Designing a far-reaching view for highway traffic scenarios with 5G-based intelligent infrastructure

Gereon Hinz, Martin Buechel,
Frederik Diehl, Guang Chen,
Annkathrin Kraemmer, Juri Kuhn
and Venkatnarayanan Lakshminarasimhan
Autonomous Systems and Sensor Systems
fortiss GmbH, Munich, Germany

Malte Schellmann
Applied Communication Techn. Lab
Huawei German Research Center
Munich, Germany

Uwe Baumgarten, Alois Knoll
Operating Systems
Robotics and Embedded Systems
Department of Computer Science
Technische Universität München
Munich, Germany

***Abstract*—Cooperative vehicle infrastructure systems offer significant potential for improved traffic safety, throughput and improved energy efficiency. Infrastructure sensors along the road can substitute vehicular sensor-sets, providing improved robustness and performance through different mounting positions and orientations, reducing occlusions, stationary locations, facilitating system-wide calibration, optimization for the specific traffic area in view, and vastly increase perception range by combining multiple measurement points. Communication via fifth Generation (5G) networks offers solutions to the corresponding substantial requirements for high bandwidth, low latency and high reliability for data and information communication. We propose a concept, which aims to provide a far-reaching view to (self-driving) vehicles and drivers with infrastructure sensors and 5G communication, as a cognitive system. The system detects and localizes traffic objects and predicts their future movements. The resulting information will be provided to traffic participants allowing for safer, more proactive and comfortable driving.**

***Index Terms*—V2X, autonomous driving, connected cars, 5G, sensor-sets, far-reaching view, big data, environmental perception**

I. INTRODUCTION

Cooperative vehicle infrastructure systems combine vehicles and roadside units in hybrid networks, utilizing the available data from all devices to provide improved information for traffic participants and active infrastructure elements [1]. As each available sensor comes with its own set of challenges, the combination of both vehicle and infrastructure sensors makes it possible to substitute sensor weaknesses and achieve an improved environmental perception. Additionally, road side infrastructure sensors often have better access to high bandwidth connections (e.g. glass fiber), electricity, and access to more computational power, enabling more sophisticated detection and fusion algorithms. Vehicle sensors can likewise provide valuable information about the vehicle's state (position, velocity, road condition) and their immediate environment.

An increasing range of promising use-cases for cooperative vehicle infrastructure systems have been identified [2], including road side animal detection, traffic monitoring, stopped vehicle detection, wrong-way driver detection, road debris detection, cooperative overtaking, cooperative collision avoidance, intersection management and automated valet parking. Roads with cooperative vehicle infrastructure systems also provide valuable testing opportunities for autonomous vehicle technologies and can quickly accumulate massive data sets for offline-testing, scenario virtualization, and training of deep neural networks.

Recent progress in sensor, detection and tracking technologies, improved computational power for reduced cost, and paradigm shifts in mobile communication (from a network of connected people to a network of connected things) resulting in high performance 5G networks, offer the potential for significantly improved performance of cooperative vehicle infrastructure systems. New systems will provide more valuable and reliable information to traffic participants, at higher frequencies, with lower latencies for reduced costs.

After the introduction, we present related work, followed by an overview of the Providentia architecture. We then explore the implications of 5G-based communication for intelligent V2X systems, and present strategies for object detection, fusion and scene prediction, as well as experimental validation.

II. RELATED WORK

Cooperative vehicle infrastructure systems provide the potential for much more extensive environmental perception. To access this potential, a number of related challenges need to be addressed concerning detection, fusion, communication and prediction. Fusion faces challenges such as bandwidth limitations, data association uncertainties, dynamic coordinate transformations and system scalability. Multiple promising approaches have gained traction in current research, including approaches based on factor graphs [3], filters [4] and deep learning [5].

5G communication offers promising improvements over prior communication technologies (e.g. LTE). However, challenges such as performing fast, seamless and reliable handover, maintaining QoS requirements, disseminating data while ensuring communication security, and the upper layer communication protocols, need to be addressed. A comprehensive overview of current Vehicle-to-Infrastructure (V2I) communication technologies, related projects and a discussion of the mentioned communication challenges is provided in [6].

A number of infrastructure systems supporting different use-cases have been built in the recent years. There is an ongoing effort for improved detection performance and reliability at reduced costs, increasing performance parameters for sensing and communication devices, increasing distances between measurement points, reducing maintenance efforts, facilitating access to more complex traffic areas (e.g. overpasses, trees) and improving auto-calibration [7][8][9]. However, many challenges still persist, when the goal is to provide (self-driven) vehicles with a real time capable far-reaching view during all lighting and weather conditions that can directly contribute to vehicle motion planning and decision making.

