

Sensor system for development of perception systems for ATO

Rustam Tagiew¹ · Dirk Leinhos² · Henrik von der Haar² · Christian Klotz¹ · Dennis Sprute³ · Jens Ziehn³ · Andreas Schmelter⁴ · Stefan Witte⁴ · Pavel Klasek¹

Received: 28 February 2023 / Accepted: 8 May 2023

Published online: 08 June 2023

© The Author(s) 2023 [OPEN](#)

Abstract

Developing AI systems for automatic train operation (ATO) requires developers to have a deep understanding of the human tasks they are trying to replace. This paper fills this gap and translates the regulatory requirements from the context of German railways for the AI developer community. As a result, tasks such as train's path monitoring for collision prediction, signal detection, door operation, etc. are identified. Based on this analysis, a functionally justified sensor setup with detailed configuration requirements is presented. This setup was also evaluated by a survey within the railway industry. The evaluated sensors include RGB/IR cameras, LIDARs, radars and ultrasonic sensors. Calculations and estimates for the evaluated sensors are presented graphically and included in this paper. However, the ultimate sensor setup is still a subject of research. The results of this paper also address the lack of training and test datasets for railway AI systems. It is proposed to acquire research datasets that will allow the training of domain adaptation algorithms to transform other datasets, thus increasing the number of available datasets. The sensor setup is also recommended for such research datasets.

Keywords Automatic train operation · ATO · GoA3 · GoA4 · Perception · AI

1 Introduction

ATO is already demonstrating its potential to improve punctuality, flexibility and capacity, for example in metro services on isolated tracks. The benefits and importance of driverless operation for the rail industry are well known [1]. In order for rail traffic to make the necessary contributions to meet climate change targets in the future, while at the same time becoming more efficient and flexible and able to carry more passengers and freight, the introduction of ATO is also necessary for mainline rail traffic on open tracks [1]. The core aspect of driverless ATO systems is that today's tasks of the train driver, including the detection of obstacles during journeys as well as the detection of operational hazards, are transferred to a technical system. The development process of such perception systems requires appropriate data collection and (evaluation) algorithms, which in turn are obtained from real test and training data.

The classification of ATO systems is based on the Grade of Automation (GoA). This indicates which tasks are automated and which are not, i.e. are performed by a human. GoA4 requires neither a train driver nor attendant, so their tasks are performed by technical systems. At level GoA3, there is a train attendant, but no permanent train driver on duty. In GoA2, the driver is still responsible for the control and safety of the vehicle and monitors the route for obstacles and operational

✉ Rustam Tagiew, tagiewr@dzsf.bund.de | ¹German Centre for Rail Traffic Research at Federal Railway Authority, 01219 Dresden, Germany. ²DB Systemtechnik GmbH, 32423 Minden, Germany. ³Fraunhofer IOSB, 76131 Karlsruhe, Germany. ⁴Technische Hochschule OWL, 32657 Lemgo, Germany.



hazards, while the driving itself is largely automated. This mode of operation is already being tested under ETCS in real operation as ATO-over-ETCS. Technical possibilities for the realization of higher GoA in future railway systems are currently being researched and implemented in various projects [2].

At higher levels of automation from GoA3 on (GoA3+), traction units for mainline railroad must be equipped with technical systems for the perception of their environment, i.e. perception systems. Such systems use cameras, LiDAR, radar sensors and their combinations whose (raw) data are algorithmically processed by appropriate methods. These methods also include machine learning (ML) or artificial intelligence (AI) approaches [3]. The development, evaluation, and safety verification of perception systems is carried out on the basis of real sensor data from rail operations. The requirements for such data, with regard to a subsequent safety approval, are high. This applies, among other things, to the coverage of operational areas as well as environmental and ambient conditions. In the automotive industry, extensive publicly available perception datasets from real traffic are accessible for research and development; for example, [4] cites about 60 perception datasets in 2020 for road traffic [5–7, e.g.]. In 2020, there have been only 3 railroad perception datasets, i.e. RailSem19 [8], FRSign [9] and RAWPED [10]. The German Centre for Rail Traffic Research (DZSF) published the first open multi-sensory perception dataset OSDaR23 [11] for railroad in 2023, about a decade later than the first multi-sensory dataset was published for the automotive industry.

This paper is an extended version of the paper presented at the AI4Rails workshop [4] and is based on results of the DZSF-funded project “Sensor Technology as a Technical Prerequisite for ATO Functions” (German: “Sensorik als technische Voraussetzung für ATO-Funktionen”) [12]. The additional material extending the paper from the AI4Rails workshop includes a short summary of related work in Sec. 2, more graphical content and tabular explanatory content as well as results of the survey. In particular, this paper provides an in-depth description of the locomotive and passenger car setups. Furthermore, the content of this article overlaps with a publication in a German magazine related to railroad technology, which is available only in German [13].

We start the main contribution in this paper with a brief overview of the key results of a requirement analysis of existing regulations for perception tasks of train drivers in Germany in Sec. 3. Subsequently, Sec. 4 presents how future AI applications introduce additional requirements for the diversity of an adequate sensor setup including research on domain gaps and their mitigation. Moreover, we summarize the results of an extensive railway industry survey on the expectations for perception systems in Sec. 5. These requirements and expectations culminate in a sensor setup that is proposed in Sec. 6 that is discussed with respect to various considered use cases for ATO, some of which are shown in Fig. 1. Finally, Sec. 7 draws the main conclusions from the results and provides an outlook on the necessary steps towards the provision of large scale and high quality open datasets for rail applications that meet the needs of modern AI and ML systems.

2 Related work

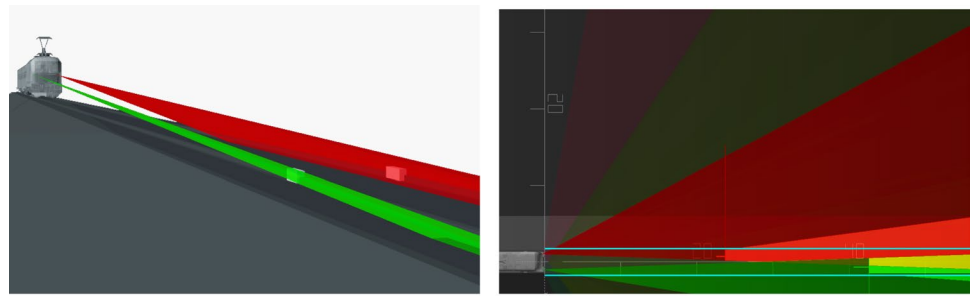
There are several projects related to the development of GoA3+ systems. The most relevant related projects are “Advanced integrated obstacle and track intrusion detection system for smart automation of rail transport” (SMART2) [16] and “Technologies for the AUtonomous Rail Operation” (TAURO) [17]. SMART2 focuses on research of on-board, trackside and airborne obstacle detection and track intrusion systems for railway trains. The aim of TAURO is to identify, analyze and finally propose suitable enabling technologies for the future European automated and autonomous rail transport beyond automated metros.

3 Requirements from existing regulations for perception tasks of train drivers in Germany

The analysis of the operational tasks of a train driver was based as much as possible on the operational rules of Deutsche Bahn AG, due to their easy availability. For the sake of simplicity, it was assumed that other railway companies operating in Germany on the network of the railway infrastructure company DB Netz AG use a similar set of operational rules and that the basic principles (in terms of protection objectives) should be the same. The following guidelines and manuals have been used:

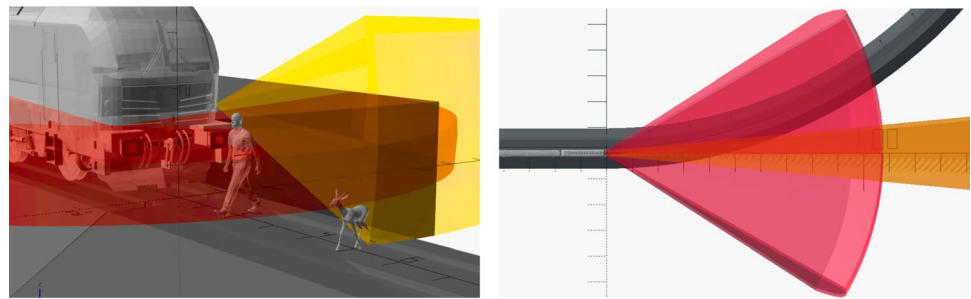
1. Driving Service Regulations [18]
2. Driver’s Manual of DB Fernverkehr AG [19]
3. Rule Book – basic part for employees in railway operations (incl. driving of traction units); DBREGIO-003 [20].

Fig. 1 Examples of perception tasks considered. Localization of brake shoes at low speeds or during shunting (a). Detection of fouling point indicators (b). Comparison of flash LiDAR and rotating LiDAR for close to medium range perception during shunting (c). Object detection at high speed for different track geometries (b), where the indicated camera ranges refer to estimated limit of 32 pixels per object of size 1 m for robust detection, as trainable on the CIFAR-10 [14] or ImageNet32 [15] datasets



(a) LiDAR single ray resolution of Valeo Scala Gen. 2 (red, rotating mirror) and Ibeo NEXT (green, MEMS) for detection of brake shoes / rail skids in a distance of 85 m [4]. Depending on the reflectivity of the brake shoe, even the wider Scala footprint could enable detection on a flat, low-reflective rail and ground.

(b) Comparison of ranges for fouling point indicator detection for LiDARs as in (a) [12]. Cyan lines show indicator positions, bright areas indicate where the narrow high-resolution FOVs hit the ground, enabling the detection in, e.g., 40 m distance at an angular size of 0.13° (approx. 10 cm on the ground).



(c) LiDARs for shunting maneuvers, allowing a comparison a flash system (yellow) and a rotating system (red) [12].

(d) Front cameras with focal lengths 12 mm (red) and 75 mm (orange) shown at a minimum German railway curve radius of 180 m [4].

4. Rule Book – basic part for employees in railway operation (incl. driving of traction units); DBCDE-003 [21].

The Driver's Manual and the Rule Books contain the necessary supplements to the generally applicable train operation regulations for the respective railway companies, i.e. also the operational tasks for the driver. For the derivation of further requirements for the specification of the sensor system, in addition to the Railway Construction and Operation Regulations (German: "Eisenbahn-Bau- und Betriebsordnung") [22] and the Railway Signaling Regulations (German: "Eisenbahnsignalordnung") [23], technical guidelines of the Deutsche Bahn AG as well as special standards were used.

The identified standard tasks or the smaller subtasks of a standard task, which require sensory perception by means of suitable ATO sensors, essentially determine the components of the measurement system and their spatially distributed installation on the rail vehicle. The tasks are classified according to whether they can be fulfilled by conventional technical systems or whether ATO sensor technology (e.g. cameras, radar, GNSS) including corresponding perception algorithms are used. The focus is on enabling the most comprehensive data collection possible of real-world operational data via the sensor system to derive and provide test and training data for a perception system. The sensor setup has been evaluated on relevant use cases within these tasks and operational situations, as shown in Table 1.

