Digital Signal Processing 119 (2021) 103193

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Positioning and perception in LIDAR point clouds

Csaba Benedek, Andras Majdik, Balazs Nagy, Zoltan Rozsa, Tamas Sziranyi *

Machine Perception Research Laboratory (MPLab), Institute for Computer Science and Control (SZTAKI), Eötvös Loránd Research Network (ELKH), Kende u. 13-17, H-1111 Budapest, Hungary

ARTICLE INFO

Article history:
Available online 3 August 2021

Keywords:
Lidar
Object detection
SLAM
Change detection
Navigation

ABSTRACT

In the last decade, Light Detection and Ranging (LIDAR) became a leading technology of detailed and reliable 3D environment perception. This paper gives an overview of the wide applicability of LIDAR sensors from the perspective of signal processing for autonomous driving, including dynamic and static scene analysis, mapping, situation awareness which functions significantly point beyond the role of a safe obstacle detector, which was the sole typical function for LIDARs in the pioneer years of driver-less vehicles. The paper focuses on a wide range of LIDAR data analysis applications of the last decade, and in addition to the presentation of a state-of-the-art survey, the article also summarizes some issues and expected directions of the development in this field, and the future perspectives of LIDAR systems and intelligent LIDAR based information processing.

1. Introduction

This paper gives an overview of the rich applicability of LIDAR sensors from the perspective of signal processing for autonomous driving, including dynamic and static scene analysis. We focus on a wide range of LIDAR data analysis applications, giving a first hand experience about the state-of-the-art and the challenges of a new depth mapping device category.

1.1. Motivation and significance

In recent decades, remarkable progress has been made in sensor development for environment analysis, which greatly influences the scientific progress in the fields of object detection and classification, scene segmentation, and understanding. Light Detection and Ranging (LIDAR) sensors became one of the most widely used sensing technologies in various applications of geo-data analysis, including perception, mapping and localization.

LIDAR is an active remote sensing technology that uses electromagnetic waves in the optical range to detect an object (target), determines the distance between the target and the sensor (range), and measures further physical properties of the target surface such as scattering and reflection [1]. The sensor calculates the distance of the target objects from the echo time of the emitted and the detected laser beam where the beam spreads with the speed of light. The result of the measurement is a highly accurate 3D point cloud where the coordinates of the points are given in a local or global coordinate system depending on the type of the LIDAR system and the application area.

LIDAR scanners can be mounted either to static terrestrial stations or to ground based and aerial moving vehicles. By using terrestrial LIDAR sensors, high density point clouds and notably accurate and largely detailed 3D models can be created, which properties are required in architectural and engineering applications. Mobile laser scanning (MLS) allows quick surveys of the road network and environment, furthermore, it can contribute to the localization and control of mobile robots and autonomous vehicles.

This paper addresses the main aspects of the broad application area of mobile LIDAR sensors in autonomous driving related fields. LIDARs have special roles in autonomous driving and transportation and special vehicle based intelligent control systems, as they are used parallel to the camera systems. Novel high-resolution pieces can be built in the car's body, and they give an important accessory to the on-board safety. Although LIDARs are currently the most expensive pieces of the on-board sensor systems, the prices are going down quickly, while the application areas are rapidly expanded. The authors, working in the Machine Perception Research Laboratory of SZTAKI, Hungary, have focused on a wide range of LIDAR data analysis applications for several years, thus in addition to the presentation of a state-of-the-art survey, this article also summarizes their first hand experiences in the field.

Today the LIDAR itself is still most frequently considered as the sensor of safety, since its usage is mainly limited to reliable free

* Corresponding author.
E-mail address: sziranyi.tamas@sztaki.hu (T. Sziranyi).

https://doi.org/10.1016/j.dsp.2021.103193

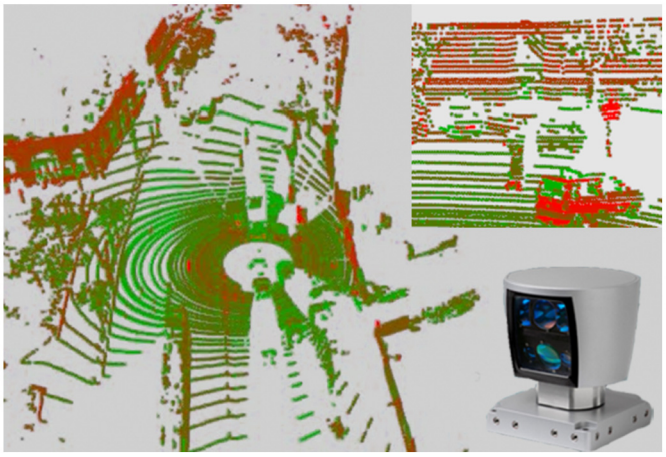


Fig. 1. Data sample of a Velodyne HDL-64 RMB Lidar.

Fig. 2. Data of a Riegl VMX-450 Mobile Laser Scanning (MLS) system.

space verification. However, as we will demonstrate in this study, the potential of this technology goes far beyond simple obstacle detection, since the development of LIDAR technology in terms of temporal and spatial resolution and noise elimination led us to more sophisticated 3D measurements for various real-time perception, navigation and mapping problems.

Next, we introduce the reader to the latest exciting results and their background in real-life LIDAR applications.

1.2. Outline of the paper

First, we show the diversity of laser scanner devices to get a point cloud of the 3D environment. Next, we present an overview of a wide range of LIDAR-based application modules built on each other, which implement various functions, including object perception, classification, mapping and localization. We also discuss the opportunities in challenging situations such as extreme weather conditions or the availability of low-range one-beam (plane) sensors only. Finally, we show the on-the-fly calibration of the LIDAR and camera system.

2. LIDAR sensors and resources

2.1. LIDAR sensor types

LIDAR equipments give us a versatile application and operational richness: static/mobile, 360°/wide angle/narrow scan, equidistant scanning resolution/special beam-patterns, single echo/multiple echos. We will see that LIDARS can be used in any area of imaging the world.

Rotating Multi-beam (RMB) Lidar systems provide a 360° field of view of the scene, with a vertical resolution equal to the number of the sensors, while the horizontal angle resolution depends on the speed of rotation. Although RMB Lidars can produce high frame-rate *point cloud videos* enabling dynamic event analysis in the 3D space, the measurements have a low spatial density, which quickly decreases as a function of the distance from the sensor, and the point clouds may exhibit particular patterns typical to sensor characteristic (see Fig. 1). In special cases, only one or a few beams are available.

Mobile laser scanning (MLS) platforms equipped with time synchronized Lidar sensors and navigation units can rapidly provide very dense and accurate *static* point clouds from large environments, where the 3D spatial measurements are accurately registered to a geo-referenced global coordinate system (Fig. 2). These point clouds may act as a basis for detailed and up-to-date 3D High Definition (HD) maps of the cities, which can be utilized by self-driving vehicles for navigation.

Another recently emerging new technology is the Doppler-LIDAR (First mention is [2] for wind measurements): e.g., Blackmore[1] has just introduced a LIDAR for autonomous driving with velocity or rotation speed data output. Very recent models, such as the Livox sensors, use advanced non-repetitive scanning patterns to deliver highly-accurate details. These scanning patterns even provide relatively high point density in a short period of time, which can build up a higher density as the duration increases. The actually available models can achieve the same or greater point density as conventional 32-line RMB LIDAR sensors.

2.2. LIDAR resources

Numerous autonomous driving datasets have been released in the recent years with LIDAR data. The most important ones are listed in Table 1 with their purpose. As we can observe, for typical benchmark problems, such as object detection, tracking or Simultaneous Localization and Mapping (SLAM), one can choose between various public resources, corresponding to different sensor characteristic and scenario types. A main challenge in the future, however, will be the timely completion of the available benchmark datasets with reliable measurement and ground truth information, following the appearance of newer and newer LIDAR sensors technologies.

3. LIDAR based object perception

Object perception and recognition is a central objective in LIDAR based 3D point cloud processing. Though several 3D object detection and classification approaches can be found in the literature, due to the large differences in data characteristics obtained by different LIDAR sensors, object perception methods are still strongly sensor dependent making very challenging to adopt them between different types of LIDAR data.

Since LIDAR sensors provide very accurate 3D geometric information, the localization and shape recognition of the objects can be more intuitively compared to 2D image processing. However, beyond the different sensor data characteristics, several challenges occur in automatic LIDAR-based object detection and classification, such as the sparsity of the data, variable point density, non-uniform sampling and in addition, in cluttered scenes objects often

[1] https://blackmoreinc.com.

Table 1
LIDAR datasets with different purposes.

Name of the dataset	Lane detection	Object detection/tracking	Segmentation	Localization and mapping
Standford Track [3]			X	
Ford [4]				X
KITTI [5]	X	X	X	X
Málaga urban [6]				X
Oxford RobotCar [7]				X
Apolloscape [8]	X	X	X	X
KAIST Urban dataset [9]				X
KAIST Multispectral [10]		X		
Multivehicle Stereo Event [11]		X		
UTBM RoboCar [12]				X
Unsupervised Llamas (Bosch) [13]	X			
PandaSet*		X	X	
BLVD [14]		X		
H3D (Honda) [15]		X		
Lyft level 5 [16]		X		
NuScenes [17]		X	X	
Waymo Open [18]		X		
Argoverse [19]	X	X		
SZTAKI-Velo64Road [20]		X		
SZTAKI CityMLS [21]		X	X	

* https://scale.com/open-datasets/pandaset

occlude each other causing partially extracted object blobs in the measurements.

