Article

Demonstrations of Cooperative Perception: Safety and Robustness in Connected and Automated Vehicle Operations

Mao Shan [1,*] , **Karan Narula** [1] , **Yung Fei Wong** [2] , **Stewart Worrall** [1] , **Malik Khan** [2] , **Paul Alexander** [2] **and Eduardo Nebot** [1]

[1] Australian Centre for Field Robotics, The University of Sydney, Sydney, NSW 2006, Australia; karan.narula@sydney.edu.au (K.N.); stewart.worrall@sydney.edu.au (S.W.); eduardo.nebot@sydney.edu.au (E.N.)

[2] Cohda Wireless, 27 Greenhill Road, Wayville, SA 5034, Australia; ricky.wong@cohdawireless.com (Y.F.W.); malik.khan@cohdawireless.com (M.K.); paul.alexander@cohdawireless.com (P.A.)

* Correspondence: mao.shan@sydney.edu.au; Tel.: +61-2-8627-4496

Abstract: Cooperative perception, or collective perception (CP), is an emerging and promising technology for intelligent transportation systems (ITS). It enables an ITS station (ITS-S) to share its local perception information with others by means of vehicle-to-X (V2X) communication, thereby achieving improved efficiency and safety in road transportation. In this paper, we present our recent progress on the development of a connected and automated vehicle (CAV) and intelligent roadside unit (IRSU). The main contribution of the work lies in investigating and demonstrating the use of CP service within intelligent infrastructure to improve awareness of vulnerable road users (VRU) and thus safety for CAVs in various traffic scenarios. We demonstrate in experiments that a connected vehicle (CV) can "see" a pedestrian around the corners. More importantly, we demonstrate how CAVs can autonomously and safely interact with walking and running pedestrians, relying only on the CP information from the IRSU through vehicle-to-infrastructure (V2I) communication. This is one of the first demonstrations of urban vehicle automation using only CP information. We also address in the paper the handling of collective perception messages (CPMs) received from the IRSU, and passing them through a pipeline of CP information coordinate transformation with uncertainty, multiple road user tracking, and eventually path planning/decision-making within the CAV. The experimental results were obtained with manually driven CV, fully autonomous CAV, and an IRSU retrofitted with vision and laser sensors and a road user tracking system.

Keywords: collective perception; cooperative perception; collective perception service; V2X communication; intelligent roadside unit; connected and automated vehicle

1. Introduction

Autonomous vehicles (AVs) have received extensive attention in recent years as a rapidly emerging and disruptive technology to improve safety and efficiency of current road transportation systems. Most of the existing and under development AVs rely on local sensors, such as cameras and lidars, to perceive the environment and interact with other road users. Despite significant advances in sensor technology in recent years, the perception capability of these local sensors is ultimately bounded in range and field of view (FOV) due to their physical constraints. In addition, occluding objects in urban traffic environments such as buildings, trees, and other road users impose challenges in perception. There are also robustness related concerns, for instance, sensor degradation in adverse weather conditions, sensor interference, hardware malfunction and failure. Unfortunately, failing to maintain sufficient awareness of other road users, vulnerable road users (VRU), in particular, can cause catastrophic safety consequences for AVs.

In recent years, V2X communication has garnered increasing popularity among researchers in the field of intelligent transportation system (ITS) and with automobile man-

ufacturers, as it enables a vehicle to share essential information with other road users in a V2X network. This can be a game changer for both human operated and autonomous vehicles, which would be referred to as connected vehicles (CVs) and connected and automated vehicles (CAVs), respectively. It will also open many doors to new possibilities with peer-to-peer connectivity. The connected agents within the cooperative ITS (C-ITS) network will be able to exploit the significant benefits that come from sharing information amongst the network. For instance, the standardised Cooperative Awareness Messages (CAMs) enable mutual awareness between connected agents. Nevertheless, there are other types of road users such as non-connected vehicles, pedestrians, and cyclists that have not been included in the C-ITS services yet. The detection of these non-connected road users in this case becomes an important task for road safety.

The major standardisation organisations such as European Telecommunications Standard Institute (ETSI), SAE and IEEE have made a significant effort to standardise specifications regarding C-ITS services, V2X communication protocols, and security. This is essential to facilitate the deployment of C-ITS in road transportation network globally. The collective perception (CP) service is among those C-ITS services that are currently being standardised by ETSI. The CP service enables an ITS station (ITS-S), for instance, a CAV or an intelligent roadside unit (IRSU) to share its perception information with adjacent ITS-Ss by exchanging Collective Perception Messages (CPMs) via V2X communication. The ETSI CPMs convey abstract representations of perceived objects instead of raw sensory data, facilitating the interoperability between ITS-Ss of different types and from different manufacturers. A CAV can benefit from the CP service in terms of improved awareness of surrounding road users, which is essential for ensuring road safety. Specifically, it facilitates a CAV to extend its sensing range and improve sensing quality, redundancy, and robustness through cross-platform sensor fusion, i.e., fusing its local sensory data with other CAV and IRSU information. In addition, the improved perception quality as a result of the sensor fusion potentially relaxes the accuracy and reliability requirements of onboard sensors. This could lower per vehicle cost to facilitate the massive deployment of CAV technology. As for traditional vehicles, CP also brings an attractive advantage of enabling perception capability without retrofitting the vehicle with perception sensors and the associated processing unit.

Over the last two years, we have been working with Cohda Wireless on CP and CAV. We are particularly interested in the safety implications the CP service is bringing into the current and future transportation network, and how the CP service will potentially shape the development of intelligent vehicles. To this end, we have developed an IRSU and tested it with the Australian Centre for Field Robotics (ACFR) CAV platforms in different traffic scenarios. The IRSU is equipped with a multi-beam laser scanner and multiple cameras for road user detection and tracking, and the vehicles are retrofitted with a suite of local sensors to implement full autonomy. The IRSU and CAV have the V2X communication capability through Cohda Wireless MK5 Road-Side Unit (RSU) and On-Board Unit (OBU), respectively.

The first and also the main contribution of the paper is to investigate and demonstrate through three representative experiments how a CV and a CAV achieve improved safety and robustness when perceiving and interacting with VRU, using the CP information from an intelligent infrastructure in different traffic environments and with different setups. These experiments do not use the CAV's internal perception capabilities so as to highlight the benefits of using intelligent infrastructure and CP service in the traffic environments. The received perception data from the IRSU are used as the only source of information for multiple road users tracking and path planning within the smart vehicle we tested with. The first experiment was conducted on a public road in an urban traffic environment, and the CV was able to "see" a visually obstructed pedestrian before making a turn into an alleyway. It is demonstrated in the next two experiments that a CAV navigated autonomously and safely when interacting with walking and running pedestrians in proximity in simulated and real lab traffic environments, respectively. The second contribution

of the paper is to address the coordinate transformation of perception information considering the respective uncertainties, and to analyse the influence of ITS-S self-localisation accuracy through numerical simulations. The uncertainties associated with the perceived object information are essential for performing consistent state estimation for road users, and probabilistic fusion of perception information from multiple ITS-Ss in CP.

The remainder of the paper is organised as follows. Section 2 will focus on the related work on CP and its use cases for CAV. Section 3 presents the IRSU and CAV platforms developed and used in the experiments, and the coordinate transformation of perception information with uncertainty is addressed in Section 4. The results from the three experiments are presented in Section 5, followed by conclusions drawn in Section 6.

2. Related Work

The concept of CP has been extensively studied in the ITS research community over the last two decades. Initial CP related work proposes to share raw sensory data between two mobile agents, such as images [1], lidar point clouds [2], both combined [3–5], location and relative range measurements [6,7]. Those approaches however tend to require prohibitively high bandwidth for existing V2X communication technologies in a dense environment. In addition, raw sensor data are often vendor dependent and proprietary, causing interoperability issues among communicating ITS-Ss.

More theoretical and experimental work on CP was conducted as part of the Ko-FAS [8] research initiative. These include [9–11] based on the Ko-PER specified Cooperative Perception Message (CPM). It is a supplementary message to the standard ETSI ITS G5 CAMs, to support the abstract description of perceived dynamic and static objects. Experimental studies are conducted in [9] on the Ko-PER CPM transmission latency and range. In [10], a high level object fusion framework in CP is proposed, which combines the local sensor information with the perception data received from other V2X enabled vehicles or roadside units (RSUs). The authors of [11] investigate the inter-vehicle data association and fusion for CP. More recent work in [12] proposes a variant of Ko-PER CPM and analyses the trade-off between message size as a result of enabling optional data fields in the CPM and global fusion accuracy.

Based on the work in [9], the authors of [13] propose Environmental Perception Message (EPM) for CP with different information containers specifying sensor characteristics, originating station state, and parameters of perceived objects. High level object fusion using the perceived information in received EPMs is also addressed in [13]. Both EPM and the earlier Ko-PER CPM are a separate message that contains all CP related data elements (DEs) and data frames (DFs), and has to be transmitted in parallel with an ETSI CAM. There is also work towards extended CAM. For instance, Cooperative Sensing Message (CSM) from AutoNet2030 [14–16] extends CAM specifications to include a description of objects perceived by local or remote sensors. Following a similar concept of CP, Proxy CAM is presented in [17–19], where intelligent infrastructure generates standard CAMs for those perceived road users, while the work in [20] proposes a CPM comprised of a collection of CAMs, each describing a perceived object. The work in [21,22] evaluates different EPM dissemination variants under low and high traffic densities and proposes to attach the CP relevant DFs in EPM to CAM to minimise communication overhead. The CPM currently being specified at ETSI, as in [23], is derived from optimising the EPM and combining it with CAM. It is therefore more self-contained, no longer dependent on the reception of CAMs.

Similarly, there are early stage standardisation activities in SAE advanced application technical committee to standardise messages and protocols for sensor data sharing in SAE J3224 [24]. These messages and protocols are not yet defined and are thus not considered in this work.

Considering the limited communication bandwidth and avoiding congestion in the wireless channel, more recent studies in the CP area tend to focus on the communication aspect of the technology, weighing up provided CP service quality and the V2X network

resources. The work in [25] investigates ETSI CPM generation rules that balance provided service quality and the V2X channel load. It is provided in [26] an in-depth study on the impact of different CPM generation rules from the perspectives of V2X communication performance and perception capabilities in low and high density traffic scenarios. The authors of [27] raise the concern of redundant data sharing in V2V based CP with the increase of CAV penetration rate. To tackle the redundant transmission issue, a probabilistic data selection approach is presented in [28]. The authors of [29] propose an adaptive CPM generation rule considering change in perceived object's state, and the authors of [30] propose to employ object filtering schemes in CPM, to improve communication performance while minimising the detriment to perception quality. Similarly, the work in [31] presents a deep reinforcement learning based approach that a vehicle can employ when selecting data to transmit in CP to alleviate communication congestion.

There is also work conducted to explore the use cases, benefits, and challenges of CP. The authors of [32] provide early study of CP illustrating its potential in terms of improved vehicle awareness and extended perception range and field of view. The work presented in [13] evaluates EPM for obstacle avoidance of two manually driven CAVs, showing that the CP helps gain extra reaction time for the vehicles to avoid obstacles. It is analysed in [27] the performance gain in extending horizon of CAVs by leveraging V2V based CP. The work in [33,34] analytically evaluates the enhancement of environmental perception for CVs at different CP service penetration rates and with different traffic densities. The authors of [35] discuss the security threats in CP and propose possible countermeasures in V2X network protocols, while the work in [36] focuses on using CP for detecting vehicle misbehaviour due to adversarial attacks in V2X communication. Most of the CP related use cases studied in the literature are safety related, including cooperative driving [5,16], cooperative advisory warnings [37,38], cooperative collision avoidance [4,13,39], intersection assistance [17,40], and vehicle misbehaviour detection [41], to name a few. It is presented in [42] quantitative comparison of V2V and V2I connectivity on improving sensing redundancy and collaborative sensing coverage for CAV applications. The work concludes that infrastructure support is crucial for safety related services such as CP, especially when the penetration rate of sensing vehicles is low. The authors of [17] demonstrate improved awareness of approaching vehicles at an intersection using the CP information from an IRSU. The CP in the work is achieved through Proxy CAM. The authors of [43] compare CAM and CPM and demonstrate IRSU assisted augmented perception through simulations. Recent work in [44] demonstrates the IRSU assisted CP for extending perception of CAVs on open roads. Infrastructure-assisted CP is also part of the scope of Managing Automated Vehicles Enhances Network (MAVEN) [45], an EU funded project targeting traffic management solutions where CAVs are guided at signalised cooperative intersections in urban traffic environments [40]. Other CP related joint research projects include TransAID [46] and IMAGinE [47].

A significant proportion of the existing work conducts the analyses of V2X communication and CP in simulated environments. For instance, the work in [27,28] is carried out in an open source microscopic road traffic simulation package SUMO (Simulation of Urban Mobility) [48]. Another commonly used network and mobility simulator is *Veins* (Vehicles in Network Simulation) [49,50], which integrates SUMO and a discrete-event simulator OMNeT++ (Open Modular Network Testbed in C++) [51] for modelling realistic communication patterns. The authors of [21,22,25,30,32,52] conduct work in *Artery* [53,54] framework, which wraps SUMO and OMNet++, and enables V2X simulations based on ETSI ITS G5 protocol stack. There are also simulators for advanced driving assistance systems (ADASs) and autonomous driving systems, which provide more realistic sensory level perception information. In recent years, they start to show their potential in testing and validating CP with CAVs. For instance, Pro-SiVIC is employed in [43] for CP related simulations, and CARLA [55,56] is combined with SUMO in a simulation platform developed for CP in [31]. Despite the notable progress on the study of CP achieved in recent

years, the CP for CAV operations has not been extensively tested using real platforms and in real traffic environments.

3. Platforms

3.1. Intelligent Roadside Unit

The IRSU developed is comprised of a sensor head, a processing workstation, and a Cohda Wireless MK5 RSU (Cohda Wireless, Wayville, Australia). The sensor head is mounted on a tripod for easy deployment in different testing environments, as shown in Figure 1. Specifically, two Pointgrey Blackfly BFLY-PGE-23S6C-C cameras (FLIR Systems, Wilsonville, OR, USA) are mounted with an angle separation of 45°. The camera with its lens Fujinon CF12.5HA-1 has a horizontal field-of-view (FOV) of about 54°, and a vertical FOV of 42°. The setup of dual cameras achieves a combined FOV of approximately 100°, and the FOV can be further augmented by adding more cameras to the sensor head. A RoboSense 16-beam lidar (Suteng Innovation Technology, Shenzhen, China) is also installed into the sensor head, performing 360-degree scanning of the environment at 10 Hz. The workstation has AMD Ryzen Threadripper 2950X CPU, 32 GB memory, RTX2080Ti GPU, running Robot Operation System (ROS) Melodic on Ubuntu 18.04.2 long-term support (LTS). The Cohda Wireless MK5 RSU is a rugged unit with dual antennas, housed in a weatherproof enclosure, making it ready for outdoor deployment.
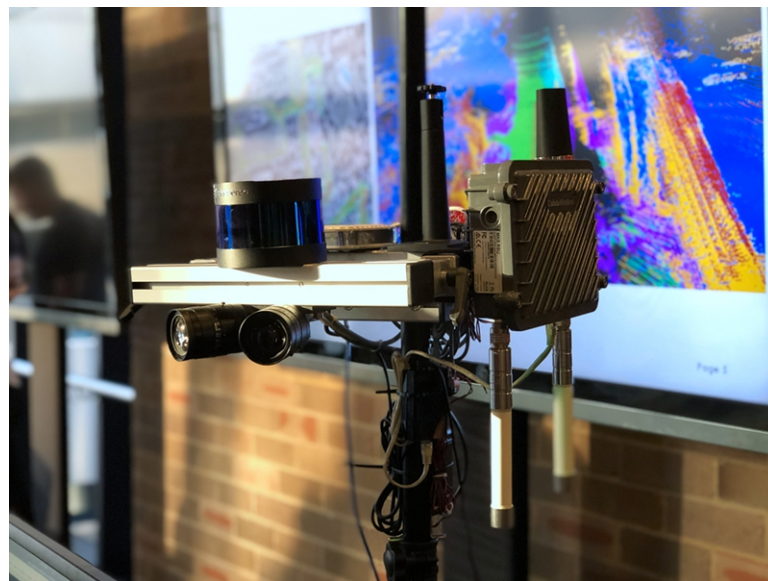

Figure 1. The developed IRSU is equipped with multiple sensors including dual cameras and a 16-beam lidar. The sensors sit on a tripod for easy deployment in the field.

