Object Detection and Distance Estimation via Lidar and Camera Fusion for Autonomous Driving

Salma Ariche, Zakaria Boulghasoul, Abdelilah Haijoub, Abdelouahed Tajer, Hafid Griguer, and Abdelhafid El Ouardi

Abstract One of the most main perception challenges for autonomous vehicles is cars detection. Classic vision-based cars identification approaches are insufficiently accurate, particularly for small objects, whereas sensors such as Lidars help in detecting objects in all shapes and sizes but still limited in classifying and recognizing detected obstacles. To fully exploit the benefits of Lidar's depth information and vision's obstacle classification capabilities, this paper presents an object detection and distance estimation via Lidar and camera fusion. Both sensors have varied different characteristics and must be aligned by performing a geometrical transformation and projection to fuse the sensor's data. The main purpose of the conducted research is to fuse sensor data to estimate the distance of objects detected using Tiny YOLOv4. Finally, the results of the evaluations on the KITTI datasets show that the proposed approach enables both object detection and distance estimation.

Keywords Autonomous driving · Vehicle perception · Sensor fusion · Distance estimation · Object detection

1 Introduction

Improving the safety of road users has been a significant challenge for societies for many decades. Researchers and engineers are undertaking several efforts to propose innovative solutions for intelligent transportation systems, which will allow, through control and optimization strategies, to improve the traffic safety.

Interest in self-driving vehicles has grown in recent years because they provide comfort and safety for drivers by relying on three major technological components:

S. Ariche (✉) · Z. Boulghasoul · A. Tajer
System Engineering and Applications Laboratory, Cadi Ayyad University, Marrakech, Morocco
e-mail: salma.ariche@edu.uca.ma

A. Haijoub · H. Griguer
Innovation Laboratory for Operation, Mohammed VI Polytechnic University, Benguerir, Morocco

A. El Ouardi
SATIE, ENS, Paris-Saclay University, 91190 Gif-sur-Yvette, France

M. Bendaoud et al. (eds.), *The Proceedings of the International Conference on Electrical Systems & Automation*, https://doi.org/10.1007/978-981-19-0039-6_4

sensing and perception—a system that collects information and understands the surrounding environment—localization and mapping—allowing the vehicle to map its environment and locate itself at any given time—and finally, a control system—which is primarily responsible for decision making in various situations.

A key aspect of driving autonomous vehicles is the detection of obstacles and other cars through data fusion of several sensors. Cameras, light detection and ranging (LiDAR), and radar sensors are mainly developed for environment perception allowing an excellent geometric and semantic modeling of the vehicle's environment. Each sensor modality observes the environment in its way but is confined to detecting object attribute information.

LiDAR sensor readings, for example, can generate a three-dimensional frame of the surrounding environment. However, they are frequently impacted by severe weather conditions and are limited in their object classification capacity, whereas image sensing technologies are generally used for object detection and recognition but unreliable when it comes to estimating distance or velocity.

Although many research works outline a highly accurate object recognition system based on camera sensors only, there are still substantial challenges to overcome in the context of autonomous driving. Most research work has provided many approaches and algorithms for road object identification and recognition depending on the modality of each sensor, but only a few have addressed the issue of estimating the object distance [1, 2].

In this paper, we present a method for object detection with Tiny YOLOv4 and distance estimation relying on Lidar and camera data provided by the KITTI datasets.

The rest of this paper is structured as follows: Sect. 2 presents an overview of related work of object distance estimation; Sect. 3 describes our suggested approach for fusing data from a self-driving vehicle's camera and LiDAR. Section 4 presents evaluation results and discussions. Finally, in Sect. 5, a conclusion with some suggestions for further research and perspectives.

2 Related Work

Nowadays, sensors are essential devices that are embedded in vehicles, equipped with driver assistance technologies. For self-driving cars primarily, to observe their surroundings, numerous cameras, radar sensors, LiDAR sensors, and ultrasonic sensors are employed [3, 4]. As stated in [5] image processing is a critical component of vision sensing technology since it makes use of an automatic system to comprehend complicated environmental scenes. While a laser radar sensor is used to detect, track the target, a Lidar sensor can be classified based on function, line number, and laser emission waveform. Finally, conventional ultrasonic radars are classified into two types:

1. short-range ultrasonography, which helps detect impediments in the front and back.