Table 1 Standard tasks and perception tasks [4, 12]

| Perception tasks | Recommended sensors | | | | |
|---|---------------------|----|----|----|----|
| | RGB | IR | LD | RD | US |
| Standard task: shunting train | | | | | |
| Compliance with the permissible speed depending on the minimum visual range | ✓ | | ✓ | ✓ | |
| Continuous determination of current location | ✓ | | ✓ | | |
| Approaching a vehicle to be coupled | | | ✓ | | ✓ |
| Detect end of own track (e.g. buffer stop) | ✓ | | ✓ | ✓ | ✓ |
| Standard task: observing track | | | | | |
| Identify path to be traveled on | ✓ | ✓ | ✓ | ✓ | |
| Identify the minimum clearance outline of the own track within the path to be traveled on | ✓ | ✓ | ✓ | ✓ | |
| Detecting objects on or next to the track | ✓ | ✓ | ✓ | ✓ | |
| Detecting collision with objects classified as obstacles | ✓ | ✓ | ✓ | ✓ | |
| Detecting collision with persons on the own track or the adjacent track for another vehicle | ✓ | ✓ | ✓ | ✓ | |
| Standard task: observing railway systems | | | | | |
| Observe signals—fixed signals in signal-guided train operation | ✓ | | ✓ | | |
| Observe signals—observe signaling and react accordingly (to the rules and regulations) | ✓ | | ✓ | | |
| Observe the track—detect damage and irregularities on the superstructure | ✓ | | ✓ | | |
| Observe neighboring tracks—detect damage or irregularities in oncoming trains | ✓ | ✓ | ✓ | | |
| Observe neighboring tracks—detect potential obstacles in the directly neighboring tracks | ✓ | ✓ | ✓ | ✓ | |
| Observe catenary—detect damage and irregularities of the catenary | ✓ | ✓ | ✓ | | |
| Standard task: controlling passenger exchange | | | | | |
| Monitor conditions for the door opening and closing procedure | ✓ | ✓ | ✓ | ✓ | ✓ |
| Prevent injuries to persons both between the individual vehicles in the train formation and between the vehicles and the platform edge | ✓ | ✓ | ✓ | ✓ | ✓ |
| Detect persons and objects on the platform that are at an unacceptable distance to the passenger outer doors when the doors are closed, as well as when the train is dispatched and departing | ✓ | ✓ | ✓ | ✓ | ✓ |
| Standard task: observing trains | | | | | |
| Detect displaced load | ✓ | | ✓ | | |

Checkmarks (✓) denote the applicability of a sensor to a task

RGB Color camera, IR infrared camera, LD LiDAR, RD radar, US ultrasonic sensor

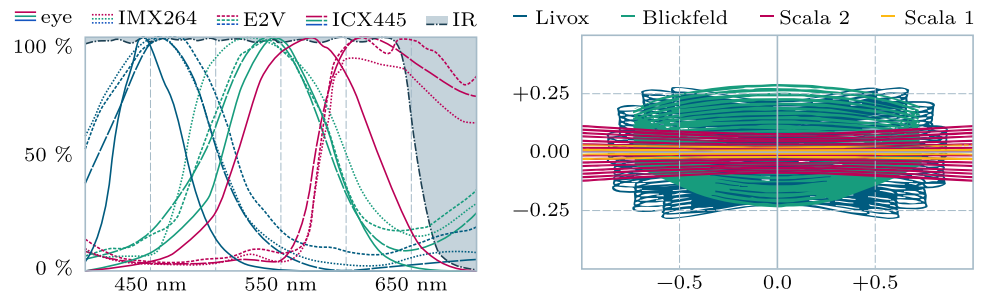
4 AI-Specific requirements for training and testing data

After presenting requirements from existing regulations for train drivers as described in Sec. 3, AI-based applications, which will be highly relevant for the implementation of perception tasks, also introduce additional considerations. These applications are typically data-driven and hence require large amounts of data, which have to meet certain requirements. For example, the European AI Act states for high-risk AI systems:

Training, validation and testing data sets shall be relevant, representative, free of errors and complete. They shall have the appropriate statistical properties, [and] shall take into account, to the extent required by the intended purpose, the characteristics or elements that are particular to the specific geographical, behavioural or functional setting within which the high-risk AI system is intended to be used.— [24, p. 48]

These requirements for data-driven perception applications must be satisfied with respect to the observed scene and sensor characteristics. For example, different environmental conditions, such as lighting and weather, must be present in the datasets to allow a technical system to perform well in a diversity of situations. Similarly, different sensor characteristics influence the appearance of data, e.g. different camera spectral sensitivities or different LiDAR scan patterns as illustrated in Fig. 2 inhibit the direct transfer of ML models, that are trained on data with different sensor characteristics. This *domain gap* is addressed in several works to facilitate the portability of ML models by *domain adaptation* techniques [25, 26]. The same challenge occurs for camera sensors with different wavelengths,

Fig. 2 Examples of domain variations [4, 12]: Differences in spectral sensitivity in RGB camera sensors (a) and beam steering patterns in LiDAR sensors (b)



(a) Spectral sensitivity of the human eye vs. the camera sensors Sony IMX264 (CMOS), E2V EV76C560 (CMOS) and Sony ICX445 (CCD) (with IR cut-off filter indicated). CMOS stands for complementary metal oxide semiconductor.

(b) Various beam steering patterns for automotive LiDAR sensors projected onto a normalized pinhole camera plane: Livox HAP, Blickfeld Cube, Valeo Scala 2 and Valeo Scala 1.

such as visible and infrared spectrum. In this case, captured images look different and domain adaptation methods try to bridge this gap between both modalities [27, 28]. These techniques for domain adaptation cannot only be employed for different sensor characteristics but also to reduce the gap under different environmental conditions. For example, deep learning-based generative adversarial networks (GANs) are used to transform images captured by a camera at night and during the day to adapt the images to the different environmental conditions [29, 30]. Moreover, domain adaptation methods are used when working with real and synthetic data to cope with the different appearances [31, 32].

Although there is already work in the field of domain adaptation to deal with data from different sensor types, sensor poses, and environmental conditions, these approaches are currently insufficient for a safety-critical application such as automatic train operation. This is mainly due to the fact that these methods have been developed for specific use cases and have not been evaluated on large datasets in the railway context. Moreover, state-of-the-art results as presented in the previous paragraph are often not optimal. Therefore, we propose to build a large dataset including annotation of data, which is collected by different sensor types and characteristics under different environmental conditions. The annotation of data, which is necessary for supervised ML tasks (cf. also Sec. 5 and Fig. 3), poses another challenge due to the high time effort. For example, in the case of semantic segmentation, which requires annotation on pixel level, the time required can average up to 1½ h per frame for annotation and quality control [5].

5 Survey on expectations to perception systems

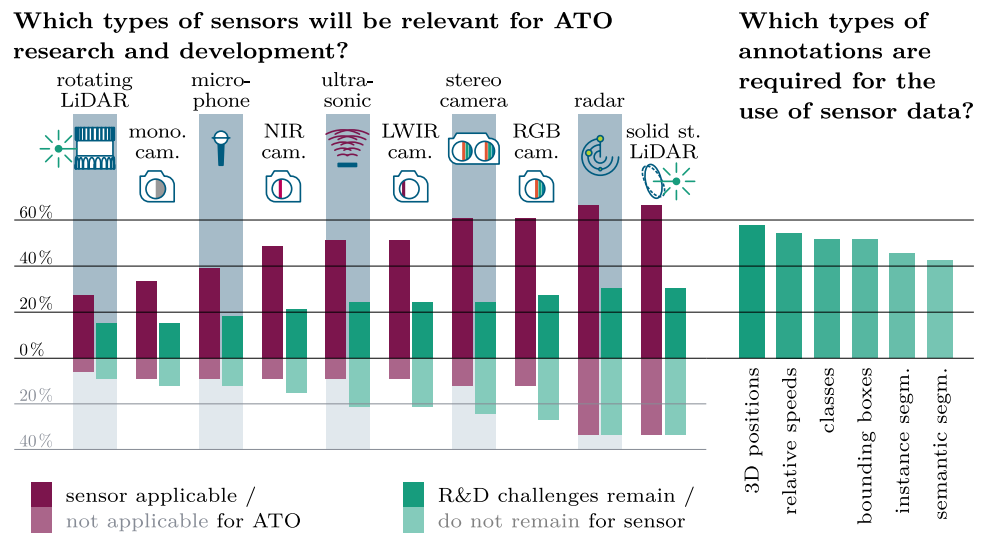
The design of sensor systems and the development and training of perception functions are subject to ongoing extensive research. In various projects and prototypes, sensor systems are tested and the data is evaluated. A consensus on the type, amount, positions and specifications of the applied sensors is not yet established in the railway sector.

Hence the expectations of the sector and the experience have been gathered by means of a sector survey. A summary of this survey can also be found in [13] in German language. A total of 53 questions have been compiled concerning sensor types, applications, data management as well as safety and validation questions, and were kindly answered by members of 30 institutions, among them vehicle and sensor manufacturers, railway companies and R & D service providers. Also, a preliminary specification of the sensor system presented in this article was provided with a request for comments. The sensor system specification was then revised and enhanced based on the received remarks. A complete and detailed summary and evaluation of the survey can be found in [12].

The use of ATO GoA3+ is seen as having great potential due to the expected improvements in capacity, flexibility and safety. Applications in shunting and train composition were of particular interest. The perception tasks under different environmental conditions are generally considered to be the most challenging for sensors and algorithms.

LiDARs and radars are considered essential by the majority of the respondents, while further research is still necessary, as shown in Fig. 3. Solid state LiDARs are the preferred solution, while for LiDARs with rotating mirrors concerns were voiced about the reliability in the harsh railway environment. More generally, conformity issues to

Fig. 3 Summary of the sector survey as presented in [13]



railway standards regarding shock, vibration, electromagnetic emissions or temperatures are considered a limiting factor for various sensors.

RGB cameras, infrared cameras, and even stereo cameras are also widely accepted. Camera data can help identify the detected objects if properly annotated training data is available. The right part of Fig. 3 shows that locating the objects is even more important to the community than identifying them. Most of the responding institutions do not have the capacity to collect and annotate the necessary data themselves. This underscores the need for a collaborative database of ATO perception data.

6 Specification of a sensor system for ATO research and development

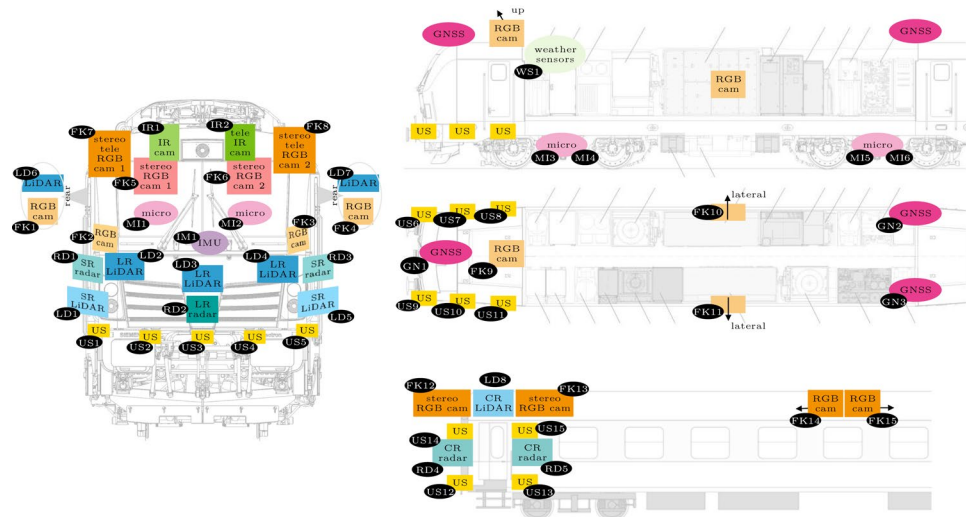
Based on the identified requirements from existing regulations for train drivers in Germany and from AI applications as well as the expectations from sector experts, we propose a sensor system for the development of perception systems for ATO functions. This sensor setup comprises a locomotive-hauled train and a passenger car. It primarily comprises ultrasonic sensors, cameras, LiDARs and radars to enable a perception of the train's state and environment. In case of the locomotive, the sensors are positioned at the front, roof, sides, chassis and in the cabin, while one side of the passenger car is used for installation to observe the lateral space of the passenger car. A first version of the described sensor system was developed, and later discussed and refined based on the results of the sector survey discussed in Sec. 5. Figure 4 shows the resulting sensor system, along with the coordinate system used to specify the sensor setup in the following sections.

Ultrasonic sensors (US1–US5) are used to monitor the space immediately in front of the locomotive for obstacles or people when the train is moving slowly (e.g., shunting) and can be used to determine the distance during coupling.