In terms of information processing, the workstation first performs sensory data fusion of images and lidar point clouds for pedestrian and vehicle detection. Specifically, the raw images from cameras are first rectified using camera intrinsic calibration parameters. As illustrated in Figure 2, the road users within the images are classified/detected using YOLOv3 that runs on GPU. At the same time, the lidar point clouds are projected to the image coordinate system with proper extrinsic sensor calibration parameters. The lidar points are then segmented, clustered and labelled by fusing the visual classifier results (in the form of bounding boxes in the images) and the projected lidar points. Given two images from the cameras and a lidar point cloud at a time, the camera-lidar sensor fusion for road user detection takes approximately 70 ms within the workstation.

The detection results are then encoded into ETSI CPMs and broadcast by the Cohda Wireless MK5 RSU at 10 Hz. Details are available in Section 3.3. We tested the working range of the developed IRSU for detecting common road users, such as pedestrians and vehicles. The maximum detection range is approximately 20 m for pedestrians and 40 m

for cars. In addition, in the IRSU, a variant of Gaussian mixture probability hypothesis density (GMPHD) filter [57] is employed to track multiple road users and has its tracking results visualised in real time within the workstation. The same tracking algorithm is also employed on the CAV side. Details will be given in Section 3.2.
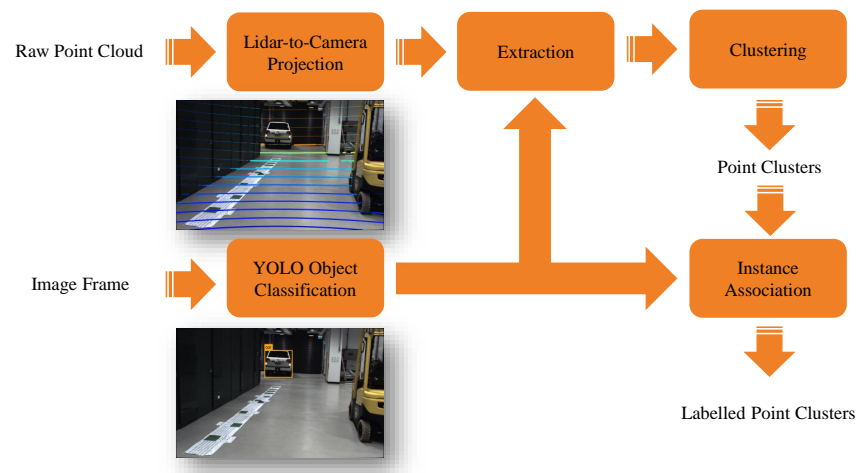

Figure 2. The sensory data processing pipeline within the IRSU.

To assess the position tracking accuracy of the IRSU, an outdoor test was conducted at the Shepherd Street car park at the University of Sydney. Figure 3 illustrates the setup of the test at the car park, where a pedestrian was walking in front of the IRSU for approximately one minute. The ground truth positions of the target pedestrian were obtained at 1 Hz by an u-blox C94-M8P RTK receiver, which reports a standard deviation of 1.4 cm of positioning when in the RTK fixed mode.
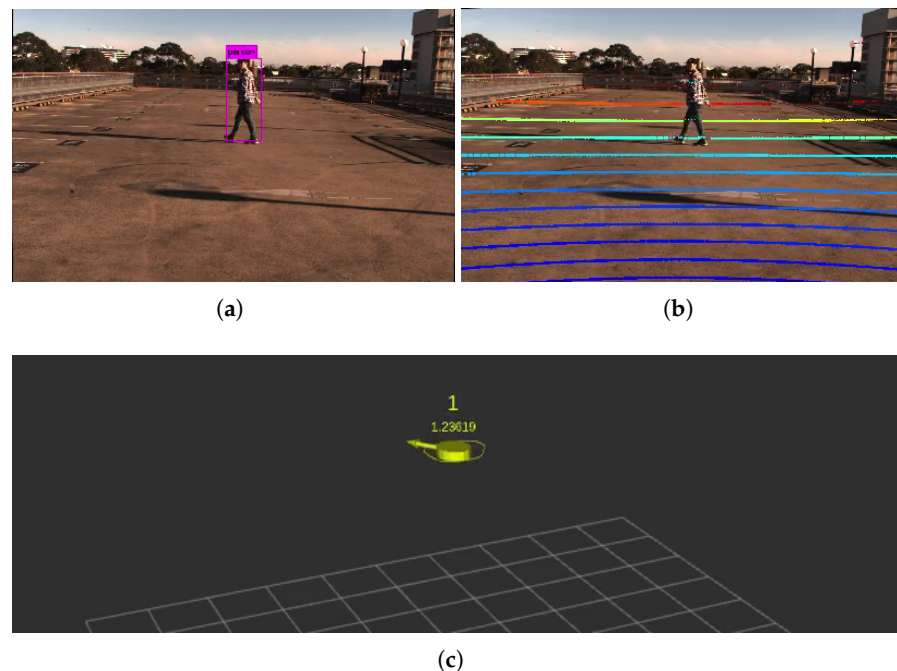

(a)

(b)

(c)

Figure 3. Pedestrian tracking setup for the IRSU. (**a**) shows the detection of the target pedestrian within a camera image using YOLOv3. The lidar points are projected to the image frame with extrinsic calibration parameters. The projected points in (**b**) are colour coded based on their range to the sensor in 3D space. Those bold points indicate those hitting the ground plane. (**c**) demonstrates the tracking of the pedestrian in 3D space. The RTK antenna was hidden within the cap of the pedestrian to log GNSS positions as the ground truth.

As mentioned previously, the road user detection in the IRSU is achieved through camera-lidar sensor fusion. Despite a centimeter level of ranging accuracy obtained from the multi-beam lidar, the bounding box accuracy in visual object detection decreases dramatically as the target moves further from the cameras. In addition to the noises from the sensors themselves, there are errors introduced in the data processing and sensor fusion, such as lidar-to-camera projection, point cloud segmentation, and clustering. Therefore, based on the comparison with ground truth data and considering the detection range of the IRSU, the overall position noise contained in the pedestrian detection is modelled as a zero-mean Gaussian variable with a standard deviation of 0.2 m in both x- and y-directions.

As presented in Figure 4, the trajectory of the target reported by the local tracker is found close to the ground truth points. The root mean squared error (RMSE) in position is calculated by comparing the ground truth positions and the corresponding estimates from the tracker. As the two sources of positions were obtained at different rates (10 Hz from the tracker versus 1 Hz from RTK), each RTK reading is compared with the position estimate that has the nearest timestamp. It can be seen from Figure 4b that the distance of the target pedestrian to the IRSU varies from 5 to 22 metres, and the position RMSE values are less than 0.3 m for most of the time in the test, as presented in Figure 4c.
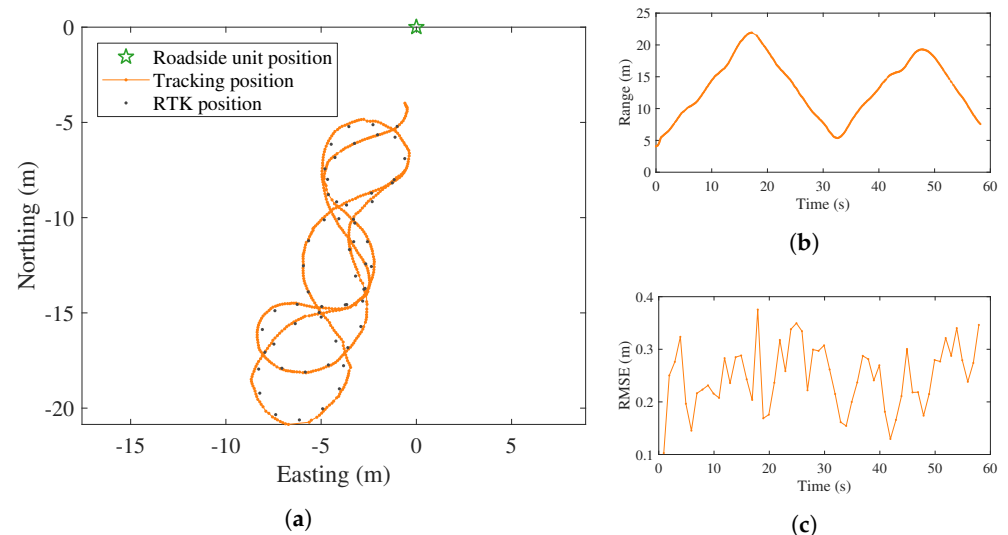

Figure 4. Pedestrian tracking results for the IRSU. It can be seen in (**a**) that, when the target pedestrian was walking in a figure eight pattern in front of the IRSU, the tracked positions are found to be well aligned with the ground truth path obtained from an RTK receiver. The distance of the pedestrian relative to the IRSU changes between 5 and 22 m, as depicted in (**b**). In (**c**), the position RMSE is found to be lower than 0.3 m for the majority of the time in the test.

3.2. Connected and Automated Vehicle

Figure 5 presents an overview of hardware configuration of the CAV platform built by the ACFR ITS group. Images are captured onboard at 30 FPS by a NVIDIA Drive PX2 automotive computer with six gigabit multimedia serial link (GMSL) cameras with 1080p resolution and a 100° horizontal FOV each. They are arranged to cover a 360° horizontal FOV around the vehicle. The vehicle also has a 32-beam scanning lidar with 30° vertical FOV and 360° horizontal FOV for scanning the surrounding at 10 Hz. Both the cameras and the lidar have been calibrated to the local coordinate system of the platform, using the automatic extrinsic calibration toolkit presented in [58]. In addition, the vehicle has a GNSS receiver, a six- degrees of freedom IMU, and four wheel encoders for odometry and localisation. The onboard Intel next unit of computing (NUC) has 32 GB of memory and a quad-core Intel i7-6670HQ processor, serving as the main processing computer within the CAV. The NUC is running ROS Melodic on Ubuntu 18.04.2 LTS. Last but not least, the CAV platform is equipped with a Cohda Wireless MK5 OBU to enable the V2X communication

capability. Built with the same chipset as the MK5 RSU, the MK5 OBU is a small module that can be retrofitted to vehicles for aftermarket deployment. Please refer to [59] for more details on the CAV platform and the USyd Campus data set collected using the platform.
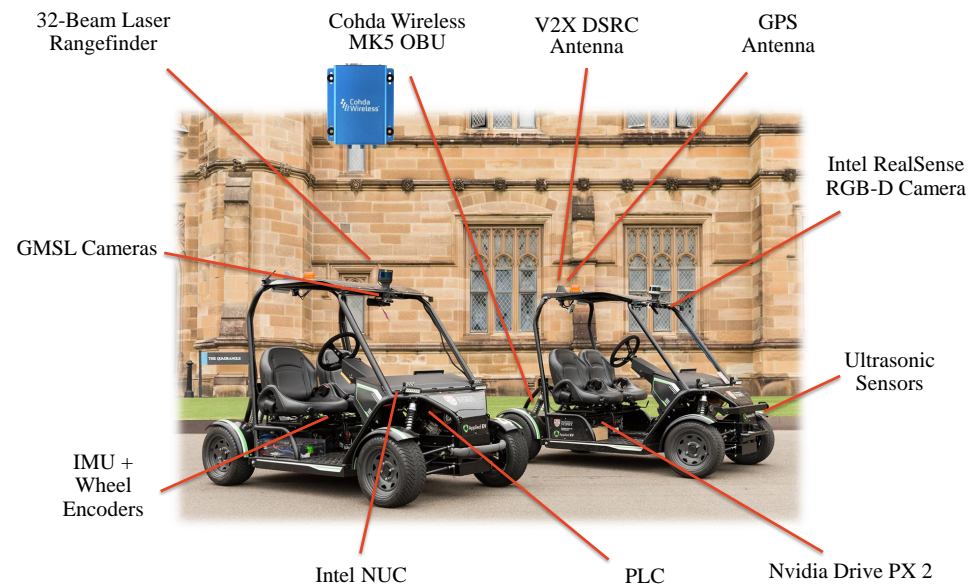

Figure 5. The CAV platform and onboard sensors.

Although the CAV can achieve full autonomy using its various built-in sensors, its perception sensors were not used for road user detection in the experiments presented in the paper, in which case, the CAV had to rely on the CP information from the IRSU when interacting with pedestrians in the traffic environments. This is to highlight the benefits of using the CP service in CAV operations. The multiple cameras were used for video recording purposes, and the multi-beam lidar was enabled only for aiding self-localisation within the map. Lidar feature maps of the experiment sites were built using a simultaneous localisation and mapping (SLAM) algorithm. The maps are based on pole and building corner features extracted from lidar point clouds, which are essential for localisation since GNSS alone cannot provide the desired level of accuracy in the experiment environments. Interested readers can refer to [60] for more information. In the meantime, a *Lanelet2* map is built for every experiment site, which includes road network, lane layout and traffic rules such as speed limits, traffic lights and right-of-way rules.

When the CAV receives an ETSI CPM through the onboard Cohda Wireless MK5 OBU, the received perceived objects information is first decoded from binary ASN.1 encoding, and transformed with its uncertainty to the local frame of reference of the CAV, as presented in Section 4, which also takes into account the estimated egocentric pose of the CAV in self-localisation.

Following the coordinate transformation with uncertainty, the perceived objects information from the IRSU are fused into a multiple road user tracking algorithm that is a variant of the GMPHD filter running within the local frame of the receiving CAV. The tracked states of road users include their position, heading, and speed. The general formulation of the GMPHD tracker is to use a Gaussian mixture to represent the joint distribution of the group of tracked targets. The GMPHD approach is considered attractive due to its inherent convenience in handling track initiation, track termination, probabilistic data association, and clutter. Compared with the naive GMPHD algorithm, the road user tracker running in both the IRSU and CAV is improved with measurement-driven initiation of new tracks, and track identity management. In addition, note that an instance of the tracker is required for each type of road user, which effectively reduces the overall computational cost.

