Final Report of Project NKFIH KH-17 #125681

Principal Investigator: Dr. Csaba Benedek, Institute for Computer Science and Control (SZTAKI)
Hungarian title: Változásdetekció és eseményfelismerés képi és Lidar mérések fúziójával
English title: Change detection and event recognition with fusion of images and Lidar measurements
Duration: 1. October 2017 – 30. September 2019 (24 Months)

The project team has accomplished research work in the addressed area of multisensory data processing and event analysis. In this report, we summarize the main contributions.

Topic 1: SFM based automatic camera - Lidar calibration

Autonomous driving systems, equipped with 3D LIDAR sensors and electro-optical cameras can achieve accurate and comprehensive environment perception. Accurate LIDAR and camera calibration is essential for robust data fusion. Existing calibration techniques can be grouped based on various aspects: the necessity of user interaction, specific environmental conditions, operational requirements, semi- or fully automatic, target-based (Pusztai, et al. 2018) or targetless (Scaramuzza et al. 2009), offline or online (Moghadam et al. 2013). In self driving applications, however, even a well calibrated system needs some re-calibration due to vibration on the roads and some sensor artifacts, calling for robust online registration techniques, which are able to precisely calibrate LIDAR and camera sensors on the fly.
In the project we proposed and published [C1, C3] a novel targetless fully automatic extrinsic calibration method between a camera and a Rotating Multi-Beam (RMB) LIDAR mounted on a moving car. As pre-condition, we only have to fix the sensors on the vehicle and start driving in a typical urban environment, and the method calculates all necessary registration parameters in situ, online. State-of-the-art competing approaches extract features for correspondence calculation from the observed natural environment without calibration objects. For example, (Scaramuzza et al. 2009), transforms the range sensor's 3D measurement into a so called Bearing Angle (BA) image, and identifies point correspondences between the BA and the camera image. Alternatively, mutual Information was used in (Wang et al. 2012) to calibrate different range sensors with cameras. However, experiments show that the above techniques require a critical point density of the point cloud for reliable operation, which is not ensured at the single RMB LIDAR frames provided by a car during self-driving operation. The correspondences in (Moghadam et al. 2013) are detected based on automatically extracted sets of lines both in the 2D images and in the 3D point clouds, thus this method is preferably used indoors, where the required number of line correspondences can be often observed. However, such conditions cannot be guaranteed in RMB LIDAR point cloud frames recorded in outdoor urban environments, which are notably sparse and their density rapidly decreases as a function of the distance from the sensor. In summary, finding meaningful feature correspondences between the 3D point cloud and the 2D image domain is the main challenge in online, targetless calibration, which we aim to overcome here in a novel way.

To avoid feature (2/3D interest points, line and planar segments) detection we turn to a structure from motion (SfM) based technique to generate point clouds from the image sequences recorded by the moving vehicle (Fig. 1.1a,b), and we perform an alignment between the LIDAR and the generated point clouds. In this way, our main task can be interpreted as a point cloud registration problem (Fig. 1.1 b,e). Most of the conventional point level iterative registration techniques, such as variants of ICP or NDT (Magnusson et al. 2009), may fail when the density characteristic is quite different between the point clouds, and in our case, they can also be misled by false correspondences on the ground caused by the typical ring patterns of RMB LIDAR data. To avoid such artifacts we proposed a robust object level alignment approach between sparse RMB LIDAR point clouds and a dense reference point map [C4]. This technique extracts object blob centers in both point cloud frames, which are matched in the Hough domain, based on the idea of a fingerprint minutiae matching algorithm. Although that approach is able to find a robust transformation between two point sets even if the number of points are different, it becomes sensitive to several false or inaccurate hits of the object detector, which are present in our case since both the RMB LIDAR and the SfM point clouds are quite sparse and noisy.