Based on the object perception literature, we can define two main groups: traditional geometry based methods and deep learning based approaches. To handle the expensive calculations between huge amount of 3D points geometry based methods usually adopt some space partitioning techniques such as Kd-tree, Octree [22,23] or 3D voxel [24] and 2D grid based methods [25]. Some approaches apply different region growing techniques over tree-based structures to obtain coherent objects. The authors of [26] present an Octree based occupancy grid representation to model the dynamic environment surrounding the vehicle and to detect moving objects based on inconsistencies between scans.

In general, building and maintaining a tree-based structure is expensive, so usually, some kind of 3D voxel or 2D grid approaches are applied for streaming data. In [25] the authors propose a fast segmentation of point clouds into objects, which is accomplished by a standard connected component algorithm in a 2D occupancy grid, and object classification is done on the raw point cloud segments with 3D shape descriptors and a SVM classifier. Different voxel grid structures are also widely used to complete various scene understanding tasks, including segmentation, detection and recognition [24]. The data is stored here in cubic voxels for efficient retrieval of the 3D points. Among geometry based 2D grid approaches [20] implements a pipeline of a geometry based ground separation step, a two-layer grid structure based an efficient object extraction, and a deep learning based object classification which represents the extracted objects in the range image domain.

Other recent techniques focus on deep learning based object detection and classification in 3D point clouds. VoxelNet [28] is able to predict accurate bounding boxes utilizing discriminative feature learning. PointPillars [27] (Fig. 3) is a state-of-the-art real-time object detection method, which can predict object-candidates from multiple classes, together with their 3D oriented bounding boxes and class confidence values.

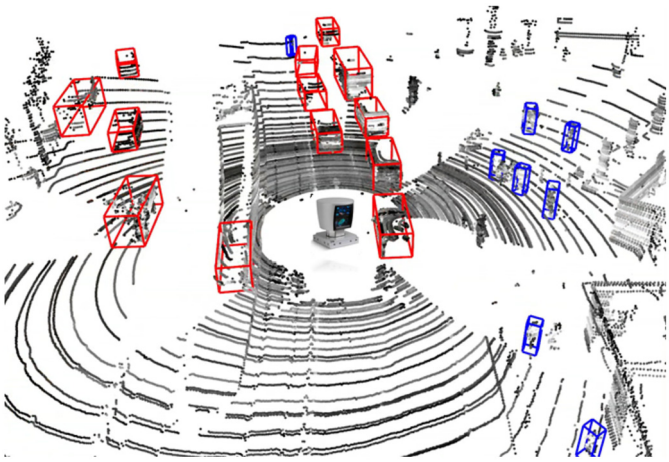


Fig. 3. LIDAR Object detection results with deep learning based PointPillars approach [27]. Red boxes show detected vehicles, blue boxes pedestrians. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

4. The limits of usage: low-resolution LIDAR perception and extreme circumstances

The previous section (Sect. 3) shows that Lidar pattern evaluation can result in semantic interpretation; now we see that a very limited (diluted) information of LIDAR scans can also be used for accurate perception. In this section, limitations of LIDAR sensors, resulted challenges, and current solutions are discussed. Besides the developing high-end 3D LIDAR sensors, it is also worth investigating the capability of sensors of lower or extremely low resolution (equipped with a few or even only one laser channel) because of cost-efficiency and increasing robustness. As one of the main effects of extreme circumstances (e.g., harsh weather) is information loss; and installing more than one planar or a few-layer LIDARs in task-based optimized positions [29] may result in a better alternative (in some point of view) than using one with high-resolution. The limited information content of scanners with low vertical resolution makes a high-level semantic interpretation of the data more challenging and makes the easier real-time running of the algorithms. Naturally, machine vision in this subfield has gone through rapid development in recent years as well.

4.1. Vision with LIDARs of extremely low resolutions

2D range scanners and their applications have a relatively long history in robotics [30]. Automated Guided Vehicles (AGVs) have been using these sensors for decades for safety and navigation purposes. Today, there are products available in the market with extremely high horizontal angular resolution, high scanning frequency, and safety guarantee of the manufacturer. Also, fully devel-

(a) SICK LMS500 sensor (b) Frame acquired by 2D LIDAR sensor

Fig. 4. Examples of planar LIDAR sensor and point cloud acquired by a planar LIDAR.

oped, real-time scan matching and Simultaneous Localization and Mapping algorithms [31] available in industrial and market products based on only 2D laser scanner data; making localization and mapping one of their primary application area.

Besides navigation, we would like to extract as much information from the available data as possible about the environment. The point cloud processing algorithms can directly utilize point clouds of these 2D scanners. So there are various solutions for object detection and recognition from this type of data. One can use handcrafted general [32], data specific features [33], image descriptors [34] or neural networks [35]. Applying LIDARs with (still very) narrow vertical field of view [36] makes the information content richer and semantic interpretation easier.

We can distinguish two different reasons that result in dealing with low (vertical) resolution LIDAR point clouds. The first case is when the hardware limits the resolution (LIDAR layers) because we measure with a planar LIDAR or one with 4-8 layers. The second case when our LIDAR has a sufficient number of layers (16 and above), but its usage scenario limits the acquired point cloud's vertical resolution. A typical example of this kind of scenario is a highway. We would like to look far (because of the high speed and straight road sections), but distant objects will occupy only a very few layers of our LIDAR (with high vertical resolution) measurement. We will show what kind of solutions have been developed to these two particular cases of low-resolution LIDAR perception recently.

4.1.1. Hardware limited low-resolution perception

As mentioned earlier, planar and narrow vertical field LIDARs (a 2D LIDAR sensor and frame acquired by it in tilted position is illustrated in Fig. 4[2]) are frequently used in logistic transportation systems (on AGVs), not just for navigation but for specific purposes (e.g., overhang detection, see Fig. 5). The sensors with specific positions, the speed of the transport vehicle (about 1 m/s), and the presence of positioning sensor [37] in the vehicle (for navigation purpose) are making adequate to use the 2D sensor data in a 3D reconstruction. This will result in a very special partial point cloud data, incrementally giving more and more information about objects. We proposed a solution to deal with this specific kind of LIDAR data in [38].

The proposed method's main idea to deal with partial clouds to compare statistics of local structures. The steps of it are summarized below: First, local surface definition around each point is needed. We measure the saliency of the point by 3D Harris [39] operator. Next, to determine a repeatable number of keypoints, a local scale is assigned to significant points, and a local surface descriptor characterizes keypoints. After it, we define local patterns as graphs of keypoints. In the last step, the frequency of local patterns is compared.

[2] https://www.sick.com.

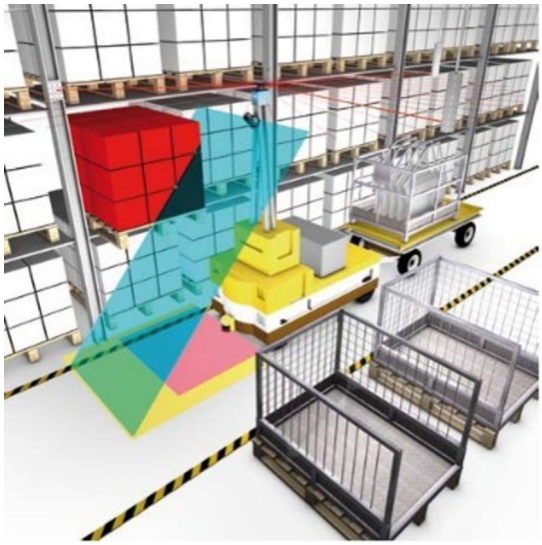


Fig. 5. Tilted sensor installation for overhang detection. Photo source: SICK - Efficient solutions for material transport vehicles in factory and logistics automation.

Besides our measurements, we used an MLS database for real-life testing. (These types of point clouds are acquired similarly as described above.) In this database, we measured 73.3% classification accuracy for 5 classes when only about 20% of the 3D object was visible, and 80.0% accuracy for 30% visibility which results in a usable and safe incremental prediction for early decisions. For more details, see [38].

4.1.2. Scenario limited low resolution perception

It follows from the LIDAR measurement principle that the density of the acquired point cloud decreases with distance from the sensor. Resulting in the phenomenon of measuring a high-resolution sensor but distant objects are not observable in sufficient resolution. In the case of a sensor with a lower number of channels, it happens in the near-field too. (The case when we perceive an object only in one layer is also not rare.)

In this type of object, local surface information is not extractable, so we cannot expect methods based on that, designed for 2.5D point clouds, to work. To solve this problem, we relied on methods designed for 2D point clouds and extended them to utilize all the information available [40]; we proposed a method to classify objects with this point cloud characteristic constructed from the steps below: First, generating a shape descriptor for object segments using low-frequency components to be robust against angular resolution drop; Then, we extract statistical measures of geometries coding the 3D location of the (approximately) 2D curve. After it, we group tracklets (tracks up to 5 frames) of segments (if there are any available). The next step is classifying of segments (or tracklets of segments) with CNN (Convolutional Neural Network). Finally, an object-level decision is made with maximum likelihood aggregation of segment class probabilities (if more than one segment is available from an object).

With the proposed method, we reached 96,6% classification accuracy (and even better if we could perceive and track and object segment more than one frame) for 6 categories on such point clouds of the KITTI [5] database where objects were present at maximum 4 LIDAR layers (with 41 m average distance to the sensor). These point clouds cannot be handled with conventional methods (and so ignored in most cases). Illustration of object classification with the proposed method of a typical scenario (highway observed with relatively narrow vertical field LIDAR) is shown in Fig. 6. For further details and experimental proof see [40].