2. long-distance ultrasonography, that measures the distance between the side barrier and the vehicle.

Autonomous driving researchers employ various sensor combinations to compensate the limits of each sensor as already indicated. Authors in [6] describe Sensor data fusion as the process of manipulating data and information from heterogeneous sensors to improve particular criteria and data elements for decision tasks. A normal camera's resolution, for example, is far greater than that of a detection and ranging sensor, but the camera has a restricted field of view and cannot give correct or precise distance and velocity information of detected objects in contrast to LiDAR, which is also limited by its inability to recognize color and classify items. Therefore sensor fusion approaches are unavoidably required for the safety and reliability of an autonomous vehicle. Authors in [6] list the 3 most used schemes of sensor fusion as illustrated in Fig. 1. Early fusion or raw data level, feature fusion (feature extraction then halfway fusion across the network), and finally decision fusion or late fusion which involves making final decisions. An overview of sensors and sensor fusion technologies in self-driving cars is provided by [5–7].

The fusion approach makes a correspondence between the 3D points from LiDAR and the RGB images of a camera. Authors in [8] reviewed environment perception algorithms for intelligent vehicles, with emphasis on lane and road, traffic sign detection, recognition, and scene comprehension. Multi-sensor approaches and a single fusion algorithm were conducted by [9, 10]. Paper [11] employed several fully convolutional neural networks and three different fusion techniques to detect roads using camera pictures and LiDAR point clouds. For advanced driver assistance systems (ADAS), authors in [12] propose a high-level sensor data fusion architecture. Reference [13] provides a hybrid multi-sensor fusion architecture that performs

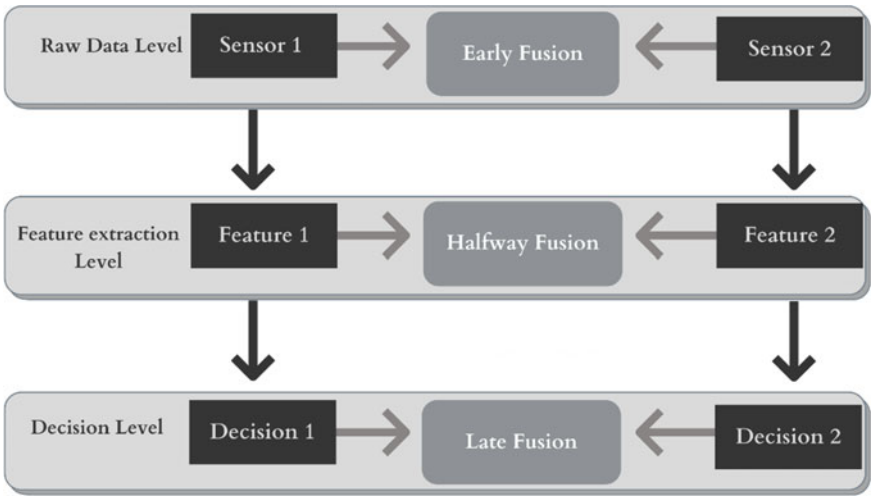

Fig. 1 Three known levels of sensor data fusion [5]

environment perception tasks such as road segmentation, obstacle identification, and tracking. In [14] a fusion method is presented based on fuzzy logic to calculate the object's distance by separately parsing image and point cloud data.

Lidar and camera fusion should be seen as an extrinsic calibration process to achieve low-level sensor fusion. This means that the geometrical properties of each sensor, such as its location and orientation, should be taken into consideration [15]. Usually, an external object is used like a trihedral rig [16], a circle [17], or a checkerboard pattern that serves as a target to align the characteristics of the two sensors [18, 19]. In [20], a black circular plane board is employed, to eliminate the checkerboard pattern's numerous inaccuracies. Automatic calibration methods also exist as described in reference [21]. Although many researchers have tried to conduct great calibration and data fusion work with high accuracy detecting objects forms, by putting sensors next to each other. Yet these methods are unsuitable for practical vehicle experiments where the vehicle has to detect recognize and estimate object distance in real-time. However, the authors in [1] make use of a 3D marker to fuse data of a Lidar and Camera with high precision and detect remote regardless of their position on the self-driving vehicle, with further work on distance estimation for object detection by data fusion on real road.