Knowledge about the intentions of other vehicles in a traffic scenario greatly simplifies decision making and planning. A cooperative vehicle infrastructure system is ideally suited to provide traffic participants with long-term and short-term traffic predictions, including prediction of lane changes, acceleration, braking, and swerving maneuvers at real time. Many of the current driver models like the Intelligent Driver Model (IDM) [10] or physics-based models can neither predict these actions nor create predictions far into the future. Newer approaches based on Deep Learning have been proposed recently [11, 12, 13] with good results and can now be extended as an integral part of cooperative vehicle infrastructure systems, fully utilizing vehicle and infrastructure perception.

III. Providentia

The Providentia system aims to provide (autonomous) vehicles with a real time far-reaching view during all weather and lighting conditions. Each measurement point combines several sensors with redundant measurement principles (e.g. cameras and radars). Vehicles can contribute knowledge about their own state and about their immediate environment to the system. The detection of all sensors are fused and combined with a high-resolution map to create a detailed model of the environment in the infrastructure's computer system - a "real time digital twin". Based on the fused information, long- and short-term traffic movements are predicted. The fused and predicted information can be provided to self-driving vehicles, drivers or passengers. Similarly, the digital twin can be used for the virtualization of traffic scenarios for simulation tests. One of the research goals is to investigate the interplay of various information streams in highly automated vehicles, and the communication with back end infrastructure. Another goal is to enable accurate tests of autonomous driving technologies in the harsh reality of a heavily travelled motorway, instead of an isolated test cell with very few vehicles. The Providentia system aims to support mixed traffic scenarios, including both vehicles with a variety of sensor sets and degrees of automation. Such mixed scenarios will be of major importance for years to come. Therefore, this scenario will have a strong impact on the upgrade of the 4.5G communications infrastructure towards 5G in future [14].

IV. System Architecture

Figure 1 illustrates the targeted Providentia system architecture. It consists of the following subsystems and interconnecting networks.

Vehicles The vehicles that are part of the system play a dual role - first as an information source capable of feeding reliable data about themselves and their environment into the various algorithms that create and maintain the real time digital twin, and second as users of the digital twin data. Vehicle-2-Everything communication (V2I/V2N/V2V) is realized through 5G mobile networks, accessed with on-board 5G modems. Applications and services can be visualized with suitable computational and display units. In order to support mixed traffic scenarios with legacy vehicles, a communication link based on 4G LTE is also included. A variety of vehicle sensor information is supported for communication and fusion. Supported sensor types include radars, lidars, in-vehicle cameras and ultrasonic sensors. Additionally, information about current and planned vehicle states can be transmitted, e.g. steering angles, indication light status, desired velocities or accelerations or planned trajectories of automated vehicles or ADAS functions.

Measurement Points The measurement points capture the traffic flow in their local regions, extract useful representational information, and transfer it to neighboring measurement points and the back end for a more comprehensive and wider representation. They are equipped with a variety of sensors that are in combination capable of capturing traffic flow reliably, a powerful local data fusion unit that operates on the information provided by all the sensors, and on the information provided by the neighboring measurement points to generate the real time digital replica within its geographical area of relevance. This information is consequently made available to the back end in real time over the 5G communication network. In order to sense traffic participants reliably, multiple measurement principles are combined. Cameras have a high accuracy in angular resolution and image processing algorithms allow to perform good object classification. Multiple cameras with optimized lenses can increase camera detection range. Radar sensor readings are very accurate in object distance and speed, while angular accuracy is relatively poor and provide good performance during bad weather conditions. Both radar and thermal sensors allow for improved functionality during night. The combination of the sensors promises to compensate the negative effects of each measurement technique.

Base Stations The base stations are in charge of establishing the physical 5G radio interface required to connect the rest of the system participants. Apart from performing the radio access duty, the base stations are also equipped with Mobile Edge Computing units, which can provide