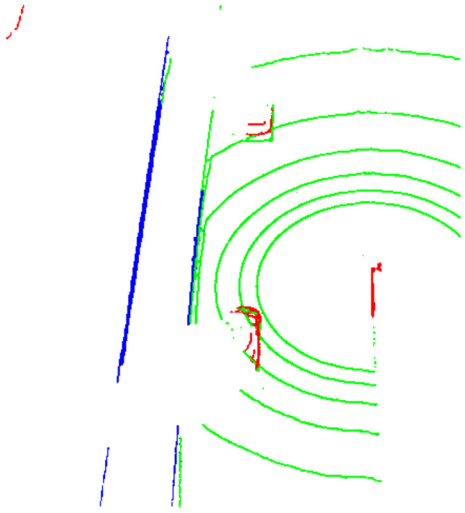


Fig. 6. Classification of objects observed at max 4 layers. Colormap: Red - vehicle, Blue - guard rail, Green - ground.

4.2. Robustness in harsh weather conditions

Harsh weather conditions call for challenging problems: LIDARs have decreased performance in snow/rain or fog. This kind of limitation has to be addressed by semantic-based methods or physically modeled filters.

Recent researches target hardware [41] and software [42] developments to eliminate this effect. To avoid the problem above, alternative devices can be used, noisy measurement has to be filtered [43] and incomplete data has to be completed. Researchers have just started developing the first stages of the solution, recognizing the given weather conditions [44], and examining the influence of different ones [45]. To support that, pursuit datasets in adverse weather have been released latterly [46].

5. LIDAR based localization and mapping

The capability of recognizing patterns in LIDAR point clouds led to high precision odometry techniques in SLAM and similar methods. Next, we will briefly summarize the state-of-the-art algorithms and current trends in LIDAR-based ego-motion estimation, 3D mapping, and localization.

5.1. Visual-odometry using LIDARs

Recently, several visual-odometry algorithms were proposed to compute the motion of a vehicle in real time using only the continuously streaming data gathered by the LIDAR sensor. Creating thus LIDAR-only odometry methods eliminates the need for any other supplementary sensor, e.g., Inertial Measurement Unit (IMU), wheel encoder, and satellite based Global Positioning System (GPS). One of the best performing algorithms in terms of translational and rotational errors on the KITTI [47] dataset is the LOAM [47] algorithm, which estimates the six DoF (Degree of Freedom) displacement of the vehicle on short trajectories with very low drift in scenes of high-density feature points and available reference ground planes. The algorithm can process the measurements robustly for different LIDAR sensors with varying point cloud densities. However, in the case of long trajectories and since the drift is continuously accumulated, a significant error could build up in the position estimation.

5.2. Simultaneous localization and mapping with LIDARs

In order to correct the accumulated error in the odometry backend, loop-closure situations can be detected by place recognition algorithms whenever the vehicle returns to previously visited places in the navigation area. In the case of Simultaneous Localization and Mapping (SLAM) algorithms, it is assumed that the vehicle explores for the first time the given environment, and therefore there is no a priori map to localize itself against. Recently, the SegMap [48] algorithm was proposed to extract and match LIDAR segments in 3D point clouds. SegMap computes a data-driven compact descriptor to extract distinctive and meaningful features from point cloud segments in order to identify loop-closure situations along the trajectory.

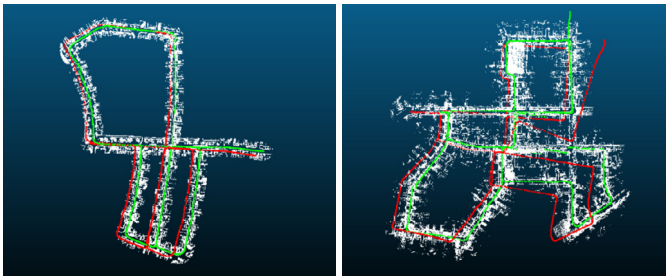
In order to increase the robustness and precision of the localization algorithm in feature-poor environments, a framework was proposed in LIO-SAM [49] to tightly couple LIDAR and inertial measurements obtained from an IMU. Also, the proposed architecture allows the integration of GPS measurements in case these are available. Further on, by adapting the factor graph optimization framework the LIDAR Inertial Sub-system (LIS) was fused with a traditional monocular-based Visual Inertial Sub-system (VIS) to create a Lidar-Visual-Inertial (LVI-SAM) localization and mapping system [50]. Conversely to these methods, next we will show the outcomes of a LIDAR-only odometry and localization method for urban environments where a target map exists to localize within.

5.3. LIDAR-only odometry and localization in 3D point cloud maps

Accurate 3D city models and high-definition maps are becoming increasingly available with recent mapping technology advancements. In addition, in many real-world applications, maps are available to localize against. Therefore, these should be utilized to correct the accumulated drift along the vehicle's trajectory whenever a geometrically similar location is detected between the online 3D point cloud and the offline map.

In [51] we proposed LOL, a LIDAR-only Odometry and Localization algorithm that integrates the advantages of the LOAM [47] odometry and the SegMap [48] algorithm. In the odometry backend, the LOAM algorithm estimates the six DoF odometry in real time based only on the continuously streaming point cloud data from a Velodyne LIDAR sensor. In a scene of high-density feature points and available reference ground planes, the algorithm computes the displacement of the vehicle on short trajectories with very low drift using only the consecutive Lidar measurements. The algorithm can process the measurements robustly for different Velodyne sensors with varying point cloud densities. On the other hand, in the case of long trajectories and since the drift is continuously accumulated, a significant error could be accumulated in the estimation that needs to be canceled by a localization method whenever a correct match is detected between the online Lidar stream and the offline reference map. Therefore, for the localization frontend, we integrated the SegMap method that is a state-of-the-art algorithm for the extraction and matching of 3D point cloud segments.

We also included some additional improvements in the acceptance of correct matches by applying further geometrical constraints complementing the feature similarity ones. Namely, once a good match is detected between the online measurements and the target map, we only search for similar 3D Lidar segments (with relaxed similarity constraints) in the neighborhood of the current location defined by the location uncertainty. In addition, we only use the shift between the target map and the online source segments centroids as a prior, and we refine the final transformation by applying a fine-grained ICP matching between the two point clouds.



(a) Drive 18, approx. 2,2km (b) Drive 27, approx. 3,6m

Fig. 7. Results of the LOL localization algorithm with respect to the ground truth map on various length KITTI [5] datasets: LOL algorithm (green line), LOAM trajectory (red line) with respect to the ground truth point cloud map.

We tested the proposed algorithm on several Kitti [5] datasets, cf. Fig. 7, and found a considerable improvement in term of precision without adding a significant computational cost increase.

5.4. Localization in dense LIDAR maps

LIDAR measurements can also be utilized for accurate self localization of self-driving vehicles (SDV) in high resolution 3D point cloud maps of the environment. A solution provided in [52] can robustly register the sparse RMB Lidar point clouds of the SDVs to the dense Mobile Laser Scanning (MLS) point cloud data, starting from a GPS based initial position estimation of the vehicle. 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 previously segmented MLS based map.

6. Semantic segmentation of MLS point clouds

Dense MLS point clouds can act as a basis for detailed and up-to-date 3D High Definition (HD) maps of the cities, which can be utilized by self-driving vehicles for navigation, or by city authorities for road network management and surveillance, architecture, or urban planning. All of these applications require semantic labeling of the data (Fig. 8). While the high speed of point cloud acquisition is a clear advantage of MLS, due to the huge data size yielded by each daily mission, applying efficient automated data filtering and interpretation algorithms in the processing side is crucially needed, which steps still introduce a number a key challenges.

Taking the raw MLS measurements, one of the critical issues is the *phantom* effect caused by independent object motions [21]. Due to the sequential nature of the environment scanning process, scene objects moving concurrently with the MLS platform (such as passing vehicles and walking pedestrians) appear as phantom-like longdrawn, distorted structures in the resulting point clouds [53]. It is also necessary to recognize and mark all movable scene elements such as pedestrians, parking vehicles [54] or trams from the MLS scene. On the one hand, they are not part of the reference background model, thus these regions must be eliminated from the HD maps. On the other hand, the presence of these objects may indicate locations of sidewalks, parking places, etc. Column-shaped objects, such as poles, traffic sign bars [55], tree trunks are usually good landmark points for navigation. Finally, vegetation areas (bushes, tree foliage) should also be specifically labeled [56]: since they are dynamically changing over the whole year, object level change detection algorithms should not take them into account.

While a number of various approaches have already been proposed for general point cloud scene classification, they are not focusing on all practical challenges of the above introduced workflow of 3D map generation from raw MLS data. In particularly, only a few related works have discussed the problem of *phantom* removing. Point-level and statistical feature based methods such as [57] and [58] examine the local density of a point neighborhood, but as noted in [59] they do not take into account higher level structural information, limiting the detection rate of *phantoms*. The task is significantly facilitated if the scanning position (e.g., by tripod based scanning [60]) or a relative time stamp (e.g., using a rotating multi-beam Lidar [61]) can be assigned to the individual points or point cloud frames, which enables the exploitation of multi-temporal feature comparison. However, in the case of our examined MLS point clouds, no such information is available, and all points are represented in the same global coordinate system.