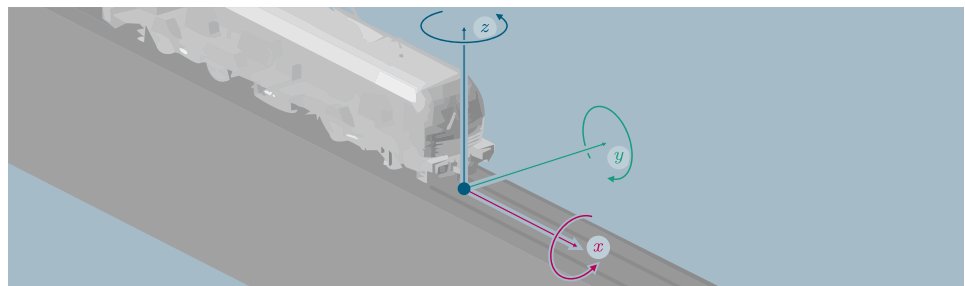
The long range radar (RD2) and technological variants of long range LiDARs (LD2–LD4) monitor the track ahead. The sensors focus on obstacle detection of possible objects in front of the vehicle and distance determination (e.g. to a buffer stop). In addition, short range radars (RD1; RD3) and short range LiDARs (LD1; LD5) are used to monitor the frontal area, but the main task of these systems is the lateral areas of the track. Furthermore, the area ahead is monitored by four RGB color cameras (FK5–FK8) in the upper area of the locomotive with different fields of view. These can be used individually or combined to stereo pairs. In the latter case, depth information can be obtained. Furthermore, a supporting inspection of the catenaries is made possible, which is also the focus of the dedicated color camera (FK9) on the vehicle roof.

The detection of heat signatures (e.g. humans or animals) is enabled by two thermal / long-wave infrared (LWIR) cameras (IR1; IR2). They are used to detect other vehicles and obstacles during shunting, persons next to the tracks, as well as irregularities and imminent dangers on the neighboring track (humans) or at the edge of the track (e.g.

Fig. 4 Front, side and roof views of the locomotive sensor system along with a side view of the passenger car setup (a) and coordinate system used to denote sensor orientations (b) based on ISO 8855, denoted as r_x , r_y and r_z



(a) Sensor positions [4]



(b) Coordinate system for Secs. 6.1–6.9 [12]

fallen trees). The RGB cameras (FK2; FK3) are intended to assist in these tasks and are mounted on the locomotive with lateral orientation. Furthermore, they can be used for the inspection for potential irregularities on passing trains.

The inertial measurement unit (IMU) (IM1) can be used to improve self-localization (GNSS, ETCS balises, wheel impulse generator) and detect damage to the infrastructure. Three GNSS antennas (GN1–GN3) are positioned on the vehicle roof at a large baselines in between and as far as possible clear of shadows; the three GNSS antennas are further used to detect rotational movements (roll, pitch, yaw) around the spatial vehicle axes. Additional localization support is provided by a RGB camera (FK9), enabling the detection of landmarks (e.g., church spires). In addition, the latter sensors result in an extended observation of the surroundings.

Further detection of damage to the track, the train itself and any acoustic signals is possible thanks to the installed microphones (MI1; MI2); the use of two microphones in the train even enables spatial localization of acoustic signals.

The rear of the train can be monitored by RGB cameras (FK1; FK4) and LiDARs (LD6; LD7). The focus is on passenger exchange and any hazards on a train station's platform; it is also possible to detect damage to passing trains. In addition, ultrasonic sensors (US6–US11) and color cameras (FK12; FK13) are used in the lateral area of the train. Just as the front ultrasonic sensors (US1–US5), the side sensors are used to monitor areas that are difficult to see during shunting. In addition, the color cameras provide detailed information about the situation on the platform. However, they also enable the detection of obstacles on the neighboring track and irregularities on other trains, e.g. open doors.

Finally, the locomotive is equipped with a weather station (WS1), which collects information about the environmental conditions (temperatures, pressures, humidity), to enrich the sensor data.

In case of the passenger car, we propose to install ultrasonic sensors (US12–US15), (stereo) RGB cameras (FK12–15), radars (RD4; RD5) and a LiDAR sensor (LD8) with the main purpose to observe the passenger exchange.

In the following subsections, we give details about the various sensor types and how they contribute to the different use cases.

6.1 Ultrasonic sensors

Ultrasonic sensors are used in the setup at the locomotive for distance measurements during coupling, obstacle detection during shunting and the detection of persons in the space immediately in front of the locomotive, e.g. when approaching the platform. Additionally, ultrasonic sensors are installed at a passenger car to monitor and control the opening and closing of doors during passenger changes and to detect objects at a train station's platform.

This type of sensor emits an ultrasonic pulse which is reflected by a measuring object and received by the sensor. By measuring the time of flight, the distance between sensor and object can be determined. Due to the use of sound waves, the measurement accuracy is influenced by environmental conditions, such as air temperature or air flow. For this reason, this type of sensor should only be used at low speeds (max. 10 km/h) and short distances (max. 5 m). Therefore, we propose a detection range of 15 cm to 500 cm for the described short-distance perception tasks. This is the range in which state-of-the-art ultrasonic sensors can perform accurate distance measurements [33]. Detection ranges smaller than 15 cm are difficult to realize due to the measurement principle based on the travel time of sound waves. However, detection ranges smaller than 15 cm are not mandatory for the considered use cases. Moreover, with respect to the availability of these sensors on the market and their specifications, we propose a horizontal and vertical field of view of 120° and 60°, respectively. Additionally, ultrasonic sensors can influence each other if they are installed incorrectly. Thus, minimum distances between the sensors must be respected during installation or a multiplex operation must be supported, in which individual sensors are controlled sequentially by a central processing unit.

6.1.1 Setup locomotive

The specific setup of the ultrasonic sensors at the locomotive is visualized in Fig. 5. It consists of eleven ultrasonic sensors which are installed at the front and side of the locomotive with the goal to measure distances and detect obstacles in the vicinity of the locomotive at slow speeds (similar to a parking assistant in the automotive area). Five sensors are mounted at the front (US1–US5), while the other six sensors are placed at both sides (right: US6–US8 and left: US9–US11). These should be installed as low as possible considering the available installation space, and their fields of view should not be occluded by vehicle parts. Their orientations are parallel to the ground and oriented in the direction of travel (US2–US4) and at right angles to the left (US9–US11) and right (US6–US8). An exception are the two ultrasonic sensors US1 and US5 as they are mounted at the buffer beam with an angle of 45° to the right and left to better cover the corner areas. In addition, the centrally placed sensor US3 is elevated 20 cm and 15° tilted downwards to better cover the front area.

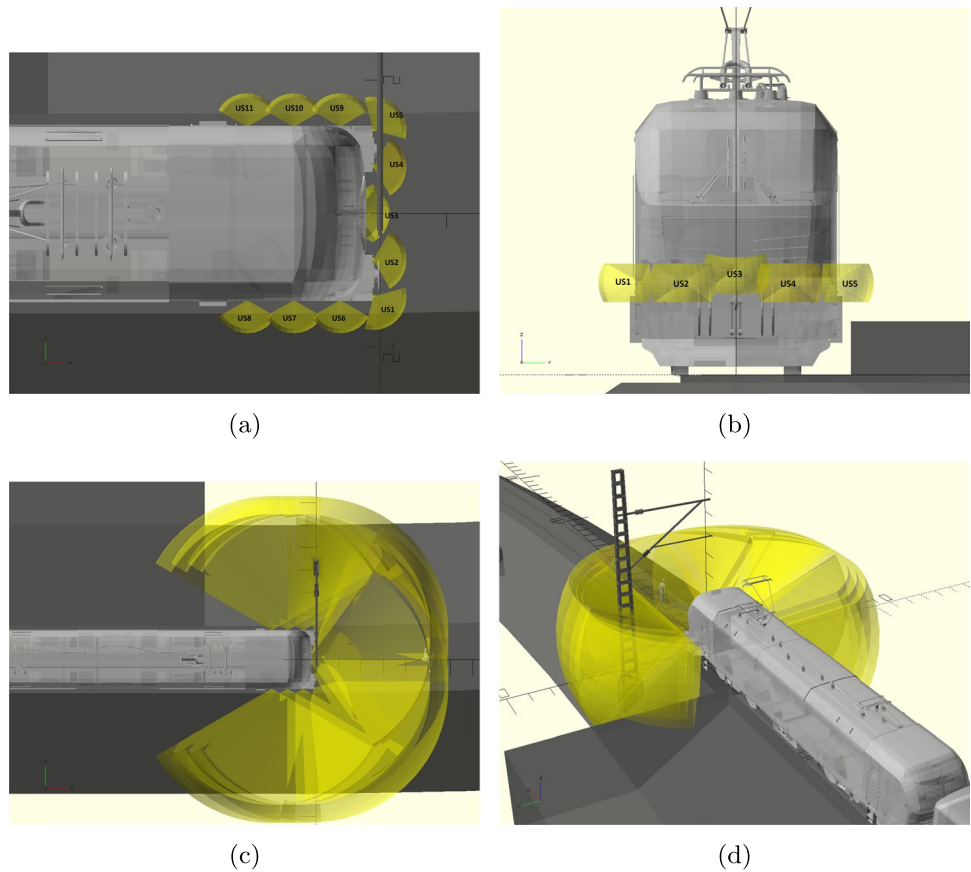
The positions and orientations of the ultrasonic sensors are chosen so that their detection ranges can detect objects up to a distance of 5 m in front and laterally of the locomotive. This is visualized in Fig. 5c, d, which shows the detection area of a specific ultrasonic sensor type with horizontal and vertical fields of view of 120° and 60°, respectively, with a detection range of 6 m [34]. The figure also visualizes the specific use case in which a person is standing on the track in front of the locomotive at a distance of 5 m. It is clear that the detection ranges of the ultrasonic sensors, especially sensors US2–US4, can perceive the person at this distance.

6.1.2 Setup passenger car

The setup of the ultrasonic sensors at the passenger car is illustrated in Fig. 6. Its main task is to allow the detection of objects in front of the door area during passenger change and in the area leading between two vehicles in the formation, e.g. passenger car and locomotive. For this purpose, four ultrasonic sensors are proposed to be installed around the door area. All sensors are rotated by 60° around the z-axis. This allows the sensors US13 and US15 to observe the area in front of the passenger door to a maximum extent since their horizontal field of view is aligned with the passenger car's side. Similarly, sensors US12 and US14 are installed with a focus to observe the area leading between two vehicles and detect intrusions into this area by persons.

The detection areas of the ultrasonic sensors are the same as the sensors at the locomotive, i.e. a detection range of 6 m and a horizontal and vertical field of view of 120° and 60°, respectively. These detection areas are visualized in Fig. 6c, d as an example on a train station's platform with passengers. These visualizations demonstrate that the area in front of the passenger door and the area leading between the locomotive and the passenger car can be monitored by the sensor setup. It is noted that this setup is not intended to detect persons standing close to the train in the entire

Fig. 5 Installation positions and detection ranges of the ultrasonic sensors at the locomotive [12]. For reasons of readability, the sensors' detection ranges in (a) and (b) are shortened to 0.5 m. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