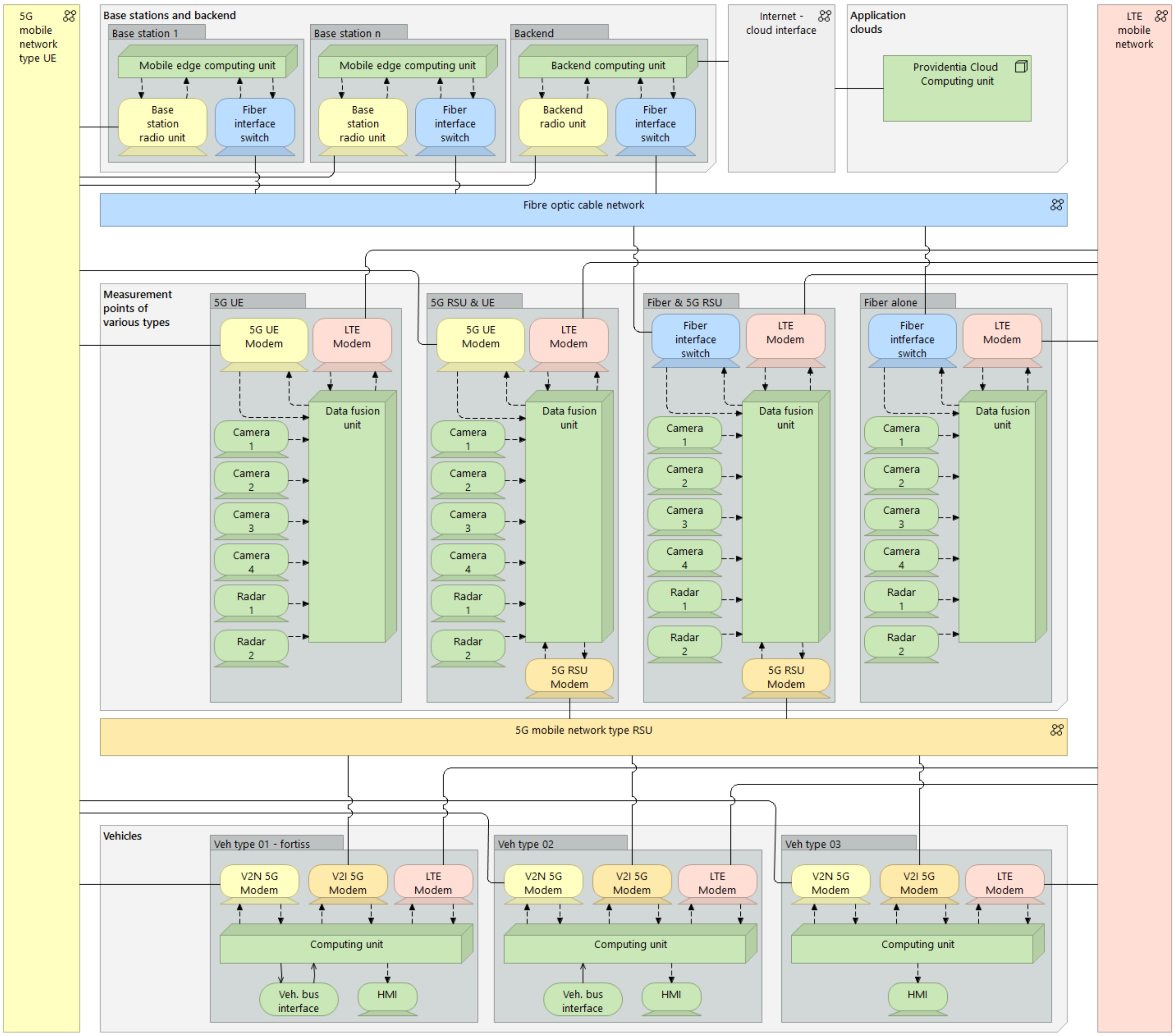

Fig. 1. Providentia system architecture illustration showing the system participants - vehicles that feed information into the system and utilize its services; measurement points equipped with sensors that digitally capture the traffic flow; back end infrastructure that hosts the overall digital twin information; and interconnecting 5G mobile networks (UE and RSU types) established with the help of base stations and 5G modems.

additional computational power to system processes, such as data fusion.

Backend The back end generates and hosts the comprehensive real time digital twin based on the extracted information received from all measurement points. This data is then made available to traffic participants.

Communication Network The system is enabled by two 5G communication networks - the UE (User Equipment) network for V2V communication that is deployed for the localized communication between the vehicles themselves and the measurement points, and the RSU (Road Side Unit) network for V2I communication between the central back end and the local vehicles or measurement points. Additionally, optical fiber networks, if available, could be used for communication between measurement points and back end.

Application Cloud The Providentia system also includes an interface by which the real time digital twin data from the back end can be made available over the internet to a wide array of application services that are deployed on the cloud.

To achieve real time capability, the Providentia system requires an effective communication architecture. It needs to provide an efficient and reliable way to exchange data between multiple processes on one machine, by using classical inter process communication (IPC) approaches (e.g. shared memory). It also needs to support fast, reliable data exchange between multiple physical endpoints, using network-based communication. A possible solution could be provided by the new Time Sensitive Networking Standard (TSN), in order to schedule critical, as well as non-critical traffic within our network. The system concept is based on a decentralized communication approach, meaning that it does not require a central master, who has overall knowledge about the whole system and its connected clients. A related approach has been implemented for IPC communication on one single physical endpoint. A service provider and requester were attached to a local multicaster group, where control information, as how can specific information get accessed (e.g. shared memory, pipes, etc.) and by whom the data gets provided, get exchanged to make the local system self-organized. A benefit of this attempt is added redundancy within the system, since every client has knowledge about the current local system's state, without adding further complexity to the architecture.