Several techniques extract various object blob candidates by geometric scene segmentation [55,20], then the blobs are classified using shape descriptors, or deep neural networks [20]. Although this process can be notably fast, the main bottleneck of the approach is that it largely depends on the quality of the object detection step.

Alternative methods implement a voxel level segmentation of the scene, where a regular 3D voxel grid is fit to the point cloud, and the voxels are classified into various semantic categories such as roads, vehicles, pole like objects, etc. [56,62,63]. Here a critical issue is feature selection for classification, which has a wide bibliography. Handcrafted features are efficiently applied by a maximum-margin learning approach for indoor object recognition in [64]. Covariance, point density, and structural appearance information is adopted in [65] by a random forest classifier to segment MLS data with varying density. However, as the number and complexity of the recognizable classes increase, finding the best feature set by hand induces challenges.

Deep learning techniques have been widely used for point cloud scene classification in recent years, following either *global* or *local* (window based) approaches. *Global* approaches consider information from the complete 3D scene for classification of the individual voxels, thus the main challenge is to keep the time and memory requirements tractable in large scenes. The OctNet method implements a new complex data structure for efficient 3D scene representation, which enables the utilization of deep and high resolution 3D convolutional networks [66]. From a practical point of view, by OctNet's training data annotation operators should fully label complete point cloud scenes, which might be an expensive process.

Sliding window based techniques are usually computationally cheaper, as they move a 3D box over the scene, using locally available information for the classification of each point cloud segment. The *Vote3Deep* [62] assumes a fixed-size object bounding box for each class to be recognized, which might be less efficient if the possible size range of certain objects is wide. A CNN based voxel classification method has recently been proposed in [63], which uses purely local features, coded in a 3D occupancy grid as the input of the network. Nevertheless, they did not demonstrate the performance in the presence of strong *phantom* effects, which require accurate local density modeling [58,59].

The multi-view technique [67] projects the point cloud from several (twelve) different viewpoints to 2D planes, and trains 2D CNN models for the classification. Finally, the obtained labels are backprojected to the 3D point cloud. This approach presents high quality results on synthetic datasets and in point clouds from factory environments, where due to careful scanning, complete 3D point cloud models of the scene objects are available. Application for MLS data containing partially scanned objects is also possible, but the advantages over competing approaches are reduced here [67].

PointNet++ [68] introduces a hierarchical neural network for point set classification. The method takes random samples within a given radius of the examined point, so it does not exploit density

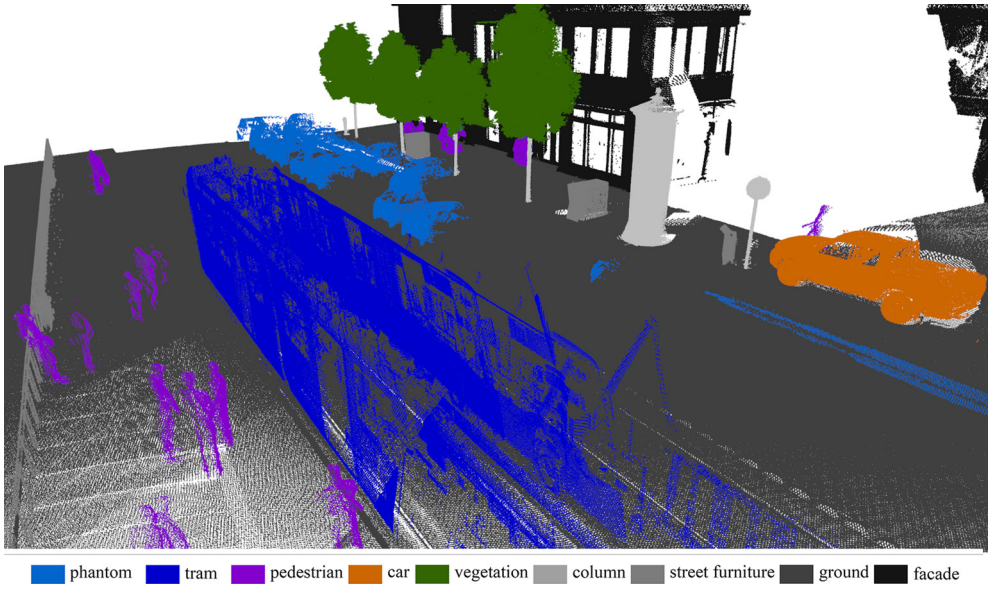


Fig. 8. Labeling result of the proposed 3D CNN based scene segmentation approach (test data provided by Budapest Közút Zrt.)

features. Results are demonstrated on synthetic and indoor data samples, with dense and accurate spatial data and RGB color information.

The *Similarity Group Proposal Network* (SGPN) [69] uses PointNet++ as a backbone feature extractor, and presents performance improvement by adding several extra layers to the top of the network structure. However, as noted by the authors, SGPN cannot process large scenes on the order 10^5 or more points [69], due to using a similarity matrix whose size scales quadratically as the number of points increases. This property is disadvantageous for MLS data processing, where a typical scene may contain over 10^7 points.

The *Sparse Lattice Network* (SPLATNet$_{3D}$) [70] is a recent technique which able to deal with large point cloud scenes efficiently by using a Bilateral Convolution Layer (BCL). SPLATNet$_{3D}$ [70] projects the extracted features to a lattice structure, and it applies sparse convolution operations. Similarly to voxel based approaches, the lattice structure implements a discrete scene representation, where one should address under- and oversegmentation problems depending on the lattice scales.

The C^{2}CNN technique introduced in [21] is based on two-channel 3D convolutional neural network (CNN), and is specifically improved to segment MLS point clouds into nine different semantic classes, which can be used for high definition city map generation. The main purpose of semantic point labeling is to provide a detailed and reliable background map for self-driving vehicles (SDV), which indicates the roads and various landmark objects for navigation and decision support of SDVs. This approach considers several practical aspects of raw MLS sensor data processing, including the presence of diverse urban objects, varying point density, and strong measurement noise of phantom effects caused by objects moving concurrently with the scanning platform. We evaluate the proposed approach on a manually annotated new MLS benchmark set, and compare our solution to three general reference techniques proposed for semantic point cloud segmentation.

A numerical comparison between many of the above mentioned methods is shown in Table 2, using the `SZTAKI CityMLS` Benchmark Set [21].[3]

[3] url: `http://mplab.sztaki.hu/geocomp/SZTAKI-CityMLS-DB.html`.

7. Change detection using onboard Lidar and MLS maps

For self-driving car navigation and environment perception, *change detection* between the instantly sensed RMB Lidar measurements and the MLS based reference environment model appears as a crucial task, which indicates a number of key challenges. Particularly, there is a significant difference in the quality and the density characteristics of the i3D and MLS point clouds, due to a trade-off between temporal and spatial resolution of the available 3D sensors.

In recent years various techniques have been published for change detection in point clouds, however, the majority of the approaches rely on dense terrestrial laser scanning (TLS) data recorded from static tripod platforms [71,72]. As explained in [71], classification based on calculation of point-to-point distances may be useful for homogeneous TLS and MLS data, where changes can be detected directly in 3D. However, the point-to-point distance is very sensitive to varying point density, causing degradation in our addressed i3D/MLS cross-platform scenario. Instead, [71] follows a ray tracing and occupancy map based approach with estimated normals for efficient occlusion detection, and point-to-triangle distances for more robust calculation of the changes. Here the Delaunay triangulation step may mean a critical point, especially in noisy and cluttered segments of the MLS point cloud, which are unavoidably present in a city-scale project. [72] uses a nearest neighbor search across segments of scans: for every point of a segment they perform a fixed radius search of 15 cm in the reference cloud. If for a certain percentage of segment points no neighboring points could be found for at least one segment-to-cloud comparison, the object is labeled there as moving entity. A method for change detection between MLS point clouds and 2D terrestrial images is discussed in [73]. An approach dealing with purely RMB Lidar measurements is presented in [74], which uses a ray tracing approach with nearest neighbor search. A voxel based occupancy technique is applied in [75], where the authors focus on detecting changes in point clouds captured with different MLS systems. However, the differences in data quality of the inputs are less significant than in our discussed case.

In [76] authors proved that change detection can be accelerated if they compare only keyframes to the map or previous frames. Here, keyframes are the ones that contain changes with high probability. [76] proposed a solution to find these keyframes by exploiting mapping residuals. The authors demonstrated the

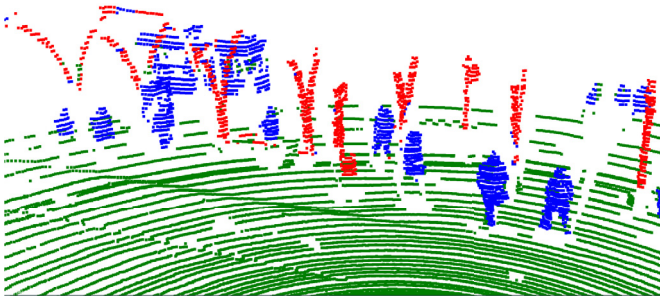
Table 2
Quantitative comparison of various point cloud segmentation techniques [63], [67], [68], [70] and [21] on the SZTAKI CityMLS benchmark set.