In this paper, we propose a method of fusing data from Lidar and camera to measure the distance of the detected object, relying on a robust and accurate detection algorithm like Tiny YOLOv4. Finally with the provided data from the KITTI datasets we evaluate our algorithm on real-life driving scenarios.

3 Sensor Fusion for Autonomous Driving

3.1 Perception Sensors

We focus on two key sensors in this paper: camera and LiDAR. As previously stated, both sensors have many limitations and shortcomings, combining the data provided by these two sensors with the appropriate fusion approach enhances detection robustness and performance.

- **Camera:** Equipping two lenses next to each other enables 3D vision, much like in humans. This sort of sensor may give 3D information at a cheap cost, in a small size, and with little power usage. Technologies (CCD, CMOS), resolutions (HD, etc.), and frame rates (up to 100 frames/s) are being continually developed. Furthermore, cameras enable the accurate extraction of geometric and photometric information, paving the way for higher level approaches of scene analysis and interpretation. Obstacle detection, parking assistance, road detection, traffic light and sign detection and identification are some examples.

  Before using camera information, an intrinsic and extrinsic calibration should be performed. Intrinsic parameters are concerned with the camera's intrinsic features (focal length, distortion, and picture center), they reflect a projective

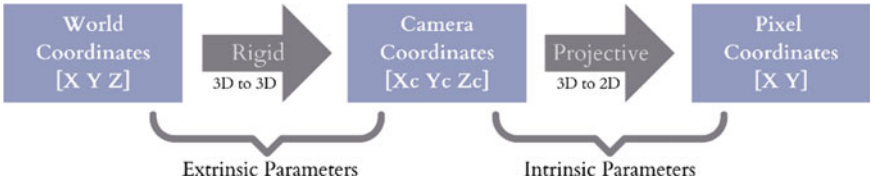

Fig. 2 Calibration with intrinsic and extrinsic parameters [22]

Table 1 Lidar parameters

Parameter	Value
Range length	40–100 m
Resolution accuracy	1.5–10 cm
Vertical angular resolution	0.35°–2°
Horizontal angular resolution	0.2°
Operating frequency	10–20 Hz

translation from the coordinates of the 3-D camera to the coordinates of the 2-D image. While extrinsic parameters refer to a rigid transformation from 3-D world coordinate system to the 3-D camera's coordinate system. Figure 2 highlights the calibration process.

- **Lidar:** Light (or Laser Imaging) Detection and Ranging is an abbreviation for the detection and measuring of distance by light. It is a time-of-flight (ToF) technology that analyzes the properties of an infrared laser (IR) returned by the target to its transmitter. Knowing the speed of light, the Lidar sensor can determine the distance to each object from the time between the laser pulse's emission and return pulse and provides angular resolution (horizontal and vertical). Every second, the Lidar sensor collects millions of accurate distance measurement points, which may be used to generate a 3D matrix of its surroundings. This detailed mapping can offer information on an object's position, shape, and behavior (Table 1).

3.2 The Proposed Sensor Fusion System Overview

This section describes the suggested raw sensor fusion approach for self-driving automobiles. The approach observes its surroundings by employing LiDAR and camera sensors to capture the many physical aspects of the environment (Fig. 3). Our main contributions in this paper are explained as follows:

(a) The first step of the process involves performing a calibration of both sensors using extrinsic and intrinsic matrices. Given the camera intrinsic matrices: only projection matrix if the camera images are rectified, otherwise matrices like S (1*2 size of the image before rectification), K-×× (3*3 calibration of the camera before rectification), and D-×× (1 × 5 distortion coefficients