V. 5G-based Communication

One of the key elements of the next generation radio system (5G) is the native support for ultra-reliable and low latency communication (URLLC), which can be considered a crucial enabler for safety-critical services and thus an important driver for future V2X communications. Compared to current 4G Long Term Evolution (LTE) networks, 5G introduces a new frame structure tailored for URLLC that facilitates much faster response times for the transmission of single data packets. The key idea here is on the one hand to use more bandwidth for the transmission of a data packet and thus reduce the time required for its transmission, and on the other hand to multiplex control information supporting the transmission of the data packet into the same transmission slot. This way, transmission time for a single data packet becomes short, and even retransmissions of the packet can be facilitated without having a significant impact on the constrained delay budget.

Moreover, 5G will come with a bunch of technologies that help to increase the reliability of data transmissions, such as beamforming techniques based on multiple antennas at transmitter and receiver, aiming to increase the robustness of the communication link, and network multi-connectivity, where a terminal is maintaining multiple links to the network to exploit the statistical independence of these links (also called Ïink diversity) for attaining reliable transmission and reception of data. In more detail, beamforming techniques allow to direct the radio wave towards the mobile terminal and thus achieve a higher signal quality at the receiver for constant transmit power. If there are sufficient reflectors in the propagation environment, different beams can be formed reaching the receiver through independent propagation paths, and if these are used for transmitting the same data packet, the provided link diversity can be beneficially exploited for improving the reliability of the transmission. As the vehicular communication node is usually moving at higher speed, a transmission beam needs to be continuously modified to follow the vehicle properly and avoid any signal loss. Since the

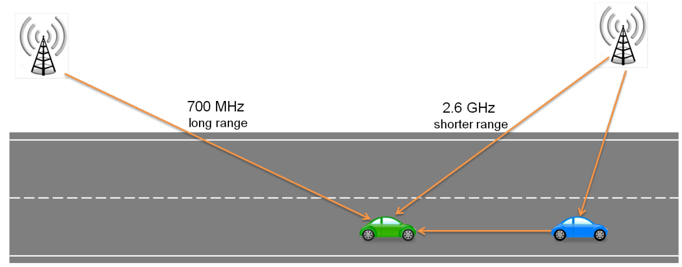

Fig. 2. Network multi-connectivity in 5G networks for providing link diversity

trajectory for a moving vehicle (in particular a car moving along a road) is known for short-term, beam tracking and beam prediction is facilitated if information on this trajectory is exchanged between the involved communication nodes.

Network multi-connectivity, the other technology mentioned above to leverage reliability, can be provided to a vehicle by several means, as illustrated in Figure 2: The vehicle can be connected to different base stations at the same time, where these connections may even be established on different carrier frequencies. Since the coverage of a carrier frequency scales inverse-proportionally with the carrier frequency itself, i.e. a low carrier frequency has a longer reach, a vehicle can be easily connected to base stations being located at significantly different distance. Moreover, an additional link to the vehicle can also be established through another vehicle, which may relay the signal from the base station using a V2V (vehicle to vehicle) connection. For all these multi-connectivity cases, it can be assumed that the propagation paths for the links are statistically independent, thus providing the link diversity that finally can be exploited to yield a highly reliable transmission. All the technologies mentioned in this section will be applied in the 5G prototype deployed in the Providentia project for the reliable and delay-critical data exchange between the vehicles on the highway and the back end.

VI. Object Detection

To cover all lighting and weather conditions at high range and acceptable cost with infrastructure sensors, cameras, radars and thermal sensors are evaluated. Position, velocity, and class of each object should be identified. While the utilized radars provide object detection, a deep learning based object detection approach is adapted for the camera. Both online and offline traffic object detection are of interest, because while real time capable online detection with good performance is needed to create the live digital twin, adapted data sets for deep neural network optimization can be created with offline detection, which should provide highest performance, but does not need to run at real time. For this we evaluated Faster R-CNN [15], which consists of two advanced modules: a region proposal network (RPN) and Fast R-CNN [16] detector. The entire system is a single, unified network for object detection.