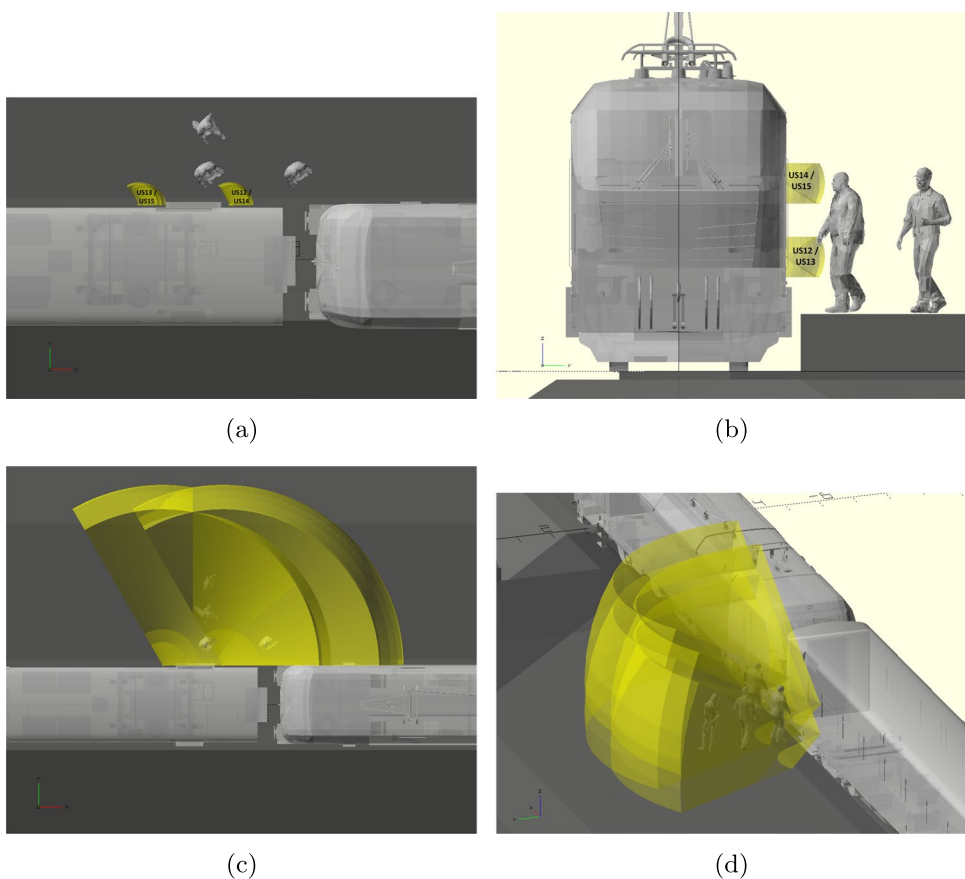
| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| | US1 | 0 | 0 | -45 | 120 | 60 | 6 |
| | US2 | 0 | 0 | 0 | 120 | 60 | 6 |
| | US3 | 0 | 15 | 0 | 120 | 60 | 6 |
| | US4 | 0 | 0 | 0 | 120 | 60 | 6 |
| | US5 | 0 | 0 | 45 | 120 | 60 | 6 |
| | US6 | 0 | 0 | -90 | 120 | 60 | 6 |
| | US7 | 0 | 0 | -90 | 120 | 60 | 6 |
| | US8 | 0 | 0 | -90 | 120 | 60 | 6 |
| | US9 | 0 | 0 | 90 | 120 | 60 | 6 |
| | US10 | 0 | 0 | 90 | 120 | 60 | 6 |
| | US11 | 0 | 0 | 90 | 120 | 60 | 6 |

detection range of the sensors. This is due to the measurement principle of ultrasonic sensors, which return a single distance value instead of higher-dimensional spatial values, such as cameras or LiDARs. For example, if a high distance value is returned by a sensors, a person can still be near the train, e.g. far right at the end of the detection range near the locomotive as shown in Fig. 6c, d. If it would be a use case to ensure that there are no people in the direct vicinity of the entire passenger car, a setup of ultrasonic sensors similar to that of the locomotive would be adequate.

6.2 Camera

High-quality camera systems are widely used in industry and science resulting in a large and diverse range of products. The core of a camera system is the actual image sensor. It is characterized by various properties, such as the

Fig. 6 Installation positions and detection ranges of the ultrasonic sensors at the passenger car [12]. For reasons of readability, the sensors' detection ranges in (a) and (b) are shortened to 0.5 m. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| | US12 | 0 | 0 | 60 | 120 | 60 | 6 |
| | US13 | 0 | 0 | 60 | 120 | 60 | 6 |
| | US14 | 0 | 0 | 60 | 120 | 60 | 6 |
| | US15 | 0 | 0 | 60 | 120 | 60 | 6 |

active sensor area, the pixel size and thus the resulting resolution, the dynamic range and the recording rate, as well as (to a slighter degree) the spectral sensitivity as shown in Fig. 2a. A camera's horizontal and vertical field of view result from the combination of sensor size and focal length of the lens used.

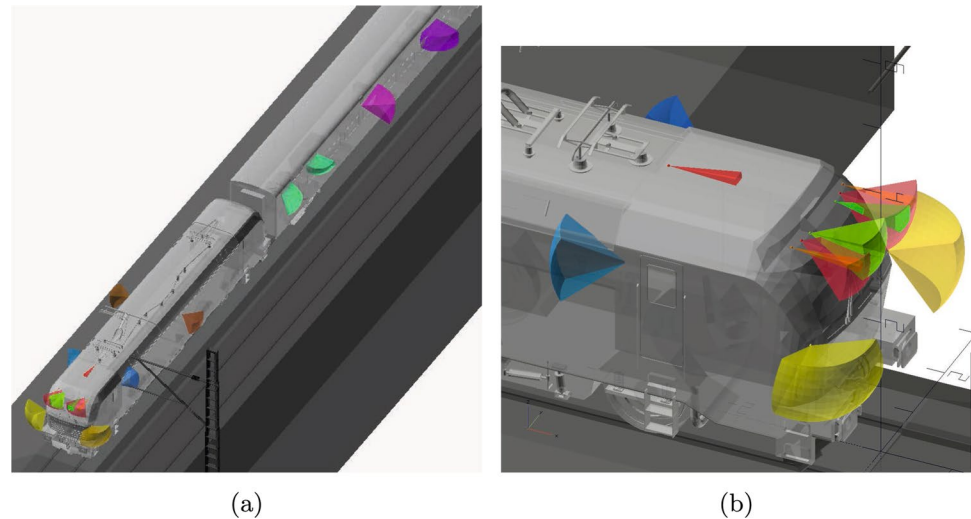
For a specification of the used cameras, the project team decided to define the image sensors of the color cameras in advance. The focal length of the lens can be calculated on the basis of the chip size and the fields of view, which are derived from use cases and perception tasks of a train driver.

An exemplary image sensor for further, geometrical considerations:

Color camera: Sony IMX255 1ε – CMOS-Sensor with global shutter and Bayer filter
4112(H) × 2176(V) pixels(9 MP); pixel size: 3.45μm

Industrial color cameras with 1" CMOS sensors (Sony) are widely used and can be combined with numerous lenses on the market. Due to the large sensor area, 4k resolutions can be realized without having to significantly reduce the pixel size and thus the light sensitivity. Nevertheless, the cameras remain compact and are suitable for installation in existing vehicle structures. Manufacturers specify recording rates of up to approx. 90 frames per second for their cameras with this sensor resolution. The resulting data volumes of up to 5 Gbit/s per camera (uncompressed) can only be processed with a great deal of computing effort and represent a limit with regard to resolution and recording rate. These high

Fig. 7 Overview of the position and viewing direction of the cameras [12]



resolutions and recording rates offer research facilities the opportunity to evaluate future technologies and algorithms. Reducing the resolution and compressing the video recording is always possible.

Figure 7 provides an overview of the proposed positioning and orientation of 17 cameras. The fields of view are shortened in the figure (here 1 m) for a better overview. The position and orientation of the cameras as well as the lenses' fields of view were determined in such a way that a large number of the perception tasks can be performed by cameras. In addition, the fields of view of the different cameras partially overlap and complement each other. In this way, possible gaps in the observation space are avoided, information (e.g. light signals) can be detected redundantly and the detected areas can be compared with each other.

In the following sections, a selection of different camera positions are described in detail. The lengths of the field of view in the figures indicate the distance at which an object of size $1\text{ m} \times 1\text{ m}$ is projected onto $32\text{ px} \times 32\text{ px}$ of the image sensor, taking into account the recommended focal lengths. This size of $32\text{ px} \times 32\text{ px}$ is based on ML methods in the state of the art, which already classify a large number of classes in various contexts at this image resolution, such as with the CIFAR-10 dataset [14] or the ImageNet-32 dataset [15]. This value is explicitly used only as a provable reference for an approximate magnitude in order to roughly forecast recognition potentials of future algorithm developments. With this information the remaining parameters of the field of view can be obtained. The general lens equation and the imaging scale A are used to calculate the distance (= object distance g):

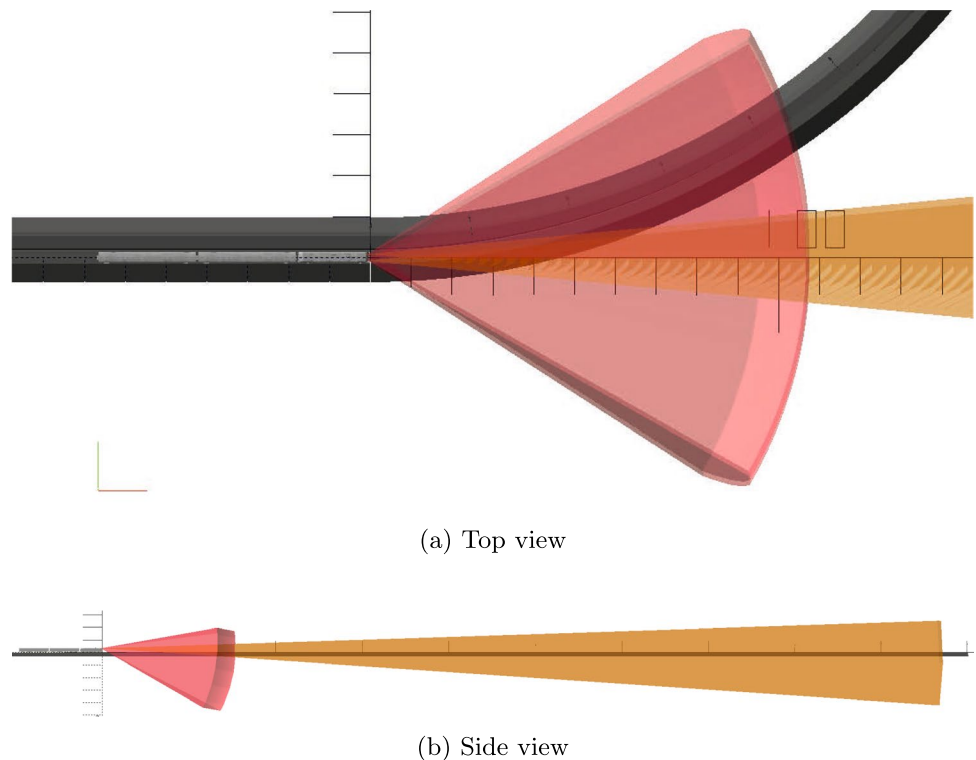
$$g = \left(\frac{1}{A} + 1 \right) \cdot f \quad \text{with } A = \frac{B}{G} \quad (1)$$

where the focal length f of the lens, the object size G ($= 1\text{ m}$) and image size B ($= 32 \times \text{pixel size}$) are used. A simple linear relationship is obtained. For an already selected camera sensor, the focal length is defined as a function of the required field of view, which is defined based on the requirements of the use cases.

