

Lidar for Autonomous Driving: The principles, challenges, and trends for automotive lidar and perception systems

You Li, Javier Ibanez-Guzman

Abstract

Autonomous vehicles rely on their perception systems to acquire information about their immediate surroundings. It is necessary to detect the presence of other vehicles, pedestrians and other relevant entities. Safety concerns and the need for accurate estimations have led to the introduction of Light Detection and Ranging (LiDAR) systems in complement to camera or radar based perception systems. This article presents a review of state-of-the-art automotive LiDAR technologies and the perception algorithms used with those technologies. LiDAR systems are introduced first by analyzing the main components, from laser transmitter to its beam scanning mechanism. Advantages/disadvantages and the current status of various solutions are introduced and compared. Then, the specific perception pipeline for LiDAR data processing, from an autonomous vehicle perspective is detailed. The model driven approaches and the emerging deep learning solutions are reviewed. Finally, we provide an overview of the limitations, challenges and trends for automotive LiDARs and perception systems.

I. INTRODUCTION

Autonomous driving is entering a pre-industrialization phase with significant progress attained over the past years. Sensors initially capture data representation of the environment that are processed by perception algorithms to build the vehicle's immediate environment used for autonomous vehicle navigation. A perception system for autonomous vehicle navigation would consist of a combination of active and passive sensors, namely, cameras, radars, and LiDARs (Light detection and ranging) [1]. LiDARs are active sensors that illuminate the surroundings by emitting lasers. Ranges are measured precisely by processing the received laser returns from the reflecting surfaces. As stated in [2], LiDARs *"poised to significantly alter the balance in commercial, military, and intelligence operations, as radar has done over the past seven decades"*. Despite much progresses in camera-based perception, image processing methods estimate distances. This approach poses difficulties when estimating distances for cross-traffic entities,

You Li, Javier Ibanez-Guzman are with research division at RENAULT S.A.S, 1 Avenue du Golf, 78280 Guyancourt, France
you.li@renault.com, javier.ibanez-guzman@renault.com



Fig. 1: The autonomous vehicle prototype developed by Groupe RENAULT for EVAPS project. The most evident sensor is the Velodyne UltraPuck LiDAR on top.

in particular for monocular solutions. The 2007 DARPA Grand Challenge, a milestone in autonomous driving demonstrated the potential of LiDAR perception systems. The top 3 teams were all equipped with multiple LiDARs. A 64 layers LiDAR – Velodyne HDL64 ¹, played as a critical central role for the champion team and runner-up [3], [4]. Currently, most high level autonomous vehicles use LiDARs as part of their perception systems despite their high cost and moving parts. A typical example is their incorporation into the autonomous vehicles being tested as robo-vehicles in different countries, e.g. the EVAPS field operational test in France², several participants (e.g. Renault, Transdev, Vedecom, etc) collaborate together to exploit mobility services based on autonomous vehicles in Saclay area in Paris. One of the prototypes developed by Renault is shown in Fig. 1. As a result, there are more than 20 companies developing distinctive LiDAR systems for autonomous driving systems ranging from low level to high end, a sort of "big bang". Which LiDAR type(s) will dominate autonomous driving in the future is still in the mist.

On the other hand, LiDAR based algorithms also entered into a fast-track. For an autonomous vehicle, LiDARs are mainly used for *perception* and *localization*. Due to the page limit, this paper focuses only on perception usages. In the context of autonomous driving, a perception system provides a machine interpretable representation of the environment around the vehicle. From a user's perspective, the output of a perception system comprises 3 levels of information:

- *Physical description*: pose, velocity, shape of objects.
- *Semantic description*: the categories of objects.
- *Intention prediction*: likelihood of an object's behavior.

¹<https://www.velodynelidar.com/hdl-64e.html>

²<https://pole-moveo.org/projets/evaps/>

Therefore, the LiDAR outputs are used for *object detection*, *classification*, *tracking* and *intention prediction*, corresponding to the various layers of information. Due to LiDAR's superiority in ranging accuracy, the provided physical information is highly reliable. While the semantic information carried by LiDAR is less, or more difficult acquired than camera – a contextual sensor good at object recognition. In practice, LiDARs are combined with cameras to complement each other [5]: camera is poor in distance estimation while LiDAR is not-satisfying for object recognition. Precise physical and semantic information, together with map information will improve the intention prediction without doubts. With many years of progresses, LiDAR-centric perception system become mature for model based processing algorithms. While emerging deep learning (DL) methods are changing this domain. Traditional model based LiDAR data processing methods are computation-friendly, explicable. While data-driven DL methods have demonstrated extraordinary capabilities in providing semantic information, which is the weak point of traditional methods.

In this article, section II firstly introduce different ranging principles for a LiDAR, and dismantles its principal components. Several representative LiDAR products or manufactures are listed and classified according to the different technical solutions. Section III surveys the LiDAR data processing algorithms, including classic methods and the emerging deep learning approaches, to provide different level of information required from a perception system. Section IV covers open challenges for the current LiDARs and LiDAR based perception systems for the automotive industry, and their future trends according to the authors' opinions.

II. LiDAR TECHNOLOGIES

A typical LiDAR operates by scanning its field of view with one or several laser beams. This is done through delicately designed beam steering system. The laser beam is generated by an amplitude modulated laser diode that emits at near infrared wavelength. The laser beam is reflected by the environment back to the scanner. The returned signal is received by photodetector. Fast electronics filter the signal and measure the difference between the transmitted and received signals that proportional to the distance. Range is estimated from the sensor model based on this difference. Difference in variations of reflected energy due to surface materials as well as state of the milieu between transmitter and receiver are compensated through signal processing. The LiDAR outputs comprise both 3D point clouds that corresponds to the scanned environments and the intensities that correspond to the reflected laser energies. Figure 2 shows a conceptual representation of this operating principle.

A LiDAR system can be partitioned into: the *laser rangefinder* system and the *scanning system*. The laser rangefinder comprises the *laser transmitter* that illuminates the target via a modulated wave; the *photodetector* that generates electronic signal from the reflected photons after optical processing and

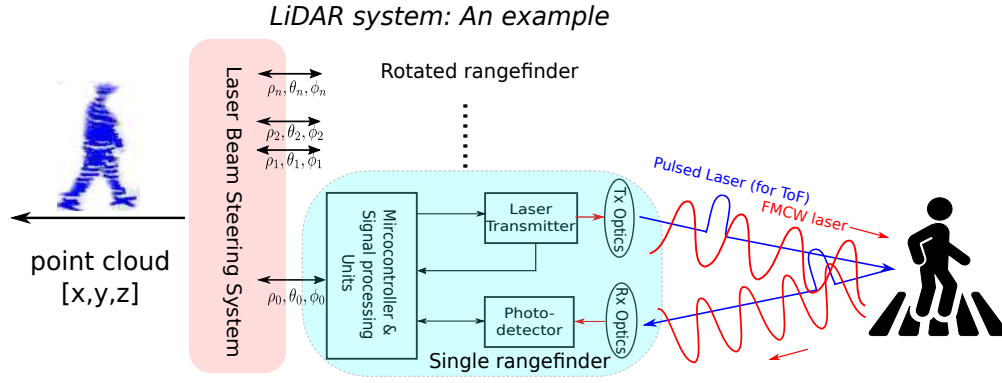


Fig. 2: An example of ToF laser rangefinder. The rangefinder uses either direct or coherent method to measure the distance at a certain direction controlled by the scanning system.

photoelectric conversion; *optics* which collimate the emitted laser and focus the reflected signal onto the photodetector, and *signal processing electronics* to estimate the distance between the laser source and the reflecting surface, based on the received signal. The *scanning system* typically will steer laser beams at different azimuths and vertical angles, denoted by ϕ_i, θ_i , where i is an index that determines the direction at which the beam is being pointed.