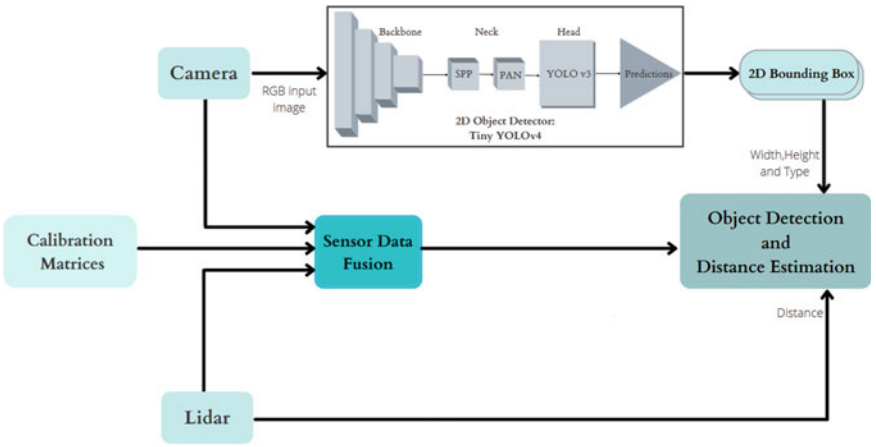

Fig. 3 Flow of the proposed method's process

of camera ×× before rectification) will be needed. Extrinsic parameters like rotation (R: 3 × 3), translation (T: 3 × 1) to convert Velodyne coordinates to camera coordinates. We can then project each 3D LiDAR point onto the camera image plane.

(b) A compressed version of YOLOv4 called Tiny YOLOv4 is utilized in our model to detect existing objects on RGB camera images that are provided as input. The detection algorithms consist of CSPDarknet53 as a backbone, spatial pyramid pooling additional module, PANet path-aggregation neck, and YOLOv3 head. In our study, we use tiny YOLOv4 for faster training and real-time object detection [23, 24].

(c) After calibrating both sensors and performing object detection, the LiDAR points are then projected onto the 2D bounding box. Each detected object is represented with bounding box coordinates (x, y, w, h) and 2D projected LiDAR points [Lidar]_2D(x, y, id), every Lidar point with 2D coordinates is presented as P_{2D}. Retrieved indices of 2D Lidar points that exist inside the 2D bounding Box, are calculated using Eq. (1):

$$\begin{aligned} &\text{Extracted id} \\ &= \left\{ \text{index}_i \in [\text{Lidar}]_{\_2D}(x, y, id), \right. \\ &\left. \text{if } P_{2D,i}[\text{Lidar}]_{\_2D}(x, y, id) \text{inside box}(x, y, w, h) \right\} \end{aligned} \tag{1}$$

(d) Finally, we measure the distance separating the self-driving vehicle and other obstacles relying on the fusion of two perception sensors: camera and LiDAR. We calculate the distance based on the minimum value of each 3D Lidar point which index corresponds to the extracted id calculated in the previous section.

Table 2 Tiny YOLOv4 parameters

Parameter	Value
Batch size	64
Learning rate	0.001
Channels	3
Number of epochs	100

4 Experimental Results and Discussion

4.1 KITTI Datasets and Processing

In this work, a dataset derived from KITTI driving sequences is utilized to verify the validity of our proposed fusion approach, highlighting the benefits of merging LIDAR data with camera pictures for object recognition and distance calculation [25].

For computer vision applications in an autonomous driving context such as perception and localization, the KITTI open-Source dataset provides a collected data by a 1.4 Megapixel color camera synced with a Velodyne Lidar HDL-64E from different scenarios that include eight categories of obstacles: vehicles, vans, trucks, standing and seated people, bicycles, trams, and others.

We generated the training and testing sets by relying on a random selection approach to divide all 7481 photos into the training and testing sets in a 7:3 ratio.

The experiments were carried out using an Intel (R) Core (TM) i7-4600U 2.7 GHz processor, NVIDIA-GPU Tesla K80 and 12G RAM. The size of input images is 416 $\times$ 416. Table 2 presents the training parameters of Tiny YOLOv4.

4.2 Results and Discussion

The KITTI supplied test set was used in the test section. Figures 4, 5 and 6 illustrate some of the evaluation's situations: 1-Object detection via Tiny YOLO v4 (Fig. 4a–c), 2-Lidar point cloud projection onto the 2D Bounding Boxes (Fig. 5a–c), 3-estimating the distance separating the self-driving vehicle and detected Objects (Fig. 6a–c). We chose three random scenarios containing obstacles such as vehicles, pedestrians, cyclists.

Based on the results observed in each figure, it is clear that using the Tiny YOLOv4 algorithm aided in the detection of obstacles surrounding the autonomous vehicle. However, due to the extreme occlusion, notably in scenes 4(a) and 4(b), the model was unable to recognize most of the existing obstacles. Because the Lidar points will only be projected onto the detected 2D bounding boxes, the model will only estimate the distance of identifiable objects within the visual field (Table 3).