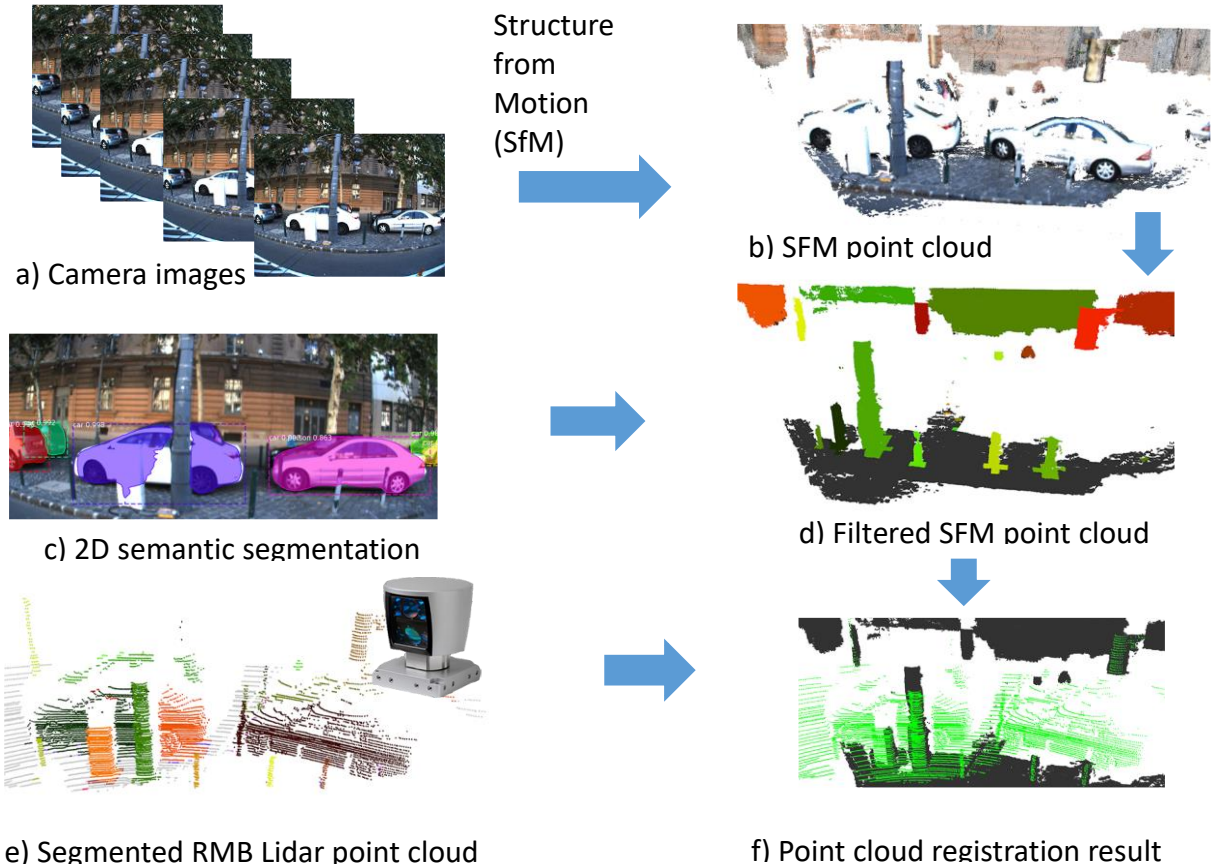

Fig. 1.1 Results of main steps in the proposed object based alignment method. In subfigures (h) and (i) RMB LIDAR data is displayed with green, while the generated SfM point cloud is shown with dark grey

In particularly, we observed that vehicles in the SfM clouds often fall into several pieces due to their homogeneous surfaces (Fig. 1.1 b), causing false matches to the Hough- based estimator [C4]. The next key step is to use *semantic information* for eliminating many of the false object candidates. While object segment classification in sparse point clouds is often unreliable due to occlusion, we can robustly detect vehicle instances in the original camera images with deep neural networks such as *Mask R-CNN* (He et. Al 2017) (see Fig. 1.1 c). Even in deficient SfM clouds, by projecting the 2D class labels into 3D the vehicle points can be efficiently identified) and removed (Fig. 1.1 b,d) helping registration enhancement.