Class	OG-CNN [63]			Multi-view [67]			PointNet++ [68]			SPLATNet$^{xyz}_{rgb}$ [70]			C^{2}CNN [21]		
	Pr	Rc	F-r	Pr	Rc	F-r	Pr	Rc	F-r	Pr	Rc	F-r	Pr	Rc	F-r
Phantom	85.3	34.7	49.3	76.5	45.3	56.9	82.3	76.5	79.3	83.4	78.2	80.7	84.3	85.9	**85.1**
Pedestrian	61.2	82.4	70.2	57.2	66.8	61.6	86.1	81.2	83.6	80.4	78.6	79.5	85.2	85.3	**85.2**
Car	56.4	89.5	69.2	60.2	73.3	66.1	80.6	92.7	86.2	81.1	89.4	85.0	86.4	88.7	**87.5**
Vegetation	72.4	83.4	77.5	71.7	78.4	74.9	91.4	89.7	90.5	86.4	87.3	86.8	98.2	95.5	**96.8**
Column	88.6	74.3	80.8	83.4	76.8	80.0	83.4	93.6	**88.2**	84.1	89.2	86.6	86.5	89.2	87.8
Tram/Bus	91.4	81.6	86.2	85.7	83.2	84.4	83.1	89.7	86.3	79.3	82.1	80.7	89.5	96.9	**93.0**
Furniture	72.1	82.4	76.9	57.2	89.3	69.7	84.8	82.9	83.8	82.6	81.3	81.9	**88.8**	78.8	83.5
Overall	76.9	74.2	75.5	72.5	73.4	72.9	85.6	87.5	86.5	82.5	83.7	83.0	**90.4**	**90.2**	**90.3**

Note: Voxel level Precision (Pr), Recall (Rc) and F-rates (F-r) are given in percent (overall values weighted with class significance).

performance of the proposed method in real-life experiments with an AGV equipped with a 2D LIDAR sensor.

In [77] the authors introduced a new technique for change detection in urban environment based on the comparison of 3D point clouds with significantly different density characteristics. This approach extracts moving objects and environmental changes from sparse and inhomogeneous instant 3D (i3D) measurements, using as reference background model dense and regular point clouds captured by mobile laser scanning (MLS) systems (see Fig. 9). The introduced workflow consists of consecutive steps of point cloud classification, crossmodal measurement registration, Markov Random Field (MRF) based change extraction in the range image domain and label back projection to 3D. Experimental evaluation has been conducted in four different urban scenes, and the advantage of the proposed change detection step is demonstrated against a reference voxel based approach.



(a) Detected changes at Kálvin tér, Budapest downtown

(b) MLS scan from a tram stop at Kálvin tér

Fig. 9. *Top:* Detected changes at a tram stop in Kálvin tér, Budapest using [77]. Red, blue and green points represent background objects, foreground objects and ground regions, respectively. *Bottom:* MLS laser scan of the tram stop.

8. Camera-Lidar calibration

Nowadays, state-of-the-art autonomous systems rely on a wide range of sensors for environment perception such as optical cameras, radars and Lidars, therefore efficient sensor fusion is a highly focused research topic in the fields of self-driving vehicles and robotics. Though the resolution and the operation speed of these sensors have significantly improved in recent years, and their prices have become affordable in mass production, their measurements have highly diverse characteristics, which makes the efficient exploitation of the multimodal data challenging. While real time Lidars, such as Velodyne's rotating multi-beam (RMB) sensors provide accurate 3D geometric information with relatively low vertical resolution, optical cameras capture high resolution and high quality image sequences enabling to perceive low level details from the scene. A common problem with optical cameras is that extreme lighting conditions (such as dark, or strong sunlight) largely influence the captured image data, while Lidars are able to provide reliable information much less depending on external illumination and weather conditions. On the other hand, by simultaneous utilization of Lidar and camera sensors, accurate depth with detailed texture and color information can be obtained in parallel from the scenes.

Accurate Lidar and camera calibration is an essential step to implement robust data fusion, thus, related issues are extensively studied in the literature [78–80]. Existing calibration techniques can be grouped based on a variety of aspects [78]: based on the level of user interaction, they can be semi- or fully automatic, methodologically we can distinguish target-based and target-less approaches, and in the term of operational requirements offline and online approaches can be defined.

As their main characteristics, target-based methods use special calibration targets such as 3D boxes [79], checkerboard patterns [81], a simple printed circle [82], or a unique polygonal planar board [83] during the calibration process. In the level of user interactions, we can subdivide target-based methods into semi-automatic and fully-automatic techniques. Semi-automatic methods may consist of many manual steps, such as moving the calibration patterns in different positions, manually localizing the target objects both in the Lidar and in the camera frames, and adjusting the parameters of the calibration algorithms. Though semi-automatic methods may yield very accurate calibration, these approaches are very time consuming and the calibration results highly depend on the skills of the operators. Moreover, even a well calibrated system may periodically need re-calibration due to artifacts caused by vibration and sensor deformation effects.

Fully-automatic target-based methods attempt to automatically detect previously defined target objects, then they extract and match features without user intervention: Velas et al. [84] detect circular holes on planar targets, Park et al. [83] calibrate Lidar and camera by using white homogeneous target objects, Geiger et al. [81] use corner detectors on multiple checkerboards and Rodriguez et al. [85] detect ellipse patterns automatically. Though the mentioned approaches do not need operator interac-

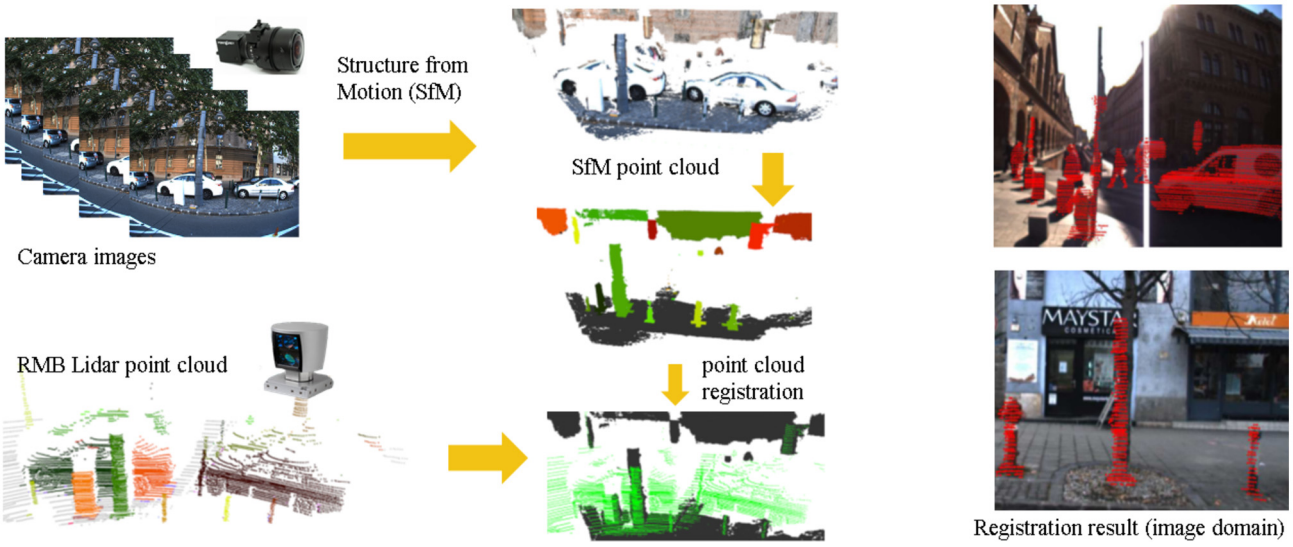


Fig. 10. Workflow of the on-the-fly Lidar-camera registration technique [95].

tions, they still rely on the presence of calibration targets, which often should be arranged in complex setups (i.e., [81] uses 12 checkerboards). Furthermore, during the calibration both the platform and the targets must be motionless.

On the contrary, target-less approaches rely on features extracted from the observed scene without using any calibration objects. Some of these methods use motion-based [86–88] information to calibrate the Lidar and camera, while alternative techniques [78,89] attempt to minimize the calibration errors using only static features.

Among motion-based approaches, Huang and Stachniss [87] improve the accuracy of extrinsic calibration by the estimation of the motion errors, Shiu and Ahmad [86] approximate the relative motion parameters between the consecutive frames, and Shi et al. [90] calculate sensor motion by jointly minimizing the projection error between the Lidar and the camera residuals. These methods first estimate the trajectories of the camera and Lidar sensors either by visual odometry and scan matching techniques, or by exploiting IMU and GNSS measurements. Thereafter they match the recorded camera and Lidar measurement sequences assuming that the sensors are rigidly mounted to the platform. However, the accuracy of these techniques strongly depends on the performance of trajectory estimation, which may suffer from visual featureless regions, low resolution scans [91], lack of hardware trigger based synchronization between the camera and the Lidar [90], or urban scenes without sufficient GPS coverage.

We continue the discussion with single frame target-less and feature-based methods. Moghadam et al. [89] attempt to detect correspondences by extracting lines both from the 3D Lidar point cloud and the 2D image data. While this method proved to be efficient in indoor environments, it requires a large number of line correspondences, a condition that cannot often be satisfied in outdoor scenes. A mutual information based approach has been introduced in [92] to calibrate different range sensors with cameras. Pandey et al. [78] attempt to maximize the mutual information using the camera's grayscale pixel intensities and the Lidar reflectivity values. Based on Lidar reflectivity values and grayscale images Napier et al. [93] minimize the correlation error between the Lidar and the camera frames. Scaramuzza et al. [94] introduce a new data representation called the Bearing Angle image (BA) which is generated from the Lidar's range measurements. Using conventional image processing operations, the method searches for correspondences between the BA and the camera image. As a limitation, target-less feature-based methods require a reasonable initial transformation estimation between the different sensors measurement [90], and mutual information based matching is sensitive to inhomogeneous point cloud inputs and illumination artifacts, which are frequently occurring problems when using RMB Lidars [78].