The navigation subsystem in the CAV is responsible for path planning, monitoring and controlling the motion of the vehicle from the current position to the goal. A hybrid A*

path planner presented in our recent work [61] is running within the CAV to plan a path navigating around obstacles and the tracked pedestrians. It maintains a moving grid-based local cost map that considers (1) structural constraints, such as road and lane boundaries, present in the *Lanelet2* map, (2) obstacles picked up by local perception sensors, and (3) current and future estimates of road users detected and broadcast by IRSU through V2I communication, such that it can plan a smooth, safe, and kinematically feasible path to avoid collision with any other road users the CAV becomes aware of. It should be noted that, in the experiments presented in the paper, the use of local perception information in the navigation was minimised or disabled. Currently, the path planner is designed for low-speed scenarios, as our CAV platforms are intended for low-speed operations in shared traffic environments. However, this does not affect the demonstration of CP.

3.3. Handling of ETSI CPMs on IRSU and CAV Platforms

The overall system diagram of the IRSU and the CAV in a C-ITS configuration is illustrated in Figure 6. The detected road user descriptions are encoded to ETSI CPMs and transmitted from the IRSU to the CAV through the Cohda Wireless MK5 devices at 10 Hz. Each ETSI CPM consists of an ITS PDU header and five types of information containers accommodating mandatory and optional DEs and DFs. These containers are:

1. A *CPM Management Container*, which indicates station type, such as a vehicle or an IRSU, and reference position of the transmitting ITS-S.
2. An optional *Station Data Container*, which provides additional information about the originating station. This includes the heading, speed, and dimensions when the originating station is a vehicle.
3. Optional *Sensor Information Containers*, which describe the type and specifications of the equipped sensors of the transmitting ITS-S, including sensor IDs, types, and detection areas.
4. Optional *Perceived Object Containers*, each of which describes the dynamics and properties of a perceived object, such as type, position, speed, heading, and dimensions. These perceived object descriptions are registered in the coordinate system of the originating station.
5. Optional *Free Space Addendum Containers*, which describe different confidence levels for certain areas within the sensor detection areas.
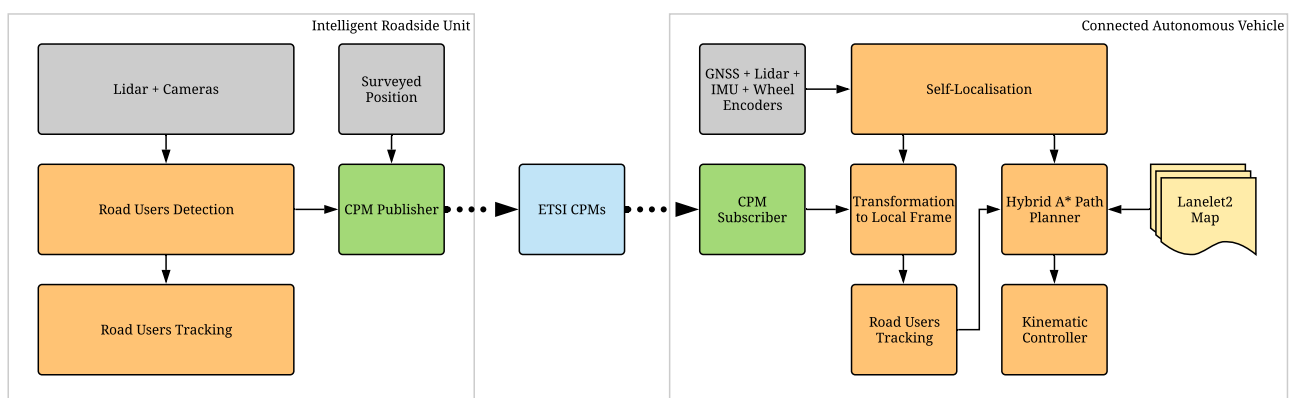
Figure 6. System diagram of the IRSU and the CAV platforms in the experiments.

The CPM generation rules support different abstraction levels for perceived object descriptions for the implementation flexibility, which can derive from low-level detections made by individual sensors, or the results of the high-level data fusion. In the developed IRSU, the road user detections as a result of camera-lidar sensor fusion are considered in the encoding of CPMs, as depicted in Figure 6. The road user tracking results within the IRSU can be used for different purposes, such as visualisation, traffic monitoring and

management. Sharing the tracking results in V2X communication for CP will require track-to-track fusion on the receiving ITS-S side, in which case special care needs to be taken to avoid double counting of common prior information, which causes inconsistency in state estimation.

The overall perception information flow from the IRSU to CAV is illustrated in Figure 7. Specifically, the handling of the perception information within the system is mostly carried out in ROS realm. The ROS Kinetic is installed in both Cohda Wireless MK5 devices, with the workstation/NUC set as the ROS master to the transmitting/receiving MK5, respectively. The perceived road users on the workstation side are first described in the form of a ROS CPM, which is used internally within the ROS system and brings convenience of message handling and diagnostics in ROS. Each ROS CPM is a self-contained message that contains all information required for an ETSI CPM. It is the ETSI CPM that is eventually transmitted in the V2X communication.
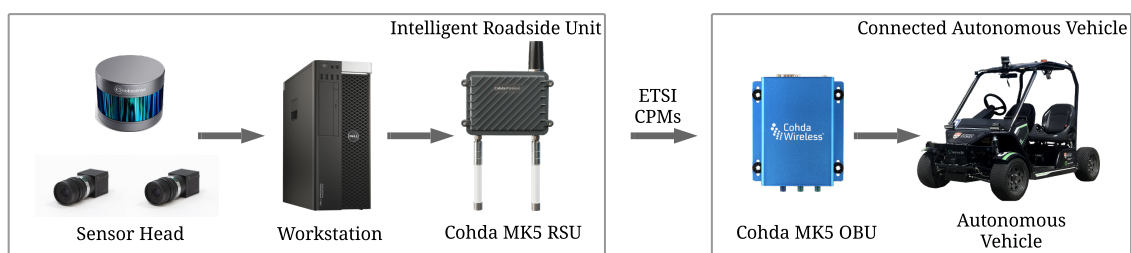

Figure 7. Perception information flow from the IRSU to the CAV. The IRSU broadcasts perceived objects information in the form of ETSI CPMs through the Cohda Wireless MK5 RSU. As the receiving agent, the CAV has V2X communication capability through the Cohda Wireless MK5 OBU.

A ROS *msg_bridge* node running within each MK5 serves as a bridge between ROS CPMs and ETSI CPMs. It populates an ETSI CPM given information from an ROS CPM and based on the ASN.1 definition of the ETSI CPM. In the meantime, it decodes an ETSI CPM received from other ITS-Ss back into the ROS CPM format. The transmission and reception of ETSI CPM payload are handled by Cohda's V2X stack in MK5. We keep upgrading the ROS *msg_bridge* node to support the features required in the CP demonstrations in accordance with the status quo of ETSI CPM standardisation.

4. Coordinate Transformation of Perceived Objects with Uncertainty

The perceived objects information broadcast through CP service is produced from the perspective of the sensing ITS-S. When a piece of perceived object information is received, it is not usable for the receiving agent until it is transformed into its local coordinate system. In [13], the coordinate transformation of perceived objects does not explicitly incorporate the accuracy fields accompanied with DEs in the EPMs. We stress that uncertainty bounds associated with the perception information are indispensable in successive data fusion and thus have to be considered. In addition, it is essential for the coordinate transformation to also incorporate the uncertainty contained in the pose estimation of both ITS-Ss. These are discussed in detail as follows.

1. Perception uncertainty in the sensing ITS-S. Every commonly used perception sensor in ITS area has its own strengths and limitations. For instance, RGB images are useful in detecting object instances and classification of road users with an estimated confidence level using a visual classifier algorithm. Doppler RADARs produce both position and velocity measurements but are prone to noise and interference from the environment. Lidars have a high range resolution to observe physical extent and shape of objects. Nevertheless, the point density decreases dramatically along with range. These sensors all produce measurements corrupted by noise, and thus should be modelled with uncertainty. Combining multi-modal sensory information not only improves robustness of the perception system, but also increases accuracy,

which means a lower level of uncertainty. The produced estimates of perceived objects with uncertainty have been represented in *Perceived Object Containers* of the CPM specification.

2.　Self-localisation uncertainty of sensing and receiving ITS-Ss. A vehicle ITS-S has to constantly localise itself within a global frame of reference such as a map or UTM frame for navigation and safety reasons. However, accurate self-localisation for a moving platform is known to be one of the existing challenges in ITS applications such as urban navigation. Using GNSS as the only source for localisation often cannot achieve satisfactory accuracy, in particular in GNSS-degraded or even GNSS-denied environments. There are various existing solutions that can provide higher localisation accuracy such as RTK, GNSS and inertial/encoder data fusion, feature based localisation based on existing map, etc. Nevertheless, even with the same localisation approach, the uncertainty magnitude in the localisation can vary significantly depending on certain external conditions, such as GNSS satellite visibility in the sky, and the quantity of observable features and their qualities in the surroundings. The location of an IRSU, although deployed static in a traffic environment, is not immune from localisation error either when set either by GNSS or through a surveying process. The localisation uncertainties of both sensing and receiving ITS-Ss therefore have to be considered in the perceived objects coordinate transformation since it cannot be completely eliminated regardless of the self-localisation means employed. The originating ITS-S information including its pose with associated uncertainty has been contained within the *CPM Management Container* and *Station Data Container* in the CPM definition.

Consequently, the transformed object state estimates contain uncertainty that is mainly a combined result of uncertainty in the localisation of both ITS-Ss and that in the sensory perception. As illustrated in Figure 8, both sensing and receiving ITS-Ss contain uncertainties in their estimated egocentric states within the global frame *G*. The sensing ITS-S *S* observes multiple road users within its local frame with some level of sensory measurement uncertainty labelled with purple dashed lines. The road users measurement uncertainties grow when transformed to a global frame *G* due to self-localisation uncertainty of ITS-S *S*, as labelled with gray dotted lines. The road users measurement uncertainties grow further, as labelled by orange solid lines, when transformed to the local frame of the receiving ITS-S *R* as its self-localisation uncertainty is also incorporated.
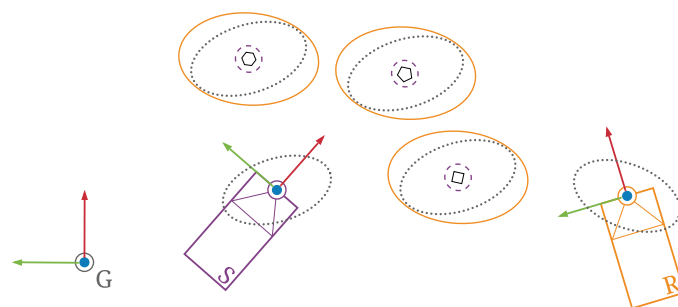
Figure 8. Coordinate transformation of perception information with uncertainty considered from the sensing ITS-S *S* to the receiving ITS-S *R*.

Both works [10,62] perform the coordinate transformation of perceived object states with their uncertainty through temporal and spatial alignment. The temporal alignment is introduced in the work mainly due to the time difference between the reception of a CAM and a Ko-PER CPM. It is, however, less of a concern for an ETSI CPM, which now can contain the originating station information required from a CAM.

4.1. Problem Formulation

This section presents mathematical formulation of the coordinate transformation of perceived objects information a recipient ITS-S receives from a sensing ITS-S. We propose to use unscented transform (UT) in the coordinate transformation with uncertainty, which here is recognised as a nonlinear and non-deterministic process. The formulation presented is intended for 2D transformation, as it is reasonable to assume a planar road surface for the road user detection and tracking problem discussed in the paper. Nevertheless, the formulation can be easily extended to transformation in 3D space. It is important to note that although the paper is mainly focused on CP service provided by intelligent infrastructure, the formulation presented here is intended to provide a generic form for two arbitrary ITS-Ss in a V2X network, which includes both V2V and V2I scenarios. Before proceeding further, some definitions about coordinate systems are first given to facilitate discussions.

The global frame, which is the fixed coordinate system attached on the ground, is represented as $\{G\} = \{\overrightarrow{x}_G, \overrightarrow{y}_G\}$. The local frame of the sensing ITS-S S is attached on the platform body and is represented as $\{S\} = \{\overrightarrow{x}_S, \overrightarrow{y}_S\}$, with $\overrightarrow{x}_S$ pointing to the east direction if it is an IRSU and pointing towards its moving direction for a vehicle. Local frame of the recipient ITS-S R is attached on the platform body and moving with the platform. It is denoted as $\{R\} = \{\overrightarrow{x}_R, \overrightarrow{y}_R\}$ with $\overrightarrow{x}_R$ pointing towards its moving direction.

The pre-superscript is employed to describe the coordinate frame in which the corresponding variable is expressed. For instance, $^V x_t^p$ denotes the position of an object p in the x-direction with respect to $\{V\}$ at time t. A variable without pre-superscript is defined in a generic way to not specify a particular coordinate frame.

The state of the receiving ITS-S R at time t and its estimate are denoted as

$$^G\mathbf{x}_t^R = \begin{bmatrix} ^G x_t^R & ^G y_t^R & ^G \theta_t^R \end{bmatrix}^T, \tag{1}$$

where $^G x_t^R$, $^G y_t^R$, and $^G \theta_t^R$ denote 2D coordinates and the heading, respectively, in $\{G\}$ at time t. The state estimate is represented by a multivariate Gaussian as

$$^G\mathbf{x}_t^R \sim \mathcal{N}\left(^G\bar{\mathbf{x}}_t^R, ^G\mathbf{\Sigma}_t^R \right), \tag{2}$$

where $^G\bar{\mathbf{x}}_t^R$ and $^G\mathbf{\Sigma}_t^R$ denote the mean vector and covariance matrix, respectively.

Likewise, the estimated pose of the perceiving ITS-S S at time t is also represented as a Gaussian variable:

$$^G\mathbf{x}_t^S = \begin{bmatrix} ^G x_t^S & ^G y_t^S & ^G \theta_t^S \end{bmatrix}^T \sim \mathcal{N}\left(^G\bar{\mathbf{x}}_t^S, ^G\mathbf{\Sigma}_t^S \right), \tag{3}$$

where $^G x_t^S$ and $^G y_t^S$ are the location of S, $^G \theta_t^S$ is its heading with respect to $\{G\}$ at time t, $^G\bar{\mathbf{x}}_t^S$ and $^G\mathbf{\Sigma}_t^S$ denote the mean vector and covariance matrix, respectively. In the special case of an IRSU serving as the perceiving ITS-S, the subscription t in (3) can be dropped as $^G x_t^S$ and $^G y_t^S$ are time invariant and $^G \theta_t^S$ is towards due east for an IRSU. Nevertheless, $^G\mathbf{\Sigma}_t^S$ is still applicable to an IRSU due to the presence of uncertainty in the surveying process.

The state vector of a perceived object p at time t is represented from the sensing ITS-S's perspective as

$$^S\mathbf{x}_t^p = \begin{bmatrix} ^S x_t^p & ^S y_t^p & ^S \theta_t^p \end{bmatrix}^T \sim \mathcal{N}\left(^S\bar{\mathbf{x}}_t^p, ^S\mathbf{\Sigma}_t^p \right), \tag{4}$$

where $^S x_t^p$ and $^S y_t^p$ are object coordinates and $^S \theta_t^p$ is heading within the sensing ITS-S's local frame at time t.