In the first stage, RPN, images are processed by a feature extractor, and features at some selected level are used to predict class-agnostic box proposals. The RPN is trained

Fig. 3. View of the short-range camera on the left and of the long-range camera on the right with camera detections (bounding boxes) and radar measurements (blue cuboids). The image illustrates fields of views and detection ranges of the 2 cameras and the radar. The sensors complement each other. The short-range camera detects objects closer than the radar, the far range camera detects objects beyond the radar range. There is overlap between each sensor.

end-to-end to generate high-quality region proposals. In the second stage, Fast R-CNN, box proposals are used to crop features from the same intermediate feature map, which are subsequently fed to the remainder of the feature extractor in order to predict a class and class-specific box refinement for each proposal. Faster R-CNN merged RPN and Fast R-CNN into a single network by sharing convolutional features, using 'attention' mechanisms, the RPN component tells the unified network where to search [17]. The adaption of networks pretrained on general images enables good initial detection results before time consuming labeling of specialized data sets becomes necessary.

We show exemplary detection results of sample images in Figure 3. Most vehicles get detected. However, detection range is limited and occlusion due to adjacent vehicles can still lead to errors. In general, our test vehicle detection system shows reasonable and high accuracy detection results, but it also illustrates the value of fusing camera detections with additional sensor types to increase robustness, especially when high performance during a multitude of lighting and weather conditions is required.

VII. Reliable Multi-Object Multi-Sensor Data-Fusion

Fusion in the Providentia system concept is performed first locally on the measurement points (infrastructure sensors and sensors from connected nearby vehicles), and in a second step with the combined information of multiple measurement points in the back end, where they are fused to form the comprehensive and real-time digital twin. Thereby, Providentia combines a number of filtering, tracking and fusion challenges, all of which have to be solved in real time:

Multi-object: In comparison to fusion tasks for autonomous vehicles, the term multi-object for filtering algorithms takes on new meaning when it is applied to large scale infrastructure systems. During traffic jams, one measurement point's overall detection range can contain 800 vehicles (assuming a traffic jam, $10\,m$ per vehicle, four lanes in each direction, and $500\,m$ field of view in each direction). At this scale, the association problem of assigning each measurement to the corresponding object can become a computational bottleneck when using classical filters like the Kalman filter (see [18]), Multiple Hypotheses Tracking (MHT) (see [19]) or the Joint Probabilistic Data Association filter (JPDA) (see [20]). Therefore, the current concept utilizes filters based on the random finite set (RFS) framework developed by Mahler (see e.g. [4]), as those do not need to solve the association problem explicitly. Further advantages of RFS based filtering methods for highway scenarios are that they can handle varying and unknown numbers of objects and achieve accurate and robust estimates despite clutter and detection failures.

Sensor dependent characteristics: A good fusion algorithm incorporates the characteristics of each sensor in the system and treats their measurements according to their individual strengths and limitations. For example, cameras are well suited for classifying objects while radars can determine velocities and angles accurately. Furthermore, the algorithm should eliminate sensor dependent ghost objects. In our test field, there is a wall next to the most left highway lane reflecting radar signals especially from trucks driving by and leading the radar to erroneously assume the existence another vehicle. This can be seen in the left picture of Figure 3. Each stationary measurement point can be optimized for its specific location. Therefore, we take such knowledge into account when weighting each measurement in the fusion procedure. Furthermore, every sensor type provides measurements at a different frequency and requires a different processing time, i.e. adds a different delay to the system. These delays together with those stemming from the communication have to be determined and incorporated appropriately in the fusion algorithm, even though we assume the latter to be quite small thanks to 5G.

The combination of the individual sensors' fields of view constitute another challenge. Regarding one Measurement Point in one direction, the short range camera and the far range camera cover different areas with overlapping field of views while the radar's field of view starts some meters into that of the short range camera but continues into that of the long rang camera and also depends on the specific lane, as can be seen in Figure 3. Additionally, these field of views intersect with the corresponding ones from the neighboring Measurement Point as well and the sensors in the communicating cars have yet others, completely different perspectives. Therefore, the area covered by one Measurement Point contains several parts that are observed with different numbers of sensors, i.e. our algorithm needs to be able to fuse measurements from an adaptive number of different sensors. However, the different field of views have the advantage of providing more information and redundancy. The Providentia system sees the same object from different angles and sides and can dissolve occlusions resulting into an overall improved tracking performance.