In [95], authors proposed a new fully automatic and target-less extrinsic calibration approach between a camera and a rotating multi-beam (RMB) Lidar mounted on a moving car. This technique adopts a structure from motion (SfM) method to generate 3D point clouds from the camera data, which can be matched to the Lidar point clouds; thus, the extrinsic calibration problem is addressed as a registration task in the 3D domain (see Fig. 10). The method consists of two main steps: an object level matching algorithm performing a coarse alignment of the camera and Lidar data, and a fine alignment step that implements a control point based point level registration refinement. The superiority of the method is that it relies on only the raw camera and Lidar sensor streams without using any external Global Navigation Satellite System (GNSS) or Inertial Measurement Unit (IMU) sensors. Moreover, it is able to automatically calculate the extrinsic calibration parameters between the Lidar and camera sensors on-the-fly which means we only have to mount the sensors on the top of the vehicle and start driving in a typical urban environment.

Note that there exist a few end-to-end deep learning based camera and Lidar calibration methods [80,96] in the literature, which can automatically estimate the calibration parameters within a bounded parameter range based on a sufficiently large training dataset. However, the trained models cannot be applied for arbitrary configurations, and re-training is often more resource intensive than applying a conventional calibration approach. In addition, the failure case analysis and analytical estimation of the limits of operations are highly challenging for black box deep learning approaches.

9. Conclusion and future directions

When LIDARs started in autonomous driving systems, it was mainly a part of the supporting development toolkit. For a long time up to now, it was seriously considered that LIDAR was not needed as a traffic controller sensor, since (i) it cannot see through bad weather (ii) vulnerable opto-mechanics and (iii) its relative high price. However, today we can get high quality but much cheaper LIDAR sensors, with more robust opto-mechanical solutions; bad weather problems can be partly eliminated and so on.

This technology cannot be stopped, the intense research carried out in the past decade has shown significant progress. Today LIDARs are indispensable sensors in many autonomous robotic applications, and they will have a definite place in autonomous driving. Nevertheless, there are still many opportunities to further enhance their capabilities in the future by exploiting, e.g., multiple reflection patterns, the Doppler effect, reflection intensity. Its main advantage that LiDAR is the only sensor that gives high resolution at range: the power to recognize objects very fine and very accurately in space, even from afar.

In the methodology of the above tasks, based on huge testing databases, Deep Learning methods are widely used for detection and classification. In 3D calibration and positioning tasks, the classical 3D geometry based mathematical framework, e.g., [97], is still used. However, the results of direct mathematical methods can also be adopted to the data feeding step of Deep Learning training methods, especially in 3D geometry tasks, to obtain a more complete set of input variations.

As the industrial tendency of LIDAR sensor development offers in parallel quick increment of spatial and temporal resolution of the 3D measurement sequences and reduced measurement noise, LIDAR sensors in the near future may act as general tools to reliably capture dense and accurate 3D environmental information for various real-time perception, navigation and mapping problems.

Over the past decade, another defining trend is noticeable towards producing smaller and more compact LIDAR devices. These tendencies facilitate the use of LIDAR sensors on small Unmanned Aerial Vehicles (UAVs) [98], enabling further directions of research and many novel applications [99].

CRediT authorship contribution statement

All authors contributed equally to this article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The research was supported by the Ministry of Innovation and Technology NRDI Office within the framework of the Autonomous Systems National Laboratory Program and by the Hungarian National Science Foudation (NKFIH OTKA) No. K 139485.

References

[1] J.C.F. Diaz, W.E. Carter, R.L. Shrestha, C.L. Glennie, LiDAR Remote Sensing, Springer International Publishing, Cham, 2017, pp. 929–980.
[2] M.L. Chanin, A. Garnier, A. Hauchecorne, J. Porteneuve, A Doppler Lidar for measuring winds in the middle atmosphere, Geophys. Res. Lett. 16 (11) (1989) 1273–1276, https://doi.org/10.1029/GL016i011p01273.
[3] A. Teichman, J. Levinson, S. Thrun, Towards 3D object recognition via classification of arbitrary object tracks, in: 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 4034–4041.
[4] G. Pandey, J.R. McBride, R.M. Eustice, Ford campus vision and Lidar data set, Int. J. Robot. Res. 30 (13) (2011) 1543–1552, https://doi.org/10.1177/0278364911400640.
[5] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? The Kitti vision benchmark suite, in: IEEE CVPR, 2012.
[6] J.-L. Blanco, F.-A. Moreno, J. Gonzalez-Jimenez, The Málaga urban dataset: high-rate stereo and Lidars in a realistic urban scenario, Int. J. Robot. Res. 33 (2) (2014) 207–214, https://doi.org/10.1177/0278364913507326, http://www.mrpt.org/MalagaUrbanDataset.
[7] W. Maddern, G. Pascoe, C. Linegar, P. Newman, 1 year, 1000 km: the Oxford robotcar dataset, Int. J. Robot. Res. 36 (1) (2017) 3–15, https://doi.org/10.1177/0278364916679498.
[8] Y. Ma, X. Zhu, S. Zhang, R. Yang, W. Wang, D. Manocha, Trafficpredict: trajectory prediction for heterogeneous traffic-agents, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 6120–6127.
[9] J. Jeong, Y. Cho, Y.-S. Shin, H. Roh, A. Kim, Complex urban dataset with multi-level sensors from highly diverse urban environments, Int. J. Robot. Res. (2019) 0278364919843996.
[10] Y. Choi, N. Kim, S. Hwang, K. Park, J.S. Yoon, K. An, I.S. Kweon, Kaist multi-spectral day/night dataset for autonomous and assisted driving.
[11] A.Z. Zhu, D. Thakur, T. Özaslan, B. Pfrommer, V. Kumar, K. Daniilidis, The multivehicle stereo event camera dataset: an event camera dataset for 3D perception, IEEE Robot. Autom. Lett. 3 (3) (2018) 2032–2039, https://doi.org/10.1109/LRA.2018.2800793.
[12] Z. Yan, L. Sun, T. Krajnik, Y. Ruichek, EU long-term dataset with multiple sensors for autonomous driving, in: Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.
[13] K. Behrendt, R. Soussan, Unsupervised labeled lane marker dataset generation using maps, in: Proceedings of the IEEE International Conference on Computer Vision, 2019.
[14] J. Xue, J. Fang, T. Li, B. Zhang, P. Zhang, Z. Ye, J. Dou, Blvd: building a large-scale 5D semantics benchmark for autonomous driving, in: 2019 International Conference on Robotics and Automation (ICRA), 2019, pp. 6685–6691.
[15] A. Patil, S. Malla, H. Gang, Y.-T. Chen, The H3D dataset for full-surround 3D multi-object detection and tracking in crowded urban scenes, in: International Conference on Robotics and Automation, 2019.
[16] J. Houston, G. Zuidhof, L. Bergamini, Y. Ye, A. Jain, S. Omari, V. Iglovikov, P. Ondruska, One thousand and one hours: self-driving motion prediction dataset, https://level5.lyft.com/dataset/, 2020.
[17] H. Caesar, V. Bankiti, A.H. Lang, S. Vora, V.E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom, Nuscenes: a multimodal dataset for autonomous driving, arXiv preprint, arXiv:1903.11027.
[18] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, D. Anguelov, Scalability in perception for autonomous driving: Waymo open dataset, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[19] M. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, J. Hays, Argoverse: 3D tracking and forecasting with rich maps, CoRR, arXiv:1911.02620.
[20] A. Börcs, B. Nagy, C. Benedek, Instant object detection in Lidar point clouds, in: IEEE Geosci. Remote Sens. Lett, 2017, pp. 992–996.
[21] B. Nagy, C. Benedek, 3D CNN-based semantic labeling approach for mobile laser scanning data, IEEE Sens. J. 19 (21) (2019) 10034–10045, https://doi.org/10.1109/JSEN.2019.2927269.
[22] C. Benedek, D. Molnár, T. Szirányi, A dynamic MRF model for foreground detection on range data sequences of rotating multi-beam Lidar, in: X. Jiang, O.R.P. Bellon, D. Goldgof, T. Oishi (Eds.), Advances in Depth Image Analysis and Applications, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 87–96.
[23] R.B. Rusu, S. Cousins, 3D is here: point cloud library (PCL), in: 2011 IEEE International Conference on Robotics and Automation, 2011, pp. 1–4.
[24] J.-F. Lalonde, N. Vandapel, M. Hebert, Data structures for efficient dynamic processing in 3-d, Int. J. Robot. Res. 26 (8) (2007) 777–796, https://doi.org/10.1177/0278364907079265.
[25] M. Himmelsbach, A. Müller, T. Luettel, H.-J. Wuensche, Lidar-based 3D object perception, in: Proceedings of 1st International Workshop on Cognition for Technical Systems, 2008.
[26] A. Azim, O. Aycard, Detection, classification and tracking of moving objects in a 3D environment, in: 2012 IEEE Intelligent Vehicles Symposium, 2012, pp. 802–807.
[27] A. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, O. Beijbom, Pointpillars: fast encoders for object detection from point clouds, in: Conference on Computer Vision and Pattern Recognition, 2019, pp. 12689–12697.
[28] Y. Zhou, O. Tuzel, Voxelnet: end-to-end learning for point cloud based 3D object detection, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4490–4499.
[29] T.-H. Kim, T.-H. Park, Placement optimization of multiple Lidar sensors for autonomous vehicles, IEEE Trans. Intell. Transp. Syst. 21 (5) (2020) 2139–2145, https://doi.org/10.1109/TITS.2019.2915087.
[30] K.O. Arras, O.M. Mozos, W. Burgard, Using boosted features for the detection of people in 2D range data, in: IEEE ICRA, 2007.
[31] W. Hess, D. Kohler, H. Rapp, D. Andor, Real-time loop closure in 2D LIDAR SLAM, in: IEEE ICRA, 2016.
[32] L. Kurnianggoro, K.H. Jo, Object classification for LIDAR data using encoded features, in: 2017 10th International Conference on Human System Interactions (HSI), 2017, pp. 49–53.
[33] C. Weinrich, T. Wengefeld, M. Volkhardt, A. Scheidig, H.-M. Gross, Generic Distance-Invariant Features for Detecting People with Walking Aid in 2D Laser Range Data, Springer International Publishing, Cham, 2016, pp. 735–747.
[34] F. Galip, M.H. Sharif, M. Caputcu, S. Uyaver, Recognition of objects from laser scanned data points using SVM, in: 2016 First International Conference on Multimedia and Image Processing (ICMIP), 2016, pp. 28–35.