We obtain the transformed state of perceived object p with respect to $\{R\}$ by

$$^R\mathbf{x}_t^p = trans\left(^G\mathbf{x}_t^R, ^G\mathbf{x}_t^S, ^S\mathbf{x}_t^p \right). \tag{5}$$

Specifically, in $trans(\cdot)$, we have

$$
\begin{bmatrix} {}^{R}x_t^p \\ {}^{R}y_t^p \\ 1 \end{bmatrix} = T\left({}^{G}\mathbf{x}_t^R\right)^{-1} T\left({}^{G}\mathbf{x}_t^S\right) \begin{bmatrix} {}^{S}x_t^p \\ {}^{S}y_t^p \\ 1 \end{bmatrix},
$$
$$
{}^{R}\theta_t^p = {}^{S}\theta_t^p + {}^{G}\theta_t^S - {}^{G}\theta_t^R
$$

(6)

where $T\left(\begin{bmatrix} x & y & \theta \end{bmatrix}^T\right) = \begin{bmatrix} \cos\theta & -\sin\theta & x \\ \sin\theta & \cos\theta & y \\ 0 & 0 & 1 \end{bmatrix}$ is the homogeneous transformation matrix.

Given a Gaussian representation of the perceived object state probabilistic density function in the form of mean vector $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{\Sigma}$ with respect to $\{S\}$, the non-linear transformation to receiving ITS-S's frame $\{R\}$ with uncertainty is achieved through the UT process.

An augmented state vector is constructed by concatenating the state vector of the receiving and sensing ITS-Ss and that of the perceived object p before transformation. Its Gaussian estimate is written as

$$
\mathbf{x}_t^a \sim \mathcal{N}(\bar{\mathbf{x}}_t^a, \mathbf{\Sigma}_t^a),
$$

(7)

where $\bar{\mathbf{x}}_t^a = \left[\left({}^{G}\bar{\mathbf{x}}_t^R\right)^T \quad \left({}^{G}\bar{\mathbf{x}}_t^S\right)^T \quad \left({}^{S}\bar{\mathbf{x}}_t^p\right)^T \right]^T$, and $\mathbf{\Sigma}_t^a = \text{blkdiag}\left\{ {}^{G}\mathbf{\Sigma}_t^R, {}^{G}\mathbf{\Sigma}_t^S, {}^{S}\mathbf{\Sigma}_t^p \right\}$.

A collection of sigma points $\{\boldsymbol{\mathcal{X}}_i, w_i^m, w_i^c\}_{i=0}^{2d}$ are obtained based on the augmented state estimate prior to the transformation:

$$
\begin{aligned}
\boldsymbol{\mathcal{X}}_0 &= \bar{\mathbf{x}}_t^a \\
\boldsymbol{\mathcal{X}}_i &= \bar{\mathbf{x}}_t^a + \left(\sqrt{(d+\lambda)\mathbf{\Sigma}_t^a}\right)_i \quad \text{for } i = 1, \cdots, d \\
\boldsymbol{\mathcal{X}}_i &= \bar{\mathbf{x}}_t^a - \left(\sqrt{(d+\lambda)\mathbf{\Sigma}_t^a}\right)_i \quad \text{for } i = d+1, \cdots, 2d \\
w_0^m &= \frac{\lambda}{d+\lambda} \\
w_0^c &= \frac{\lambda}{d+\lambda} + \left(1 - \alpha^2 + \beta\right) \\
w_i^m &= w_i^c = \frac{1}{2(d+\lambda)} \quad \text{for } i = 1, \cdots, 2d
\end{aligned}
$$

(8)

where $\lambda = \alpha^2(d+\kappa) - d$, $d = dim(\mathbf{x})$ is the dimension of state $\mathbf{x}$, scaling parameters $\kappa \geq 0$, $\alpha \in (0,1]$, and $\beta = 2$ is optimal for Gaussian distributions, $\left(\sqrt{\mathbf{\Sigma}_t^a}\right)_i$ is to obtain the ith column of the matrix square root $\mathbf{R} = \sqrt{\mathbf{\Sigma}_t^a}$, which can be computed by Cholesky decomposition such that we have $\mathbf{\Sigma}_t^a = \mathbf{R}\mathbf{R}^T$. Note that each sigma point can be decomposed in accordance with the concatenation sequence in (7), i.e., $\boldsymbol{\mathcal{X}}_i = \left[\left(\mathcal{X}_i^R\right)^T \quad \left(\mathcal{X}_i^S\right)^T \quad \left(\mathcal{X}_i^p\right)^T \right]$.

This is followed by passing each sigma point through the frame transformation function $trans(\cdot)$ in (5), which yields a set of transformed sigma points. For $i = 0, 1, \cdots, 2d$,

$$
\boldsymbol{\mathcal{Y}}_i = trans\left(\mathcal{X}_i^R, \mathcal{X}_i^S, \mathcal{X}_i^p\right).
$$

(9)

Lastly, the transformed state of perceived object p in $\{R\}$ is recovered by

$$
{}^{R}\bar{\mathbf{x}}_{t}^{p} = \sum_{i=0}^{2d} w_i^m \boldsymbol{\mathcal{Y}}_i
$$

$$
{}^{R}\boldsymbol{\Sigma}_{t}^{p} = \sum_{i=0}^{2d} w_i^c \left(\boldsymbol{\mathcal{Y}}_i - {}^{R}\bar{\mathbf{x}}_{t}^{p} \right) \left(\boldsymbol{\mathcal{Y}}_i - {}^{R}\bar{\mathbf{x}}_{t}^{p} \right)^T .
$$

(10)

4.2. Numerical Simulation

The coordinate transformation of perceived objects in ETSI CPMs is validated through numerical simulation. As illustrated in Figure 9, the simulation is setup with two ITS-Ss acting as a publishing/sensing ITS-S and a receiving ITS-S of CPMs, respectively. The sensing ITS-S is static at a position of (100 m, 100 m) on the map, facing the east direction. It is perceiving road users within its local frame with some level of uncertainty in its sensory measurements. For the convenience of analysis, a group of 20 static road users are positioned in front of the sensing ITS-S in a line with identical perception uncertainty parameters, as demonstrated in Figure 10. The receiving ITS-S in the simulation is a CAV moving at 2 m/s from its initial position at (0 m, 75 m) with the same heading of the sensing ITS-S, which can be either an IRSU (V2I case) or another CAV (V2V case). Both V2V and V2I scenarios are considered in the simulation for a comparison. Both ITS-Ss are assumed to contain uncertainty in their self-localisation.
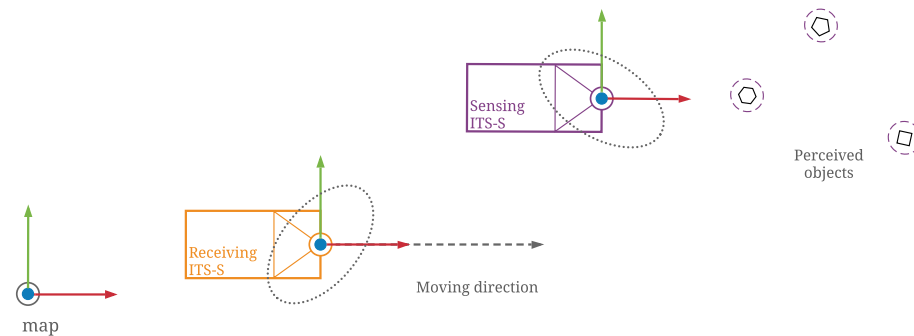

Figure 9. Setup of the sensing and receiving ITS-Ss in the simulation.

Within the sensing ITS-S, the perceived objects information and egocentric pose estimate are encoded into CPMs in the form of binary payloads and published at 10 Hz. The receiving CAV, as it moves on the map, decodes the CPMs received and transforms the perceived objects into its local coordinate system. Through coordinate transformation, the uncertainty in the transformed perceived information is a combined result of the sensing uncertainty and the self-positioning uncertainties of both ITS-Ss.
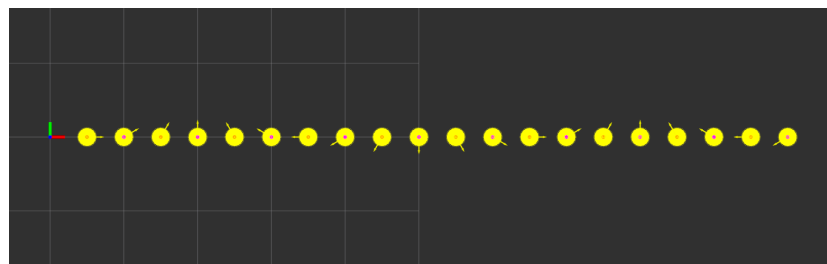

Figure 10. Perceived objects within the local frame of the sensing ITS-S. The objects are a mixture of static pedestrians and vehicles that are placed in a line along with the x direction of the sensing ITS-S. They are perceived with the same location and heading uncertainties in the simulation. The 95% confidence ellipse for the 2D position estimate of each perceived object is shown in yellow. Each grid in every figure represents an area of 10×10 m.

Other parameters adopted in the simulation can be found in Table 1. The covariance matrix in the self-localisation estimate of the receiving CAV is denoted as ${}^G\Sigma_t^R = \text{blkdiag}\left\{ (\sigma_{pos})^2, (\sigma_{pos})^2, (\sigma_\theta)^2 \right\}$, where σ_{pos} and σ_θ denote the standard deviation of position and heading estimate, respectively. The simulation consists of two tests for evaluating the effect of different σ_θ and σ_{pos} values in coordinate transformation. Specifically, σ_θ is varied in *Test 1* with σ_{pos} kept constant, while a range of σ_{pos} are tested for the CAV in *Test 2*. Each of the tests further includes both V2V and V2I cases. In the V2V case, the sensing CAV is setup with the same covariance matrix as for the receiving CAV, i.e., ${}^G\Sigma_t^S = {}^G\Sigma_t^R$. In the V2I case, where the sensing ITS-S is an IRSU, the localisation noise is assumed small yet non-zero. Please note that standard deviations as low as 0.05° and 0.005 m are tested in the simulation as they are roughly the minimum values supported in CPMs for representing the uncertainty of heading and position estimates, respectively, for the originating ITS-S. This is mainly due to discretisation of confidence levels of the corresponding DEs defined in CPMs. Likewise, 0.005 m is adopted as the standard deviation for the position estimate of the IRSU, and the heading estimate standard deviation is set to a close-to-zero value ϵ for preventing numerical errors in the transformation.

Table 1. Simulation Parameters.

		Standard Deviation in State Estimate	
		Position	**Heading**
Perceived Objects		0.5 m	6°
Sensing ITS-S	IRSU	0.005 m	ϵ
	CAV	Same as Receiving CAV	
Receiving CAV	Test 1	0.25 m	0.05°, 0.5°, 1.0°, 1.5°, 2.0°
	Test 2	0.005 m, 0.25 m, 0.5 m, 0.75 m, 1.0 m	0.5°

Figure 11 reveals the result of *Test 1*, i.e., the effect of different uncertainty levels in heading estimates of the CAV(s), along with the movement of the receiving CAV. Figure 12 depicts the result of *Test 2*, which is with different uncertainty in position estimates of the CAV(s). It can be seen from both figures that the confidence ellipses of those perceived objects, after being transformed into the local coordinate system of the receiving CAV, are bloated and distorted to different extents, depending on their relative poses with respect to both ITS-Ss. Please note that, in both figures, the first set of ellipses on the left in each subfigure represents the uncertainties in the transformed position estimate of the sensing ITS, while the receiving CAV is located at the origin, facing the *x*-direction.

In Figure 11, the bloating effect is found to be sensitive to the uncertainty contained in the heading estimate of both ITS-Ss. In addition, the bloated ellipses are slanted along the tangential directions of both ITS-Ss, and the bloating is found to be more serious for ellipses that are further from both ITS-Ss. In theory, these thin and long confidence ellipses caused by the heading uncertainty are expected to be banana shaped due to the nonlinear nature of the coordinate transformation. The result therefore indicates that, with a larger uncertainty level in heading estimate and for further perceived objects, the Gaussian assumption starts to show its limitation for representing the perceived object estimates. One can choose to use Gaussian mixture or non-Gaussian representations in the transformation to alleviate the issue. As a comparison, the bloating effect as a result of the uncertainty in position estimates of the ITS-Ss is found to be less correlated to the relative distance, and the bloating happens in both *x*- and *y*-directions, as shown in Figure 12. This shows that reducing the estimate uncertainty of heading in ITS-S self-localisation is more effective than that of position, in suppressing the bloating of uncertainty during the coordinate transformation of perceived objects information.

In addition, the two scenarios of an IRSU and a CAV acting as the sensing ITS-S are compared in each of Figures 11 and 12. When the sensing ITS-S is an IRSU, it can be seen

that the transformed confidence ellipses are still slanted but in general less bloated due to very low uncertainty in the IRSU's position.

It should be noted that the perceived objects and the sensing ITS-S are set to be static in the simulation just for the convenience of analysis. In fact, the bloating of the position uncertainty in the coordinate transformation is independent of moving speeds of the participants in the simulation, mainly because only pose estimates of the perceived objects and ITS-Ss are involved in the transformation formulated in Section 4.1. Their kinematic states would be considered when the transformed perceived object states need to be predicted to a different time within the receiving ITS-S.
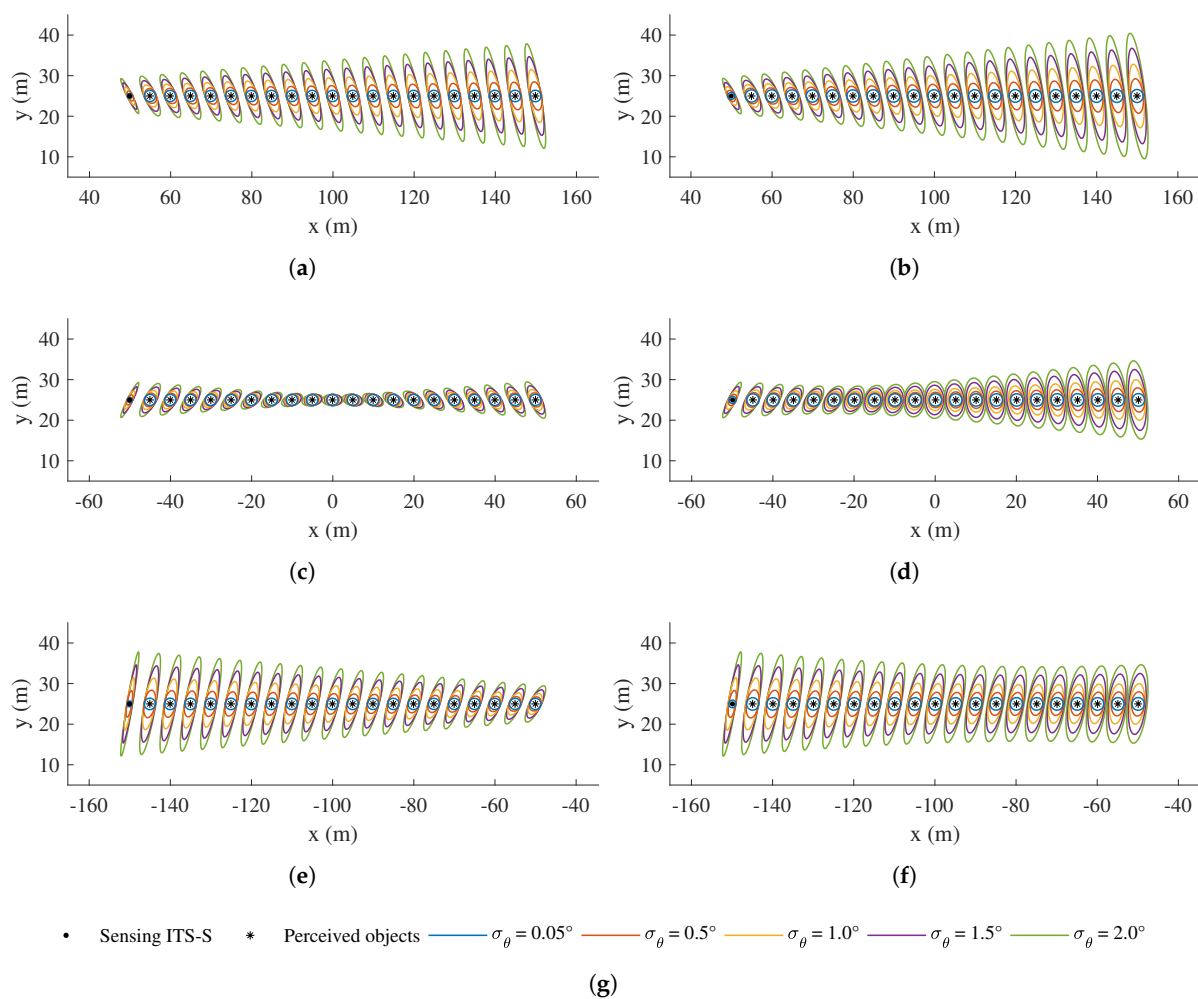

