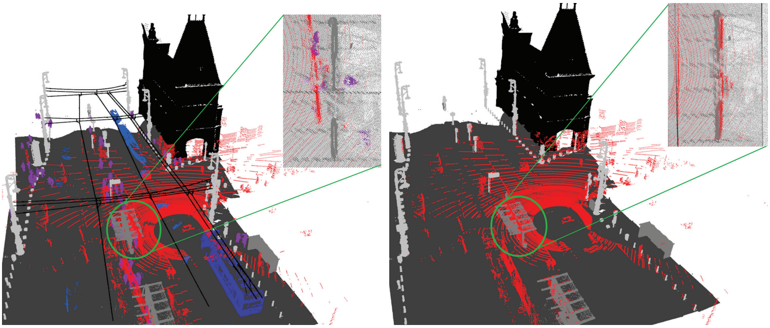
> *1.1. I have proposed a voxel-grid representation of the point cloud data containing two feature channels derived from the density and height properties of the given point cloud segments.*

> *I have shown that the proposed voxelized data is a compact representation of the point cloud which can be used as a direct training input of CNN architectures. To experimentally validate the proposed approach and to show the benefits of the introduced voxel data representation I have created a new manually labeled mobile laser scanning (MLS) database and I have made it publicly available.*

Several point cloud segmentation approaches exist in the literature [15, 16, 17, 18], however, most of them perform weakly in cases of inhomogeneous point density and most of them do not consider the global position of the given point cloud segments, but treat them just local independent training samples. The proposed data representation, before voxelizing the data, takes a local point neighborhood of the given sample, then it assigns a density and the global position value to each voxel of the sample referred to as a first and second channel. I have shown that the two-channel voxelized data is a compact representation of the raw point cloud and it can be efficiently learned using CNN architectures. To train and test deep neural networks I have implemented an efficient tool for point cloud annotation and I have created a large dataset for point cloud segmentation which contains around 500 million manually labeled points. I have made the dataset publicly available (`http://mplab.sztaki.hu/geocomp/SZTAKI-CityMLS-DB.html`).

> *1.2. For segmenting 3D point cloud data, I have proposed a new LeNet-style 3D CNN architecture which is able to efficiently learn the 2-channel voxelized data structure proposed in the previous thesis. I have evaluated the proposed method both on the proposed manually annotated dataset and on various well-known databases. I have shown the advantages of the proposed approach versus different state-of-the-art deep learning based models published in top journals and conferences in the last 5 years.*

I have proposed a new 3D convolutional neural network which is able



(a) Registration result [9] based on the complete map (b) Registration result [9] on CNN filtered map

Figure 2: Application of the proposed CNN classification approach for point cloud registration enhancement. Automatic registration results of a sparse Lidar point cloud (shown with red in all images), to the dense MLS measurements (remaining colors). Figure (a) shows the registration results on the raw point cloud with notable inaccuracies (different class colors in MLS only serve better visibility). Figure (b) demonstrates the output of successful registration based on removing the dynamic objects from the MLS point cloud using the proposed CNN method.

to efficiently extract different features from the voxelized data representation of dense MLS point clouds. I have quantitatively shown that considering the density and global position feature channels, the method is able to overcome other state-of-the-art methods in an inhomogeneous point cloud labeling task.

> *1.3. I have proposed a new real-time point cloud alignment method for point clouds taken with different sensors exhibiting significantly different density properties and I have proposed a vehicle localization technique using the segmented dense point cloud as a high-resolution map. I have experimentally evaluated the proposed method in various urban scenarios.*

Due to the lack of signals, GPS based localization is often unreliable, thus the accuracy of it usually fluctuates between a wide range. According to our experiments, GPS position error can be larger to 10 meters in a dense urban environment such as the streets of Budapest, Hungary. I have proposed an object-based coarse alignment procedure for point cloud registration and I showed that the registration result of the method can be further refined with a point level registration step. I have proposed a new robust key-points extracting method, which is able to extract robust registration key-points from the 3D point cloud segments to make the coarse alignment process more reliable. I have experimentally shown that the proposed coarse alignment method is able to reduce the positioning error of the GPS measurement below 0.4 meter which is a significant improvement against the initial 10 meters translation error.