6.2.1 Setup locomotive, front cameras (stereo setup)

It is proposed to install two stereo camera pairs with different focal lengths at the front of the locomotive in the area above the windshield. One camera from each of the two systems is located on the left and one on the right of the locomotive's centerline (see Fig. 7b). For reasons of flexibility in terms of positioning, spacing and future research tasks, it is not recommended to install a system in one component, but rather two separate cameras. The stereo setup is used to determine the distance of objects and is a complement to the LiDAR and radar sensors. Figure 8a shows an example of the smallest approved radius of a curve in the German rail network (180 m) and the field of view of all four cameras. The cameras FK5 and FK6 of the first stereo system (red), which correspond to the estimation of a train driver's vision, are used with lenses of the focal length $f = 12\text{ mm}$. The horizontal fields of view in this case are 61.2° , which means that a wide area in front of the locomotive is covered. The effective range is approx. 108 m. The focal length is selected so that the two immediately adjacent tracks (left/right) as well as the signals

Fig. 8 Field of view of the stereo cameras for lenses with focal lengths of 12 mm (red) and 75 mm (orange) in a curve with a radius of 180 m [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



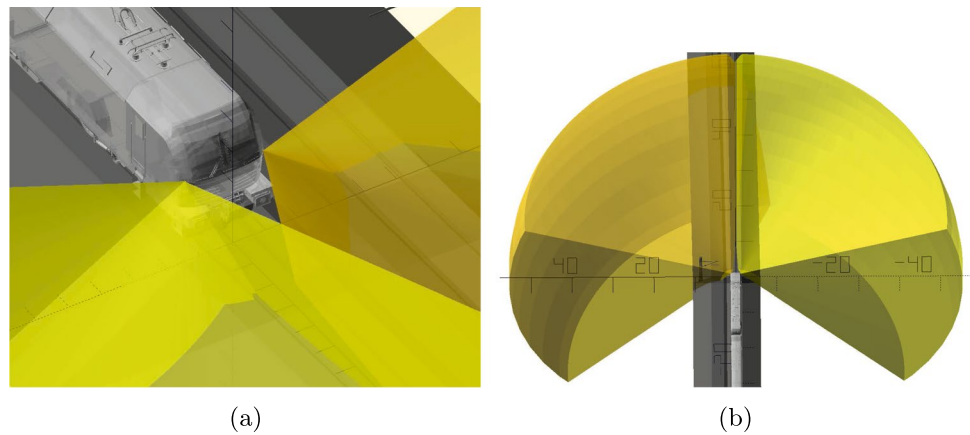
| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|------|------------|
| | | r_x | r_y | r_z | H | V | |
| | FK5 | 0 | 2 | 0 | 61.2 | 34.7 | 108 |
| | FK6 | 0 | 2 | 0 | 61.2 | 34.7 | 108 |
| | FK7 | 0 | 1 | 0 | 10.8 | 5.7 | 679 |
| | FK8 | 0 | 1 | 0 | 10.8 | 5.7 | 679 |

next to and above the tracks can be reliably perceived. In order to detect and classify objects (< 1 m) at a greater distance, a stereo system with the cameras FK7 and FK8 (orange) is recommended. With a 75 mm lens, this has an effective range of up to 679 m and thus clearly surpasses the first camera system in terms of range (see Fig. 8b). However, due to the resulting small field of view of 10.8°, this system can only be used as a supplement. The stereo system (FK7; FK8) can still detect the neighboring tracks and signal areas on a high-speed line (increased track spacing of 4.5 m) with a minimum radius of 5000 m. It is recommended to tilt both systems slightly downwards to the ground. This field of view corresponds more to that of a train driver.

6.2.2 Setup locomotive, shunting cameras

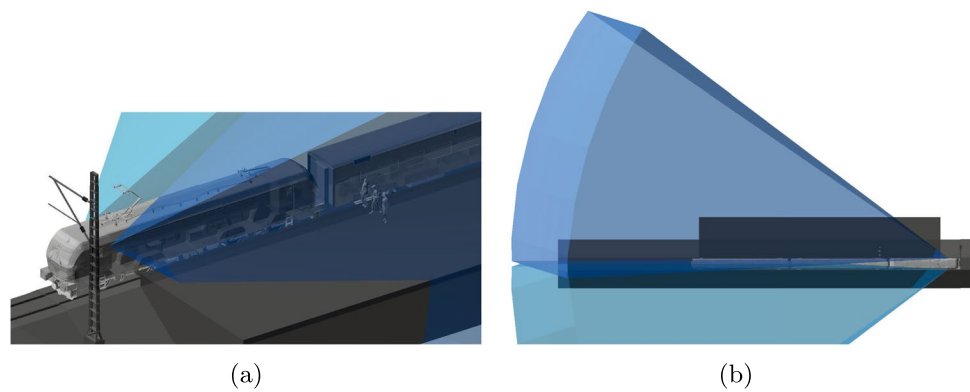
The shunting cameras FK2 and FK3 (yellow) close the observation area between the front cameras FK5 and FK6 (red) and the side cameras FK10 and FK11 (brown) of the locomotive (see Fig. 7). They are located on the left and right of the vehicle front and are rotated counterclockwise and clockwise by an angle of 45° (see Fig. 9). Using lenses with a focal length of 6 mm results in horizontal and vertical fields of view of almost 100° and 64°, respectively. Due to the large fields of view, the cameras are appropriate for shunting and driving on sight to observe the near surroundings. It is also possible to inspect other rail vehicles for damages or a slipping load. Both cameras can observe the area in front of the train and to the side. However, the effective range is limited due to the low focal length to 54 m. By tilting the sensors by 15°, the immediate area in front of the locomotive, that cannot be viewed, is significantly reduced.

Fig. 9 Field of view of the rear cameras for lenses with a focal length of 12 mm [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| | FK2 | 0 | 15 | -45 | 99.6 | 64 | 54 |
| | FK3 | 0 | 15 | 45 | 99.6 | 64 | 54 |

Fig. 10 Field of view of the rear cameras for lenses with a focal length of 12 mm [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|------|------------|
| | | r_x | r_y | r_z | H | V | |
| | FK1 | 90 | 0 | -163 | 34.7 | 61.2 | 108 |
| | FK4 | 90 | 0 | 163 | 34.7 | 61.2 | 108 |

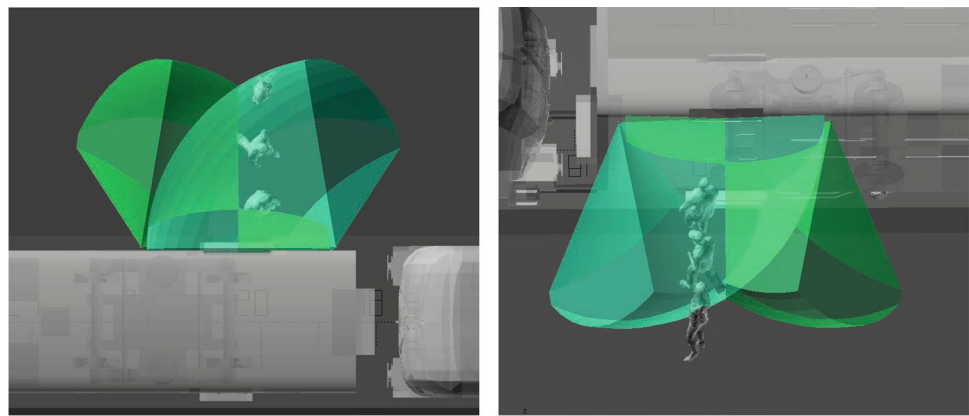
6.2.3 Setup locomotive, rear cameras

The rear cameras FK1 and FK4 (blue) are mounted on the sides in the upper front area of the locomotive and look in the opposite direction of travel (see Fig. 10). 12 mm lenses are recommended for this setup. The fields of view correspond to those of the stereo cameras at the front of the vehicle. In this setup, the image is taken in portrait format. In this way, the areas directly next to the locomotive and passing rail vehicles can be captured by the camera. The fields of view of the cameras are aligned with the vehicle side to monitor passenger movements on a train station's platform.

6.2.4 Setup passenger car, door cameras (stereo setup)

In case of the camera sensors at the passenger car, special attention must be paid to the monitoring area in front of the door. Before departure, it must be ensured that no persons or objects (e.g. umbrellas) are trapped in the door area.

Fig. 11 Field of view of the door cameras on the passenger car with lenses with a focal length of 6 mm [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



(a)

(b)

| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| | FK12 | 0 | 40 | 140 | 99.6 | 64 | 54 |
| | FK13 | 0 | 40 | 40 | 99.6 | 64 | 54 |

The cameras FK12 and FK13 (cyan) are used in a stereo setup with 6 mm lenses to address this use case. The cameras are installed above the door area with a distance of 3.4 m (Fig. 11). The cameras are tilted downwards by 40° and oriented towards the door. Immediately in front of the door, there is a large overlap area of the two color cameras.

6.3 Radar

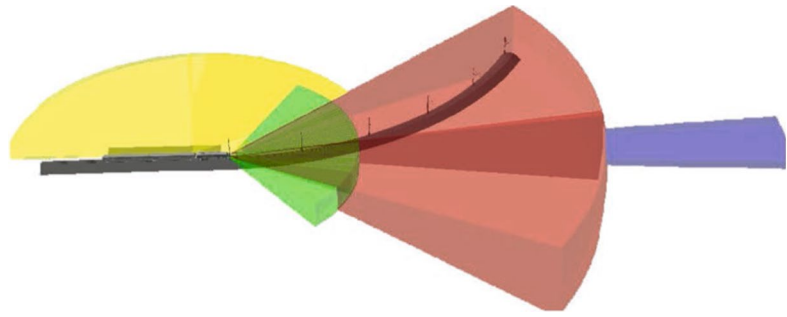
Radar systems are based on transmitting an electromagnetic wave in the radar frequency range and receiving the reflected signal. Typical frequency ranges are 77 GHz (primarily long-range) and 24 GHz (primarily short-range). Basically, a distinction is made between impulse and frequency-modulated continuous wave (FMCW) radar. An important feature of FMCW radar is the ability to measure relative velocity in addition to distance to an object. Besides this information, additional angular information to the object is needed, which can be obtained by means of antenna arrays (for transmitting and/or receiving antennas). Reflected signals with different complex amplitudes (magnitude, phase) are received and the derivation of an angle estimate is made possible. This basic principle is mapped in today's systems with a larger number of antennas and used for a horizontal angular resolution; a vertical resolution can be realized similarly. Such sensors are referred to as 4D radar sensors because they can measure distance, relative velocity, and horizontal and vertical angles. In the proposed setup, radar systems are used frontally on the locomotive and laterally on the passenger car. Figure 12 shows the areas illustrated by the installed radar sensors (locomotive, passenger car).

6.3.1 Setup locomotive, front

In this setup, three radar sensors are installed at the front of the locomotive: Centrally, a system with the longest possible range (300 m) and a relatively small field of view (in the proposal, a maximum of 18°) to detect objects at great distances (including oncoming traffic) as early as possible is installed. A combined short-range / long-range radar is installed at each side with a field of view of 80° for the short range (below 100 ms) and a field of view of 36° for the long range (up to 180 ms). It is mounted on the right and left side of the locomotive with a small overlap of the long-range components. This ensures that curve illumination (up to 180-meter arcs) is possible to the extent that at least driving on sight requires. At the same time, the overlap in the central area allows investigations on fusion with the long-range radar.

Figure 12 shows the general conditions for driving in curves: on the one hand the two combined long-range/short-range sensors (green, red) and on the other hand the long-range radar (blue). This variant is characterized by the fact that the necessary visibility range can be achieved for the curve shown. It should be noted that radar systems currently on the market do not fully cover the parameters actually required by the specification. Figure 12 contains the specification

Fig. 12 Radar setup using the example of a curve drive [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| Red | RD1 | 0 | 0 | -9 | 80 | 12 | 70 |
| Green | RD1 | 0 | 0 | -9 | 36 | 12 | 200 |
| Red | RD3 | 0 | 0 | 9 | 80 | 12 | 70 |
| Green | RD3 | 0 | 0 | 9 | 36 | 12 | 200 |
| Blue | RD2 | 0 | 0 | 0 | 18 | 1 | 300 |
| Yellow | RD4 | 75 | 0 | -9 | 150 | 30 | 100 |
| Yellow | RD5 | 105 | 0 | -9 | 150 | 30 | 100 |

of these radars. It should be noted that for presentation reasons, the sensor is shown as separate devices for the short and far range respectively. In reality, only one physical device is required.

The setup mainly covers the detection of large animals and persons on the track as well as shunting and driving on sight. Assuming direct visibility of the obstacle and achieving a detection range of 300 m, the requirements of stopping in front of an object are possible in curves and on the track at the maximum permissible speeds for automated driving on sight.

6.3.2 Setup passenger car, lateral

Two radar sensors on the sides of the passenger door of a passenger train car are recommended. This makes it possible to map the task of entry monitoring (detection of trapped objects, counting of entering and exiting passengers). For this purpose, a close-range sensor is proposed that has a horizontal field of view wider than 120°. With two sensors, this results in a complete overlap area in front of the entrance. With a vertical field of view of 30°, the sensors should be tilted downwards by approx. 25° at a mounting height of one meter above the upper edge of the platform, so that objects with a height of 40 cm can still be detected by one sensor from a distance of approx. 20 cm from the edge of the platform. Higher objects are then assumed to be detectable. Persons are assumed to be more than 80 cm high and more than 30 cm in diameter. Figure 12 shows the specification of these radars.