Figure 11. Confidence ellipses of perceived objects transformed to the local frame of the receiving CAV in *Test 1*, where different heading estimate standard deviations are tested for CAV(s). As the receiving CAV moves, (**a**,**c**,**e**) show the results at different relative positions between IRSU and the receiving CAV, while (**b**,**d**,**f**) present the results for the V2V case. In each subfigure, the first object on the left represents the position of the sensing ITS (visualised as a black dot) after being transformed to the receiving CAV's local frame, while its perceived objects are denoted by black star markers. The receiving CAV sits at the origin in each subfigure. Each ellipse represents 95% of position estimate confidence.

Figure 12. Confidence ellipses of perceived objects transformed to the local frame of the receiving CAV in *Test 2*, with different standard deviations in position estimate of CAV(s). (**a**,**c**,**e**) demonstrate the results for the V2I case, while (**b**,**d**,**f**) illustrate the results for the V2V case at different relative positions. In each subfigure, the first object on the left represents the position of the sensing ITS (visualised as a black dot) after being transformed to the receiving CAV's local frame, while its perceived objects are denoted by black star markers. The receiving CAV sits at the origin in each subfigure. Each ellipse represents 95% of position estimate confidence.
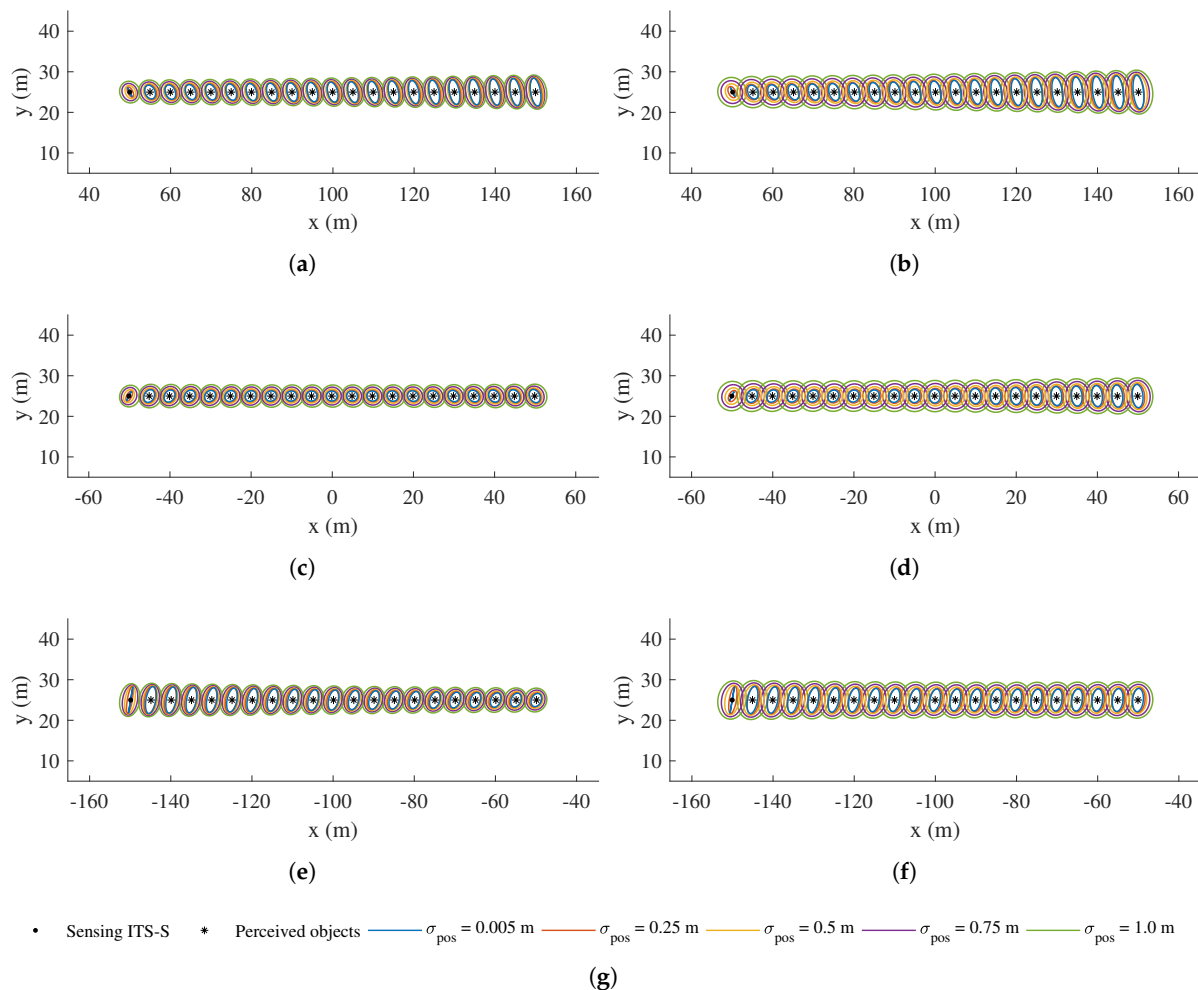
5. Demonstrations

The demonstrations consist of three experiments conducted in different traffic environments, namely, an urban traffic environment, the CARLA simulator, and a lab traffic environment. The configurations of these experiments are compared in Table 2. The developed IRSU and CAV platforms, as previously described in Section 3, were used in the real-world experiments. The CAV was however configured differently according to its operation mode and safety requirements in the two real traffic environments. In addition, the CAV onboard perception capability was disabled in the experiments, and the lidar was used in the last experiment only for automatic emergency braking, which acts like a virtual bumper, in autonomous driving mode. In the experiment setup in the CARLA simulator, realistic sensory data were obtained from simulated sensors, and the same perception data processing pipeline presented in Figure 2 was employed in the simulated IRSU.

Table 2. An overview of experiment configurations.

Configuration	Experiment No.		
	1	2	3
Environment	Real Urban Traffic Env.	CARLA Simulator	Real Lab Traffic Env.
IRSU Perception Sensors	Real Lidar + Cameras	Simulated Lidar + Cameras	Real Lidar + Cameras
Pedestrian(s) Activity	Walking + Standing	Walking	Running
V2I Communication	Cohda Wireless MK5	Simulated	Cohda Wireless MK5
CAV Driving Mode	Manual	Autonomous	Autonomous
CAV Perception Sensors	Disabled	Disabled	Real Lidar [1]
CAV Self-Localisation [2]	Lidar Features + GNSS	Simulated GNSS	Lidar Features + GNSS
Multi-Target Tracking [3]	GMPHD Tracker	GMPHD Tracker	GMPHD Tracker
CAV Navigation	N.A.	Hybrid A* Path Planning on *Lanelet2* Map	

[1] Used only for the virtual bumper subsystem as part of safety measures. [2] Onboard IMU and wheel encoders are used for odometry. [3] Running in both IRSU and CAV.

5.1. Experiment in an Urban Traffic Environment

The experiment was conducted in a real urban traffic environment next to the ACFR building located at Chippendale, Sydney. In the experiment, the IRSU was deployed near the intersection of Abercrombie St and Little Queen St, monitoring the traffic activity at the intersection and broadcasting the perception information in the form of ETSI CPMs in real time, as depicted in Figure 13. The intersection was chosen for the experiment mainly because, turning from Abercrombie St to Little Queen St, which is a side road, requires extra attention from vehicle drivers due to the lack of traffic control and poor visibility of road users behind the corner building. The main purpose of the experiment is to demonstrate the improved VRU awareness and thus road safety for a vehicle ITS-S when it is able to learn pedestrian activity in blind spots through the CP information provided by IRSU.

Instead of autonomous driving, the CAV in the experiment was driven manually for ensuring public road safety in the real traffic scenario. The CAV platform therefore practically functioned as a CV in the experiment and is referred to as the CV in the remainder of the section. While the CV was performing self-localisation within the experiment area, it was able to transform the perceived objects contained in CPMs into its local coordinate system with uncertainty and have the road users tracked in real time. The CV identifies itself and its driver based on the position of the transformed perceived objects in its local frame to avoid self-tracking.

In the experiment, as labelled in Figure 13, the CV first made a right turn from Meagher St to Abercrombie St, drove for about 60 m before turning to Little Queen St. As illustrated in Figure 14, the CV learnt the ongoing traffic activity at the intersection with the IRSU deployed far beyond the reach of its onboard perception sensors. This illustrates the extended sensing range for smart vehicles through the CP service. At a later time shown in Figure 15a,b, when the CV was about to make a right turn to Little Queen St, the vehicle could "see" a visually occluded pedestrian behind a building with the information coming through the IRSU. Please note that it is achieved seconds before the pedestrian could actually be visually picked up from the CV perspective in Figure 15c,d, which is crucial for safety in both manual driving and autonomous driving scenarios. This is considered another benefit of the CP service demonstrated in the experiment.

(a)　　　　　　　　　　　　　　　　　　　　**(b)**

Figure 13. Experiment setup at the intersection of Abercrombie St and Little Queen St. In (**a**), the IRSU was setup facing south of Abercrombie St. The car traffic flow was going north on this one-way street. The trajectory of the CV is denoted by a dashed red line, which indicates that the CV turned from Meagher St to Abercrombie St, followed by Little Queen St. In (**b**), CPMs were transmitted from the IRSU to the moving CV through Cohda Wireless MK5 devices.
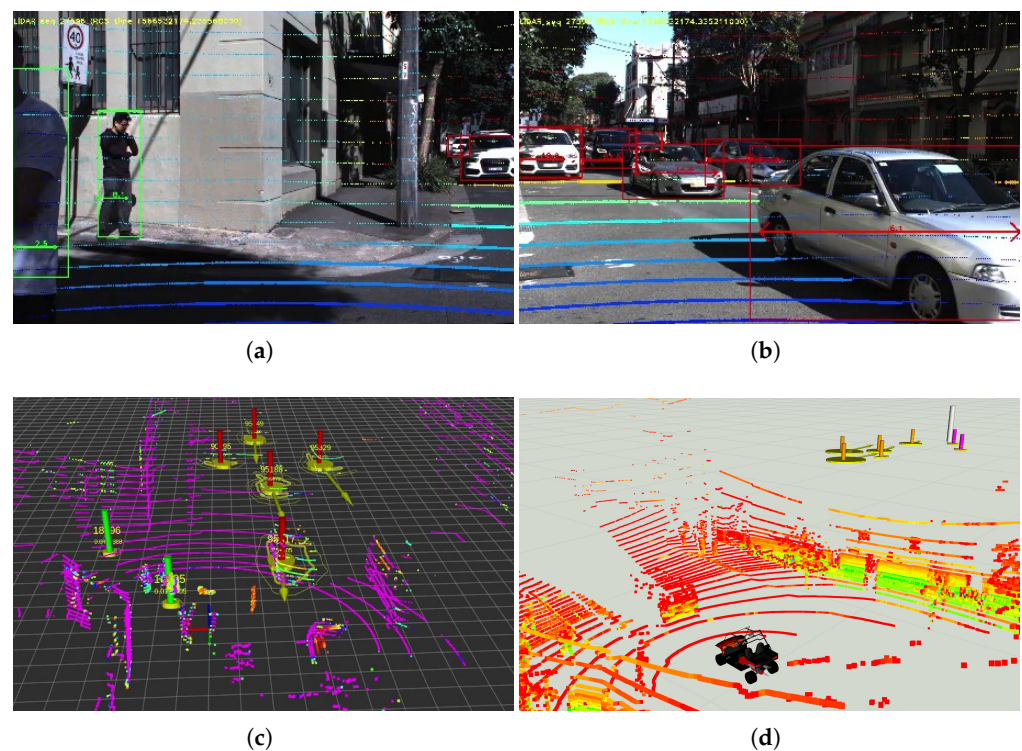

