Multi-Sensor Data Fusion for Checking Plausibility of V2V Communications by Vision-based Multiple-Object Tracking

Marcus Obst  
BASELABS GmbH  
Ebertstr. 10  
Germany, Chemnitz 09126  
Email: marcus.obst@baselabs.de

Laurens Hobert  
Hitachi Europe SAS  
955, Route des Lucioles  
France, Sophia Antipolis 06560  
Email: laurens.hobert@hitachi-eu.com

Pierre Reisdorf  
Technische Universität Chemnitz  
Reichenhainer Str. 70  
Germany, Chemnitz 09126  
Email: pierre.reisdorf@etit.tu-chemnitz.de

Abstract—Automated vehicles require an adequate and reliable perception of the surrounding world in order to make good decisions. Using vehicle-to-vehicle (V2V) communication to exchange location data (i.e. time, position, heading and speed) can improve the perception beyond the capabilities of traditional on-board sensors (e.g. radar, lidar). However, it is vital to trust the data before it is being used. Cryptographic mechanisms can protect the exchange and authenticity of data but do not guarantee the correctness of the content. In this paper we present a vision-based multi-object tracking system for checking the plausibility of V2V communication. The system is addressing the challenge of fusing relative sensor observations as provided by a MobilEye vision-system with time-delayed absolute GNSS-based measurements from Cooperative Awareness Messages (CAMs) as provided by V2V. The plausibility check is implemented in a prototype and based on a state-of-the-art multiple-object tracking algorithm. The proposed system is evaluated and validated under real-world conditions by conducting several test drives under urban conditions.

I. INTRODUCTION

A clear trend in the automotive industry is the automation of driver tasks. Early technologies such as ABS and ESP are supporting the driver respectively in braking and unstable situations whereas more recent technologies such as the lane-keep assistant and adaptive cruise control go one step further by taking over driver tasks. Considering the intention of many vehicle manufacturers to introduce fully automated vehicles by the year 2025, we may expect a continuation of this trend in the future.

Many of these technologies depend on an adequate and reliable perception of the surrounding world. For example, the lane-keep assistant needs to know the vehicle’s position in a lane and a fully automated vehicle requires a view of surrounding obstacles in order to navigate around them. Traditional automotive sensors (e.g. radars, lidars, vision-systems) can sense the environment with a high frequency and high reliability. However, they are limited in range, field of view (FOV) and target measurements (e.g. distance and angle).

With the introduction of vehicle-to-vehicle (V2V) communication—that is the exchange of data between vehicles—and standards like the Cooperative Awareness Message (CAM), the perception of vehicles may be improved beyond the capabilities of traditional sensors by the exchange of location data (i.e. time, position, heading and speed), vehicle dimensions, vehicle type, yaw rate, acceleration and many other parameters. However, V2V communication involves by definition a sending and receiving vehicle and an important question is whether the received data can be trusted; especially if the data is used to control the vehicle. For example, an attacker might have compromised the transmitting system or the system is unintentionally sending wrong data due to a broken sensor. Cryptographic mechanisms can protect the exchange and authenticity of data but do not guarantee the correctness of the content so we need another way of checking the plausibility of V2V data.

In this paper, we develop and verify a system to check the plausibility of received CAM data by cross correlating it with data obtained from a MobilEye camera. This system is implemented as a probabilistic multi-sensor multi-object data fusion algorithm and is addressing the challenges of fusing global and vehicle-centric position coordinates as provided respectively by CAMs and the MobilEye camera and out-of-sequence measurements due to V2V communication delays. The presented work deliberately builds on existing technology such as commercial sensors. Furthermore, the efforts of implementing the probabilistic multi-sensor data fusion algorithms are limited by using and relying on a generic data fusion framework BASELABS Create. By that approach, a multi-sensor data fusion system which comprises the state-of-the-art performance of single sensors can be developed and implemented in a prototypical way1 rather efficiently. Consequently, this paper will not present theoretical details of Bayesian data fusion nor implementation details of typical multi-object tracking (MOT) approaches. In [2] the whole development cycle for a probabilistic vehicle localization system, which was implemented with the aforementioned tools, is presented and elaborated in detail.