High reliability: The basis for a highly reliable digital twin are good models of the reality within the fusion algorithms like suitable dynamic and observation models, good birth densities and appropriate survival probabilities. In this context, one challenge is to differ between vehicles that have vanished from the field of view and those that are still there but are now occluded. Therefore, a good way to represent occlusions in the filtering needs to be developed so that they get resolved in the back end fusion as mentioned above. Additionally, good models of the clutter process are required. Conventional filtering and tracking methods assume it to follow a spatially uniform distribution although it is unknown and dynamically changing over time. Therefore, an algorithm not requiring a priori clutter models like the RFS based CPHD filter described in [21] seems more promising (see also Section VII in [22]). Representing the vehicles as extended objects also increases reliability. Classical filters and basic RFS based ones like the GM-PHD and SMC-PHD assume point objects, leading to a mismatch between filtering results and reality. For example, shapes cannot be taken into account and only one measurement per object is assumed. This is contradictory to the measurements in the Providentia project like (2D or 3D) bounding boxes resulting from the camera object detection or raw reflection data from the radars. Furthermore, it is essential for autonomous driving algorithms to know the extension of other traffic participants. There already exist extended object filtering methods, also for multiple objects within the RFS filter framework (see [23]). However, they focus on single sensor systems or such with very similar sensors.

The fused measurements are used for the virtualization of recorded traffic scenarios, which is an important part of simulation testing for autonomous vehicles. Effective generation of test scenarios that can be rerun in different scenario variations can greatly improve testing capabilities. The digital twin that results as the output of the Providentia system aims to be compatible with simulation software. Based on high resolution maps, including precise spatial and semantic information about all lanes, lane markers, road signs and further infrastructure, a traffic simulation can both generate synthetic measurement data to test and further develop fusion algorithms, and import measurement data from the Providentia system, making it possible to experience the full scope of the digital twin in a physical test drive simulator. The Providentia system works with IPG CarMaker, which provides the required capabilities for map import, sensor simulation and traffic visualization (see Figure 5).

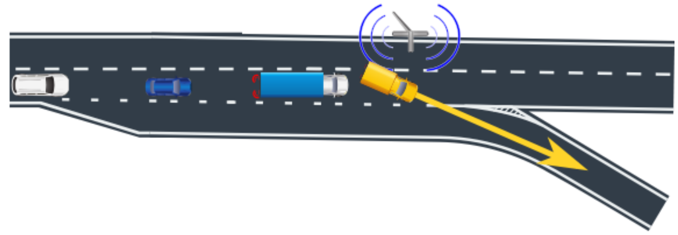

Fig. 4. The Providentia system, a measurement point visible as a stylized pole, providing a warning of an impending braking cascade. The yellow car suddenly changes lanes to exit the highway. This causes the blue truck to brake. Due to the early warning for the blue and white car, even before the truck brakes, the cascade can be prevented.

Fig. 5. The CarMaker as a dynamic virtualization platform for the real time digital twin model, showing two infrastructure measurement points in a highway scenario perceiving a number of vehicles *Source: IPG CarMaker*.

VIII. SCENE PREDICTION

A key strength of attentive drivers is good scene understanding and anticipation of upcoming traffic events. Accurately predicting traffic events allows for preventing actions and better structured reactions. An example of preventing a braking cascade through prediction and early warning for participating vehicles, is shown in Figure 4). This short-term prediction of traffic events, on the order of magnitude of about ten seconds, is significantly different from long-term traffic state prediction (e.g. predict traffic jams), which has a time horizon of tens of minutes. The advantages of using an infrastructure system for this prediction are manifold and include a significantly longer perception horizon, strongly reduced occlusion and the possibility for optimization for the infrastructure system's spe-

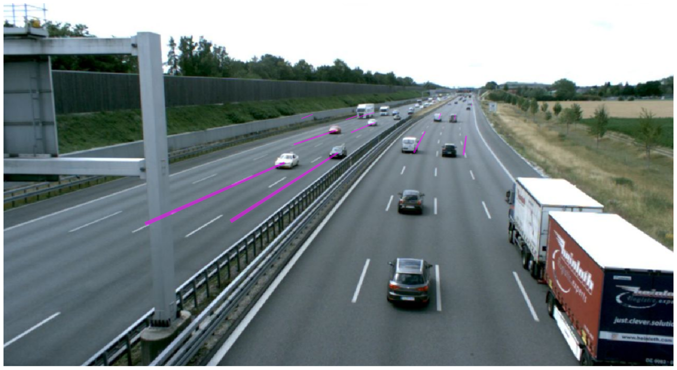

Fig. 6. Short-term prediction for a traffic scene, based on radar detections. The current prediction system predicts the next seconds of a car's motion. Prediction performance, in distance, time and accuracy, depends on the detection performance.

cific location. Furthermore, more efficient computation can be achieved, since many cars are interested in the same stretch of highway, and once calculated, the prediction can be distributed to every participant. The computational complexity for the infrastructure based prediction scales with the length of the perceived highway stretch. Contrary to that, prediction running on every car's computer will require redundant computing power for every car.