(a)　　　　　　　　　　　　　　　　　　　　**(b)**

(c)　　　　　　　　　　　　　　　　　　　　**(d)**

Figure 14. The road user detection and tracking in the IRSU and CV in the Abercrombie St experiment. (**a**,**b**) show the image frames from Cam #1 and #2, respectively, overlaid with pedestrians and vehicles bounding boxes and projected lidar point cloud in the IRSU. The detection results from the dual cameras and the lidar were then fused within the IRSU for tracking in 3D space, as visualised in (**c**). The detection results were encoded into CPMs and transmitted to the CV, where they were decoded, transformed, and used in the CV's local frame. In the particular traffic scene shown in (**c**), the generated CPM contains one *Sensor Information Container* and seven *Perceived Object Containers*, two for pedestrians and five for vehicles. (**d**) illustrates the tracking results of the same group of road users within the CV at the same time. Due to the change of visualisation colours in IRSU and CV, the vehicles in orange pillars in (**d**) correspond to red ones in (**c**), while the pedestrians in magenta pillars correspond to green ones in (**c**).
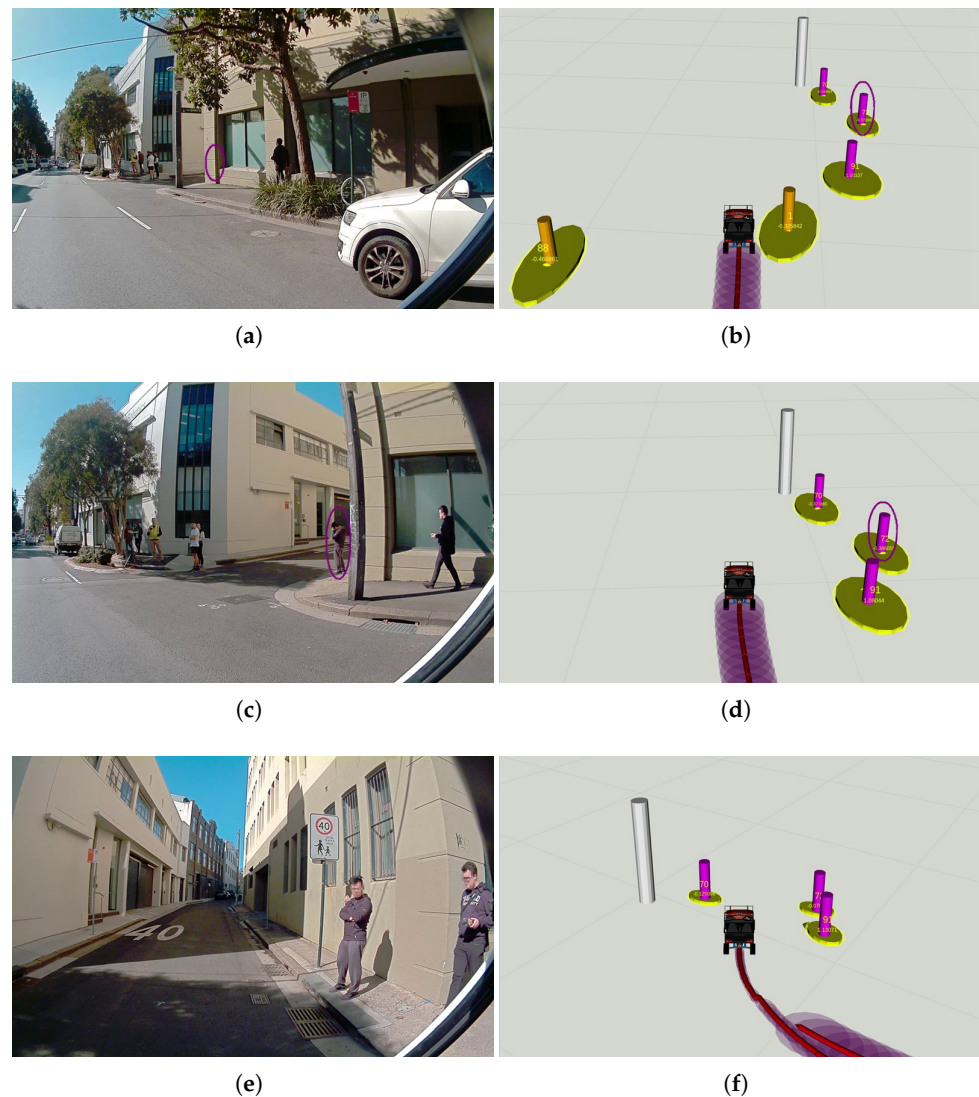
Figure 15. The CV became aware of a hidden pedestrian before turning from the main street to Little Queen St. The left column shows the images captured by the front-right camera on the CV at different times, the right column illustrates the corresponding road user tracking results within the CV. In addition, in the right column, the tracked pedestrians are represented as magenta pillars with their 2-sigma estimation confidence areas denoted as yellow ellipses, the IRSU is visualised in the CV as a tall white pillar. In (**a**,**b**), the CV could "see" through the building a visually occluded pedestrian (labelled with purple circles) around the corner with the help of CPMs received from the IRSU. This is seconds before the CV drove closer and could observe by itself the previously occluded pedestrian on Little Queen St, as depicted in (**c**,**d**). The localisation covariance is presented as purple disks along the vehicle trajectory in every right column figure. It can be seen in (**e**,**f**) that the uncertainty in road user tracking reduced after the CV received a correction in self-localisation.

In addition, the CV localisation quality varied depending on the lidar feature quality at different locations of the experiment environment. It is illustrated by comparing Figure 15d,f that the refinement in egocentric pose estimate as a result of a correction in the self-localisation shown in the latter contributes to the improvement of tracking accuracy, which is consistent with the finding from the Section 4.2 that the level of ITS-S localisation estimate uncertainty correlates to the perceived objects detection uncertainty after coordinate transformation and thus tracking uncertainty.

5.2. Experiment in the CARLA Simulator

The experiment demonstrates the autonomous operation of a CAV using the perception information received from an IRSU in CARLA simulator as the only source for pedestrian estimation. CARLA [56] is an open source game engine based simulator, which supports flexible configurations of sensors, road users, and urban traffic environments. One of the highlights of the CARLA simulator is the various and realistic sensory data it provides, including RGB images, lidar point clouds, IMU, and GNSS measurements. Another highlight is its ROS integration with the *ros-bridge* package. It is considered a safe, realistic, and repeatable environment for experimenting with autonomous driving systems, in particular, for conducting CAV operations that have safety concerns to test in the real world.

The experiment is setup with the IRSU detecting multiple walking pedestrians within its sensor range. The information of perceived pedestrians is then broadcast in the form of ETSI CPMs and is received by the CAV. The information is then transformed into its local frame of reference, and taken into account as the only perception data in the autonomous navigation of the CAV. The experiment is to demonstrate how a CAV interacts with VRU at non-designated crossing areas with the perception information from an IRSU. It is strongly recommended to conduct these experiments with higher risk CAV maneuvers in a simulator before performing them in the real world.

The experiment was setup in the *Town01* map of CARLA 0.9.8 simulator with participants labelled in Figure 16. These include an IRSU deployed static next to a straight road, a CAV navigating autonomously on the same road towards east, and two walking pedestrians. One of the pedestrians was crossing the road, while the other is walking along the road in the west direction. A *Lanelet2* map was built for the experiment environment in the simulator, which is essential for the CAV to stay on the road and plan drivable paths for autonomous navigation. The map includes the lane markings and the network of drivable regions as per right hand traffic rules. Accordingly, the CAV respects right hand traffic rules in the demonstration. On the road where the CAV is navigating, the CAV has right of way over pedestrians and is allowed to cross the broken dividing line to overtake other road users when it is safe to do so.

Figure 16. Experiment setup in the CARLA simulator. The CAV, a red car, moved to the east direction towards two walking pedestrians and an IRSU deployed on the roadside. The IRSU was spawn in CARLA as a mini car parked on the roadside.

The IRSU in the simulator is configured as close as possible to the real IRSU developed, which is equipped with two cameras and a 16-beam lidar, as revealed in Figure 16. It is visualised as a mini car parked on the roadside in the simulator. The same suite of sensory

data processing algorithms for the real IRSU is adopted in the simulated IRSU. The road user perception information is encoded to ETSI CPMs in the form of binary payloads and broadcast by the IRSU. The CAV employed in the experiment is as also shown in Figure 16, which is able to position itself within the map using a combination of measurements from GNSS, IMU, and wheel encoder. The self-localization, however, is not assumed to be perfect but with noise added to the GNSS, IMU, and encoder measurements. An unscented Kalman filter (UKF) is used to achieve sub-meter accuracy. In the experiment, the CAV is also set up to receive ETSI CPMs from the IRSU. It should be noted that the CAV does not have any onboard sensors for detecting road users; instead, it becomes aware of the road users and infer their states purely relying on information received from the IRSU. The CAV navigates at a speed up to 18 km/h in the simulator.

In the experiment, one of the pedestrians was crossing the road at a non-designated crossing area while the other is walking along the road in the opposite direction to the CAV. Figure 17 demonstrates the road user detection and tracking in the IRSU at the beginning of the experiment. Specifically, Figure 17a shows the detection of the two pedestrians and the CAV by fusing the lidar point cloud and image frame from Cam #1 of the IRSU. The tracking of the same three road users in 3D space is shown in Figure 17b. The tracking of the walking pedestrians in the CAV is illustrated in Figure 18. Figure 18c depicts the planned path in the CAV to navigate around the pedestrians when approaching them. The tracked pedestrians are visualised with 95% confidence ellipses and arrows indicating their moving direction. When the grid-based cost map is updated in the hybrid A* path planner in the CAV, the current and future estimates of the pedestrian states account for a higher cost (darker colour) in the grid such that the path planner would consider them when searching for an optimal and kinematically feasible path to its goal. Please note that the physical dimension of the CAV has been considered by bloating the confidence ellipses in the cost map.

In Figure 18, it is shown that the CAV responded to the pedestrian walking along the road by not altering its planned path as expected. Although the pedestrian was walking on the road, it was blocking only the opposite lane of the road. In Figure 18a, the CAV became aware of the crossing pedestrian even when it was approximately 35 m away, with the information coming through the IRSU. This was not taken into account by the path planning since the CAV was still far from the pedestrian. As the CAV drove closer, it reacted to the event by planning a path that can safely navigate around the pedestrian while the pedestrian was crossing the road at a non-designated crossing area. As revealed in Figure 19, when the CAV was close to the pedestrian, it crossed the broken dividing lane to avoid the crossing pedestrian from behind. Following the avoidance maneuver, the CAV switched back to the right lane.
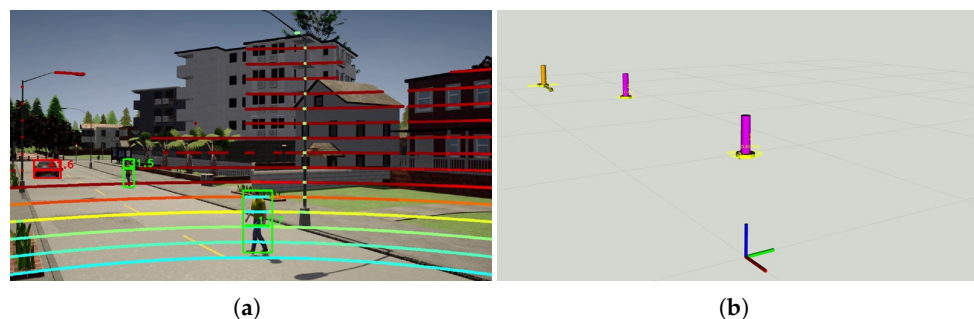

(a)　　　　　　　　　　　　　　　　　　　　　　(b)

Figure 17. Pedestrian detection and tracking in the IRSU in the CARLA simulator. Despite dual cameras enabled for detecting road users in the IRSU, the Cam #1 captured all the road users at the particular moment in the experiment. Therefore, only the image frame from that camera is shown in (**a**). In (**b**), where the road users are tracked in 3D space, the CAV itself is labelled as an orange pillar with a confidence ellipse and an arrow indicating its moving direction. The two pedestrians are denoted by magenta pillars.
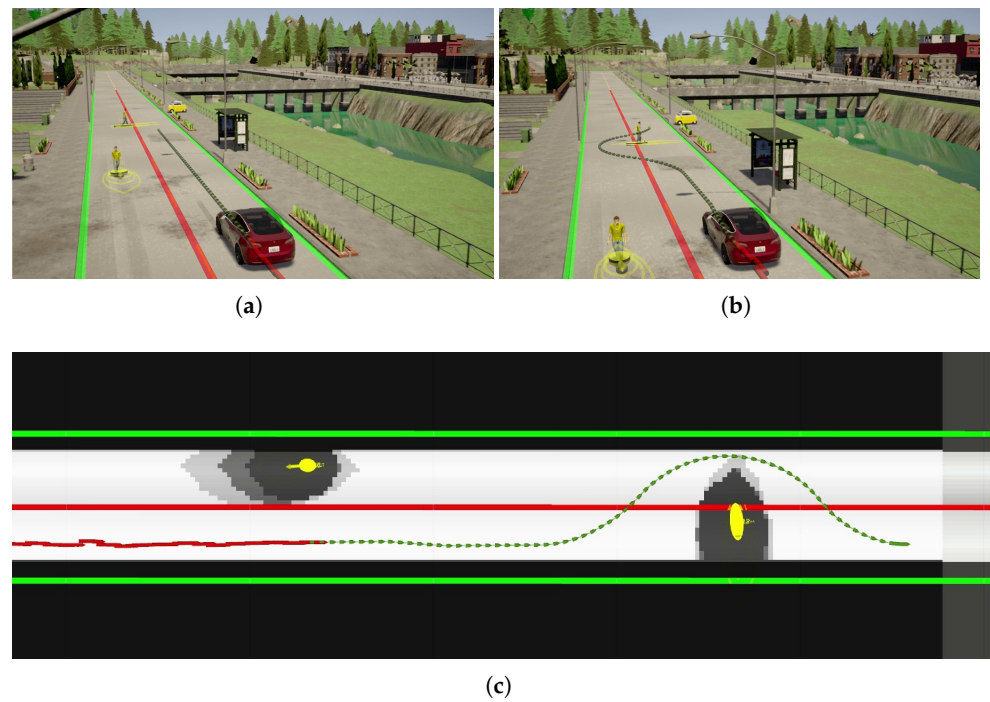
(**a**)

(**b**)

(**c**)

Figure 18. The pedestrian tracking in the CAV relying only on CPMs from the IRSU in the CARLA simulator. (**a**) and (**b**) show the two walking pedestrians and the planned path in the CAV at different time instances. In (**c**), the grid-based cost map maintained by the path planner in the CAV is presented with white as free space and black as occupied areas, the tracked pedestrians are visualised with 95% confidence ellipses and arrows indicating their moving direction, and the odometry trajectory is shown as a red line, and the planned path is presented as a dotted green line.
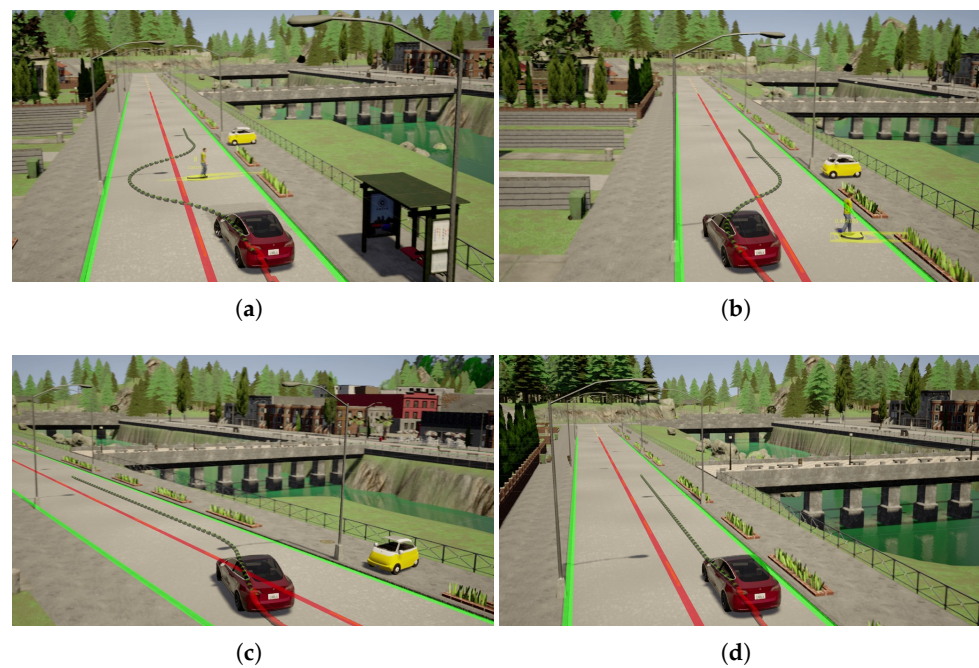

(**a**)

(**b**)

(**c**)

(**d**)

Figure 19. The CAV safely navigated around the crossing pedestrian autonomously based on the CPMs from the IRSU in the CARLA simulator. (**a**–**d**) sequentially show the complete pedestrian avoidance maneuver of the CAV.