The radar setup on the passenger car can be used to automate use cases around passenger boarding. The horizontal fields of view are aligned in such a way that they run in line with the passenger car and can cover the passenger door area and its surroundings through their detection range. With this, it should be possible to detect objects, e.g. persons, in the close door area. The extent to which smaller objects trapped in the passenger door can be detected by radar sensors is a question for future research and development tasks. In addition, the use case includes the control of the side of the train over its entire length. Radar sensors with a range of 100 m can cover the entire side of the train car. Furthermore, the setup can detect intrusions into the gap between two vehicles in the train formation, e.g. locomotive and passenger car.

6.4 LiDAR

LiDAR sensors, or *laser scanners*, have been widely used in robotics and automated driving for road vehicles, where the latter application provided a particular thrust in technological developments in the past years. Through this, LiDAR has evolved considerably both in terms of quantitative performance, but also in the breadth of technological variations.

Today, commercial systems exist that differ in nearly any aspect of the measurement principle except for the commonality of each emitting laser light towards the target / scene and measuring the response to obtain, at least, an object range per ray.

When considering LiDAR for ATO applications, it must be noted that the application differs considerably from robotics and automated road vehicles. Typical ranges of up to a few hundred meters suffice for a wide range of applications even for highway driving on motorways / interstate highways. For ATO however, stopping distances at typical traveling speeds exceed these ranges by far, and (unlike for road traffic) unexpected obstacles will typically be stationary and can only be avoided through braking. Thus, a LiDAR sensor that can provide considerable benefit for adaptive cruise control (ACC) keeping the distance to a forward road vehicle, or allowing to navigate on urban roads and intersections, can still be unable to assure safety in ATO due to traveling speeds.

On the other hand, the capability of LiDAR to measure ranges robustly under a wide range of conditions (including at night) can provide considerable benefit for the detection of obstacles at low to medium speeds, for example during shunting, and the classification of objects based on their shape. In the following, we first discuss aspects of technological variations (Sec. 6.4.1) before presenting some configurations that enable a comparative evaluation of different LiDAR technologies (Sec. 6.5–6.7).

6.4.1 Technological variations and considerations

6.4.1.1 Beam steering and resolution Beam steering refers to technologies determining the direction of emitted laser beams, leading to different scan patterns and a different ordering of point acquisition time (Fig. 2b), which may lead to considerable distortions if not accounted for in processing. In classical rotating and rotating *mirror* scanners (for example Velodyne, Ouster or Valeo Scala), beams follow circular patterns with a regular angle spacing. For MEMS scanners, the beam is steered by displacements of micro-scanning mirrors and can be set to follow variable vibration patterns.

Scanners by the Livox company steer the beams through a pair of rotating prisms leading to a highly complex pattern, with Livox stating that the scanners can be operated such that, over a long exposure period, any angular volume within the FOV is sampled at high precision.

Flash scanners such as the Continental High Resolution 3D Flash LiDAR illuminate the FOV simultaneously and measure the echos in a regular pixel pattern, very similar to a global-shutter camera.

Depending on the processing application, these differences may have a significant impact beyond performance values such as field of view, resolution or scan rate. For example, clustering algorithms such as DBSCAN [35] depend on accelerated queries for nearby points in large clouds. For irregular patterns, efficient queries are more difficult to define and potentially still more costly to compute. Similarly, a highly inhomogeneous spatial resolution may adversely affect approaches based on local neighborhood features, such as CNNs. To provide a basis for research in these effects, a combination of different beam steering methods is recommended for the proposed sensor setup. Resolutions considered in the evaluation ranged from 0.94° horizontally and 1.88° vertically down to 0.03° in either direction. Fields of view ranged from 18° to 360° horizontally, and from 5.6° to 90° vertically. Some scan patterns of proposed sensors are given in Fig. 2b.

6.4.1.2 Measurement principle The established measurement principle in LiDAR sensors is time of flight (ToF) measurement, where the time difference between a pulse emission and the return echo is used to determine the distance of the target. This method is highly precise under typical conditions, with speeds of light varying only slightly with air density, and main challenges lying primarily in the distinction between actual echos and environment light noise, limiting the effective range. Recently, Aeva Technologies, Inc. has presented a sensor that is based on the frequency-modulated continuous wave (FMCW) principle [36]. This principle is common in radar sensors; a modulated signal is emitted, and any Doppler shift on the echo indicates a relative motion of the target. This allows to detect the motion of the sensor relative to its environment, and speeds of other objects relative to either the sensor or the environment, and can facilitate detection and tracking of moving obstacles. It can also support the discrimination between the modulated echos and random environment noise, increasing the effective range. The specification proposes the use of both multiple ToF scanners, due to their widespread use and type variation, as well as an FMCW scanner to explore the potentials of this new principle in comparison, as discussed in Sec. 6.5.

6.4.1.3 Wavelength LiDAR sensors typically operate at wavelengths between 800 nm and 1600 nm in the near infrared domain. A particular consideration for the choice of wavelengths is eye safety. Near infrared light is invisible to the human eye but can still damage it, if it penetrates to the retina—or even especially so, since humans will not perceive

and avoid the glare. The eye, for example the vitreous body, is highly transparent still at near infrared wavelengths (e.g. 800 nm), while transmission drops considerably around 1500 nm [37] between near and short-wave infrared. Thus, longer wavelengths can increase the allowable power at the same level of eye safety, and thus increase the range. The difference in LiDAR wavelengths is wider than the width of the visible spectrum, hence differences in reflectance and transmission in objects are comparable to visible color differences. This can affect detection methods that utilize the echo intensity, which is provided by most state-of-the-art LiDAR sensors. More importantly, to install LiDAR sensors with overlapping fields of view, it is necessary to consider interference. Different scanners with a wide margin between operating wavelengths are less likely to interfere with each other. Wavelengths between 850 nm and 1550 nm were considered in variants of the specification.

6.4.1.4 Scan rate The scan rate, i.e. the time between two complete scans, affects recorded data in several ways. As with cameras, effects of lower scan rates can be simulated by dropping frames from higher frame rates (downsampling). However, since most of the currently available scanners record the points sequentially (with the exception of flash scanners), the different recording time of points within the same scan can lead to relevant effects similar to the rolling shutter effect for cameras for highly dynamic scenes. Unlike for cameras, however, scan rate and delay are usually directly linked by the beam steering principle, in the sense that the timing of the recorded points is distributed evenly over one scan period; while rolling shutter delays are technologically distinct from the frame rate limit and typically can be much shorter than the frame rate (if not avoided entirely by global shutter cameras).

Thus, scanners with a low scan rate will typically introduce stronger temporal artifacts to be compensated than scanners with a high scan rate, and downsampling from faster scanners may understate the effect, unlike for cameras. This motivates the inclusion of scanners with notably different scan rates, to enable a direct comparison of the effects. Scan rates between 10 Hz and 30 Hz were considered in the evaluation.

Based on these considerations, we denote sensor variants in the pattern of LDX_Y, where X is the mounting position as in Fig. 4 and as in other sections, while Y indicates the sensor variant: “R” for rotating or rotating mirror LiDARs, “M” for MEMS LiDARs, “F” for flash LiDARs (each time of flight-based) and “D” for FMCW.

6.5 Setup locomotive, forward perception, long range

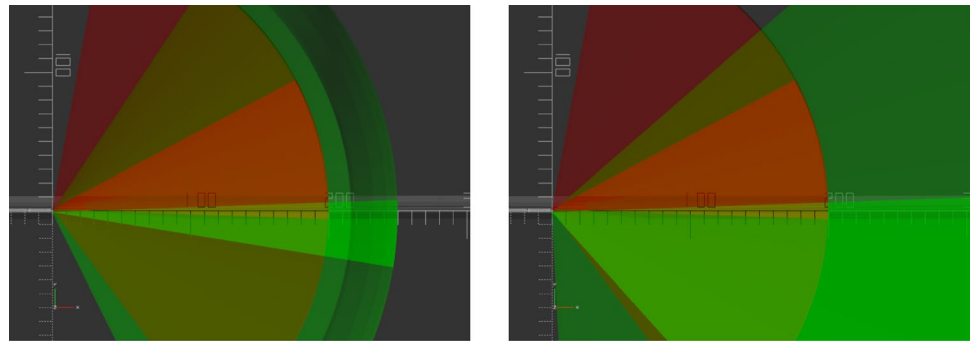
As previously indicated, forward perception using LiDAR sensors must be considered with stopping distances at traveling speeds in mind; hence, a front LiDAR will not serve the same purpose for ATO as for automated road driving, and hence the designation as “long range” (this section, several 100 ms) must be considered in contrast to “short range” (Sec. 6.6, several dozen meters), and not as “long” by the standards of ATO requirements.

Considered use cases include the detection of brake shoes (or *rail skids*) as shown in Fig. 1a or of fouling point indicators near railroad switches (Fig. 1b). In these cases, the train is operating at relatively low speeds where a detection and stopping within 100 m can be achieved. The detection and localization of large obstacles, and the differentiation between humans and animals, can be achieved with the current technology – typically not in time to avoid a collision at traveling speeds, but potentially to estimate and mitigate the consequences. In this domain, no existing flash LiDAR solutions are known to date; however various existing solutions for MEMS, rotating (and rotating mirror) as well as rotating prism LiDARs exist.

Two proposed variants are shown in Fig. 13. The first variant (Fig. 13a–c) is a combination of a ToF LiDAR with rotating mirror (specifications of LD4_R according to the Valeo Scala Gen. 2) with a ToF MEMS LiDAR [specifications of LD2_M according to the Ibeo NEXT (var. 1) or the, at this time yet unavailable, Continental HRL131 (var. 2)]. This variant addresses the property of several considered sensors (e.g. the Valeo Scala Gen. 2) to provide a narrower field of view with high-resolution, along with a wide field of view at reduced resolution. In this case, the setup proposes to assure an overlap of 2° of the high-resolution FOVs, shown yellow in the figures, while still preserving a wide combined FOV.

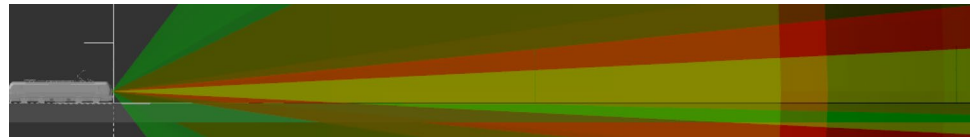
An alternative variant (Fig. 13d) uses two identically oriented LiDARs, one MEMS LiDAR (specifications of LD2_M (var. 3) according to the Blickfeld Vision SR) and one FMCW LiDAR (specifications of LD4_D according to Aeva FMCW), maximizing the overlap for comparability, at the expense of achieving a smaller combined FOV. Neither sensor is known to be equipped with a technically prescribed high-resolution narrow FOV; yet at the time of the analysis, final product specifications were not yet available.