This section addresses initially the principles of a rangefinder in order to understand the measurement process and limitations, then it introduces the scanning systems that define the sensor field of view. It is then possible to classify LiDARs based on the technologies they use. This classification is then applied to examine the commercially available automotive type LiDARs.

A. Laser Rangefinder Principles

A rangefinder that measures the distance to an object by a laser beam is known as a laser rangefinder. The manner in which they operate depends on the type of signal modulation used in the laser beam. Pulsed Laser are used so their *time-of-flight (ToF)* can be measured, these are known as *direct detection* laser rangefinders. The laser signal can be also a *frequency-modulated continuous wave (FMCV)* that indirectly measures both distance and velocity, from the Doppler effect. These are known as *coherent detection* laser rangefinders.

1) *LiDAR power equation:* The transmitted laser is firstly attenuated through the transmission medium, then diffused as it reflects from the target surface. It is partially captured by the receiving optics and finally transformed into an electrical signal by a photodetector. For a target at distance r , the amount of received power P_r by the photodetector based from a pulsed laser emitter can be approximately modeled

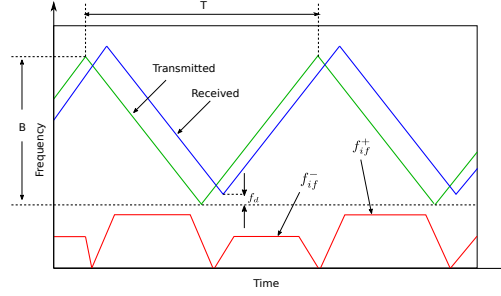


Fig. 3: Principle of coherent detection: distance is estimated by the intermediate frequency (in red trace) generated by mixing transmitted and received light waves (from [8]).

as [6]:

$$P_r = E_p \frac{c\eta A_r}{2r^2} \cdot \beta \cdot T_r \quad (1)$$

where E_p is the total energy of a transmitted pulse laser, c is light speed. A_r represents the area of receive aperture at range r . η is the overall system efficiency. β is the reflectance of the target's surface, which is decided by both surface properties and incident angle. In a simple case of Lambertian reflection with a reflectivity of $0 < \Gamma < 1$, it is given by: $\beta = \Gamma/\pi$. The final part T_r denotes the transmission loss through the transmission medium. When the LiDAR works under adverse conditions (e.g. fog, rain, dust, snow, etc), the particles within the air would scatter and absorb the photons. Eq. 1 reveals that the received power p_r decreases quadratically with respect to distance r : an object at hundreds meters away is order of magnitudes "darker" than at tens of meters. Simply increasing the power of laser transmitter is restricted by the eye-safety standard: IEC 60825 [7]. To overcome this, the overall system efficiency need to be improved through optics, photodetectors, and more advanced signal processing algorithms. For a FMCW laser, Eq. 1 still holds except slight differences.

2) *Time-of-flight (ToF)*: A ToF laser rangefinder measures the range by calculating the time difference between the transmitted and received lasers:

$$r = \frac{1}{2n} c \Delta t \quad (2)$$

where c is the light speed, n is the index of refraction of the propagation medium (≈ 1 for air). Δt is the time gap between the transmitted laser and the received laser. ToF LiDARs prevail in current automotive LiDARs market as their simple structure and signal processing methods. However, potential for increasing their maximum range is constrained by limited transmit power due to eye-safety requirements. In ToF LiDAR, the return signal can be interference from strong sunlight or laser beams from other ToF LiDARS.

3) *Coherent detection*: By mixing the local carrier signal with the received signal, it is possible to demodulate the received signal and thus it is possible to obtain the phase and frequency shift of the laser signal, hence acquire the distance and velocity from the reflecting surface. This can be regarded as the optical version of FMCW (Frequency Modulated Continuous Wave) Radars, which are popular nowadays passenger vehicles' ADAS systems. FMCW LiDARs continuously emit a FM laser signal (e.g. linearly-chirped laser) to a target while keeping a reference signal, also known as a local oscillator. The common modulation functions are sawtooth or triangle waves. Because a FMCW LiDAR continuously illuminates the objects using less emitted power for this purpose, complying with eye-safe requirements and opening the possibility to use more power to extend their field of view. The signals used for coherent detection are shown in Fig. 3. The intermediate frequency (IF, red trace) can be generated by mixing the local oscillator signal (a linearly-chirped triangular modulation function) from the laser transmitter (in green) with the laser signal reflected from observed surfaces (in blue). Processing of the resulting signal generates an intermediate frequency f_{if} shown in red in Fig. 3 (b). By assuming that the Doppler frequency shift f_d is less than the intermediate frequency f_{if} , we have:

$$f_{if} = \frac{4rB}{ct} = \frac{f_{if}^+ + f_{if}^-}{2}, \quad f_d = \frac{f_{if}^+ - f_{if}^-}{2} (\text{when } f_d < f_{if}) \quad (3)$$

where B, r, t are the modulation bandwidth, the waveform period and the light speed, respectively. The velocity is obtained as:

$$v = \frac{f_d \lambda}{2} \quad (4)$$

where λ is the laser wavelength

FMCW LiDAR is able to directly measure the distance and velocity at the same time, while for ToF LiDAR, speed is obtained from indirect estimation through several consecutive sensor readings. By using a FMCW laser signal, it is possible to reduce the interference effect from other laser sources and strong sunlight. However, FMCW LiDARs require high quality laser generators that possess long coherent distance.

B. Laser Transmission and Reception

The generation of the laser signals and their emission as well as the receiver electronics of the reflected signals also characterise the performance and cost of the laser rangefinders.

1) *Laser sources*: ToF LiDARs need a pulsed (amplitude modulated) laser signal. This is generated using a pulsed laser diode or a fiber laser. A diode laser causes laser oscillation by flowing an electric current to the diode's junction. Diode lasers can be grouped into 2 classes: *Edge-emitting laser (EEL)* and *Surface-emitting semiconductor laser (VCSEL)*. EEL have been applied in the telecommunication industry

for a long time. VCSEL outputs a circular beam, while EEL transmits an elliptical laser beam, requiring additional beam-shaping optics. In VCSEL, forming a 2D laser array on a single chip is easier than for an EEL, this is important as it increases the LiDAR resolution. By contrast, the range for VCSEL is shorter due to power limits. The pulsed laser diodes using in automotive applications are hybrid devices. That is, a laser chip is mounted with capacitors that are triggered by a MOSFET transistor. Thus at every gate opening, the electric charge accumulated in the capacitors will be discharged into the chip, this will emit the optical pulse in a controlled manner. These sources are cost-effective, as their 905 nm output can be detected by cost effective silicon detectors. However, these diodes have a limited pulse repetition rate and lower peak power and might require cooling. Laser diode sources for 3D flash LiDAR use diode stack technology with several edge-emitting bars assembled into a vertical stack. Heat dissipation becomes an issue, hence the need for heat sinks as well as the accumulation of emitted power beyond eye-safe requirements. Fiber laser can have higher output power which is useful when operating at high wavelengths. Their output beams can be split and routed to multiple sensor locations using optical fiber, they have better pulse repetition frequency, better beam quality, etc. However, they can be bulky and thus resulting in non-compact systems that are difficult to be integrated in vehicles.