[35] L. Beyer, A. Hermans, B. Leibe Drow, Real-time deep learning-based wheelchair detection in 2-D range data, IEEE Robot. Autom. Lett. 2 (2) (2017) 585–592, https://doi.org/10.1109/LRA.2016.2645131.

[36] L. Spinello, K.O. Arras, R. Triebel, R. Siegwart, A layered approach to people detection in 3D range data, in: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, AAAI'10, AAAI Press, 2010, pp. 1625–1630.

[37] V. Alvarez-Santos, A. Canedo-Rodriguez, R. Iglesias, X. Pardo, C. Regueiro, M. Fernandez-Delgado, Route learning and reproduction in a tour-guide robot, Robot. Auton. Syst. 63 (Part 2) (2015) 206–213, https://doi.org/10.1016/j.robot.2014.07.013.

[38] Z. Rozsa, T. Sziranyi, Obstacle prediction for automated guided vehicles based on point clouds measured by a tilted Lidar sensor, IEEE Trans. Intell. Transp. Syst. 19 (8) (2018) 2708–2720, https://doi.org/10.1109/TITS.2018.2790264.

[39] I. Sipiran, B. Bustos, Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes, Vis. Comput. 27 (11) (2011) 963–976, https://doi.org/10.1007/s00371-011-0610-y.

[40] Z. Rozsa, T. Sziranyi, Object detection from a few Lidar scanning planes, IEEE Trans. Intell. Veh. 4 (4) (2019) 548–560, https://doi.org/10.1109/TIV.2019.2938109.

[41] M. Kutila, P. Pyykönen, W. Ritter, O. Sawade, B. Schäufele, Automotive LIDAR sensor development scenarios for harsh weather conditions, in: IEEE ITSC, 2016.

[42] N. Charron, S. Phillips, S. Waslander, De-noising of Lidar point clouds corrupted by snowfall, https://doi.org/10.1109/CRV.2018.00043, 2018, pp. 254–261.

[43] R. Heinzler, F. Piewak, P. Schindler, W. Stork, CNN-based Lidar point cloud de-noising in adverse weather, IEEE Robot. Autom. Lett. 5 (2) (2020) 2514–2521, https://doi.org/10.1109/LRA.2020.2972865.

[44] R. Heinzler, P. Schindler, J. Seekircher, W. Ritter, W. Stork, Weather influence and classification with automotive Lidar sensors, in: 2019 IEEE Intelligent Vehicles Symposium (IV), 2019, pp. 1527–1534.

[45] M. Kutila, P. Pyykönen, H. Holzhüter, M. Colomb, P. Duthon, Automotive Lidar performance verification in fog and rain, in: 2018 IEEE Intelligent Transportation Systems Conference, ITSC 2018, IEEE Institute of Electrical and Electronic Engineers, United States, 2018, pp. 1695–1701.

[46] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, F. Heide, Seeing through fog without seeing fog: deep multimodal sensor fusion in unseen adverse weather, in: The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[47] J. Zhang, S. Singh, Loam: Lidar odometry and mapping in real-time, in: Robotics: Science and Systems Conference, 2014.

[48] R. Dube, A. Cramariuc, D. Dugas, J. Nieto, R. Siegwart, C. Cadena, SegMap: 3D segment mapping using data-driven descriptors, in: Robotics: Science and Systems Conference, 2018.

[49] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, R. Daniela, LIO-SAM: tightly-coupled Lidar inertial odometry via smoothing and mapping, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE, 2020, pp. 5135–5142.

[50] T. Shan, B. Englot, C. Ratti, R. Daniela, LVI-SAM: tightly-coupled Lidar-visual-inertial odometry via smoothing and mapping, in: IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2021, arXiv preprint, arXiv:2104.10831.

[51] D. Rozenberszki, A.L. Majdik, LOL: Lidar-only odometry and localization in 3D point cloud maps*, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), 2020, pp. 4379–4385.

[52] B. Nagy, C. Benedek, Real-time point cloud alignment for vehicle localization in a high resolution 3D map, in: ECCV Workshops, 2018.

[53] B. Nagy, C. Benedek, 3D CNN based phantom object removing from mobile laser scanning data, in: Int'l Joint Conference on Neural Networks (IJCNN), Anchorage, Alaska, USA, 2017, pp. 4429–4435.

[54] Y. Yu, J. Li, H. Guan, C. Wang, Automated detection of three-dimensional cars in mobile laser scanning point clouds using DBM-Hough-Forests, IEEE Trans. Geosci. Remote Sens. 54 (7) (2016) 4130–4142, https://doi.org/10.1109/TGRS.2016.2537830.

[55] H. Zheng, R. Wang, S. Xu, Recognizing street lighting poles from mobile LiDAR data, IEEE Trans. Geosci. Remote Sens. 55 (1) (2017) 407–420, https://doi.org/10.1109/TGRS.2016.2607521.

[56] B. Wu, B. Yu, W. Yue, S. Shu, W. Tan, C. Hu, Y. Huang, J. Wu, H. Liu, A voxel-based method for automated identification and morphological parameters estimation of individual street trees from mobile laser scanning data, Remote Sens. 5 (2) (2013) 584.

[57] S. Papadimitriou, H. Kitagawa, P.B. Gibbons, C. Faloutsos, LOCI: fast outlier detection using the local correlation integral, in: Int'l Conf. on Data Engineering, Los Alamitos, CA, USA, 2003, pp. 315–326.

[58] S. Sotoodeh, Outlier detection in laser scanner point clouds, in: ISPRS Arch. Photogramm. Remote Sens. and Spatial Inf. Sci., vol. XXXVI-5, 2006, pp. 297–302.

[59] J. Köhler, T. Nöll, G. Reis, D. Stricker, Robust outlier removal from point clouds acquired with structured light, in: Eurographics (Short Papers), Cagliari, Italy, 2012, pp. 21–24.

[60] T. Kanzok, F. Süß, L. Linsen, R. Rosenthal, Efficient removal of inconsistencies in large multi-scan point clouds, in: Int'l Conf. in Central Europe on Computer Graphics, Visualization and Computer Vision, Pilsen, Czech Rep., 2013.

[61] J. Gehrung, M. Hebel, M. Arens, U. Stilla, An approach to extract moving objects from MLS data using a volumetric background representation, in: ISPRS Ann. Photogramm. Remote Sens. and Spatial Inf. Sci., vol. IV-1, 2017.

[62] M. Engelcke, D. Rao, D.Z. Wang, C.H. Tong, I. Posner, Vote3Deep: fast object detection in 3D point clouds using efficient convolutional neural networks, in: IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, pp. 1355–1361.

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

[64] H.S. Koppula, A. Anand, T. Joachims, A. Saxena, Semantic labeling of 3D point clouds for indoor scenes, in: Int'l Conf. Neural Inf. Processing Systems (NIPS), Granada, Spain, 2011, pp. 244–252.

[65] T. Hackel, J.D. Wegner, K. Schindler, Fast semantic segmentation of 3D point clouds with strongly varying density, ISPRS Ann. Photogramm. Remote Sens. and Spatial Inf. Sci. III-3.

[66] G. Riegler, A.O. Ulusoy, A. Geiger, OctNet: learning deep 3D representations at high resolutions, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 6620–6629.

[67] G. Pang, 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.

[68] C. Qi, L. Yi, H. Su, L. Guibas, PointNet++: deep hierarchical feature learning on point sets in a metric space, in: Conf. NIPS, 2017.

[69] W. Wang, R. Yu, Q. Huang, 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, 2018, pp. 2569–2578.

[70] H. Su, V. Jampani, D. Sun, S. Maji, V. Kalogerakis, M.-H. Yang, J. Kautz, SPLATNet: Sparse lattice networks for point cloud processing, 2018, pp. 2530–2539.

[71] W. Xiao, B. Vallet, M. Brédif, N. Paparoditis, Street environment change detection from mobile laser scanning point clouds, ISPRS J. Photogramm. Remote Sens. 107 (2015) 38–49.

[72] A. Schlichting, C. Brenner, Vehicle localization by Lidar point correlation improved by change detection, in: ISPRS International Archives of the Photogrammetry, in: Remote Sensing and Spatial Information Sciences, vol. XLI-B1, 2016, pp. 703–710.