Fig. 4 Results of vehicle detection using the Tiny YOLOv4 algorithm for 3 different scenarios **a**, **b** and **c**

As previously indicated, the model allows us to obtain distance information for only detected objects inside the 2D bounding box. Although the Deep Learning algorithm is much faster than YOLO's other versions, it still has many limitations. It is then necessary to upgrade to a higher version, such as Scaled YOLOv4, Embedded YOLOv4, or even YOLOv5, for better and more accurate results in object detection and recognition, resulting in a robust sensor fusion model capable of detecting, recognizing, and estimating distance for all existing obstacles in the surrounding environment.

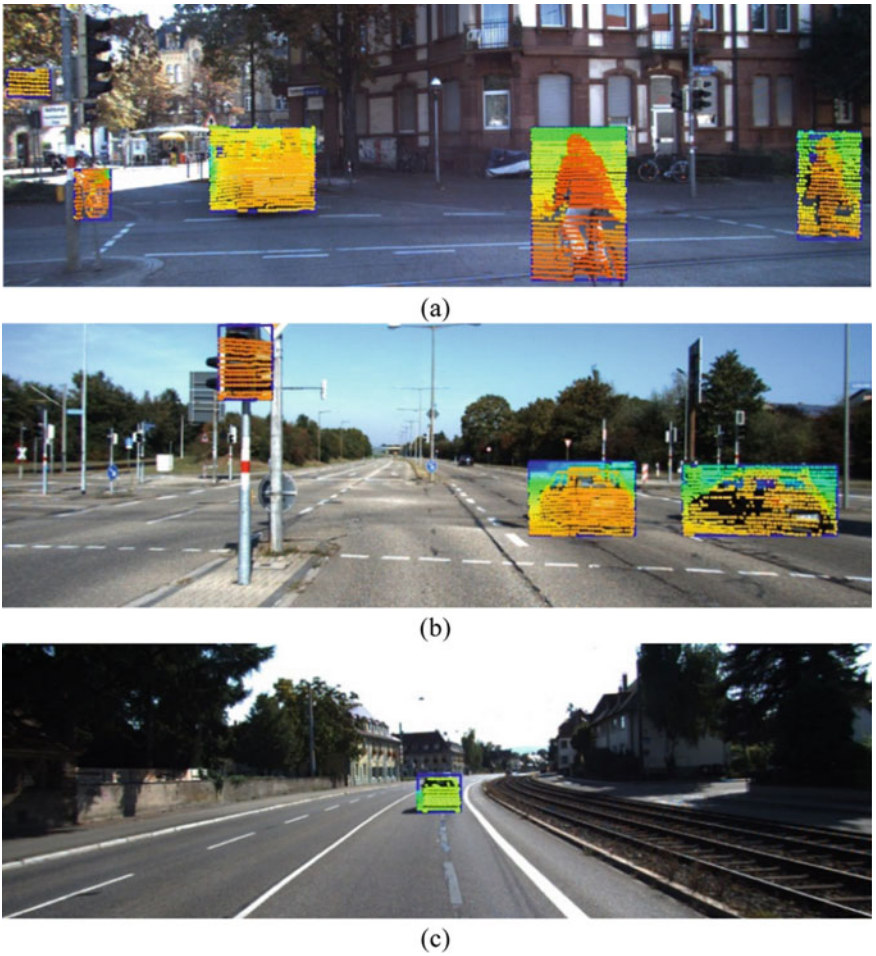

Fig. 5 Lidar points projection onto the 2D bounding box for 3 different scenarios **a**, **b** and **c**

5 Conclusions

This research work presents an object detection and distance estimation approach for self-driving vehicles. A low level real- time data fusion of 2 main perception sensors which are Lidar and Camera was conducted. First, we calibrate the LiDAR and camera sensors, relying on the extrinsic intrinsic characteristics to map the point cloud information onto the camera images. Next, Tiny YOLOv4 algorithm is implemented to detect objects and obstacles in the region of interest. Finally, we evaluate our proposed method by using different scenarios that are provided by KITTI datasets. However, the proposed algorithm has certain flaws that must be fixed. For example, Tiny YOLOv4 is unable to detect all the surrounding objects within a long or near