2) *Laser wavelength*: Selecting an appropriate wavelength of laser should have a comprehensive consideration of atmospheric windows, eye-safety requirement and the cost. The 850-950nm (near infrared (NIR)) and 1550nm (short wave infrared (SWIR)) lasers are mostly utilized because of their popularities in industry. Either a low price diode laser or a more powerful fiber laser at wavelength 850-950nm or 1550nm is easily purchased from the market. The maximum power permitted by eye-safety standard for 1550nm laser is higher than the lasers in 850-950nm, which means a larger range could be achieved. However, expensive InGaAs based photodiode is required to detector laser returns at 1550nm. The efficiency of InGaAs based photodiode is lower than the matured Silicon ones for NIR lasers. In addition, the atmospheric water absorption for 1550nm is stronger than 850-950nm. Therefore, LiDAR systems at NIR wavelength (905nm for instance) are still the mainstream.

3) *Photo-detector*: A photo-detector converts optical power to electrical power by the photoelectric effect. Photo-sensitivity that describes a photodetector's response when receiving photons is one of the most critical characteristics. The photosensitivity depends on the wavelength of the received laser. Therefore, selecting photodetector for a LiDAR system is closely related to the choice of laser wavelength. The most popular detectors are *PIN photodiode*, *Avalanche photodiode (APD)*, *Single-photon avalanche diode (SPAD)*, *Silicon Photomultiplier (SiPM)*.

PIN photodiode: is formed by a p-i-n junction that creates a depletion region that is free of mobile charge carriers. By applying a reverse bias to a photodiode, absorbing a photon will generate a current

flow in the reverse-biased photodiode.

Avalanche Photodiode (APD): is a photodiode that applies reverse voltage to multiply photocurrent through avalanche effect. The APD's ability to multiply signals reduces the effect of noise and achieves higher internal current gain (around 100) and SNR than the PIN photodiode. Therefore, APDs are quite common in contemporary LiDAR systems. Silicon based APDs are sensitive through the visible spectral region until the near infrared around 1000nm. At longer wavelengths up to 1700nm, InGaAs APDs are available, although at higher cost.

Single-photon avalanche diode (SPAD): is an APD designed to operate with a reverse-bias voltage above the breakdown voltage (Geiger-mode), which allows a detection of very few photons in very short time. SPAD can achieve a gain of 10^6 that is significantly higher than APD. This characteristic allows the SPAD detect extremely weak light at long distance. Furthermore, the CMOS technology that can be used for the SPAD fabrication enable an integrated array of photodiodes on one chip. This is desirable for increasing LiDAR's resolution while cut the cost and power consumption.

Silicon Photomultiplier (SiPM): is based on SPAD, while enable photon counting. The Geiger-mode in which a SPAD operates is a photon-trigger mode that a SPAD cannot distinguish the magnitude of received photo flux. To overcome this issue, SiPM integrates a dense array of 'microcells' (a pair of a SPAD and a quench resistor) working identically and independently. SiPM's output is in essence a combination of the photocurrents detected from each microcell. In this approach, SiPM is capable of giving information on the magnitude of an instantaneous photon flux.

C. Scanning system

A scanning system (or beam steering system) is designed to enable the transmitted lasers to rapidly explore a large area. The existing scanning approaches are usually classified either *mechanical spinning* or *solid state*. The former usually contains a bulky rotating mirror system like the Velodyne HDL64 in the early ages of autonomous driving history. The later "*solid state*" refers to scanning system without moving parts (even some are still steered by micro-mirrors), what is preferred by the automotive industry.

Mechanical Spinning: The currently most popular scanning solution for automotive LiDAR is the mechanical spinning system [10], which steers the laser beams through rotating assembly (e.g. mirror, prism, etc) controlled by a motor to create a large field of view (FoV). Conventionally, nodding-mirror system and polygonal-mirror system [11] are the main types applied. For example, as the mechanical spinning scheme shown in Fig. 4 (a), An embedded nodding-mirror system tilts the lasers to generate a vertical FoV. Then, 360° horizontal FoV is achieved by rotating the LiDAR base. State-of-the-art LiDAR use multiple beams to reduce the movable mechanism. For instance, Velodyne VLP series use arrays of

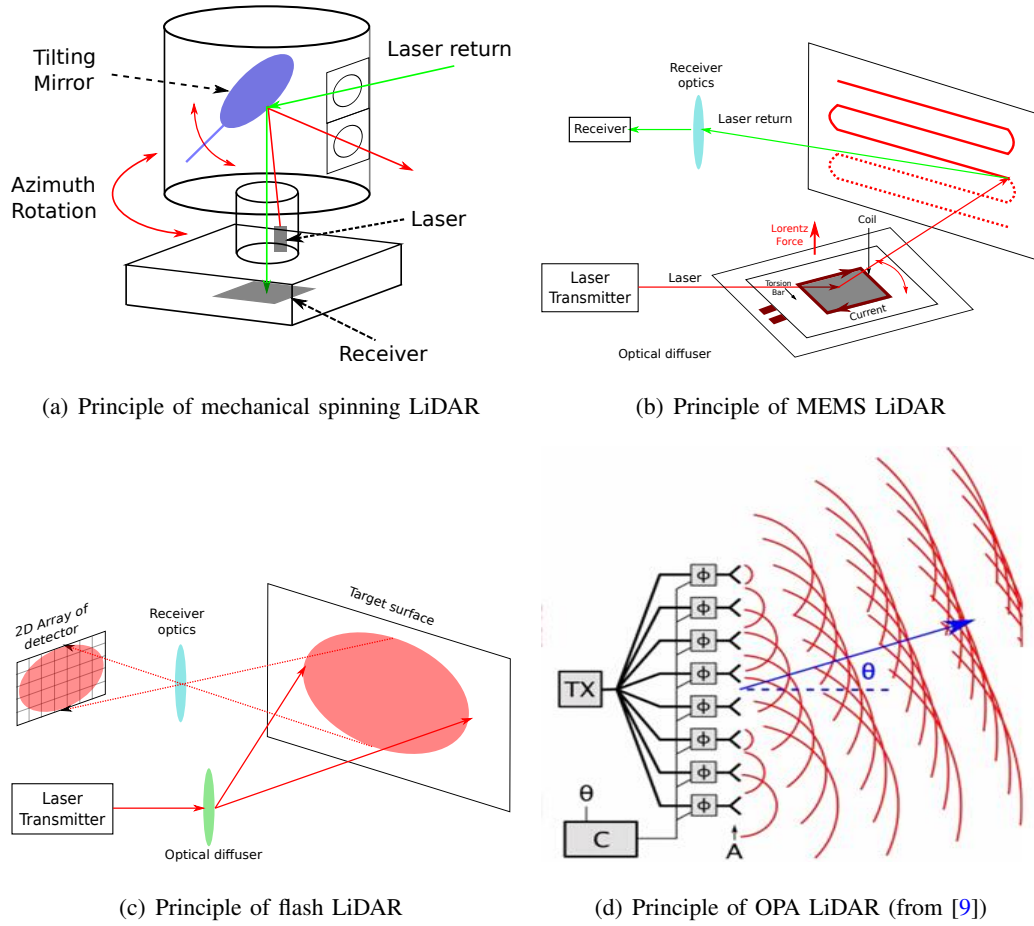


Fig. 4: LiDAR systems categorized by scanning approaches.