We evaluated the proposed method on a new manually annotated dataset containing 104 time frames of Lidar point clouds and time synchronized image sets with ground truth information. We compared our approach to a state-of-the-art target based offline calibration (Pusztai, et al. 2018) method. To demonstrate the significance of the 2D Mask R-CNN-based semantic filtering of the SfM point cloud, we also compared two variants of the proposed method: first we matched the LIDAR frame to the full SfM point cloud (see Fig. 1.1 b), second we eliminated vehicles from the generated SfM data before point cloud matching, by propagating the semantic labeling information of the Mask R-CNN through the SfM pipeline (Fig. 1.1 d).

Pixel level projection errors and standard deviations are shown in Table 1.1, and some qualitative results are in Fig. 1.2. Advantages of applying the Mask R-CNN filter are observable at each stage of the evaluation. Although numerical results show that the offline target-based calibration method can ensure higher accuracy, calibrating the camera and the LIDAR with (Pusztai, et al. 2018) is a lengthy process, taking more than one hour. When parameters change during measurements (e.g., sensor displacement) one needs to stop driving and repeat the offline calibration process.

Another artifact of conventional offline calibration (Pusztai, et al. 2018) comes from platform motion: due to the nature of the RMB scanning, as the speed of the sensor increases the shape of the point cloud gets distorted. Since offline calibration can only be performed with a static vehicle, its accuracy may decrease as the car moves with higher speed. The effect of this phenomenon is also shown in Table 1.1.

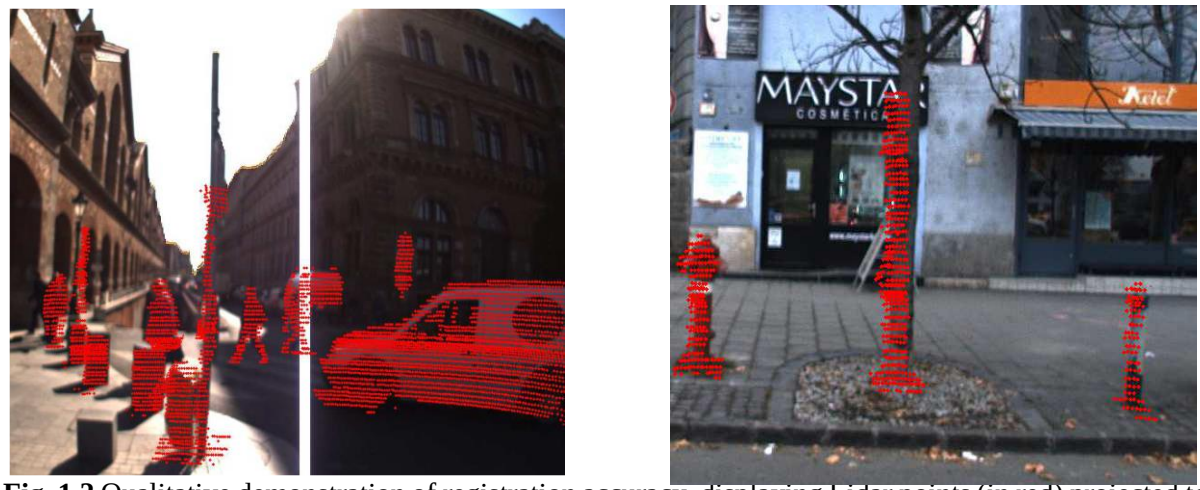

Fig. 1.2 Qualitative demonstration of registration accuracy, displaying Lidar points (in red) projected to the camera images.

Set*	Method	x-error (pixel)		y-error (pixel)	
		Avg.	**Dev.**	**Avg.**	**Dev.**
Slow	Target-based ref.	**2.87**	**0.47**	**3.57**	0.86
	Prop. on raw SfM	6.62	1.35	7.69	1.01
	Prop. by Mask R-CNN	5.35	0.98	5.97	**0.65**
Fast	Target-based ref.	**4.78**	**1.04**	6.21	1.03
	Prop. on raw SfM	6.75	1.28	7.43	0.97
	Prop. by Mask R-CNN	5.49	1.17	**5.78**	**0.87**

Table 1.1 Performance comparison of the target-based (supervised) reference technique and the proposed automatic targetless self-calibration approach without and with using the semantic segmentation (Mask R-CNN) filter

The proposed method calculates the correspondences between camera and LIDAR online during the operation of the vehicle and calculations can be repeated online periodically, thus, the average 5-6 pixel error can be acceptable considering we process camera images with relatively large resolution (1288×964). At this resolution with 5-6 pixel error we are able to robustly assign the 3D objects to the corresponding image regions using the calculated projection matrix, and this data fusion enables the autonomous vehicles to extract more visual features from the surroundings.