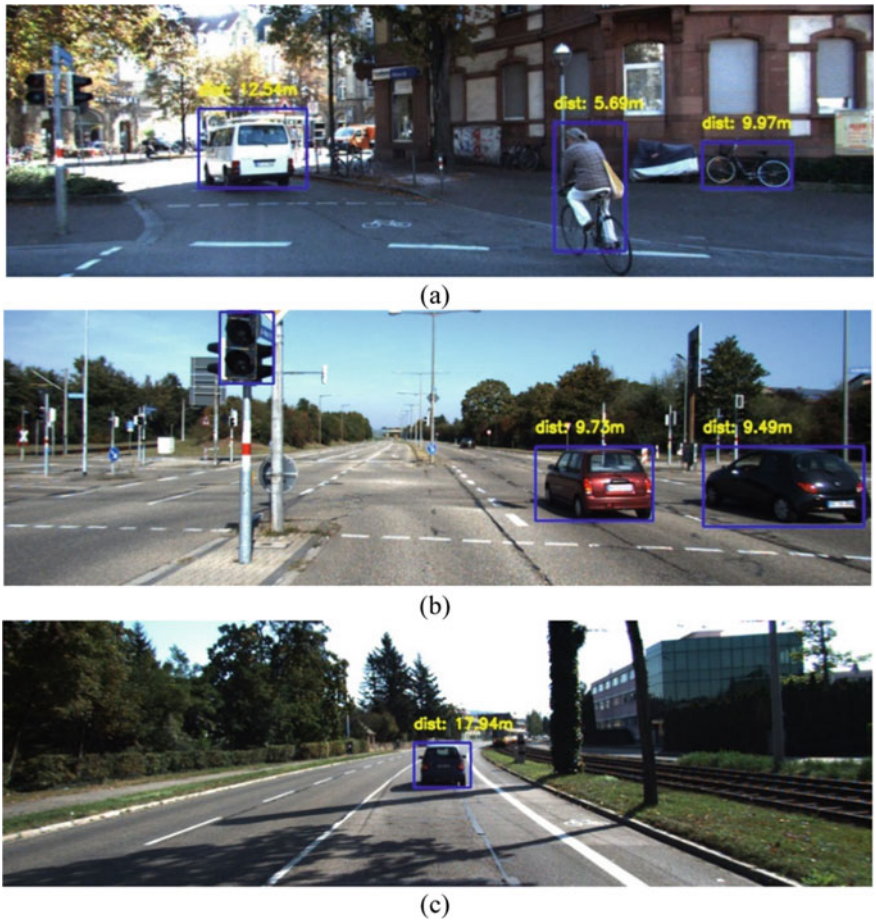

Fig. 6 Output of data fusion for distance estimation

Table 3 Results of Tiny YOLOv4

Scene	Vehicles (%)	Cyclists/pedestrians (%)	Traffic signs (%)
(a)	42	50	–
(b)	80	–	12.5
(c)	100	100	–

distance, thus the algorithm can only estimate the distance of fewer obstacles, which is still insufficient for practical tests.

In our future work, to enhance the robustness of our sensor fusion approach we will try to improve the YOLO algorithm or utilize more specialized Deep Learning models to obtain higher detection accuracy for real-time execution.