laser diodes and photodiodes to increase point cloud densities. The mechanical spinning system offers an advantage of high signal-to-noise ratio (SNR) over a wide FOV. However, the rotating mechanism is bulky for integration inside a vehicle, and is fragile in harsh conditions such as vibration, what is quite common in automotive applications. A typical product example: Velodyne's HDL64 is shown in Fig. 5 (a).

MEMS Micro-Scanning: MEMS (Micro-Electro-Mechanical Systems) technology allows the fabrication of miniature mechanical and electro-mechanical devices using silicon fabrication techniques. In essence, a MEMS mirror is a mirror embedded on a chip [12]. The MEMS mirror is rotated by balancing between two opposite forces: an electromagnetic force (Lorentz force) produced by the conductive coil around the mirror, and an elastic force from torsion bar, which serves as the axis of rotation. This principle is shown in Fig. 4 (b). The MEMS mirrors can be either single-axis for 1D movement [13], or dual-axis for 2D movement. Also, MEMS mirror can work in resonant mode at its characteristic oscillation frequency

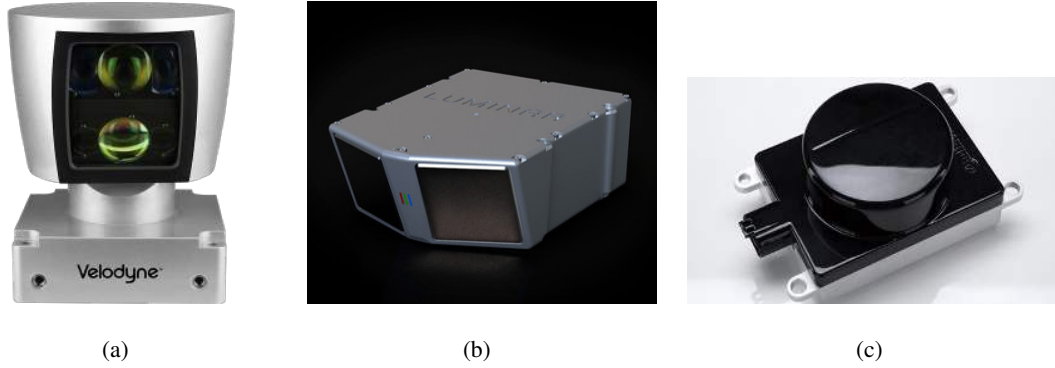


Fig. 5: LiDAR product examples: (a) mechanical spinning 905nm LiDAR from Velodyne, (b) 1550nm MEMS LiDAR from Luminar (c) Flash LiDAR from Continental

to obtain a large deflection angle and high operating frequency. In non-resonant mode, MEMS scanning mirror can be controlled to follow programmed scan trajectory. For example, for MEMS based AEYE LiDAR, the LiDAR can dynamically change the FoV and scanning path in order to focus on some critical parts. Although MEMS LiDARs still contain moving parts, this near-solid-state technology is still promising because of the mature techniques in IC industry are able to satisfy the strict cost requirements. An example of Luminar's MEMS 1550nm LiDAR is shown in Fig. 5 (b).

Flash: Originally applied for spacecraft in autonomous landing and docking with satellites, 3D flash LiDARs [14] totally remove the rotating parts within scanning systems. Hence, they are truly solid-state. A flash LiDAR behaves as a camera. A single laser that is spread by an optical diffuser to illuminate the whole scene at once. Then, it uses a 2D array of photodiodes (similar to the CMOS/CCD for camera) to capture the laser returns, which are finally processed to form a 3D point clouds, as shown in Fig. 4 (c). Since all the pixels of flash LiDAR measure the ranges simultaneously, the issue of movement compensation caused by platform motion is avoided. In addition, the semiconductor-based 3D flash LiDARs facilitate fabrication and packaging for massive production that lead to lower cost. However, the critical issue of 3D flash LiDAR is its limited detecting range (usually $< 100m$) because a single diffused laser is responsible for detecting the whole area under a small power threshold for eye-safety. Another disadvantage is the limited FoV because it cannot rotate and scan the surroundings like a scanning-type LiDAR does. A typical example of Continental's commercial flash LiDAR product for middle range perception is shown in Fig. 5

OPA (Optical Phased Array): As a type of true solid-state LiDAR, optical phased array (OPA) LiDARs [15] [16] don't comprise moving components. Similar to the phased array Radar, an OPA is able to steer

		Mechanical Spinning	MEMS	Flash	OPA	Undisclosed
ToF LiDAR	NIR	Velodyne, IBEO, Valeo, Ouster* Hesai, Robosense	Innoviz Robosense	Continental Xenomatix	Quanergy	
	SWIR	Luminar	AEYE Hesai	Argo* (Princeton Lightwave)		
FMCW LiDAR					Cruise (Strobe)	Aurora (Blackmore,1550nm)

TABLE I: Representative LiDAR manufactures and the adopted technologies. The manufactures marked by * utilize single photon Geiger-mode SPAD as photodetector.

the laser beams through various types of phase modulators. The speed of light can be changed by the optical phase modulators when the lasers are passing through the lens, as illustrated in Fig. 4 (d). Consequently, different light speeds in different paths allow control of the optical wave-front shape and hence the steering angles. Although OPA had been placed high hopes as a promising technology, there is not a commercial product yet in the market.

D. Current Status of Automotive LiDAR

Mechanical spinning LiDAR is the first stepping into mass-produced car. Announced in 2017, Audi released its latest luxury sedan A8 that equipped a Valeo's Scala LiDAR for automated driving functions – the first commercial available vehicle carrying automotive grade LiDAR in the world. Valeo's scala ³ is a 4 layer mechanical spinning LiDAR similar to its cousin, IBEO Lux4. Empowered by Scala, A8 is able to achieve L3 level automated driving functions, without the need of hands on the steering wheel (need to be allowed by legislation). In 2019, VALEO gained a €500M order of its next generation LiDAR – Scala2, from several car manufacturers.