Topic 2: Point cloud registration from different sensors, and 3D CNN-based scene segmentation

With also exploiting the co-funding of the NKFI K-120233 project (also led by Csaba Benedek), we proposed new methods for multimodal point cloud registration and point cloud scene analysis. First, we introduced a new technique for the registration of point clouds obtained by various mobile laser scanning technologies. Our new solution [C4, C5, C6] is able to robustly register the sparse point clouds of the self driving vehicles to the High Definition maps based on dense MLS point cloud data, starting from a GPS based initial position estimation of the vehicle (Fig. 2.1). The main steps of the method are robust object extraction and transformation estimation based on multiple keypoints extracted from the objects, and additional semantic information derived from the MLS based map. We tested our approach on roads with heavy traffic in the downtown of a large city with large GPS positioning errors, and showed that the proposed method enhances the matching accuracy with an order of magnitude. Comparative tests were provided with various keypoint selection strategies, and we showed the superiority of the new model against state-of-the-art techniques.

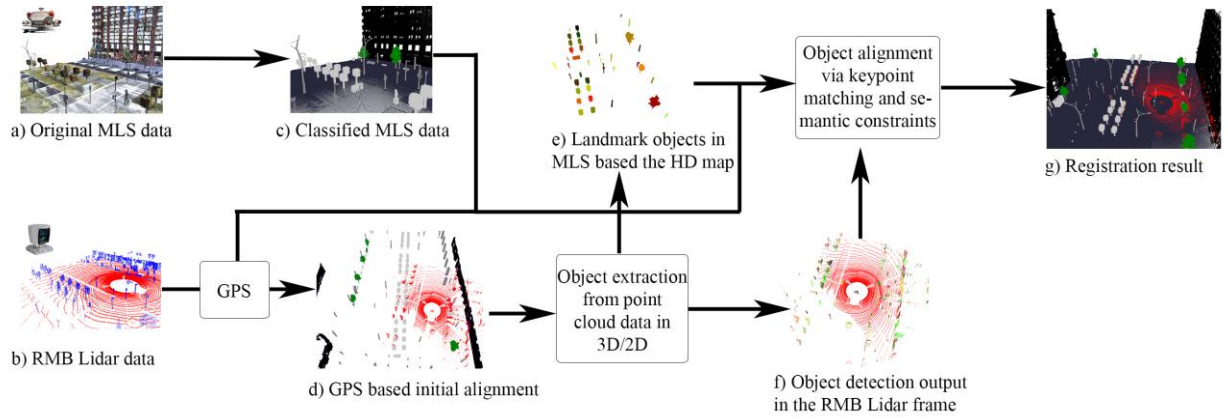

Fig. 2.1 Workflow of the proposed multi-modal point cloud registration approach [C4]

Second, we proposed a 3D convolutional neural network (CNN) based method to segment point clouds obtained by mobile laser scanning (MLS) technologies [J2, C7] enabling to recognize nine different semantic classes required for High Definition (HD) map generation: phantom, tram/bus, pedestrian, car, vegetation, column, street furniture, ground and façade (Fig. 2.2). We used a two channel data input featuring local point density and elevation; and a voxel based space representation, which can handle the separation of tree crowns or other hanging structures from ground objects more efficiently than the earlier pillar based model. To keep the computational requirements low, we implemented a sparse voxel structure avoiding unnecessary operations on empty space segments. We evaluated our proposed method against three reference techniques in qualitative and quantitative ways.