2. Thesis: I have proposed an automatic, target-less camera-Lidar extrinsic parameter calibration method and I have shown the advantages of the method versus different state-of-the-art algorithms.

Published in [1][6][8]

The main goal of the proposed method is to calibrate a camera and a Lidar sensor mounted onto the top of a moving vehicle in an automatic way. The proposed automatic method works in an end-to-end manner without any user interaction and it does not require any specific calibration objects such as 3D boxes and chessboard patterns. The main advantage of the method that it is able to periodically re-calibrate the sensors on-the-fly i.e. without stopping the vehicle.

> *2.1. I have proposed a redefinition of the camera-Lidar calibration task by generating a 3D point cloud from the consecutive camera frames using a Structure from Motion method and registering the Lidar and the SfM point cloud in the 3D domain.*

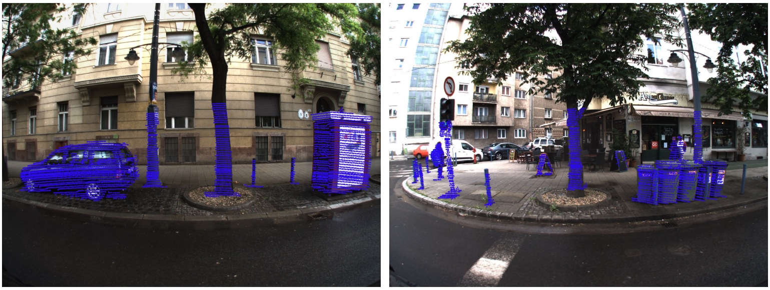


Figure 3: Projection results of the proposed target-less, automatic camera-Lidar calibration method [1].

Detecting meaningful feature correspondences between 2D and 3D domain is very challenging, since extracting the same features, points, or lines from a 2D image and a 3D point cloud domain is unreliable. To avoid this feature association problem, I have proposed a Structure from Motion pipeline to generate a 3D point cloud from the consecutive camera frames. I have formulated the camera-Lidar calibration problem as a point cloud registration task in the 3D domain so that the method aligns the Lidar point cloud to the coordinate system of the generated SfM point cloud. The point cloud registration consists of two main steps: an object-level alignment and a curve-based fine alignment process.

> *2.2. I have proposed a robust object-level registration technique between the 3D point cloud and the generated SfM point cloud data and I have proposed object filtering methods to make more robust the alignment.*

I have modified the object-based coarse alignment process proposed in (Thesis 1) [9], and I have extended it with filtering methods using two state-of-the-art deep neural networks to eliminate noisy dynamic object regions which may erroneously miss-lead the registration process. For registration, I have relied on static street furniture objects such as *columns*, *traffic signs* and *tree trunks*, furthermore based on the deep learning based object filtering the proposed method is able to analyze the scene and it only starts a new re-calibration if an adequate number of static objects are detected in the given scene.

> *2.3. I have proposed a curve-based point cloud registration refinement algorithm which is able to correct the local deformation of the SfM point cloud.*

During the SfM pipeline, local point cloud deformations and scaling errors can occur which may have a great effect on the registration process, therefore I have proposed a control curve based algorithm to eliminate these artifacts. Based on the static objects used in the coarse alignment process I have fitted a NURBS curve both to the Lidar and the SfM point cloud. The control curves describe the shape and the distortion of the point clouds and I have proposed an algorithm which is able to align the curve segments through a non-linear transformation i.e. it deforms the Lidar point cloud according to the shape of the SfM one. As a result of the refinement, the method is able to precisely register the point clouds and it is able the calculate the proper transformation matrix which projects the 3D Lidar points onto the corresponding 2D image pixels. I have quantitatively evaluated and compared the proposed method against state-of-the-art reference techniques on a 10 km long test set. I have chosen 200 different keyframes from various scenarios such as *main roads*, *narrow streets*, etc., and I have shown that the proposed approach is able to be a competitive alternative against state-of-the-art targetless methods and in some cases, it even overcomes the accuracy of target-based methods.