The paper is structured as follows: in section II a brief overview of related work is given. Afterwards, the subsequent section III introduces the principles of the used sensors and gives a glance of multi-sensor data fusion in the context of multi-object tracking. Section IV elaborates on the proposed approach to derive a decision from the tracking measures whether a CAM message is spoofed by an attacker or can be

1BASELABS provides a further tool named CODE which automatically transforms the prototyping code of CREATE to production-ready ECU source code. This process is described in more detail in [1].
assumed trustworthy. The implementation details are given in section V. Special emphasize is put on the task of transforming the absolute positions from the CAMs to the local vehicle coordinate frame. Finally, section VI presents quantitative results from real-work test drives and evaluates the capability of a multi-object tracker for plausibility checking. The paper concludes with a summary in section VII.

II. RELATED WORK

The VSC-A research project has identified in collaboration with NHTSA, 6 potential safety application concepts that can address imminent crash scenarios using V2V communication. Two of these promising concepts are the Electronic Emergency Brake Light (EEBL) and the Intersection-Movement Assist (IMA) which warn the driver respectively for hard braking vehicles and potential intersection crashes. For EEBL, Szczurek et al. have developed and compared in [3] two methods to estimate the relevancy of a hard braking vehicle. These methods depend on the exchange of position, time, vehicle speed, length, road and lane identifiers using V2V communication. In [4], a risk assessment is presented for an IMA which heavily relies on V2V communication to overcome the limitations of classical on-board vehicle sensors with limited field of view. In both aforementioned publications, the authors assume accurate and consistent data. However, no dedicated plausibility check was foreseen.

Typically, heterogeneous sensor information are combined to either implement some kind of plausibility checking or to improve the observation area and therefore provide a better perception. Apparently, both tasks are often implemented by completely different approaches. Examples for multi-sensor plausibility checking for VANETs are given in [5] and [6]. The former work introduces an integrated particle filter-based approach to implement plausibility checking in VANETs which incorporates several information such as radar observations or digital map data. The authors of the latter work propose to use multiple behavior analysis modules in order to enhance the security in VANETs. Another approach, which was developed in the sinTD project, utilizes a Kalman filter-based approach to track and assess individual CAMs. The aforementioned publications have in common that they rely on fixed station identifiers, require periodic CAMs with fixed latencies and make strong assumptions on the number of involved participants.

On the contrary, there are several studies which investigate the potential of V2V communication as part of a multi-sensor data fusion framework to extend the perception of the on-board sensors. Typically, an accurate and comprehensive knowledge of the vehicle’s environment is a crucial requirement for safety-relevant ADASs and autonomous driving in general. In [7] a high-level fusion architecture which considers CAMs as a kind of virtual sensor is presented. The authors propose to use a track-to-track fusion based data fusion scheme. In [8] a practical implementation of a tracking system which uses the position messages from V2V in combination with an on-board radar sensor to provide an adaptive cruise control application is presented. The authors conclude that the rather limited accuracy of the position messages complicates the data association step between both sensor systems. A simulative study based on artificial radar measurements and communicated vehicle positions in order to demonstrate the successful extension of the sensor field of view was performed in [9].

The paper at hand wants to demonstrate that both tasks, that is plausibility checking and extended perception, can be implemented based on a unified multi-object tracking framework. Therefore, a consumer-grade on-board perception sensor will be combined with the information coming from CAMs as communicated using ITS-G5 devices. The heterogeneous sensor information are processed by a Bayesian data fusion algorithm under real-time conditions. In addition to the dynamic parameters of the surrounding vehicles a confidence estimate is naturally provided as well. Furthermore, each object is assigned an existence probability which will be used in order to realize the plausibility checking.

III. FUNDAMENTALS

In this section, the used sensor hardware is briefly explained and a proposal how to integrate it with Bayesian multi-sensor data fusion is given.

A. Vision Based Vehicle Detection

Typically, radar or lidar sensors are used in order to perceive the local environment around the vehicle. In this paper, a commercial off-the-shelf vision-based system which comprises a mono-camera of MobilEye is used. The MobilEye system is mounted behind the windshield and provides relative measurements (distance, angle and relative speed) as well as the width of other vehicles with respect to the ego vehicle over a CAN interface. This type of detections are sometimes also called stick objects. More details on the internal image processing algorithms of this system can be found in [10]. It is worth mentioning that the detection quality is far from perfect. For example, under difficult illumination conditions or on wet roads the system is sensitive to clutter measurements. This needs to be considered inside of the data fusion.

B. Perception by Communication

With the deployment of V2V communication, a new class of sensors is available which can be used for innovative ADAS applications and intelligent vehicles. Several standards for V2V and V2I (vehicle-to-infrastructure) communication have been defined by organisations like IEEE, SAE and ISO. In Europe, ETSI and CEN have defined a set of standards for V2V and V2I (V2X) communication. This set includes the Cooperative Awareness Basic Service which mandates each road user participating in a V2X network to transmit periodically their state, time, position, velocity and many other parameters in a CAM. This rather comprehensive description of an object state in combination with a communication range of up to 1000 m [11] makes CAM data appealing candidates for multi-sensor data fusion.