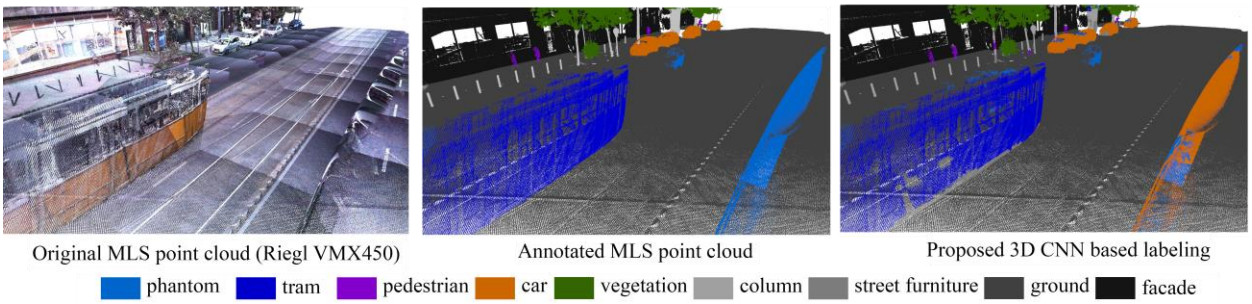

Fig. 2.2 3D CNN based MLS point cloud segmentation: comparison of manual annotation result (center) and the output of the proposed method (right).

Topic 3: Robust target classification for passive radar (ISAR) images

We presented a new approach on an automatic and robust, image feature based target extraction and classification for multistatic passive Inverse Synthetic Aperture Radar (ISAR) range/cross-range images [J3]. The method can be used as a standalone solution or for augmenting classical signal processing approaches. By extracting textural, directional and edge information as low-level features, a fused saliency map is calculated for the images and used for target detection. The proposed method uses the contour and the size of the detected targets for classification, is

lightweight, fast and easy to extend. The performance of the approach is compared with machine learning methods and extensively evaluated on real target images. For evaluation, we used a dataset of 294 images, and experienced an average recognition rate of 69%. We also performed evaluations to showcase one of the main benefits of the proposed method, that it is robust against rotation changes and it can classify images of known targets containing previously unseen distortions with higher reliability than the other approaches. As we expected, in case of the proposed approach the inclusion of the rotated versions of the dataset images did not provide considerable changes, some of the other approaches improved, but overall the proposed method remained the better performer. We also measured the time performance of the methods. Our intent was to show that the proposed method is viable from a practical usage point of view, and its processing times make it suitable for implementation in a practical system.

Topic 4: Object detection from a few LIDAR scanning planes

We have developed [J1] a 3D shape recognition algorithm that can identify objects from only one or a few planes of a LIDAR scanner (Fig. 4.1). We demonstrated the effectiveness of the method on a large public database. Vehicles are often equipped with 2D or a few layer 3D LIDAR sensors for safety considerations, and also to increase the intelligence of the machines. In case of 2D LIDARs, a very small amount of information can be obtained from 2D contour points recorded in a plane, this is hardly usable for recognition. For 3D LIDARs (Maturana and Scherer, 2015) there can be a similar situation, if the object is far from the LIDAR, the shape we want to recognize can be visible in only one plane. By examining the literature (Beyer et al., 2017) (Kurnianggoro and Jo, 2017) - in case of recognition from contour points -, we found that the available methods do not consider the fact if a particular object is examined at a certain height (this is not a criterion for our method) or the fact that due to the movement of the object, the imaged scene changes continuously. Thus, our method beside of using a descriptor not yet used for this problem is also novel in taking into account how the shape changes in consecutive frames. In addition to the above, we suggested a simple voting scheme in case there might be more than one scanned plane of the object. We tested the performance of the method on tens of thousands of samples that came from a public database and also used own measurements; and we investigated its performance as a function of how many frames we track and how far the object is located to the sensor. Compared to the previous literature, it proved to successful already in the most basic cases.

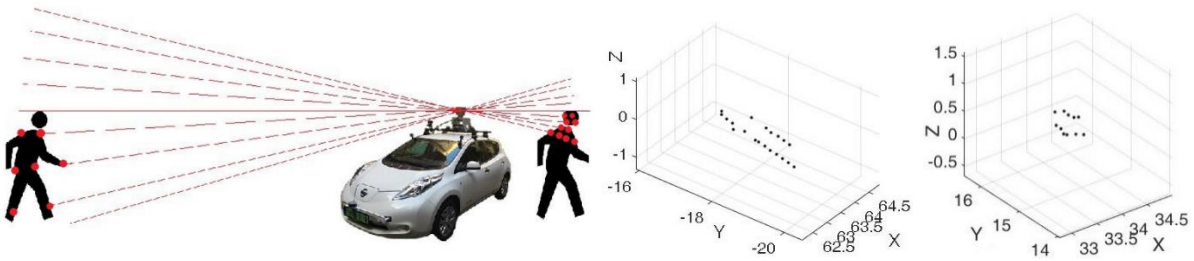

Fig. 4.1. Illustration of the far object problem for RMB Lidars (left), and the provided data from a pedestrian (center) and a car (right) in case of two scan planes