3 Application of the Results

The developed algorithms can be used by various up-to-date or future computer vision systems, especially in the application fields of autonomous driving. Many of the proposed methods directly corresponded to research projects conducted with the participation of SZTAKI and PPCU in the previous years.

Various contributions in RMB Lidar based scene analysis have been adopted in automotive industrial projects. We also integrated the person surveillance module into a real-time demonstrator, which has been introduced at the Frankfurt Motorshow 2017 in the exhibition area of sensor producer Velodyne, at the Automotive Hungary 2017 exhibition, and in multiple Researchers' Night occasions (in Hungarian: Kutatók éjszakája), which are open yearly events for the public to visit research centers in Hungary.

4 Datasets and implementing details

For training and testing I created a dense point cloud dataset called *SZTAKI CityMLS* which is publicly available. To compare and validate the proposed segmentation model I also tested it on the following datasets: *Oakland*, *Paris-rue-Madame* and *TerraMobilita*. These reference datasets show quite different characteristic than the *SZTAKI CityMLS* and they were used only for testing since they contain relatively a small amount of data (less than 10% of our publish dataset).

To evaluate the camera-Lidar calibration algorithm, I selected 200 different scenes from a database of 10 km long road sections. Each scene was associated with a camera image, a corresponding Lidar point cloud, and a reference calibration.

The main platform for point cloud handling and processing was implemented in C/C++ and OpenGL while the neural network models was implemented and train in Python3 with TensorFlow and Keras frameworks. The hardware set up for training contains two Nvidia Geforce RTX 2080 Ti GPU with 2 × 11 GB device memory and 64 GB main memory.

5 Acknowledgements

First of all I would like to express my sincere gratitude to my supervisor Csaba Benedek for his continuous support during my Ph.D study and work, for his motivation and patience. He is leading the way for me since my BSc studies and not just as a supervisor but a friend.
I have been a member of the Machine Perception Research Laboratory (MPLAB) at SZTAKI since my BSc studies, so I would like to special thank to Prof. Tamás Szirányi head of MPLAB for providing remarks and advice regarding to my research and always supporting my work.
Pázmány Péter Catholic University (PPCU) is also gratefully acknowledged, thanks to Prof. Péter Szolgay who provided me the opportunity to study here.
I thank the reviewers of my thesis, for their work and valuable comments.
I thank my closest colleagues from the SZTAKI Machine Perception Research Laboratory for their advice: Csaba Benedek, Andrea Manno-Kovács, Levente Kovács, László Tizedes, András Majdik and Attila Börcs. Special thanks to Levente Kovács and László Tizedes, who supported me their valuable advice and contributions.
Special thanks to Kálmán Tornai from PPCU who supported me during practice leading at PPCU.
Thanks to all the colleagues at PPCU and SZTAKI.
For further financial supports, thanks to the National Research, Development and Innovation Fund, under grant number K-120233, KH-125681 and KH-126688, the European Union and the Hungarian Government from the project 'Thematic Fundamental Research Collaborations Grounding Innovation in Informatics and Infocommunications' under grant number EFOP-3.6.2-16-2017-00013 and 3.6.3 VEKOP-16-2017-00002 and the ÚNKP-19-3 New National Excellence Program of the Ministry of Human Capacities.
Last but not the least, I would like to thank my family: my parents and to my brother for supporting me spiritually throughout my Ph.D years and my life in general.

6 Publications

6.1 The Author's Journal Publications

[1] **B. Nagy** and C. Benedek, "On-the-fly camera and Lidar calibration," *MDPI Remote Sensing*, 2020.

[2] ——, "3D CNN-based semantic labeling approach for mobile laser scanning data," *IEEE Sensors Journal*, vol. 19, no. 21, pp. 10 034–10 045, 2019.