However, V2V communication in general and CAM transmissions in particular have some distinctive characteristics compared to classical on-board perception sensors. The following section lists these characteristics briefly. Later in Section V, they will be addressed inside the proposed implementation of the multi-sensor data fusion system.

1) Coordinate frame: While on-board sensors usually perceive the environment in a vehicle-centric coordinate frame, CAMs describes the object’s position in absolute coordinates. It is therefore inevitable to translate to a single coordinate system in order to associate CAMs with on-board sensor detections.
2) Update rates: On-board sensors usually have a fixed and high-frequent update rate of 20 Hz-50 Hz. CAMs are however generated with a rather low dynamic rate of 1 Hz-10 Hz which depends on the vehicle dynamics. Moreover, CAM transmissions using ITS-G5 are unreliable and may get lost which leads to an even lower update rate at a receiver. It is hence considered a challenging task to observe high-dynamic maneuvers solely by using CAMs.

3) Field of view: The FOV of on-board sensors is usually rather limited. Although, long-range radars might observe the area in front or behind the ego vehicle up to 150 m, they provided only limited performance to perceive objects in the near proximity. Moreover, on-board sensors are quite sensitive to temporary occlusion which might be caused by objects which are blocking the sensor’s FOV. V2V communication based on ITS-G5 may allow omnidirectional communication with a range up to 1000 m and relative speeds up to 30 m/s [11]. Such values have been observed even under real-world conditions as reported in [12]. Furthermore, V2V communication can provide measurements even under non-line-of-sight conditions.

4) Observation latencies: The time span between the generation of a measurement and its availability to the multi-sensor data fusion is described as observation latency. The measurements of in-vehicle perception sensors are usually subject to a fixed sensor-specific observation latency of up to 100 ms for average scenarios. However, for CAMs, variable latencies of up to 1000 ms are observed. There are several reasons for this behaviour: time synchronization errors, a latency between the position measurement and the CAM packet generation, the processing latency of the sending and receiving communication stack and medium contention and transmission latencies. Without and further precautions, the CAMs will arrive out of sequence—thus, they are called out of sequence measurements (OOSM)—at the data fusion system which might lead to inconsistent tracking results [13].

5) Missing detections: The event of missed detections of on-board sensors is usually only caused by non-systematic errors. Hence, on average each object in the sensor’s FOV is detected as long as it is not occluded. For V2V this is different as the reception of CAMs is known to be more unreliable. For example, the probability of a successful CAM reception depends on the physical distance [14], package size [11], the mounting position of the antennas [15] and the link congestion. Moreover, the environmental conditions (urban, suburban) have to be considered as shown in [12] where a reliability analysis of ITS-G5 in the context of the DRIVE-C2X project was conducted.

6) Measurement refinement: The behaviour of on-board sensors, whether they provide raw data or already apply some kind of refinement (e.g. based on Bayesian filtering), is documented in the sensor manual and typically already available when designing a multi-sensor data fusion system. However, the CAM standard does not define how the position information of CAMs is generated. For example, one vendor might simply connect a low-cost GPS receiver, while another one implements a sophisticated vehicle positioning algorithm. Obviously, in this particular case the quality and statistics of the positions contained in both CAMs will be completely different and are unfortunately not known to the data fusion at the receiving vehicle. This applies to the quality indicators and confidence values transmitted in the CAMs as well.

C. Multi-Object Tracking and Multi-Sensor Data Fusion

Object tracking refers to the task of using noisy sensor measurements to derive several characteristics of objects of interest over time [16]. The characteristics typically include dynamic parameters such as position, velocity and heading as well as discrete states (e.g. the existence/relevance of an object\(^2\)). Moreover, object tracking aims to estimate the correct number of objects in the surveillance area. If the number of objects to track is larger than one and several perception sensors are utilized, a multi-object tracking (MOT) problem which naturally includes multi-sensor data fusion is present. The main challenge in MOT is to correctly perform the measurement association. In literature several approaches for multi-sensor data fusion in object tracking have been proposed in [13]. The paper at hand rigorously relies on the concept of recursive Bayesian filtering to exploit temporal correlation between the observations as provided by the modular perception framework BASELABS Create [1]. Independent of the selected MOT algorithm, the domain-specific design (e.g. the consideration of particular sensors and the targeted application) of a MOT typically involves several tasks which have to be fulfilled by the practitioner. In the subsequent text these tasks are briefly explained.