References

1. De Silva, V., Roche, J., Kondoz, A.: Fusion of Lidar and camera sensor data for environment sensing in driverless vehicles (2019)
2. Kumar, G.A., Lee, J.H., Hwang, J., et al.: LiDAR and camera fusion approach for object distance estimation in self-driving vehicles. Symmetry **12**, 324 (2020). https://doi.org/10.3390/sym12020324
3. Wang, Z., Wu, Y., Niu, Q.: Multi-sensor fusion in automated driving: a survey. IEEE Access **8**, 2847–2868 (2019). https://doi.org/10.1109/ACCESS.2962554
4. Yeong, D.J., Velasco-Hernandez, G., Barry, J., Walsh, J.: Sensor and sensor fusion technology in autonomous vehicles: a review. Sensors **21**, 2140 (2021). https://doi.org/10.3390/s21062140
5. Chen, Q., Xie, Y., Guo, S., Bai, J., Shu, Q.: Sensing system of environmental perception technologies for driverless vehicle: a review of state of the art and challenges. Sens. Actuators A Phys. **319**, 112566 (2021). https://doi.org/10.1016/j.sna.2021.112566
6. Fayyad, J., Jaradat, M.A., Gruyer, D., Najjaran, H.: Deep learning sensor fusion for autonomous vehicle perception and localization: a review. Sensors **20**, 4220 (2020). https://doi.org/10.3390/s20154220
7. Kocić, J., Jovičić, N., Drndarević, V.: Sensors and sensor fusion in autonomous vehicles. In: 2018 26th Telecommunications Forum (TELFOR), pp. 420–425 (2018). https://doi.org/10.1109/TELFOR.2018.8612054
8. Hu, H., Yuen, K.-V, Mihaylova, L. orcid.org/0000-0001-5856-2223 et al.: Overview of environment perception for intelligent vehicles. IEEE Trans. Intell. Transp. Syst. (2017). https://doi.org/10.1109/TITS.2017.2658662
9. Xiao, L., Dai, B., Liu, D., Hu, T., Wu, T.: CRF based road detection with multi-sensor fusion. In: IEEE Intelligent Vehicles Symposium (IV), pp. 192–198 (2015). https://doi.org/10.1109/IVS.2015.7225685
10. Xiao, L., Wang, R., Dai, B., Fang, Y., Liu, D., Wu, T.: Hybrid conditional random field-based camera-LIDAR fusion for road detection. Inform. Sci. **432**, 543–558 (2018). https://doi.org/10.1016/j.ins.2017.04.048
11. Caltagirone, L., Bellone, M., Svensson, L., Wahde, M.: LIDAR-camera fusion for road detection using fully convolutional neural networks. Rob. Auton. Syst. **111**, 125–131 (2019)
12. Aeberhard, M., Kaempchen, N.: High-level sensor data fusion architecture for vehicle surround environment perception. In: Proceedings of the 8th International Workshop on Intelligent Transportation (WIT 2011), Hamburg, Germany, pp. 22–23 (2011)
13. Shahian Jahromi, B., Tulabandhula, T., Cetin, S.: Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles. Sensors **19**, 4357 (2019). https://doi.org/10.3390/s19204357
14. Shi, J., Wang, W., Wang, X., Sun, H., Lan, X., Xin, J., Zheng, N.: Leveraging spatio-temporal evidence and independent vision channel to improve multi-sensor fusion for vehicle environmental perception. In: Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Changshu, China, pp. 591–596 (2018)
15. Gong, X., Lin, Y., Liu, J.: Extrinsic calibration of a 3D LIDAR and a camera using a trihedron. Opt. Lasers Eng. **51**, 394–401 (2013). https://doi.org/10.1016/j.optlaseng.2012.11.015
16. Alismail, H., Baker, D.L., Browning, B.: Automatic calibration of a range sensor and camera system. In: Proceedings of the 3DiMPVT, Seattle, WA, USA, 29 June–1 July 2013
17. Alismail, H., Baker, D.L., Browning, B.: Automatic calibration of a range sensor and camera system. In: 3DIMPVT (2012)
18. Lipu, Z.: A new minimal solution for the extrinsic calibration of a 2D LIDAR and a camera using three plane-line correspondences. IEEE Sens. J. **14**(2), 442–454 (2014)
19. Lipu, Z., Deng, Z.: A new algorithm for the extrinsic calibration of a 2D LIDAR and a camera. Meas. Sci. Technol. **25**(6) (2014)
20. Zhang, Q., Pless, R.: Extrinsic calibration of a camera and laser range finder (improves camera calibration). In: 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), vol. 3, pp. 2301–2306 (2004)

21. Geiger, A., Moosmann, F., Car, Ö., Schuster, B.: Automatic camera and range sensor calibration using a single shot. In: IEEE International Conference on Robotics and Automation, pp. 3936–3943 (2012).https://doi.org/10.1109/ICRA.2012.6224570
22. Camera Calibration. https://www.mathworks.com/help/vision/ug/camera-calibration.html. Accessed on March 2021
23. Bochkovskiy, A., Wang, C.-Y., Mark Liao, H.-Y.: YOLOv4: optimal speed and accuracy of object detection (2020)
24. Wang, H., Lou, X., Cai, Y., Li Y., Chen, L.: Real-time vehicle detection algorithm based on vision and Lidar point cloud fusion. J. Sens. **2019** (2019). Article ID 8473980. https://doi.org/10.1155/2019/8473980
25. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: the KITTI dataset. Int. J. Rob. Res. **32**(11), 1231–1237 (2013)