[3] C. Benedek, B. Gálai, **B. Nagy**, and Z. Jankó, "Lidar-based gait analysis and activity recognition in a 4D surveillance system," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 101–113, 2018.

[4] A. Börcs, **B. Nagy**, and C. Benedek, "Instant object detection in Lidar point clouds," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 7, pp. 992–996, 2017.

6.2 The author's International Conference Publications

[5] O. Zováthi, **B. Nagy**, and C. Benedek, "Exploitation of dense MLS city maps for 3D object detection," in *International Conference on Image Analysis and Recognition (ICIAR), (virtual*

conference), ser. Lecture Notes in Computer Science, Póvoa de Varzim, Portugal, 2020, pp. 1317–1321.

[6] **B. Nagy**, L. Kovács, and C. Benedek, "SFM and semantic information based online targetless camera-lidar self-calibration," in *IEEE International Conference on Image Processing, (ICIP)*, Taipei, Taiwan, September 2019, pp. 1317–1321.

[7] Y. Ibrahim, **B. Nagy**, and C. Benedek, "CNN-based watershed marker extraction for brick segmentation in masonry walls," in *16th International Conference on Image Analysis and Recognition, (ICIAR)*, ser. Lecture Notes in Computer Science, vol. 11662. Waterloo, Canada: Springer, August 2019, pp. 332–344.

[8] **B. Nagy**, L. Kovács, and C. Benedek, "Online targetless end-to-end camera-Lidar self-calibration," in *16th International Conference on Machine Vision Applications, (MVA)*. Tokyo, Japan: IEEE, May 2019, pp. 1–6.

[9] **B. Nagy** and C. Benedek, "Real-time point cloud alignment for vehicle localization in a high resolution 3D map," in *European Conference on Computer Vision (ECCV) Workshops*, ser. Lecture Notes in Computer Science, vol. 11129. Munich, Germany: Springer, September 2018, pp. 226–239.

[10] ——, "3D CNN based phantom object removing from mobile laser scanning data," in *International Joint Conference on Neural Networks, (IJCNN)*. Anchorage, USA: IEEE, May 2017, pp. 4429–4435.

[11] B. Gálai, **B. Nagy**, and C. Benedek, "Crossmodal point cloud registration in the Hough space for mobile laser scanning data," in *23rd International Conference on Pattern Recognition, (ICPR)*. Cancún, Mexico: IEEE, December 2016, pp. 3374–3379.

[12] C. Benedek, **B. Nagy**, B. Gálai, and Z. Jankó, "Lidar-based gait analysis in people tracking and 4D visualization," in *23rd European Signal Processing Conference, (EUSIPCO).* Nice, France: IEEE, August 2015, pp. 1138–1142.

[13] A. Börcs, **B. Nagy**, M. Baticz, and C. Benedek, "A model-based approach for fast vehicle detection in continuously streamed urban LIDAR point clouds," in *Asian Conference on Computer Vision, (ACCV), Workshops*, ser. Lecture Notes in Computer Science, vol. 9008. Singapore: Springer, November 2014, pp. 413–425.

[14] A. Börcs, **B. Nagy**, and C. Benedek, "Fast 3D urban object detection on streaming point clouds," in *European Conference on Computer Vision (ECCV) Workshops*, ser. Lecture Notes in Computer Science, vol. 8926. Zurich, Switzerland: Springer, September 2014, pp. 628–639.

6.3 Selected Publications Connected to the Dissertation

[15] G. Pang and U. Neumann, "3D point cloud object detection with multi-view convolutional neural network," in *International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, 2016, pp. 585–590.

[16] W. Wang, R. Yu, Q. Huang, and U. Neumann, "SGPN: Similarity group proposal network for 3D point cloud instance segmentation," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, June 2018, pp. 2569–2578.

[17] J. Huang and S. You, "Point cloud labeling using 3D convolutional neural network," in *International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, 2016, pp. 2670–2675.

[18] C. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, 2017.