The Providentia system aims to provide short-term predictions of lane changes, acceleration, braking, and swerving maneuvers, sufficiently early to allow both autonomous vehicles, as well as drivers of manual vehicles to include the prediction into the decision making and planning process. Recent approaches based on Deep Learning [11, 12, 13] have shown promising results. However, these approaches remain fundamentally limited by operating only with the ego-vehicle's perception. This results in duplicated computation and neglects the increased availability of additional measurements, arguably the main advantage of an infrastructure system. Clearly, a new approach has to be developed for infrastructure-based prediction. The concepts of an infrastructure and V2X based far-reaching view and infrastructure-based traffic prediction complement each other. Therefore the Providentia system aims to provide a far-reaching view both in distance and time. Figure 6 shows initial short-term prediction results of a traffic scene based on radar detections, predicting the next seconds of vehicle motion to illustrate the concept. Both improved detection and fusion, as well as more sophisticated prediction algorithms will further enhance the prediction capabilities.

Since Deep Learning-based approaches showed good results in the single-vehicle scenario, they are a promising avenue to pursue further. The restrictions on behaviour implied through a purely-supervised approach as in [13] suggest an approach based either on reinforcement learning or on imitation learning, as done in [12]. The latter approach using Generative Adversarial Imitation Learning (GAIL) however still suffers from restrictions from GAIL like preferring longer-running trajectories. An alternative would be to abstract further, removing the prediction of car behaviour and instead predicting the whole scene change. This has already been done on video data in the Atari domain [24]. Its advantage is an easy and clear feature representation and constant performance independent of the number of cars. For feature representation, three different approaches are possible: using a high-level abstraction as in [13], which allows modelling a wide variety of situations, but denies the main use of Deep Learning, that is learning features; using a simulated lidar-like sensor as feature representation, as in [12], which allows for feature extraction but does not allow us to include information based on the infrastructure sensors; or a grid map representation, which allows the inclusion of arbitrary sensors and allows the system to learn high-level features itself, which is a preferred approach for the Providentia system and can be combined with grid-map based event detection, as shown in [25]. The latter is also the only system in which whole-scene prediction is possible.

IX. Conclusion and outlook

Cooperative vehicle infrastructure systems offer significant potential for improved traffic safety, throughput and improved energy efficiency. The Providentia system aims to detect, localize and predict traffic objects and their movements, and make corresponding information available to traffic participants, both to drivers and to self-driving cars. The system facilitates the integration of self-driving and manual cars, a major challenge and the dominant traffic scenario of the coming decades. The combination of vehicle and infrastructure sensors promises improved environmental perception, profiting from the use of complementing sensor-sets and sense all relevant areas from multiple angles.

High resolution maps that contain geo-spatial as well as semantic information facilitate scene understanding. The combination of excellent and far-reaching perception with well understood (mapped) traffic scenarios provides the best base for scene prediction. Scene prediction performed by cooperative vehicle infrastructure systems provides a number of benefits, ranging from reduced computational cost to better prediction performance. Stationary infrastructure can be optimized to detect traffic objects in its local field of view, while sensor sets of autonomous vehicles have to provide good performance in all traffic areas. Additional computational resources in a back end help to provide situational awareness and scene prediction. Progress in each of the technological areas of sensors, detection algorithms, fusion, tracking, prediction, communication and computation power bring the provision of a far-reaching view, in both space and time, within reach. However, a number of challenges remain in each of the mentioned areas, when the goal is to provide good performance in all traffic scenarios, for all lighting and weather conditions at very low latency, with very high reliability, at acceptable cost. In this work we presented a concept for the realization of a real time digital twin based on 5G-telecommunication technology, providing this far-reaching view to self-driving cars, drivers and passengers, and demonstrated promising experimental results. Other researchers can extend the concept or evaluate different algorithms in its context for increased performance. In future work, we will investigate different vehicle detection approaches and leverage different feature fusing techniques for improved real time detection. We aim to develop reliable multi-sensor multi-object fusion algorithms for 5G communication based sensor networks. We will also strife to provide improved scene prediction approaches, because while shortest-term predictions of traffic object trajectories of a few seconds are possible with existing models, predicting trajectories and occupancies ten seconds into the future while using infrastructure sensor data remains a worthwhile and unexplored research area.

ACKNOWLEDGMENT

This research has been funded by the Federal Ministry of Transport and Digital Infrastructure in the project Providentia.

Federal Ministry
of Transport and
Digital Infrastructure

REFERENCES

[1] Xiaolan Tang, Donghui Hong, and Wenlong Chen. "Data Acquisition Based on Stable Matching of Bipartite Graph in Cooperative Vehicle–Infrastructure Systems". In: *Sensors* 17.6 (2017), p. 1327.