[73] R. Qin, A. Gruen, 3D change detection at street level using mobile laser scanning point clouds and terrestrial images, ISPRS J. Photogramm. Remote Sens. 90 (2014) 23–35.

[74] J.P. Underwood, D. Gillsjö, T. Bailey, V. Vlaskine, Explicit 3D change detection using ray-tracing in spherical coordinates, in: IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 2013, pp. 4735–4741.

[75] K. Liu, J. Boehm, C. Alis, Change detection of mobile LIDAR data using cloud computing, in: ISPRS International Archives of the Photogrammetry, in: Remote Sensing and Spatial Information Sciences, vol. XLI-B3, 2016, pp. 309–313.

[76] Z. Rozsa, M. Golarits, T. Sziranyi, Localization of map changes by exploiting slam residuals, in: J. Blanc-Talon, P. Delmas, W. Philips, D. Popescu, P. Scheunders (Eds.), Advanced Concepts for Intelligent Vision Systems, Springer International Publishing, Cham, 2020, pp. 312–324.

[77] B. Gálai, C. Benedek, Change detection in urban streets by a real time Lidar scanner and MLS reference data, in: International Conference on Image Analysis and Recognition (ICIAR), in: Lecture Notes in Computer Science, vol. 10317, Springer, Montreal, Canada, 2017, pp. 210–220.

[78] G. Pandey, J.R. McBride, S. Savarese, R.M. Eustice, Automatic extrinsic calibration of vision and Lidar by maximizing mutual information, J. Field Robot. 32 (2015) 696–722.

[79] Z. Pusztai, I. Eichhardt, L. Hajder, Accurate calibration of multi-Lidar-multi-camera systems, in: Sensors, vol. 18, 2018, pp. 119–152.

[80] G. Iyer, R.K. Ram, J.K. Murthy, K.M. Krishna, Calibnet: Geometrically supervised extrinsic calibration using 3D spatial transformer networks, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS).

[81] A. Geiger, F. Moosmann, O. Car, B. Schuster, Automatic camera and range sensor calibration using a single shot, in: IEEE International Conference on Robotics and Automation, 2012, pp. 3936–3943.

[82] H. Alismail, L. Baker, B. Browning, Automatic calibration of a range sensor and camera system, in: 2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 2012, pp. 286–292.

[83] Y. Park, S.M. Yun, C.S. Won, K. Cho, K. Um, S. Sim, Calibration between color camera and 3D LIDAR instruments with a polygonal planar board, in: Sensors, 2014.

[84] M. Velas, M. Spanel, Z. Materna, A. Herout, Calibration of RGB camera with velodyne LiDAR, 2014.

[85] S. Rodriguez-Florez, V. F., P. Bonnifait, Extrinsic calibration between a multi-layer Lidar and a camera, in: IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2008, pp. 214–219.

[86] Y.C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB, IEEE Trans. Robot. Autom. 5 (1) (1989) 16–29.

[87] K. Huang, C. Stachniss, Extrinsic multi-sensor calibration for mobile robots using the Gauss Helmert model, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017, pp. 1490–1496.
[88] K.H. Strobl, G. Hirzinger, Optimal hand-eye calibration, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2006, pp. 4647–4653.
[89] P. Moghadam, M. Bosse, R. Zlot, Line-based extrinsic calibration of range and image sensors, IEEE Int. Conf. Robot. Autom. (2013) 3685–3691.
[90] C. Shi, K. Huang, Q. Yu, J. Xiao, H. Lu, C. Xie, Extrinsic calibration and odometry for camera-Lidar systems, IEEE Access 7 (2019) 120106–120116.
[91] O. Józsa, A. Börcs, C. Benedek, Towards 4D virtual city reconstruction from Lidar point cloud sequences, in: ISPRS Workshop on 3D Virtual City Modeling, Regina, Canada, in: ISPRS Annals Photogram. Rem. Sens. and Spat. Inf. Sci., vol. II-3/W1, 2013, pp. 15–20.
[92] R. Wang, F.P. Ferrie, J. Macfarlane, Automatic registration of mobile Lidar and spherical panoramas, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2012, pp. 33–40.
[93] A. Napier, P. Corke, P. Newman, Cross-calibration of push-broom 2D LIDARs and cameras in natural scenes, in: IEEE International Conference on Robotics and Automation (ICRA), 2013, pp. 3679–3684.
[94] D. Scaramuzza, A. Harati, R. Siegwart, Extrinsic self calibration of a camera and a 3D laser range finder from natural scenes, in: IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007, pp. 4164–4169.
[95] B. Nagy, C. Benedek, On-the-fly camera and Lidar calibration, Remote Sens. 12 (7) (2020), https://doi.org/10.3390/rs12071137, https://www.mdpi.com/2072-4292/12/7/1137.
[96] N. Schneider, F. Piewak, C. Stiller, U. Franke, Regnet: multimodal sensor registration using deep neural networks, in: 2017 IEEE Intelligent Vehicles Symposium (IV), 2017, pp. 1803–1810.
[97] D. Barath, J. Matas, Graph-cut ransac: local optimization on spatially coherent structures, IEEE Trans. Pattern Anal. Mach. Intell. (2021) 1, https://doi.org/10.1109/TPAMI.2021.3071812.
[98] X. Li, C. Liu, Z. Wang, X. Xie, D. Li, L. Xu, Airborne LiDAR: state-of-the-art of system design, technology and application, Meas. Sci. Technol. 32 (3) (2020) 032002, https://doi.org/10.1088/1361-6501/abc867.
[99] D.V. Nam, K. Gon-Woo, Solid-state LiDAR based-slam: a concise review and application, in: 2021 IEEE International Conference on Big Data and Smart Computing (BigComp), 2021, pp. 302–305.

Csaba Benedek received the M.Sc. degree in computer sciences from the Budapest University of Technology and Economics (BME), in 2004, and the Ph.D. degree in image processing from Péter Pázmány Catholic University, Budapest, in 2008. From 2008 to 2009, he was a Post-Doctoral Researcher with INRIA, Sophia-Antipolis, France. He received D.Sci. degree in 2021, by the Hungarian Academy of Sciences, Budapest. He has been the manager of various national and international research projects in recent years. He is currently a Research Advisor with the Machine Perception Research Laboratory, Institute for Computer Science and Control (SZTAKI), and a Habilitated Associate Professor with Péter Pázmány Catholic University. His research interests include Bayesian image and point cloud segmentation, object extraction, change detection, and GIS data analysis. Now he is the president (2019-2023) of the Hungarian Image Processing and Pattern Recognition Society (KEPAF).

András L. Majdik received his Ph.D. degree in systems engineering from the Technical University of Cluj-Napoca, Romania in 2012 for his contribution to the simultaneous localization and mapping (SLAM) of mobile robots. While working on his Ph.D. he was a visiting scholar at Robotics, Perception and Real Time Group, University of Zaragoza, Spain and 3D Vision and Mobile Robotics Research Group, Budapest University of Technology and Economics, Hungary. Next, he was a postdoc at Robotics and Perception Group, University of Zurich, Switzerland. Currently, he is a senior research fellow with the Machine Perception Research Laboratory, Institute for Computer Science and Control (SZTAKI). His research interests include vision and LIDAR-based localization and mapping of ground and aerial vehicles, the spatial artificial intelligence of autonomous systems.

Balázs Nagy received his Ph.D. degree in computer science from the Péter Pázmány Catholic University in 2021. He is a research fellow at the Machine Perception Research Laboratory of the Institute for Computer Science and Control (SZTAKI). His research interests include sensor fusion, computer vision, deep learning, robotics and autonomous vehicles.

Zoltan Rozsa received his Ph.D. degree in vehicle and transportation engineering from Budapest University of Technology and Economics in 2020. He is currently with the Faculty of Transportation Engineering and Vehicle Engineering of Budapest University of Technology and Economics and a research fellow at the Machine Perception Research Laboratory of the Institute for Computer Science and Control (SZTAKI). His research interests include automated guided vehicles, machine vision, 3D recognition, and reconstruction.

Tamas Sziranyi received his Ph.D. and D.Sci. degrees in 1991 and 2001, by the Hungarian Academy of Sciences, Budapest. He was appointed to a Full Professor position in 2001 at Pannon University, Veszprem, Hungary, and, in 2004, at the Peter Pazmany Catholic University, Budapest. Presently he is a full professor at the Budapest University of Technology and Economics. He has been a research scientist at the Institute for Computer Science and Control (SZTAKI), since 1992, where he leads the Machine Perception Research Laboratory since 2006. His research activities include machine perception, pattern recognition, texture and motion segmentation, Markov Random Fields and stochastic optimization, remote sensing, surveillance, intelligent networked sensor systems, graph-based clustering, digital film restoration. He has participated in several prestigious international (ESA, EDA, FP6, FP7, OTKA) projects with his research laboratory.

Dr. Sziranyi was the founder and past president (1997 to 2002) of the Hungarian Image Processing and Pattern Recognition Society. He was an Associate Editor of IEEE T. Image Processing (2003-2009), and he has been an AE of Digital Signal Processing (Elsevier) since 2012. He was honored by the Master Professor award in 2001, the Szechenyi professorship, and the ProScientia (Veszprem) award in 2011. He is a Fellow both of the Int. Assoc. Pattern Recognition (IAPR) and the Hungarian Academy of Engineering from 2008. He has more than 300 publications, including 50 in major scientific journals and several international patents.