In the same time, to reduce the cost and to improve the robustness, plenty of companies focus on solid state scanning system. As shown in Tab. I, Innoviz, Continental and Quanergy are developing MEMS, Flash and OPA LiDARs, respectively. In 2018, BMW announced the collaboration with Innoviz for series production in 2021. To increase the maximum detection range, some used SPAD array working in single-photon detection mode (Geiger-mode). Ouster OS-1 64 ⁴ adopted CMOS based SPAD to detect

³<https://www.valeo.com/en/valeo-scala/>

⁴<https://www.ouster.io/blog-posts/2018/11/8/how-multi-beam-flash-lidar-works>

850nm lasers emitted by 2D VCSEL laser array. Toyota made a LiDAR prototype containing a CMOS SPAD array (202x96 pixels) for receiving 905nm lasers [17]. Princeton Lightware (acquired by Argo.ai) also realized a SPAD LiDAR prototype ⁵, while few information is disclosed. As for SiPM, a LiDAR prototype [18] has been made by SensL (acquired by OmmniVision), while the commercial products are still under development.

Some switch to SWIR laser (e.g. 1550nm) because of the higher power threshold allowed than NIR laser, such as Luminar (announced collaboration with Toyota), AEYE. Coherent detection based FMCW LiDARs are chased by car manufactures and investments as well. Strobe and Blackmore, the two representative FMCW LiDAR start-ups, are quickly acquired by Cruise and Aurora, respectively. In I, we classify and list several representative automotive LiDAR suppliers and their disclosed technologies.

Another trend of LiDAR is to overcome the adverse weather conditions, such as rain, fog, snow, dust, etc [19], [20]. According to Eq. 1, the adverse weather conditions increase the transmission loss T_r and weaken the reflectivity of an object β that the received energy becomes less. Because SWIR laser (e.g. 1550nm) can achieve higher transmission power, LiDARs belonging to this wavelength are expected to have better performances in harsh weather.

III. LiDAR PERCEPTION SYSTEM

For an autonomous vehicle, its perception system interprets the perceived environment into hierarchical object descriptions (i.e. physical, semantic, intention awareness) from the perception sensor outputs, localization and map data. As shown in Fig. 6, traditional pipeline [4], [21] of processing LiDAR data consists of 4 steps: *object detection*, *tracking*, *recognition* and *motion prediction*. The recent rise of deep learning technologies is changing this classic flow that we will introduce it after the classic approaches. Due to the popularity of Velodyne LiDARs in research communities, the reviewed data processing methods are mainly based on this mechanical spinning LiDAR.

A. Object Detection

The object detection algorithms extract the object candidates and estimate their physical information: the positions and shapes of the detected objects. Since in most traffic scenes, targets are perpendicular to a flat ground, object detection algorithms usually comprise: *Ground filtering* and *Clustering*. *Ground filtering* labels a point cloud either ground or non-ground. Then, non-ground points are grouped into different objects by *clustering* methods.

⁵<https://www.youtube.com/watch?v=K2YkrFSJc3A>

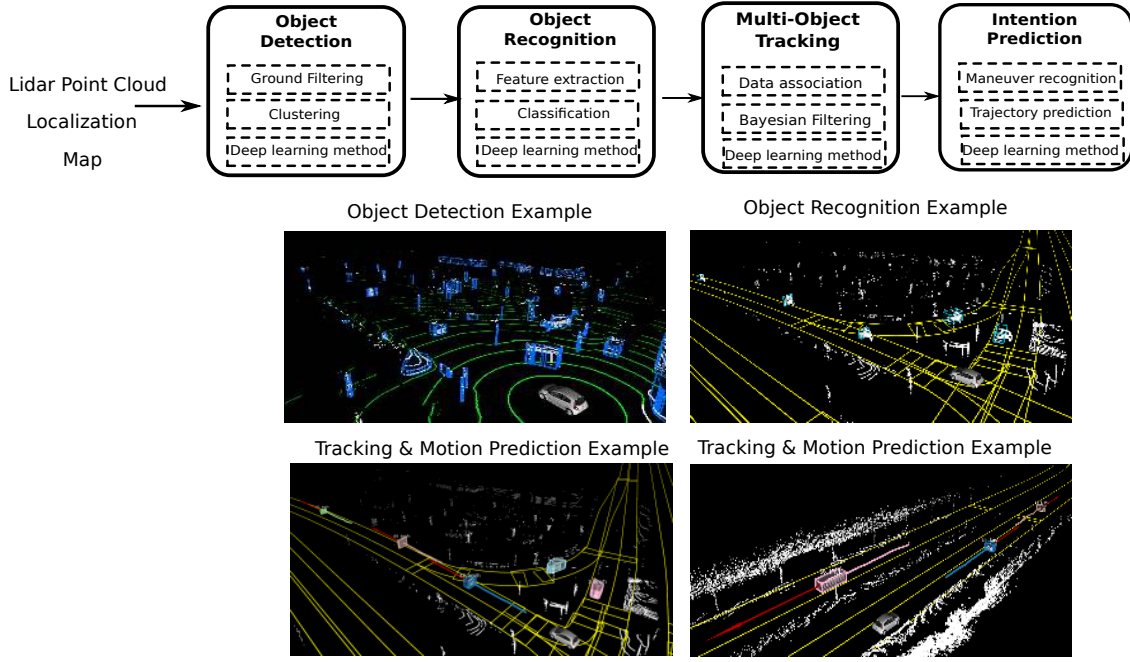


Fig. 6: Pipeline of a classic LiDAR perception system, with exemplary outputs of each step. The examples are from authors' platform as shown in Fig. 1. Please note that after object detection, we only process the objects within the road (as denoted by the yellow lines).

In an early research [4], the point clouds from LiDAR are projected into polar grids subdividing 360° around the LiDAR. The points inside each grid cell are treated consecutively to generate a virtual scan, which specifies the region of *free*, *occupied* and *occluded*. Occupied virtual scans are grouped into object clusters. [21] followed this method while utilized grid-based local plane fitting approach instead of processing every point as in [4]. The grids that are able to be fitted as a plane are classified as ground grids, the remaining non-ground grids are clustered by the connected component labeling (CCL). However, the polar grid based methods always need a projection of 3D LiDAR points into discrete grids, which loses raw information from LiDAR measurements.

Processing LiDAR signals in spherical coordinates (r, φ, θ) provides a better approach. For the Velodyne UltraPuck used by the authors, vertical angle for each laser beam is fixed, azimuth angle is decided by the scanning time and motor speed. Therefore, every range reading can be represented by $P_{i,j} = (\rho_{i,j}, \varphi_i, \theta_j)$, where i refers to a certain laser beam, j is the azimuth angle index, as seen in Fig. 7. This approach naturally fills the range readings into a predefined data buffer (range image) and hence allows fast access to a point and its neighbours. Processing LiDAR data in range view is becoming popular in recent years. For instance, based on a range image, [22] segmented the ground points in

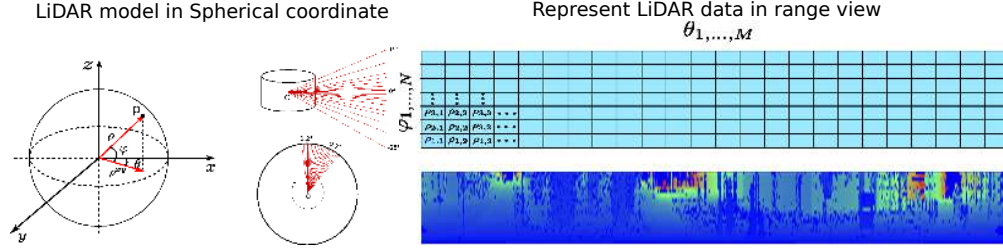


Fig. 7: Range view of spinning LiDAR (Velodyne UltraPuck) for further processing. The range image (32x1800) in pseudo-color facilitates the following processing.

each column. The remaining non-ground points are grouped easily through criterions of distance and angle. For a 32 beams LiDAR, they reached 4ms in an Intel i5 processor. [23] processed the range image row by row. They applied the clustering in each scan-line (actually the row in range image), and then merge the clusters scan-line by scan-line. Fig. 6 shows a sample result of ground filtering and clustering from our implementation based on range image as well. The green points are ground points, non-ground points are grouped into object candidates (in blue polygons). Object detection provides an initial physical information, e.g. position of the an object. The following steps such as recognition and tracking complement semantic and more physical information e.g. heading, speed, to the detected objects.