5.3. Experiment in a Lab Environment

The last experiment presented in the paper investigates the CAV interacting with a pedestrian running towards a pedestrian crossing area. It was set up in a lab traffic environment with a 55 m stretch of straight road, a virtual pedestrian crossing area, and the IRSU deployed next to the crossing, as illustrated in Figure 20. A *Lanelet2* map was built for the environment and the CAV was able to localise itself within the pre-built map of lidar features, such as poles and building corners. The pedestrian crossing area was encoded in the *Lanelet2* map, which is stored locally in the CAV. The expected behaviour of the CAV is to drive through the crossing or stop before the crossing to give way depending on whether there is any pedestrian crossing activity detected by the IRSU.

Figure 21 shows the moment the pedestrian started running at the beginning of the experiment. The IRSU captured the road users in the scene, which include the pedestrian and the CAV in 3D space. The CAV then received from the IRSU the information about the presence of a pedestrian running towards the crossing, as depicted in Figure 21d. When approaching him, the CAV kept tracking and predicting the pedestrian state based on the perception information from the IRSU. A constant velocity kinematic model was adopted for predicting the future state of the pedestrian. As shown in Figure 22b, the pedestrian was predicted by the CAV to occupy the crossing area, although he had not physically stepped onto the crossing yet. In the meantime, the CAV started to brake preemptively for the predicted crossing activity. The crossing area became a high cost region in the cost map within the path planner when the pedestrian was crossing. As a result, there was no feasible path found to drive through the crossing area for the CAV. It eventually stopped before the crossing line for the pedestrian, waiting until he finished crossing, as demonstrated in Figure 22d. When the pedestrian crossing area was clear a moment later, the CAV re-planned a path to the original goal and started to drive through the crossing at a slightly slower speed due to the road speed rule encoded in the map, as depicted in Figure 22f. As depicted in Figure 23, where the CAV speed measured by the wheel encoders is presented, the CAV managed to gain extra 1.5 s of response time to the pedestrian crossing activity. The braking behaviour was well planned and executed within the CAV for a safe and smooth deceleration.

Figure 20. Experiment setup in a lab environment. It is shown in the *Lanelet2* map and lidar feature map built for the experiment site, which contains the road boundaries and lanes information. The light blue dots distributed in the bottom right corner represent the locations of observed features in the environment (trees, poles, corners of structures). The designated pedestrian crossing area is labelled as an orange rectangle.
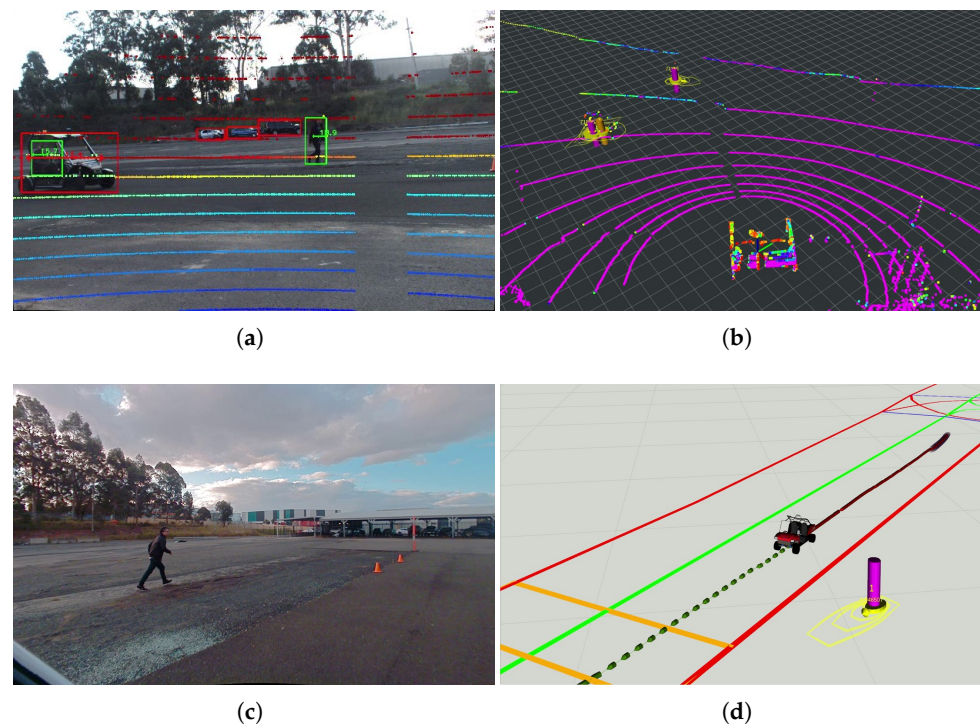
(**a**) (**b**)

(**c**) (**d**)

Figure 21. The CAV started tracking a pedestrian running toward a crossing line of a marked pedestrian crossing in a lab environment based on ETSI CPMs from an IRSU. (**a**) shows the image frame from Cam #1 of the IRSU with projected lidar point cloud. (**b**,**d**) illustrate the tracking of the road users in 3D space in the IRSU and the CAV, respectively, and (**c**) shows the image captured by the front-left camera on the CAV. The pedestrian and the safety driver of the CAV are tracked and denoted by magenta pillars with confidence ellipse and arrows indicating their moving direction, while the vehicle is labelled as an orange pillar in the IRSU's tracker. Only the running pedestrian was tracked in the GMPHD tracker running within the CAV, as depicted in (**d**), as the CAV itself and its safety driver are identified and removed from tracking based on the location of the received perceived objects after transformed into its local frame. The *Lanelet2* map is also visualised in (**d**), where the pedestrian crossing is denoted by an orange rectangle. Furthermore, the odometry trajectory of the CAV is shown as a red line, and its planned path is denoted as a dotted green line in (**d**).

The CAV onboard perception sensors were not used for pedestrian detection throughout the experiment. However, there are a few measures in place to ensure safety when conducting autonomous navigation related activities in real-world experiments. First of all, the experiment was conducted in a closed, controlled environment and the CAV speed is limited to 11 km/h when in autonomous mode. In addition, there is always a safety driver sitting in the CAV closely monitoring any hazards and has the ability to trigger an e-stop of the CAV in case of an emergency. Furthermore, the virtual bumper subsystem based on lidar points obstacle detection is enabled on the CAV, which brings the vehicle to an e-stop if any obstacle breaches any of the pre-set distance thresholds to the vehicle. However, the virtual bumper treats every surrounding object as static obstacles and therefore did not take part in the decision-making process of the CAV in Figure 22b, where it decided to give way based on the predicted future state of the pedestrian.
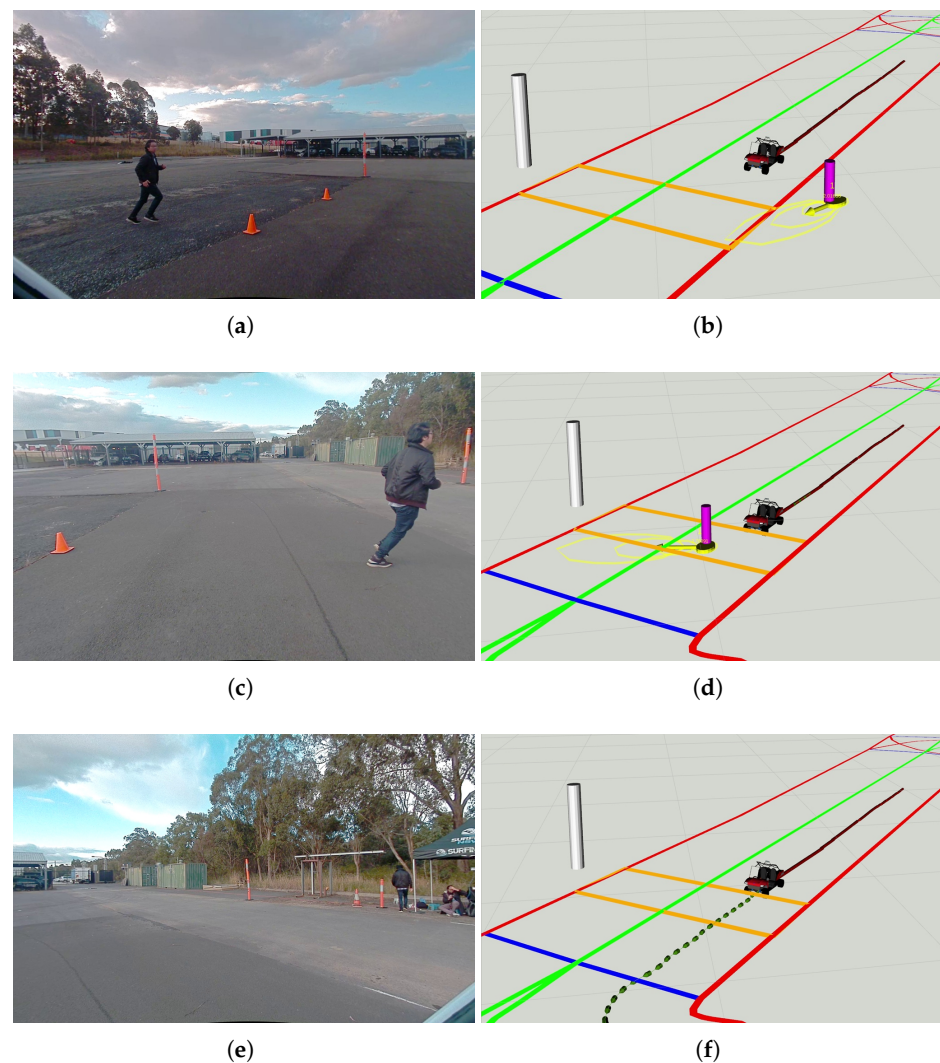
(**a**)　　　　　　　　　　(**b**)

(**c**)　　　　　　　　　　(**d**)

(**e**)　　　　　　　　　　(**f**)

Figure 22. The CAV gave way to a running pedestrian at the designated crossing area in a lab environment. Based on the ETSI CPMs broadcast by the IRSU, the CAV predicted the state of the running pedestrian up to 1.5 s into the future. The IRSU is visualised as a tall white pillar in the CAV in (**b**,**d**,**f**). It is seen in (**b**) that the predicted states of the pedestrian, visualised as yellow ellipses, intersected with the orange crossing area on the *Lanelet2* map. The CAV stopped for the pedestrian in front of the crossing in (**d**). The CAV then planned a new path forward after the pedestrian finished crossing in (**f**). In (**a**,**c**,**e**), image frames from front-left, front-centre, front-right cameras on the CAV are presented, respectively.
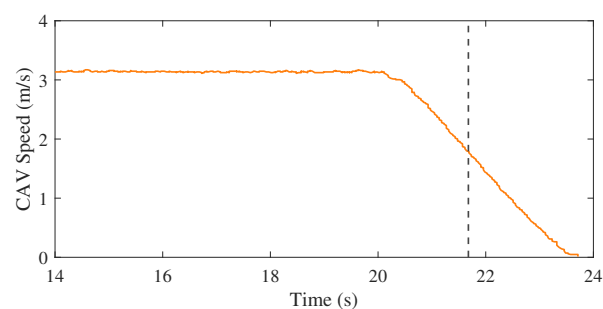

Figure 23. The CAV speed between time 14 to 24 s in the experiment. It is shown that the CAV started to decelerate about 1.5 s prior to the moment the pedestrian physically stepped onto the crossing, as indicated by the black dashed line.

6. Conclusions

The paper is mainly focused on safety and robustness implications the CP can bring to the operations of CAVs. It presents the IRSU and CAV platforms developed by the joint research team from the ACFR and Cohda Wireless over the last two years. The contributions of the paper include (1) investigating and demonstrating through three representative experiments the safety benefits associated with the use of CP service for CV and CAV operations alongside VRU in different traffic contexts, and (2) the formulation and analysis of coordinate transformation of perceived objects information with uncertainty.

In the first experiment conducted on a public road, the CV was able to track a pedestrian visually obstructed by a building with CP information from an IRSU deployed on the road side. This was achieved seconds before its local perception sensors or the driver could possibly see the same pedestrian around the corner, providing extra time for the driver or the navigation stack to react to a safety hazard. The second experiment conducted in the CARLA simulator shows the autonomy of a CAV using the CP service. The experiment highlights the autonomous navigation of the CAV and its safe interaction with walking pedestrians, purely responding to the perception information provided by the IRSU. The last experiment demonstrated the expected behaviour of the CAV when interacting with a pedestrian rushing towards the designated crossing area in a lab environment. The CAV managed to take preemptive action, that is, braking and stopping before the crossing area for the pedestrian based on the kinematic prediction of the pedestrian. The pedestrian tracking, prediction, path planning, and decision-making in the CAV in the last experiment were operating based on the perception information received from the IRSU.

We believe that these demonstrations can assist engineers and researchers in the relevant fields in achieving a better understanding of the safety implications the CP is bringing to the current and future transportation systems. The CP enables the smart vehicles to break the physical and practical limitations of onboard perception sensors, and, in the meantime, to embrace improved perception quality and robustness along with other expected benefits from the CP service and V2X communication. As importantly, the CP can also reduce the reliance on the vehicle's local perception information, thereby lowering the requirement and cost for onboard sensing systems. Furthermore, it is demonstrated that, when properly used, IRSU perception data can be used as another reliable source of information to add additional robustness and integrity to autonomous operations. These potentially trigger more in-depth discussions in the research community and industry related to the future development directions of intelligent vehicles and other C-ITS services and applications, such as maneuver intention sharing and cooperative driving.