Fig. 13 Overview of several variants considered for the forward perception at long range, including rotating and MEMS LiDARs (a–c), as well as an alternative integrating an FMCW LiDAR (d) [12]. Panels are shown in the figure that lists the angles, FOV and ranges of the sensors



(a) Top view, var. 1

(b) Top view, var. 2



(c) Side view, overlapping FOVs between rotating and MEMS LiDAR shown yellow

| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|----------------------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| Red | LD4_R | 0 | 0 | 13 | 133 | 10 | 200 |
| Green | LD2_M (var. 1 / (a)) | 0 | 0 | -3.6 | 120 | 60 | 250 |
| Green | LD2_M (var. 2 / (b)) | 0 | 0 | -23 | 128 | 28 | 1000 |



(d) Top view for setup combining a MEMS LiDAR and an FMCW LiDAR in parallel

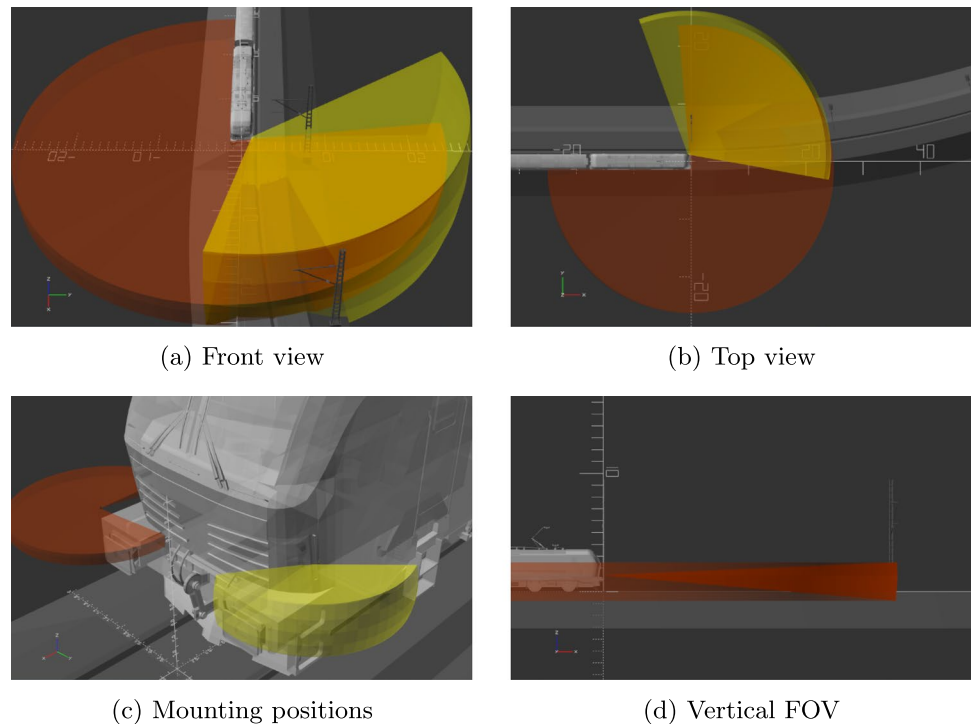
| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|----------------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| Purple | LD4_D | 0 | 0 | 0 | 120 | 30 | 400 |
| Green | LD2_M (var. 3) | 0 | 0 | 0 | 107 | 36 | 150 |

6.6 Setup locomotive, forward perception, short range

The setup shown in Fig. 14 is particularly designed to detect objects during low-speed applications, especially during shunting, as also illustrated in Fig. 1c. It compares two technological variations: A rotating LiDAR LD1_R (specifications according to the SICK MRS1000) and a flash LiDAR LD1_F (specifications according to Continental HR Flash). Both feature a relatively short range but wide field of view.

In this application, it must be considered that viewing range should be maximized while avoiding that obstacles very close to the front of the train vanish below the FOV and minimizing debris effects as far as possible. This motivates a

Fig. 14 LiDAR setup for short-range forward perception [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in $^{\circ}$ | | | FOV in $^{\circ}$ | | Range in m |
|---------------|-------|----------------------|-------|-------|-------------------|----------|------------|
| | | r_x | r_y | r_z | H | V | |
| | LD5_F | 0 | 0 | 120 | 50 | 30 | 25 |
| | LD1_R | 180 | 0 | -42.5 | 275 | -2.5...5 | 25 |

low mounting position directly above the buffers, about 1.5 m above the top of the rails and, particularly for the SICK MRS1000, an inverted mounting to maximize its asymmetric vertical FOV towards the ground.

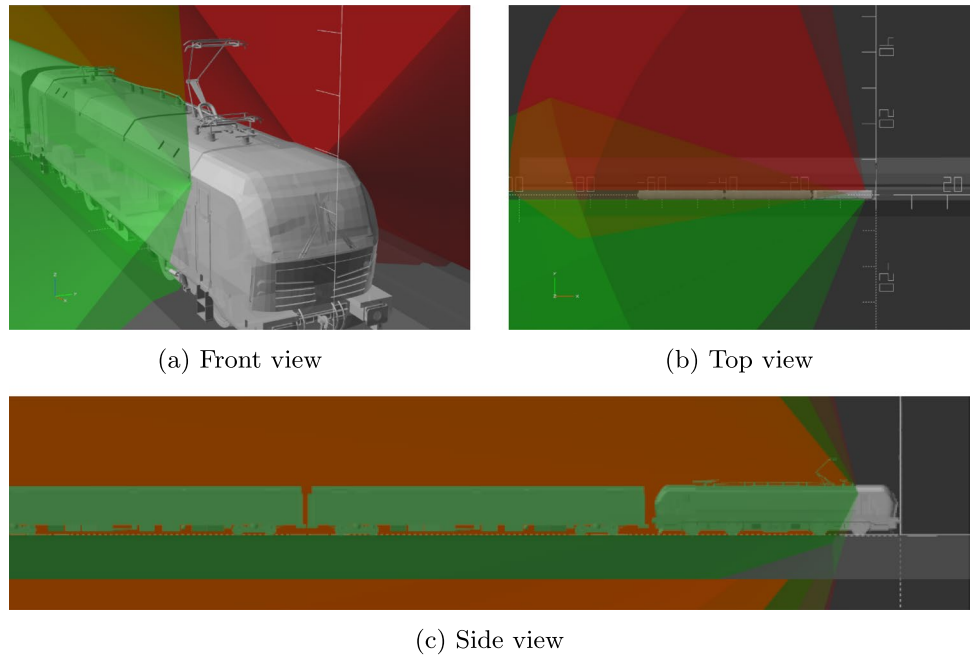
Based on their specifications, the SICK MRS1000 achieves a horizontal and vertical resolution of 0.25° and 1.88° , respectively, at 850 nm, while the Continental HR Flash provides a symmetric resolution of 0.94° in each axis, operating at 1064 nm; and while the SICK MRS1000 operates at rotations of 12.5 Hz, the Continental HR Flash exposes the entire frame simultaneously at 25 Hz. Thus, both sensors provide a considerably different set of features to be applied to the given task. Again both sensors have an overlapping region that allows to compare their perception on the same objects while keeping regions that are only observed by one sensor alone. An alternative setup (not shown) mounts the LD1_F at 120° yaw, such that both sensors cover a combined FOV of almost 360° , but the overlapping area is not in front of the train but extends to its left.

6.7 Setup locomotive, rear perception

Rear perception can enable the coverage of additional use cases by allowing to monitor the boarding of passengers close to the locomotive, to inspect passing trains on neighboring rails and neighboring catenaries, and to monitor dynamic objects approaching the front of the train in low speed situations, for example humans running towards the front of the train from behind during shunting operations, to possibly overtake it.

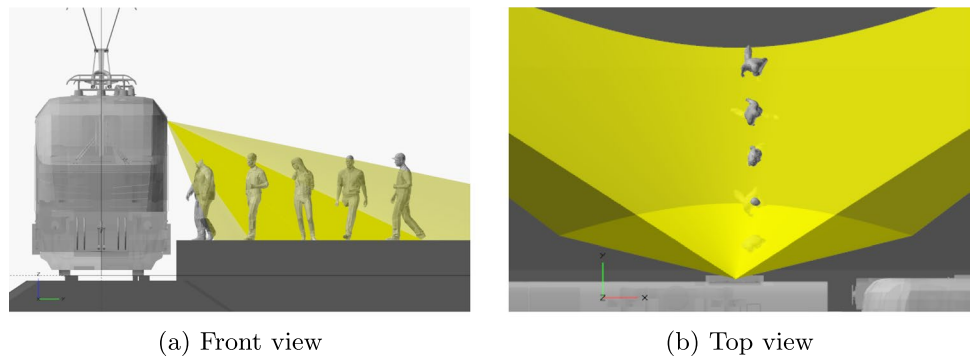
A particular requirement for this use case is to assure that the sensors are small enough to fit the allowable minimum clearance outline of the train. Two specific sensors considered that may provide sufficiently small dimensions and sufficiently similar specifications with relevant technological differences are the Velodyne Velabit (rotating LiDAR, specifications taken for the LD6_R) and the Blickfeld Vision Mini (MEMS LiDAR, specifications taken for the LD7_M), mounted on opposite sides of the locomotive, as shown in Figs. 15, 16.

Fig. 15 LiDAR setup for rear perception tasks [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| | LD6_R | 90 | 0 | 145 | 90 | 70 | 100 |
| | LD7_M | 90 | 0 | 205 | 120 | 5 | 150 |

Fig. 16 LiDAR setup on passenger car doors [12]. A panel is shown in the figure that lists the angles, FOV and ranges of the sensors



| Color in fig. | Label | Angles in ° | | | FOV in ° | | Range in m |
|---------------|-------|-------------|-------|-------|----------|----|------------|
| | | r_x | r_y | r_z | H | V | |
| | LD8 | 0 | 40 | 90 | 120 | 30 | 25 |

The sensors are located such that a continuous observed field of view is established with the front short-range LiDARS outlined in Sec. 6.6.

6.8 Setup passenger car

To monitor the boarding of passengers at passenger car doors further away from the locomotive, the use of LiDARs also provides a technological possibility; their use in a research sensor setup can furthermore contribute to automatically annotating data from cameras and radar sensors as specified in Secs. 6.2.4 and 6.3.2. In this case, ranges of several meters are sufficient, but a high resolution is desirable to recognize individual persons, while a wide field of view assures that persons close to the

side of the car can be detected robustly. Hence, a wide range of sensors can be considered, with considered specifications including the Continental HR Flash or the Luminar Hydra. The proposed mounting position considers that the presence of persons outside the near field of view of the sensor can be detected using, for example, the ultrasonic sensors described in Sec. 6.1.2.

6.9 GNSS/IMU

For self-localization, GNSS is used in addition to ETCS balises and odometer pulse generators. The antennas are positioned on the vehicle roof at a large distance and as free of shadows as possible. The use of three GNSS antennas (see Fig. 4) allows the position and orientation of the vehicle to be determined. GNSS—Global Navigation Satellite System—is the collective term for a satellite navigation system, such as GPS (USA), GLONASS (Russia), BeiDou (China) or Galileo (Europe). GNSS receivers or antennas usually offer the possibility to access several of these systems. The main challenges of these systems are the accuracy, availability, and reliability of the position determination, which depends much on the number and geometric distribution of available satellites. In addition, the accuracy of a GNSS-based localization can be improved by an inertial measurement unit (IMU). The IMU records the acceleration and angular velocity of an object (e.g. locomotive), which allows the relative movement to be determined. Overlaying this with the localization provided by the GNSS increases the accuracy and long-term stability of the positioning. In order to take advantage of the additional IMU used for localization, the use of three GNSS antennas is recommended:

- One GNSS antenna: positioning without orientation information
- Two GNSS antennas: positioning with additional information for orientation or for determining the direction of travel (yaw/pitch or yaw/roll).
- Three GNSS antennas: positioning with 3D orientation information.

6.10 Microphones

Microphones can be used to acquire internal and external acoustic input. External sounds can indicate damages or upcoming damages on the railway or train itself, while internal microphones can be used to recognize damages from the inner side of the locomotive. Due to unique issues of the railroad operation, we have to choose different types of microphones for different tasks and their appropriate mounting positions in special.

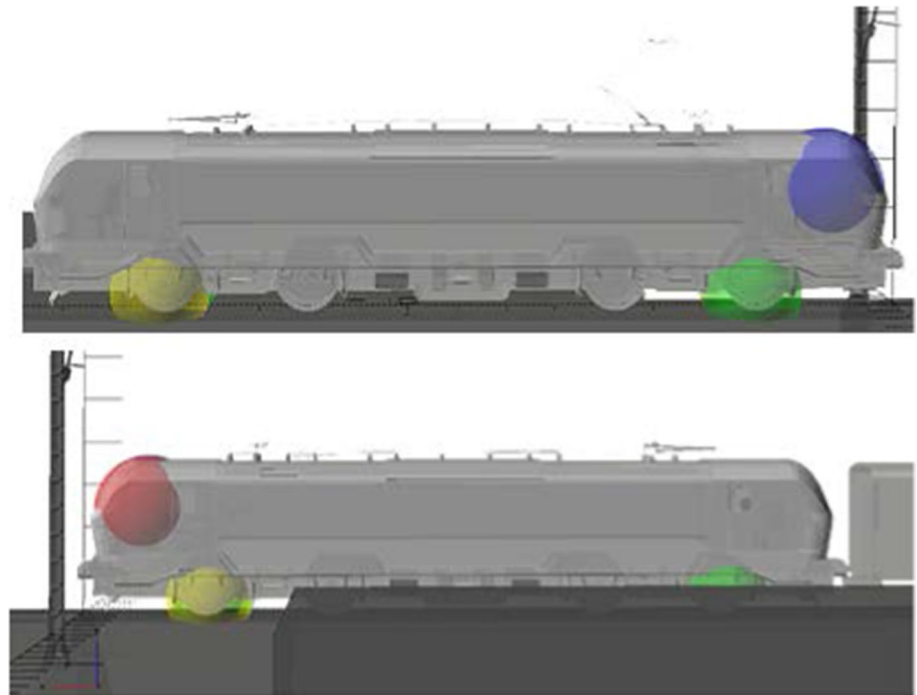
Condenser and moving coil microphones are suitable for railroad applications. Condenser microphones reproduce the recorded sound in good quality, but are sensitive to excess moisture and high sound pressure levels. They are therefore only suitable to a limited extent in the loud railroad area with its high dynamic range.