B. Object Recognition

Machine learning based object recognition methods furnish the semantic information (e.g. classes of pedestrian, vehicle, truck, tree, building, etc) to the detected objects. A typical recognition process employed in [21] comprises a *feature extraction* step calculating compact object descriptors, and a *classification* step where pre-trained classifiers predict the categories of objects based on the extracted features. As summarized in [24], the features proposed in literature can be roughly divided into two classes: the *global features* for the whole object, or the *local features* for each point. An object's size, radius, central moments or the maximum intensity [21], [25] are the most basic global features. Applying Principal Component Analysis (PCA) in 3D point clouds is another effective method to acquire global shape features. As adopted in [26], three salience features (surfaceness, linearness, scatterness) can be acquired by analyzing the eigenvalues acquired from PCA. As for local feature, [21] calculated the three salience features for each point and its neighboring points. 3 histograms, each containing 4 bins spaced between 0 and 1, for the three salience features are extracted as local features. A more complicated feature is the Spin Image (SI) introduced by [27]. A SI is created by spinning a grid around the surface normal n of a given point p . The virtual pixels of a SI are the distances either to line through the n or

to the plane defined by p and n . [26] transformed this individual point-wise feature to global feature: for an object, only the SI of its central point is utilized as object descriptor. In literature, there are more sophisticated features, such as Global Fourier Histogram (GFH) [24] descriptor. However, real-time requirement restrains the complexity of features.

After feature extraction, the following classification is a typical supervised machine learning process: a classifier trained by a ground truth dataset predict the class of input objects. Well-known datasets such as KITTI ⁶ provide abundant resources. From the arsenal of machine learning (ML), plenty of ML tools can be applied, such as Naive Bayes [25], Support Vector Machine (SVM) [21], [24], [28], KNN [26], Random Forest (RF) and Gradient Boosting Tree (GBT) [29] can be applied. SVM with RBF (radial basis function) kernel is still the most popular method, considering its speed and accuracy. Fig. 6 shows the recognition results on the detected on-road objects based on our implementations (SVM with RBF kernel). Recently, [30] applied an evidential neural network to classify the LiDAR object. Evidential classifier can better handle the *unknown* classes that are frequently encountered in practice.

C. Object Tracking

Multiple object tracking (MOT) algorithms correlate and locate the detected/recognized objects through spatio-temporal consistency. MOT maintains the identities of detected objects and yields their physical states i.e. trajectories, poses, velocities. MOT is a classic engineering problem [31] that has been researched for a long time. A basic architecture mainly comprises *single object tracker* that "optimally" estimates the state of tracked object, *data association* that assign the new detections to the trackers.

A *single object tracker* models the movement as a dynamic state-space model and estimates the state under the Bayesian filtering framework. Kalman Filter (KF) family – the classic KF under Gaussian-linear assumption and its variants Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF), is the popular toolbox. [32] employed a KF with a constant velocity model to track LiDAR detections. As a non-linear version of KF, EKF is utilized for LiDAR object tracking in [33]. Extending the single dynamic model to multiple maneuver models, the interacting multiple model (IMM) filter is able to handle more complicated cases. IMM filter consists of several filters running in parallel and each filter use different motion model. For a single object, an IMM-UKF filter was applied in [34], where three UKFs work for three motion models: Constant Velocity, Constant Turn Rate and Random Motion.

As another common approach, Particle Filter (PF) is designed for more general cases that not meet the Gaussian-linear assumption. The application of PF in LiDAR data processing can be dated from DARPA

⁶<http://www.cvlibs.net/datasets/kitti/>

Grand Challenge 2017 [4], where a Rao-Blackwellized Particle Filter is used. However, PF requires a large number of particles, especially for high dimensional state space. Hence, KF family is more popular in real-time perception systems.

Data association associates the detections with the tracks. The simplest method is the nearest neighbor (NN) filter (implemented in [32]) which assigns the detections to their closest tracks, based on Euclidean or Mahalanobis distance between the detection and the track. NN filter is insufficient for clutter scenarios. In contrast, the joint probabilistic data association filter (JPDAF) offers a soft, probabilistic way for the detection-track association. JPDAF considers all the possible detections (including no detection) in a gating window and estimates their assignment probabilities to the tracks and takes the weighted average of all the association hypotheses. In [34], JPDAF was applied for data association and an IMM-UKF tracker is to track an individual object.

In contrast to the radar based MOT in which all the detections are usually modeled as points, LiDAR based MOT is distinctive in that it should track the shapes of detections as well. The simplest shape model is a 2D bounding box [4], which assumes the detections are car-like objects. L-shape fitting [35] is the most common approach estimating the bounding box's center, width, height and heading. However, 2D bounding box is insufficient for more general objects, such as pedestrian, tree, building etc. A more sophisticated method [36] implemented multiple shape models: points, polygons, L-shape and lines for various objects. When tracking a moving object, its shape varies with the changes of pose and sensor viewpoint. [37] implemented a tracking method that simultaneously estimate the states of both poses and shapes represented by 2D polylines.

D. Object Intention Prediction

The previous introduced modules provide the past and current information of detected targets. While in autonomous driving systems, decision making and path planning algorithms require future motion of the tracked targets. Prior works based on certain kinematic models that are assumed to perfectly fit the detected objects, are not applicable for long term prediction. To address this shortcoming, maneuver or behavior recognition is proposed based on machine learning methods. Common maneuvers for a vehicle are cut-in, lane-change, brake, over-taking, etc. [38] modeled the behaviors of car-following, lane-change by GMM (Gaussian mixture model) or HMM (Hidden Markov Model). Based on the maneuver classification realized by HMM, [39] predicted the vehicles' motions by VGMMs (variational Gaussian mixture models) under the constraints of vehicle interaction models. With the success of RNN in modeling temporal sequential data, LSTM based methods are becoming popular. [40] used LSTM to classify the drivers' intentions at intersections, and the results shown that LSTM outperforms other traditional

machine learning methods. [41] proposed an encoder-decoder LSTM model to recognize maneuver and predict trajectory. Beyond recognizing the maneuver of a single object, *social LSTM* [42] was proposed to capture the interactions of all the objects. This is achieved by *social pooling*, which down-samples a target's neighboring objects' LSTM states into a *social tensor*. [43] and [44] applied and improved original social pooling part for the purpose of vehicle trajectory prediction.