Topic 5: CNN-based Watershed Marker Extraction for Brick Segmentation in Masonry Walls

Nowadays there is an increasing need for using artificial intelligence techniques in image-based documentation and survey in archeology, architecture or civil engineering applications. Brick segmentation is an important initial step in the documentation and analysis of masonry wall images. However, due to the heterogeneous material, size, shape and arrangement of the bricks, it is highly challenging to develop a widely adoptable solution for the problem via conventional geometric and radiometry based approaches.

In our work published in [C2], we proposed a new technique which combines the strength of deep learning for brick seed localization, and the Watershed algorithm for accurate instance segmentation. More specifically, we adopt a U-Net-based delineation algorithm for robust marker generation in the Watershed process, which provides as output the accurate contours of the individual bricks, and also separates them from the mortar regions. For training the network and evaluating our results, we created a new test dataset which consist of 162 hand-labeled images of various wall categories.

Quantitative evaluation is provided both at instance and at pixel level, and the results are compared to two reference methods proposed for wall delineation, and to a morphology based brick segmentation approach.

The experimental results showed the advantages of the proposed U-Net markered Watershed method, providing evaluation rates (recall, and precision, F-measure) between 70% to 97% in every test category.

Fig. 5.1. Results of the proposed brick segmentation method for three difference wall images

.

Further scientific activities and events related to the project

During the project, the PI (Csaba Benedek) prepared and submitted the **Doctor of Sciences Thesis**, which is currently under review with the opponents.

In August 2018, the PI finished his second 3-year **Bolyai Research Scholarship** project, and in connection, he was awarded **Bolyai Plaquette**, and also delivered an **invited lecture** about his research work on the Bolyai Day 2019 in the Ceremonial Hall of the Hungarian Academy of Sciences.

During the project, the PI won two further prestigious awards:

- **Imreh Csanád Plaquette**, which is the bi-annualy issued supervisor award of the National Scientific Students' Associations Conference (OTDK), Section Information Technology (2019)
- **Főtikári Kutatói Elismerés** (in Hungarian), Acknowledgment for Researchers by the Secretary-General of the Hungarian Academy of Sciences (2018)

Csaba Benedek became the **President** of the Hungarian Image Processing and Pattern Recognition Society (**KÉPAF**) and the Hungarian **Governing Board Member** of the International Association for Pattern Recognition (**IAPR**) in 2019. He was a **co-chair of the program committe**e of the KÉPAF 2019 conference.

Csaba Benedek received the **habilitation degree** from the Pázmány Péter Catholic University in 2017. He became a **Senior member of IEEE** in 2018.

The first **doctoral student** of the PI, Attila Börcs received the PhD degree in 2018. His other two students, Balázs Nagy and Yahya Ibrahim accomplished the complex exam in 2018 and 2019, respectively, and entered the second stage of their PhD studies.

Csaba Benedek was a **member of the evaluation committee** for postdoctoral project proposals of the National Research, Development and Innovation Office in 2018.

Csaba Benedek became an **Associate Editor** of *Elsevier Digital Signal Processing*, and a **Guest Editor** of *MDPI Remote Sensing*, Special Issue "3D Urban Modeling by Fusion of Lidar Point Clouds and Optical Images", the topic of the later special issue largely overlaps with the topic of the present KH17 project.

The PI has also presented the project results at the AutoSens, leading Automotive Sensor and Perception Conference & Exhibition in Brussels, 2018.