Author Contributions: Conceptualization, M.S., Y.F.W., P.A. and E.N.; methodology, M.S., K.N. and Y.F.W.; software, M.S. and K.N.; validation, M.S., K.N. and S.W.; formal analysis, M.S. and K.N.; investigation, M.S., K.N., Y.F.W. and M.K.; resources, Y.F.W., M.K., P.A. and E.N.; data curation, M.S., K.N. and S.W.; writing—original draft preparation, M.S.; writing—review and editing, K.N., Y.F.W., S.W. and M.K.; visualization, M.S. and K.N.; supervision, P.A. and E.N.; project administration, P.A. and E.N.; funding acquisition, P.A. and E.N. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACFR	Australian Centre for Field Robotics
ADAS	Advanced Driving Assistance System
AV	Autonomous Vehicle
CAM	Cooperative Awareness Message
CAV	Connected and Automated Vehicle
C-ITS	Cooperative Intelligent Transportation System
CP	Cooperative/Collective Perception
CPM	Collective Perception Message or Cooperative Perception Message
CSM	Cooperative Sensing Message
CV	Connected Vehicle
DE	Data Element
DF	Data Frames
EPM	Environmental Perception Message
ETSI	European Telecommunications Standard Institute
FOV	Field of View
GMPHD	Gaussian Mixture Probability Hypothesis Density
GMSL	Gigabit Multimedia Serial Link
GNSS	Global Navigation Satellite System
IMU	Inertial Measurement Unit
IRSU	Intelligent Roadside Unit
ITS	Intelligent Transportation System
ITS-S	Intelligent Transportation System Station
LTS	Long-Term Support
MAVEN	Managing Automated Vehicles Enhances Network
NUC	Next Unit of Computing
OMNeT++	Open Modular Network Testbed in C++
RMSE	Root Mean Squared Error
ROS	Robot Operation System
RSU	Roadside Unit
RTK	Real-Time Kinematic
SUMO	Simulation of Urban Mobility
UT	Unscented Transform
UTM	Universal Transverse Mercator
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-X
Veins	Vehicles in Network Simulation
VRU	Vulnerable Road Users

References

1. Li, H.; Nashashibi, F. Multi-vehicle Cooperative Perception and Augmented Reality for Driver Assistance: A Possibility to 'See' Through Front Vehicle. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Washington, DC, USA, 5–7 October 2011; pp. 242–247.
2. Kumar, S.; Shi, L.; Ahmed, N.; Gil, S.; Katabi, D.; Rus, D. CarSpeak: A Content-Centric Network for Autonomous Driving. *ACM SIGCOMM Comput. Commun. Rev.* **2012**, *42*, 1–12. [CrossRef]
3. Mourllion, B.; Lambert, A.; Gruyer, D.; Aubert, D. Collaborative Perception for Collision Avoidance. In Proceedings of the IEEE International Conference on Networking, Sensing & Control, Taipei, Taiwan, 21–23 March 2004; pp. 880–885.
4. Kim, S.W.; Chong, Z.J.; Qin, B.; Shen, X.; Cheng, Z.; Liu, W.; An, M.H., Jr. Cooperative Perception for Autonomous Vehicle Control on the Road: Motivation and Experimental Results. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 5059–5066.
5. Kim, S.W.; Qin, B.; Chong, Z.J.; Shen, X.; Liu, W.; Ang, M.H., Jr.; Frazzoli, E.; Rus, D. Multivehicle Cooperative Driving Using Cooperative Perception: Design and Experimental Validation. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 663–680. [CrossRef]
6. Karam, N.; Chaussel, F.; Aufrère, R.; Chapuis, R. Cooperative Multi-Vehicle Localization. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Tokyo, Japan, 13–15 June 2006; pp. 564–570.

7.   Shan, M.; Worrall, S.; Nebot, E. Nonparametric cooperative tracking in mobile Ad-Hoc networks. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 423–430.

8.   Ko-FAS. Available online: http://www.ko-fas.de (accessed on 23 October 2020).

9.   Rauch, A.; Klanner, F.; Dietmayer, K. Analysis of V2X Communication Parameters for the Development of a Fusion Architecture for Cooperative Perception Systems. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Baden-Baden, Germany, 5–9 June 2011; pp. 685–690.

10.   Rauch, A.; Klanner, F.; Rasshofer, R.; Dietmayer, K. Car2X-Based Perception in a High-Level Fusion Architecture for Cooperative Perception Systems. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Alcalá de Henares, Spain, 3–7 June 2012; pp. 270–275.

11.   Rauch, A.; Maier, S.; Klanner, F.; Dietmayer, K. Inter-Vehicle Object Association for Cooperative Perception Systems. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), The Hague, The Netherlands, 6–9 October 2013; pp. 893–898.

12.   Allig, C.; Wanielik, G. Dynamic Dissemination Method for Collective Perception. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 3756–3762.

13.   Günther, H.J.; Mennenga, B.; Trauer, O.; Riebl, R.; Wolf, L. Realizing Collective Perception in a Vehicle. In Proceedings of the IEEE Vehicular Networking Conference (VNC), Columbus, OH, USA, 8–10 December 2016.

14.   AutoNet2030. Available online: http://www.autonet2030.eu (accessed on 23 October 2020).

15.   Hobert, L. AutoNet2030—Specifications for the Enhancement to Existing LDM and Cooperative Communication Protocol Standards. In AutoNet2030 Deliverable D3.2. 2015. Available online: https://docplayer.net/32233291-Specifications-for-the-enhancement-to-existing-ldm-and-cooperative-communication-protocol-standards.html (accessed on 29 November 2020).

16.   Hobert, L.; Festag, A.; Llatser, I.; Altomare, L.; Visintainer, F.; Kovacs, A. Enhancements of V2X Communication in Support of Cooperative Autonomous Driving. *IEEE Commun. Mag.* **2015**, *53*, 64–70. doi:10.1109/MCOM.2015.7355568. [CrossRef]

17.   Kitazato, T.; Tsukada, M.; Ochiai, H.; Esaki, H. Proxy Cooperative Awareness Message: An Infrastructure-Assisted V2V Messaging. In Proceedings of the Ninth International Conference on Mobile Computing and Ubiquitous Networking (ICMU), Kaiserslautern, Germany, 4–6 October 2016.

18.   Tsukada, M. Roadside-Assisted V2V Messaging for Connected Autonomous Vehicle. In Proceedings of the Thirteenth International Conference on Wireless and Mobile Communications (ICWMC), Nice, France, 23–27 July 2017.

19.   Tsukada, M.; Oi, T.; Kitazawa, M.; Esaki, H. Networked Roadside Perception Units for Autonomous Driving. *Sensors* **2020**, *20*, 5320. [CrossRef] [PubMed]

20.   Burgstahler, D.M. Collaborative Sensing in Automotive Scenarios: Enhancement of the Vehicular Electronic Horizon through Collaboratively Sensed Knowledge. Ph.D. Thesis, Technical University of Darmstadt, Darmstadt, Germany, 2017.

21.   Günther, H.J.; Riebl, R.; Wolf, L.; Facchi, C. Collective Perception and Decentralized Congestion Control in Vehicular Ad-hoc Networks. In Proceedings of the IEEE Vehicular Networking Conference (VNC), Columbus, OH, USA, 8–10 December 2016.

22.   Günther, H.J.; Riebl, R.; Wolf, L.; Facchi, C. The Effect of Decentralized Congestion Control on Collective Perception in Dense Traffic Scenarios. *Comput. Commun.* **2018**, *122*, 76–83. doi:10.1016/j.comcom.2018.03.009. [CrossRef]

23.   ETSI ITS. *Technical Report: Intelligent Transport System (ITS); Vehicular Communications; Basic Set of Applications*; Analysis of the Collective Perception Service (CPS); Release 2. ETSI TR. 103 562 V2.1.1. 2019. Available online: https://www.etsi.org/deliver/etsi_tr/103500_103599/103562/02.01.01_60/tr_103562v020101p.pdf (accessed on 29 November 2020).

24.   SAE Advanced Applications Technical Committee. V2X Sensor-Sharing for Cooperative & Automated Driving. SAE J3224. Available online: https://www.sae.org/servlets/works/committeeHome.do?comtID=TEVCSC3 (accessed on 23 October 2020).

25.   Garlichs, K.; Günther, H.J.; Wolf, L.C. Generation Rules for the Collective Perception Service. In Proceedings of the IEEE Vehicular Networking Conference (VNC), Los Angeles, CA, USA, 4–6 December 2019.

26.   Thandavarayan, G.; Sepulcre, M.; Gozalvez, J. Analysis of Message Generation Rules for Collective Perception in Connected and Automated Driving. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019.

27.   Huang, H.; Fang, W.; Li, H. Performance Modelling of V2V based Collective Perceptions in Connected and Autonomous Vehicles. In Proceedings of the IEEE 44th Conference on Local Computer Networks (LCN), Osnabrueck, Germany, 14–17 October 2019; pp. 356–363.

28.   Huang, H.; Li, H.; Shao, C.; Sun, T.; Fang, W.; Dang, S. Data Redundancy Mitigation in V2X Based Collective Perceptions. *IEEE Access* **2020**, *8*, 13405–13418. doi:10.1109/ACCESS.2020.2965552. [CrossRef]

29.   Thandavarayan, G.; Sepulcre, M.; Gozalvez, J. Generation of Cooperative Perception Messages for Connected and Automated Vehicles. *arXiv* **2019**, arXiv:cs.NI/1908.11151.

30.   Delooz, Q.; Festag, A. Network Load Adaptation for Collective Perception in V2X Communications. In Proceedings of the IEEE International Conference on Connected Vehicles and Expo (ICCVE), Graz, Austria, 4–8 November 2019.

31.   Aoki, S.; Higuchi, T.; Altintas, O. Cooperative Perception with Deep Reinforcement Learning for Connected Vehicles. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, 20–23 October 2020.

32.   Günther, H.J.; Trauer, O.; Wolf, L. The Potential of Collective Perception in Vehicular Ad-hoc Networks. In Proceedings of the 14th International Conference on ITS Telecommunications (ITST), Copenhagen, Denmark, 2–4 December 2015; pp. 1–5.

33. Schiegg, F.A.; Brahmi, N.; Llatser, I. Analytical Performance Evaluation of the Collective Perception Service in IEEE 802.11p Networks. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 181–188.

34. Schiegg, F.A.; Bischoff, D.; Krost, J.R.; Llatser, I. Analytical Performance Evaluation of the Collective Perception Service in C-V2X Mode 4 Networks. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Seoul, Korea, 25–28 May 2020.

35. Bian, K.; Zhang, G.; Song, L. Toward Secure Crowd Sensing in Vehicle-to-Everything Networks. *IEEE Netw.* **2018**, *32*, 126–131. doi:10.1109/MNET.2017.1700098. [CrossRef]

36. Ambrosin, M.; Yang, L.L.; Liu, X.; Sastry, M.R.; Alvarez, I.J. Design of a Misbehavior Detection System for Objects Based Shared Perception V2X Applications. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 1165–1172.

37. Seeliger, F.; Weidl, G.; Petrich, D.; Naujoks, F.; Breuel, G.; Neukum, A.; Dietmayer, K. Advisory Warnings Based on Cooperative Perception. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Dearborn, Michigan, USA, 8–11 June 2014; pp. 246–252.

38. Naujoks, F.; Grattenthaler, H.; Neukum, A.; Weidl, G.; Petrich, D. Effectiveness of advisory warnings based on cooperative perception. *IET Intell. Transp. Syst.* **2015**, *9*, 606–617. [CrossRef]

39. Deng, R.; Di, B.; Song, L. Cooperative Collision Avoidance Scheme Design and Analysis in V2X-based Driving Systems. In Proceedings of the IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, UAE, 9–13 December 2018.

40. Rondinone, M.; Walter, T.; Blokpoel, R.; Schindler, J. V2X communications for infrastructure-assisted automated driving. In Proceedings of the IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), Chania, Greece, 12–15 June 2018.

41. Kamel, J.; Ansari, M.R.; Petit, J.; Kaiser, A.; Jemaa, I.B.; Urien, P. Simulation Framework for Misbehavior Detection in Vehicular Networks. *IEEE Trans. Veh. Technol.* **2020**, *69*, 6631–6643. [CrossRef]

42. Wang, Y.; de Veciana, G.; Shimizu, T.; Lu, H. Performance and Scaling of Collaborative Sensing and Networking for Automated Driving Applications. In Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops), Kansas City, MO, USA, 20–24 May 2018.

43. Merdrignac, P.; Shagdar, O.; Tohmé, S.; Franchineau, J.L. Augmented Perception by V2X Communication for Safety of Autonomous and Non-Autonomous Vehicles. In Proceedings of the 7th Transport Research Arena TRA 2018, Vienna, Austria, 16–19 April 2018.

44. Jandial, A.; Merdrignac, P.; Shagdar, O.; Fevrier, L. Implementation and Evaluation of Intelligent Roadside Infrastructure for Automated Vehicle with I2V Communication. In *Vehicular Ad-hoc Networks for Smart Cities: Third International Workshop, 2019 (Advances in Intelligent Systems and Computing 1144)*; Springer: Singapore, 2019; pp. 3–18. ISBN 978-981-15-3749-3.

45. Managing Automated Vehicles Enhances Network (MAVEN). Available online: http://www.maven-its.eu (accessed on 23 October 2020).

46. Khan, S.; Andert, F.; Wojke, N.; Schindler, J.; Correa, A.; Wijbenga, A. Towards Collaborative Perception for Automated Vehicles in Heterogeneous Traffic. In *Advanced Microsystems for Automotive Applications 2018*; Springer: Cham, Switzerland, 2019; pp. 31–42. ISBN 978-3-319-99762-9.

47. Llatser, I.; Michalke, T.; Dolgov, M.; Wildschütte, F.; Fuchs, H. Cooperative Automated Driving Use Cases for 5G V2X Communication. In Proceedings of the IEEE 2nd 5G World Forum (5GWF), Dresden, Germany, 30 September–2 October 2019; pp. 120–125.

48. Simulation of Urban MObility (SUMO). Available online: https://www.eclipse.org/sumo (accessed on 23 October 2020).

49. Vehicles in Network Simulation (VEINS): The Open Source Vehicular Network Simulation Framework. Available online: https://veins.car2x.org (accessed on 23 October 2020).

50. Sommer, C.; German, R.; Dressler, F. Bidirectionally Coupled Network and Road Traffic Simulation for Improved IVC Analysis. *IEEE Trans. Mob. Comput.* **2011**, *10*, 3–15. [CrossRef]

51. OMNet++: Discrete Event Simulator. Available online: https://omnetpp.org (accessed on 23 October 2020).

52. Garlichs, K.; Wegner, M.; Wolf, L.C. Realizing Collective Perception in the Artery Simulation Framework. In Proceedings of the IEEE Vehicular Networking Conference (VNC), Taipei, Taiwan, 5–7 December 2018.

53. Artery: V2X Simulation Framework. Available online: http://artery.v2x-research.eu (accessed on 23 October 2020).

54. Riebl, R.; Günther, H.J.; Facchi, C.; Wolf, L. Artery: Extending Veins for VANET applications. In Proceedings of the International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Budapest, Hungary, 3–5 June 2015; pp. 450–456.

55. CARLA: Open-Source Simulator for Autonomous Driving Research. Available online: https://carla.org (accessed on 23 October 2020).

56. Dosovitskiy, A.; Ros, G.; Codevilla, F.; López, A.; Koltun, V. CARLA: An Open Urban Driving Simulator. In Proceedings of the 1st Annual Conference on Robot Learning (CoRL), Mountain View, CA, USA, 13–15 November 2017; pp. 1–16.

57. Vo, B.N.; Ma, W.K. The Gaussian Mixture Probability Hypothesis Density Filter. *IEEE Trans. Signal Process.* **2006**, *54*, 4091–4104. [CrossRef]

58. Verma, S.; Berrio, J.S.; Worrall, S.; Nebot, E. Automatic extrinsic calibration between a camera and a 3D Lidar using 3D point and plane correspondences. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 3906–3912.
59. Zhou, W.; Berrio, J.S.; De Alvis, C.; Shan, M.; Worrall, S.; Ward, J.; Nebot, E. Developing and Testing Robust Autonomy: The University of Sydney Campus Data Set. *IEEE Intell. Transp. Syst. Mag.* **2020**, *12*, 23–40. [CrossRef]
60. Yi, S.; Worrall, S.; Nebot, E. Geographical Map Registration and Fusion of Lidar-Aerial Orthoimagery in GIS. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, New Zealand, 27–30 October 2019; pp. 128–134.
61. Narula, K.; Worrall, S.; Nebot, E. Two-Level Hierarchical Planning in a Known Semi-Structured Environment. In Proceedings of the IEEE Intelligent Transportation Systems Conference (ITSC), Rhodes, Greece, 20–23 September 2019; pp. 1344–1349.
62. Allig, C.; Wanielik, G. Alignment of Perception Information for Cooperative Perception. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 1849–1854.