E. Emerging Deep Learning Methods

After the huge success in computer vision and speech recognition, waves of deep learning (DL) arrived in LiDAR data processing as well. Deep learning [45] is a subset of machine learning algorithms that mainly uses multi-layer neural network. In contrast to the traditional machine learning methods such as SVM, DL technologies are able to automatically extract features from the raw input. Convolutional neural networks (CNN) and recurrent neural networks (RNN), such as LSTM (long short-term memory), are the most frequently used tools.

The basic components of perception system, *ground segmentation*, *object detection*, *tracking* and *recognition* can all be realized by deep neural networks (DNN). For instance, [46] segmented ground points by applying CNN to LiDAR points represented by multi-channel range images. In contrast to the object detection based on clustering that arbitrary targets can be detected, DNN based solutions achieve object detection by recognition, owing to the paradigm of supervised learning. As in [47], vehicles can be detected by CNN based neural network in a BEV (bird's eye view) representation of LiDAR points. A more complex neural network was proposed in [48], CNN is utilized in both range image and BEV of LiDAR data, and then fused with camera detections. However, due to the physical limits of LiDAR, only vehicles can be effectively detected by LiDAR, the best achieved results of pedestrian detection in KITTI benchmark is only 52.40% in average precision (by method DENFIDet when writing this paper) ⁷. [49] integrated the evidential theory into DL architecture for LiDAR based road segmentation and mapping. Object tracking has been realized by DL as well. In contrast to the *tracking by filtering* framework described in traditional tracking algorithm, [50] proposed a deep structure model under *tracking by detection* framework. A detection net first processes a sequence of LiDAR data and images to generate detection proposals. Then, tracks are estimated through finding best associations of detections, which is achieved by a marching net and scoring net.

Apart from improving the traditional perception components, point-wise semantic segmentation, which was hard to be realized before, is now achievable by deep learning. A well-known method PointNet was

⁷http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=bev

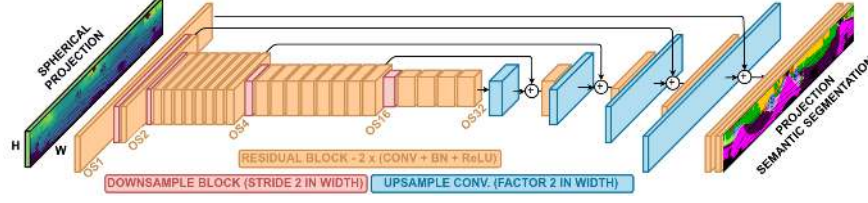


Fig. 8: The neural network structure proposed in RangeNet [51].

proposed in [52] to semantically segment 3D point clouds for general usages. However, due to sparsity of LiDAR data w.r.t distance, the method doesn't work well for autonomous driving scenarios. SqueezeSeg [53] achieved real-time segmentation by applying CNN in range view of LiDAR points. Due to a lack of massive annotated datasets, the performances of these two methods are not ready for deployment in real usage. While this situation has been changed by SemanticKITTI ⁸ [54] – the latest and the biggest point-wise annotated dataset based on KITTI. Based on this dataset, RangeNet [51] demonstrated fascinating performance and speed from a not complicated DNN structure. Fig. 8 shows the structure and a sample result of RangeNet. With more and more annotated datasets, we have sufficient reason to expect LiDAR based semantic segmentation will have better performance.

IV. CONCLUSION AND FUTURE DIRECTIONS

In this article, a review of LiDAR technologies is presented in the beginning. How a LiDAR "sees" the world and what constitutes a LiDAR are introduced. The main development directions of LiDAR technologies are analyzed as well. In summary, current automotive LiDARs face the constraints, or challenges as: 1) cost, 2) meet automotive reliability and safety standards (e.g. ISO26262, IEC61508), 3) long measuring distance (e.g. >200m for highway applications), 4) adverse weather i.e. rain, fog, snow etc, 5) image-level resolution, 6) smaller size facilitating integration. At present, all the possible solutions varying from laser sources (905nm V.S. 1550nm), scanning methods (Spinning/MEMS/OPA/Flash) or ranging principles (ToF or FMCW), are exploited to overcome several or all of these difficulties. It is really hard to predict which automotive LiDAR solution(s) will dominate the future, whereas one thing is sure: automotive LiDARs are walking out of experimental platforms, entering more and more mass-produced cars.

Then, a compact tutorial of the LiDAR based perception systems for autonomous driving is presented. Three levels of information providing by perception systems, along with typical processing pipeline are

⁸<http://www.semantic-kitti.org/>

introduced. Generally, comparing to camera or radar, LiDAR is the most precise sensor in measuring range. Therefore, the physical information (objects' positions, headings, shapes, etc.) evaluated by LiDAR based algorithms is highly reliable. However, semantic description is the shortcoming of LiDAR. This is caused by the LiDAR's poor resolution and its essence as a distance measuring sensor, not contextual sensor. Fusion with cameras remedies LiDAR's weakness in recognition. The intention prediction level is independent of specific sensor, while it will be strengthened by the precise physical information bringing by LiDAR. Applying deep learning in LiDARs' 3D data will be one of the most important directions in the future. Lacking huge amount of annotated 3D point cloud dataset was the bottleneck of successfully applying deep learning methods. However, things are changing. The aforementioned SemanticKITTI initiated a good start, and the results achieved by RangeNet++ are quite impressive. From the authors' point of view, algorithms for extracting more accurate physical information and squeezing the LiDARs' potentials in semantic estimation are the future directions. And of cause, with the fast progress of new LiDARs, new algorithms will emerge adapted to specific LiDARs.

REFERENCES

- [1] J. L. Leonard., "A perception-driven autonomous urban vehicle," *Journal of Field Robotics*, vol. 25, pp. 727–774, 2008.
- [2] National Research Council, *Laser Radar: Progress and Opportunities in Active Electro-Optical Sensing*. National Academies Press, 2014.
- [3] C. Urmson *et al.*, "Autonomous driving in urban environments: Boss and the Urban Challenge," *Journal of Field Robotics*, vol. 25, pp. 425–466, 2008.
- [4] A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Journal of Autonomous Robots, Springer*, vol. 26, pp. 123–139, 2009.
- [5] Y. Li, "Stereo vision and lidar based dynamic occupancy grid mapping: Application to scenes analysis for intelligent vehicles," Ph.D. dissertation, University of Technology of Belfort Montbeliard, 2013.
- [6] U. Wandinger, "Introduction to LiDAR," *Springer Series in Optical Sciences, Springer*, vol. 102, 2005.
- [7] *IEC 60825-1: Safety of Laser Products*, International Electrotechnical Commission Std., 2017.
- [8] D. F. Pierrottet *et al.*, "Linear FMCW Laser Radar for Precision Range and Vector Velocity Measurements," in *MRS Proceedings*, 2008.
- [9] "Lidar and other techniques," Hamamatsu Photonics, Tech. Rep., 2017.
- [10] R. Horaud, M. Hansard, G. Evangelidis, and C. Mnier, "An Overview of Depth Cameras and Range Scanners Based on Time-of-Flight Technologies," *Machine Vision and Applications*, vol. 27, pp. 1005–1020, 2016.
- [11] J. O'Neill *et al.*, "Scanning system for LiDAR," U.S. Patent US 8 072 663B2, 2011.
- [12] H. W. Yoo and G. Schitter, "Mems-based lidar for autonomous driving," *Elektrotechnik und Informationstechnik*, vol. 135, pp. 408–415, 2018.
- [13] N. Druml *et al.*, "1d mems micro-scanning lidar," in *The 9th International Conference on Sensor Device Technologies and Applications*, Venice, Italy, 2018.
- [14] F. Amzajerdian *et al.*, "Imaging Flash Lidar for Autonomous Safe Landing and Spacecraft Proximity Operation," in *AIAA SPACE*, Long Beach, U.S.A, 2016.