[2] Gwenaelle Toulminet, Jacques Boussuge, and Claude Laurgeau. "Comparative synthesis of the 3 main European projects dealing with Cooperative Systems (CVIS, SAFESPOT and COOPERS) and description of COOPERS Demonstration Site 4". In: (2008), pp. 809–814.

[3] Dhiraj Gulati et al. "Vehicle Infrastructure Cooperative Localization Using Factor Graphs*". In: *IV* (2016).

[4] Ronald P. S. Mahler. *Statistical Multisource-Multitarget Information Fusion*. Artech House, Inc., 2007.

[5] Peter Ondruska and Ingmar Posner. "Deep tracking: Seeing beyond seeing using recurrent neural networks". In: *The Thirtieth AAAI Conference on Artificial Intelligence (AAAI)* (2016).

[6] Emmanuel Ndashimye et al. "Vehicle-to-infrastructure communication over multi-tier heterogeneous networks: a survey". In: *Computer Networks* 112 (2017), pp. 144–166.

[7] Vassili Alexiadis. "Video-Based Vehicle Trajectory Data Collection". In: *Transportation Research Board 86th Annual Meeting* (2006).

[8] Ellen van Nunen et al. "Cooperative competition for future mobility". In: *IEEE Transactions on Intelligent Transportation Systems* 13.3 (2012), pp. 1018–1025.

[9] Marcel P Huijser and Patrick T McGowen. "Overview of animal detection and animal warning systems in North America and Europe". In: *Road ecology center* (2003).

[10] Martin Treiber, Ansgar Hennecke, and Dirk Helbing. "Congested Traffic States in Empirical Observations and Microscopic Simulations". In: *Physical Review E* 62.2 (2000).

[11] Jeremy Morton, Tim A Wheeler, and Mykel J Kochenderfer. "Analysis of Recurrent Neural Networks for Probabilistic Modeling of Driver Behavior". In: *IEEE Transactions on Intelligent Transportation Systems* (2016), pp. 1–10.

[12] Alex Kuefler et al. "Imitating driver behavior with generative adversarial networks". In: *arXiv preprint arXiv:1701.06699* (2017).

[13] David Lenz et al. "Deep Neural Networks for Markovian Interactive Scene Prediction in Highway Scenarios". In: *IEEE Intelligent Vehicles Symposium* (2017).

[14] Gereon Hinz et al. "Proactive Video-based Use of Telecommunications Technologies in Innovative Motorway Scenarios". In: *5th GI/ITG KuVS Fachgespräch Inter-Vehicle Communication* (2017), p. 41.

[15] Shaoqing Ren et al. "Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*. NIPS'15. Montreal, Canada: MIT Press, 2015, pp. 91–99.

[16] Ross B. Girshick. "Fast R-CNN". In: *CoRR* (2015).

[17] Jan Chorowski et al. "Attention-based Models for Speech Recognition". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems*. NIPS'15. MIT Press, 2015, pp. 577–585.

[18] Rudolf E. Kalman and Richard S. Bucy. "New Results in Linear Filtering and Prediction Theory". In: *Journal of Basic Engineering* 83 (1961), pp. 95–108.

[19] Samuel S. Blackman. "Multiple hypothesis tracking for multiple target tracking". In: *IEEE Aerospace and Electronic Systems Magazine* 19.1 (2004), pp. 5–18.

[20] Thomas E. Fortmann, Yaakov Bar-Shalom, and Molly Scheffe. "Sonar Tracking of Multiple Targets Using Joint Probabilistic Data Association". In: *IEEE Journal of Oceanic Engineering* 8.3 (1983), pp. 173–184.

[21] Ronald P. S. Mahler. "CPHD filtering with unknown clutter rate and detection profile". In: *IEEE Transactions on Signal Processing* 59.8 (2011), pp. 3497–3513.

[22] Ronald Mahler. "A Brief Survey of Advances in Radom-Set Fusion". In: *International Conference on Control, Automation and Information Sciences* (2015).

[23] Karl Granström, Marcus Baum, and Stephan Reuter. "Extended Object Tracking: Introduction, Overview and Applications". In: *ISIF Journal of Advances in Information Fusion* (2017).

[24] Junhyuk Oh et al. "Action-Conditional Video Prediction using Deep Networks in Atari Games". In: *Nips* (2015), p. 9.

[25] Richard Gruner et al. "Spatiotemporal representation of driving scenarios and classification using neural networks". In: *Intelligent Vehicles Symposium (IV)*. IEEE. 2017, pp. 1782–1788.