Publications connected to the project results (participants are underlined):

[J1] Z. Rozsa and T. Sziranyi, "Object detection from a few LIDAR scanning planes," in *IEEE Transactions on Intelligent Vehicle*, in print 2019, Open Access, https://ieeexplore.ieee.org/document/8818349

[J2] B. Nagy and Cs. Benedek: ”3D CNN Based Semantic Labeling Approach for Mobile Laser Scanning Data,” *IEEE Sensors Journal*, vol. 19, no. 21, pp 10034 – 10045, 2019, IF: 3.076*
Open Access, https://ieeexplore.ieee.org/document/8756228

[J3] A. Manno-Kovacs, E. Giusti, F. Berizzi and L. Kovács, „Image Based Robust Target Classification for Passive ISAR”, *IEEE Sensors Journal*, vol. 19 , no. 1 , pp 268 – 276, 2019, IF: 3.076*
Open Access, https://ieeexplore.ieee.org/document/8501978

[C1] B. Nagy, L. Kovács and Cs. Benedek: "SFM and Semantic Information Based Online Targetless Camera-Lidar Self-Calibration", *IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, 22-25 September, 2019

[C2] Y. Ibrahim, B. Nagy and Cs. Benedek: "CNN-based Watershed Marker Extraction for Brick Segmentation in Masonry Walls", *International Conference on Image Analysis and Recognition* (ICIAR), Waterloo, Canada, August 27-29, 2019, Lecture Notes in Computer Science, Springer, 2019

[C3] B. Nagy, L. Kovács and Cs. Benedek: "Online Targetless End-to-End Camera-LIDAR Self-calibration", *International Conference on Machine Vision Applications (MVA)*, Tokyo, Japan, 27-31 May, 2019

[C4] B. Nagy, and Cs. Benedek: "Real-time point cloud alignment for vehicle localization in a high resolution 3D map", *Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving at ECCV*, Munich, Germany, Sept 14, 2018, to appear in Lecture Notes in Computer Science, Springer, 2018

[C5] Cs. Benedek and B. Nagy, "Lidar based environment perception and mapping for autonomous vehicles", AutoSens, Brussels, Belgium, 18-20 September, 2018, Poster presentation

[C6] Ö. Zováthi, B. Nagy and Cs. Benedek: Valós idejű pontfelhőillesztés és járműlokalizáció nagy felbontású 3D térképen, Képfeldolgozók és Alakfelismerők Társaságának 12. konferenciája, 2019

[C7] B. Nagy and Cs. Benedek: 3D CNN alapú MLS pontfelhőszegmentáció, Képfeldolgozók és Alakfelismerők Társaságának 12. konferenciája, 2019

Other publications cited in the report

(Pusztai, et al. 2018) Z. Pusztai, I. Eichhardt, and L. Hajder, "Accurate calibration of multi-lidar-multi-camera systems," in *Sensors*, 2018, vol. 18, pp. 119–152.

(Scaramuzza et al. 2009) D. Scaramuzza, A. Harati, and R. Siegwart, "Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4164–4169, 2007.

(Moghadam et al. 2013) P. Moghadam, M. Bosse, and R. Zlot, "Line-based extrinsic calibration of range and image sensors," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3685–3691, 2013.

(Wang et al. 2012) R. Wang, F. P. Ferrie, and J. Macfarlane, "Automatic registration of mobile lidar and spherical panoramas," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 33– 40.

(Magnusson et al. 2009) M. Magnusson, A. Nuchter, C. Lorken, A. J. Lilienthal, and J. Hertzberg, "Evaluation of 3D registration reliability and speed - a comparison of ICP and NDT," in *IEEE International Conference on Robotics and Automation*, May 2009, pp. 3907–3912.

(He et. Al 2017) K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *IEEE International Conference on Computer Vision (ICCV)*, Oct 2017, pp. 2980–2988.

(Maturana and Scherer, 2015) D. Maturana and S. Scherer, "VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.

(Beyer et al., 2017) L. Beyer, A. Hermans, and B. Leibe, "Drow: Real-time deep learning-based wheelchair detection in 2-D range data," *IEEE Robotics and Automation Letters*, vol. 2, no. 2, pp. 585–592, April 2017

(Kurnianggoro and Jo, 2017) L. Kurnianggoro and K. H. Jo, "Object classification for LIDAR datausing encoded features," in 10th *International Conference on Human System Interactions (HSI)*, July 2017, pp. 49–53.