- [15] C. V. Poulton *et al.*, “Optical phased arrays for integrated beam steering,” in *IEEE 15th International Conference on Group IV Photonics (GFP)*, 2018.
- [16] P. F. Mcmanamon *et al.*, “Optical phased array technology,” *Proceedings of the IEEE*, vol. 84, pp. 268–298, 1996.
- [17] T. I *et al.*, “Single-Photon Avalanche Diode with Enhanced NIR-Sensitivity for Automotive LiDAR Systems,” *Sensors*, vol. 16, pp. 459–468, 2016.
- [18] S. Gneccchi and C. Jackson, “A 1 16 sipm array for automotive 3d imaging LiDAR systems,” in *International Image Sensor Workshop (IISW)*, Hiroshima, Japan, 2017, pp. 133–136.
- [19] S. Michaud, J.-F. Lalonde, and P. Gigure, “Towards characterizing the behavior of LiDARs in snowy conditions,” in *IEEE/RSJ IROS Workshop on Planning, Perception and Navigation for Intelligent Vehicles*, 2015.
- [20] M. Kutila, P. Pyykonen, H. Holzhuter, M. Colomb, and P. Duthon, “Automotive LiDAR performance verification in fog and rain,” in *21st IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2018.
- [21] M. Himmelsbach *et al.*, “Lidar based 3D Object Perception,” in *1st International Workshop on Cognition for Technical Systems*, 2008.
- [22] I. Bogoslavskyi and C. Stachniss, “Fast range image-based segmentation of sparse 3d laser scans for online operation,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016.
- [23] D. Zermas *et al.*, “Fast Segmentation of 3D Point Clouds: A Paradigm on LiDAR Data for Autonomous Vehicle Applications,” in *IEEE International Conference on Robotics and Automation (ICRA)*, Singapore, 2017.
- [24] T. Chen and D. Liu, “Performance of global descriptors for velodyne-based urban object recognition,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2014.
- [25] C. Premebida, O. Ludwig, and U. Nunes, “Exploiting LIDAR-based features on pedestrian detection in urban scenarios,” in *IEEE International Conference on Intelligent Transportation Systems (ITSC)*, 2009.
- [26] B. Douillard *et al.*, “A Pipeline for the Segmentation and Classification of 3D Point Clouds,” *Experimental Robotics, Springer Tracts in Advanced Robotics*, 2014.
- [27] A.E.Johnson and M. Hebert, “Using spin images for efficient object recognition in cluttered 3d scenes,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, pp. 433–449, 1999.
- [28] D. Z. Wang, I. Posner, and P. Newman, “What Could Move? Finding Cars, Pedestrians and Bicyclists in 3D Laser Data,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012.
- [29] Y. Li and Y. Ruichek, “Moving Objects Detection and Recognition Using Sparse Spacial Information in Urban Environments,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2012.
- [30] E. Capellier, F. Davoine, V. Cherfaoui, and Y. Li, “Evidential deep learning for arbitrary LIDAR object classification in the context of autonomous driving,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2019.
- [31] S. Challa *et al.*, *Fundamentals of Object Tracking*. Cambridge University Press, 2011.
- [32] F. Moosmann and C. Stiller, “Joint Self-Localization and Tracking of Generic Objects in 3D Range Data,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2013.
- [33] H.Cho *et al.*, “A Multi-Sensor Fusion System for Moving Object Detection and Tracking in Urban Driving Environments,” in *IEEE International Conference on Robotics & Automation (ICRA)*, 2014.
- [34] A. A. Rachman, “3d-lidar multi object tracking for autonomous driving,” Master’s thesis, Delft University of Technology, Delft, Netherlands, 2017.
- [35] X. Zhang, W. Xu, C. Dong, and J. M. Dolan, “Efficient L-Shape Fitting for Vehicle Detection Using Laser Scanners,” in *IEEE Intelligent Vehicles Symposium (IV)*, 2017.

- [36] T. Johansson, "Lidar clustering and shape extraction for automotive applications," Master's thesis, Chalmers University of Technology, Gothenburg, Sweden, 2017.
- [37] S. Kraemer *et al.*, "Lidar based object tracking and shape estimation using polylines and free-space information," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.
- [38] C. Miyajima and K. Takeda, "Driver-behavior Modeling Using On-Road Driving Data: A new application for behavior signal processing," *IEEE Signal Processing Magazine*, vol. 33, pp. 14–21, 2016.
- [39] N. Deo *et al.*, "How would surround vehicles move? A unified framework for maneuver classification and motion prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, pp. 129–140, 2018.
- [40] D. J. Phillips *et al.*, "Generalizable intention prediction of human drivers at intersections," in *IEEE Intelligent Vehicles Symposium (IV)*, 2017.
- [41] N. Deo and M. M. Trivedi, "Multi-modal Trajectory Prediction of Surrounding Vehicles with Maneuver based LSTMs," in *IEEE Intelligent Vehicles Symposium (IV)*, 2018.
- [42] A. Alahi *et al.*, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [43] N. Deo and M. M. Trivedi, "Convolutional social pooling for vehicle trajectory prediction," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018.
- [44] K. Messaoud *et al.*, "Non-local social pooling for vehicle trajectory prediction," in *IEEE Intelligent Vehicles Symposium (IV)*, 2019.
- [45] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [46] M. Velas *et al.*, "Cnn for very fast ground segmentation in velodyne LiDAR data," in *IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2018.
- [47] B. Yang, W. Luo, and R. Urtasun, "Pixor: Real-time 3D Object Detection from Point Clouds," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [48] X. Chen *et al.*, "Multi-view 3D Object Detection Network for Autonomous Driving," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [49] E. Capellier, F. Davoine, V. Cherfaoui, and Y. Li, "Transformation-adversarial network for road detection in lidar rings, and model-free evidential road grid mapping," in *11th PPNIV - IROS Workshop on Planning, Perception, Navigation for Intelligent Vehicle*, 2019.
- [50] D. Frossard and R. Urtasun, "End-to-end Learning of Multi-sensor 3D Tracking by Detection," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [51] A. Milioto *et al.*, "RangeNet++: Fast and Accurate LiDAR Semantic Segmentation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.
- [52] C. R. Qi *et al.*, "Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [53] B. Wu *et al.*, "SqueezeSeg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.
- [54] J. Behley *et al.*, "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.