Plunger coil microphones appear to be suitable, as they are very robust. They are also insensitive to high sound pressures, but do not have good sensitivity to high frequencies. The geometry to be used depends on the application. All-round observation of sound can be achieved by means of an omnidirectional characteristic. Pressure microphones can be used up to the infrasound range. For a direction-independent observation and less relevant low frequencies, pressure gradient microphones should be selected, which can be adapted for different directional characteristics. In this way, interference noise can already be geometrically prevented or reduced. Alternatively, array microphones offer the possibility of also determining the direction of the sound source by means of several built-in microphone capsules. The use of microphones in the railroad sector serves, for example, to detect anomalies on the vehicle or the tracks.

When positioning microphones, it is important that they may or may not be decoupled from the sound source. This depends on the type of microphone. Structure-borne sound sensors, which can detect anomalies in the driving position, for example, must not be decoupled. However, if noises in the passenger compartment are to be monitored, for example by (stereo) microphones, they must be decoupled from the sound source. An installation in the driver's cab (Fig. 17) can be considered for this purpose [38]. The acoustic noises or signals possibly occurring in the cabin are between 20 Hz and 16 kHz and can be perceived by average hearing.

Noise from the undercarriage can also occur and be perceived (Fig. 17). When selecting the microphones, it should be noted that the short-term g-forces occurring at the railway bogie can be comparatively high.

Fig. 17 Microphone positions with spectral sensitivity on the right and the left sides of the locomotive [12]



7 Conclusions and outlook

The sensor setup outlined in the paper from the AI4Rails workshop [4] is defined in more detail in this paper. It covers all standard tasks in connection with their perception tasks as elaborated in Sec. 3 and listed in Table 1. Each perception task requires at least two different sensor types, and where reasonable, different sensor variants are specified with pairwise partially overlapping fields of coverage are specified to allow studies on the impact of sensor characteristics on AI methods and to facilitate annotations that benefit multiple sensors at the same time.

A common data ecosystem is needed for training and testing of ATO functions, where sensor data and corresponding annotations will be the first step. The sensor system specified in Sec. 6 is designed to collect research sensor data for all perception tasks which were identified in Sec. 3. The setup has been refined according to the expectations and needs identified across the railway sector as a system which offers different sensor options for each task. The complete specification is found in [12].

The next step must be to physically implement the sensor system on a locomotive and passenger car and run extensive data collection campaigns. This must be complemented by annotation and dissemination initiatives.

As a perspective, a method must be found to define which domains, locations, objects and conditions must be represented in the data in order to cover the whole domain of ATO perception data and ensure safe operation under all conditions. Such an evaluation will most likely result in the need to incorporate augmented or synthetically generated data in the training and test data and check for relevance, representativeness and correctness.

The authors are convinced that these steps will foster and accelerate research and development of ATO functions, with a particular focus on complex AI methods in safety-critical real-time tasks.

Author contributions This research is supported by German Centre for Rail Traffic Research. This is not an official statement, guideline or directive of the German Federal Railway Authority. All authors contributed to the study conception and design. In-depth literature research and analysis were performed by the members of the contractor consortium DL, HvdH, DS, JZ, AS and SW. CK, PK and RT as members of the ordering party participated in the tender design, project supervision and editing of the project reports. The first draft of the manuscript was written by RT and all authors worked on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. BMDV. Autonomes Fahren im Schienenverkehr. FE-Nr. 97.370/2016. 2018.
2. Digitale Schiene Deutschland. Sensors4Rail testet erstmals sensorbasierte Wahrnehmungssysteme im Bahnbetrieb. 2021. <https://digitale-schiene-deutschland.de/Sensors4Rail>. Accessed 9 Feb 2022.
3. Ristić-Durrant D, Franke M, Michels K. A review of vision-based on-board obstacle detection and distance estimation in railways. *Sensors*. 2021;21(10):3452.
4. Tagiew R, Leinhos D, von der Haar H, Klotz C, Sprute D, Ziehn J, Schmelter A, Witte S, Klasek P. Onboard sensor systems for automatic train operation. In: Dependable Computing – EDCC 2022 Workshops: SERENE, DREAMS, AI4RAILS. Springer. 2022. p. 139–50.
5. Cordts M, Omran M, Ramos S, Rehfeld T, Enzweiler M, Benenson R, Franke U, Roth S, Schiele B. The cityscapes dataset for semantic urban scene understanding. In: IEEE CVPR. 2016.
6. Geiger A, Lenz P, Urtasun R. Are we ready for autonomous driving? The kitti vision benchmark suite. In: CVPR. 2012.
7. Yu F, Chen H, Wang X, Xian W, Chen Y, Liu F, Madhavan V, Darrell T. BDD100K: a diverse driving dataset for heterogeneous multitask learning. In: IEEE/CVF CVPR. 2020. p. 2636–45.
8. Zendel O, Murschitz M, Zeilinger M, Steininger D, Abbasi S, Beleznai C. Railsem19: a dataset for semantic rail scene understanding. In: IEEE/CVF CVPR. 2019.
9. Harb J, Rébéna N, Chosidow R, Roblin G, Potarusov R, Hajri H. FRSign: a large-scale traffic light dataset for autonomous trains. arXiv. 2020. <https://doi.org/10.48550/arXiv.2002.05665>.
10. Toprak T, Belenlioglu B, Aydin B, Guzelis C, Selver MA. Conditional weighted ensemble of transferred models for camera based onboard pedestrian detection in railway driver support systems. *IEEE Trans Veh Technol*. 2020;69(5):5041–54.
11. Tilly R, Neumaier P, Schwalbe K, Klasek P, Tagiew R, Denzler P, Köppel M, Klockau T. Offener Multisensordatensatz für die Entwicklung der Umfeldwahrnehmung beim vollautomatischen Fahren. *ETR – Eisenbahntechnische Rundschau*. 2023;4.
12. Leinhos D, Flatt H, Witte S. Sensorik als technische Voraussetzung für ATO-Funktionen. Deutsches Zentrum für Schienenverkehrsforschung: Technical report. 2022.
13. Klotz C, Leinhos D, Ponholzer F, Witte S, Sprute D. Test- und Trainingsdaten für ATO-Perzeptionssysteme — Anforderungen an ein Messsystem. *ETR – Eisenbahntechnische Rundschau*. 2022;9.
14. Krizhevsky A. Learning multiple layers of features from tiny images. Technical report. 2009.
15. Chrabaszcz P, Loshchilov I, Hutter F. A downsampled variant of ImageNet as an alternative to the CIFAR datasets. arXiv. 2017. <https://doi.org/10.48550/arXiv.1707.08819>.
16. Simonovic M. D1.2—analysis of requirements and definition of specifications for obstacle detection and track intrusion systems. Technical report. 2019.
17. Labayen M, de Eribe DO. D1.1—specification and design of a common database for rail AI training and testing. Technical report. 2022.
18. DB Netz AG. Richtlinie 408.20 "Fahrdienstvorschrift–Richtlinien 408.21-27 und 408.48". DB Netz AG, Frankfurt. Version 4.0, valid from Dec 12, 2021.
19. DB Fernverkehr AG. Richtlinie 418.10-90 "Triebfahrzeugführerheft der DB Fernverkehr AG". DB Fernverkehr AG, Frankfurt. Version 10, valid from Dec 15, 2019.
20. DB Regio AG. DBREGIO-003 "Regelbuch- Basisteil für Mitarbeiter im Bahnbetrieb (inkl. Führen von Triebfahrzeugen)". DB Regio AG, Frankfurt. Version 04a, valid from Dec 13, 2020.
21. DB Cargo AG, Mainz. DBCDE-003 "Regelbuch-Basisteil für Mitarbeiter im Bahnbetrieb (inkl. Führen von Triebfahrzeugen)". DB Cargo AG, Mainz. Version 04, valid from Dec 15, 2019.
22. Eisenbahn-Bau- und Betriebsordnung, originally May 8 1967 (BGBl. II p. 1563), last updated at paragraph 2 of the regulation on April 5 2019. 2019.
23. Eisenbahn-Signalordnung, originally 1959 (BGBl. III, Gliederungsnummer 933-6), last updated at paragraph 517 of the regulation on August 31 2015. 2015.
24. European Commission. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. 2021.
25. Triess LT, Dreissig M, Rist CB, Zöllner JM. A survey on deep domain adaptation for lidar perception. In: IEEE IV Workshops. IEEE. 2021. p. 350–7.
26. Wei Y, Wei Z, Rao Y, Li J, Zhou J, Lu J. LiDAR distillation: bridging the beam-induced domain gap for 3D object detection. arXiv. 2022. <https://doi.org/10.48550/arXiv.2203.14956>.
27. Herrmann C, Ruf M, Beyerer J. CNN-based thermal infrared person detection by domain adaptation. In: Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything, SPIE. 2018. vol. 10643, p. 1064308.

28. Bodensteiner C, Bullinger S, Arens M. Multispectral matching using conditional generative appearance modeling. In: IEEE AVSS, 2018. p. 1–6.
29. Ho N, Pham M, Vo ND, Nguyen K. Vehicle detection at night time. In: NAFOSTED NICS. 2020. p. 250–5.
30. Lin C-T, Huang S-W, Wu Y-Y, Lai S-H. GAN-based day-to-night image style transfer for nighttime vehicle detection. IEEE Trans Intell Transp Syst. 2021;22(2):951–63.
31. Rojtberg P, Pöllabauer T, Kuijper A. Style-transfer GANs for bridging the domain gap in synthetic pose estimator training. In: IEEE AIVR. 2020. p. 188–95.
32. Liu R, Yang C, Sun W, Wang X, Li H. StereoGAN: bridging synthetic-to-real domain gap by joint optimization of domain translation and stereo matching. In: IEEE/CVF CVPR. 2020. p. 12754–63.
33. Bosch Mobility Solutions. Ultrasonic sensor. 2023. <https://www.bosch-mobility-solutions.com/en/solutions/sensors/ultrasonic-sensor/>. Accessed 13 Feb 2023.
34. Continental. Ultrasonic parking sensor. 2021. <https://www.continental-automotive.com/en-gl/Passenger-Cars/Autonomous-Mobility/Enablers/Ultrasonic-Parking-Sensor>. Accessed 28 Oct 2021.
35. Ester M, Kriegel H-P, Sander J, Xu X, *et al.*: A density-based algorithm for discovering clusters in large spatial databases with noise. In: KDD. 1996. p. 226–31.
36. Aeva Becomes First FMCW 4D LiDAR on NVIDIA DRIVE Autonomous Vehicle Platform. Businesswire. 2022.
37. Zuclich JA, Gagliano DA, Cheney F, Stuck BE, Zwick, H, Edsall PR, Lund DJ. Ocular effects of penetrating IR laser wavelengths. In: Laser-tissue interaction VI, vol. 2391. SPIE. 1995. p. 112–25.
38. Ngigi R. Modern techniques for condition monitoring of. J Phys. 2012. <https://doi.org/10.1088/1742-6596/364/1/012016>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.