1) Identification of target variables: Besides the estimation of the total number of objects, one objective of MOT is to estimate dynamic parameters. Which particular parameters are of interest typically depends on the final application or use-case. Moreover, a coordinate frame for the estimation is selected [17].

2) Identification of support variables: In addition to the target variables which are directly required by the application, several so-called support or auxiliary variables need to be estimated as well to support the integration of observation models. Both, the target and support variables comprise the final state vector to estimate.

Additionally, for each sensor which is part of the multi-sensor data fusion, the following decisions have to be made:

3) Selection of fusion schema: For sensor observations which are directly derived from raw measurements a measurement-to-track fusion by performing a Bayes filter update step is used. If the sensor observations are already preprocessed and refined inside of the sensor a track-to-track fusion approach is more appropriate [17].

4) Implementation of observation model: The Bayesian filtering framework incorporates new observations at time \(k \) by a probabilistic transformation of the predicted (time synchronized) estimate \(x_k\), which is described by the target and support variables, to the measurement space \(z_k\). Hence, this sensor-specific transformation function \(h(\cdot)\) needs to be defined.

5) Design and implementation of detection model: The detection model is the conditional distribution

\[
P(n_x|x, \exists x = \exists),
\]

(1)

where \(n_x\) denotes the number of measurements that are expected from an object and \(x\) is the state of that object. Simply stated, the detection model gives the probability that an existing object with state \(x\) generates 0, 1, \ldots, \(n\) measurements. Typically, the

\(^2\)This work assumes that a MOT algorithm is used which estimates the existence of an object along with the dynamic states.
6) Track management: The initialization of new object/track hypotheses is tightly coupled with the addressed use-case. For example, a typical choice is to propose new tracks based on the not associated measurements after each sensor’s duty cycle. The same applies to the deletion of already existing tracks.

IV. Plausibility Checking of CAMs

The main objective of this work is to demonstrate how multi-sensor data fusion can leverage consistency and plausibility checking for perception sensor data. In particular, we are interested in independently validating the contents of CAMs as delivered by connected vehicles with on-board perception sensors. Hence, our approach does not rely on a cooperative reputation system to assess individual V2V entities. Instead we aim to derive a significant measure of trustworthiness solely based on the sensor information available in the host vehicle. Consequently, the proposed approach only delivers representative results when the V2V entity which should be validated is inside of the surveillance area of the on-board perception sensors. Thus, the system is able to identify three different states of a V2V entity: neutral which indicates that the V2V entity is not inside of the sensor’s FOV and thus no plausibility checking can be performed. A valid state implies that the sensor information transmitted by the V2V entity does comply with the observations of the in-vehicle perception sensors. Finally, the invalid state signals that a V2V entity’s sensor information is not consistent with the on-board observations.

In MOT there is already a concept called track score of assigning a probabilistic confidence measure to a track hypotheses which is usually computed over time and considers sensor characteristics [17]. According to different authors, practical implementations use this measure to derive the existence or relevance of tracks. The principle idea is that successful consecutive assignments of sensor measurements to a track hypotheses should lead to a high confidence and finally confirm a track hypotheses to represent a real object. Dependent on the selected implementation of a MOT, the existence is either estimated separately from the dynamic parameters or in an integrated manner. Basically, the existence measure is expected to increase in the surveillance area of a sensor by considering the sensor characteristic such as the detection probability $P_D$. On the other side, if an expected sensor observation of a track hypotheses is not available over a period of time, the hypotheses is likely to be not relevant and should be removed.

This concept naturally applies to the generic case of multi-sensor data fusion as well. Here, the existence of a track is updated by all available sensors which are observing the area the target is supposed to be. Again, if a particular sensor observation is not available (considering $P_D$) the existence/relevance of the track will be lowered. Consequently, this can be used to perform a plausibility check of sensor observations.

As we are assessing the plausibility of the V2V messages, only the CAMs are used for proposing new track hypotheses inside of the MOT. Afterwards, the local perception system is in charge of validating these newly created objects. Of course, subsequent CAMs will be used to update the existing track hypotheses as well. Apparently, this only works as long as these object are within the FOV of the front camera. Consequently, the plausibility checking concept implemented in the paper at hand is restricted to a rather limited FOV. However, this could be easily extended by plugging-in addition perceptions sensors such as rear or blind sport radars to achieve a 360 degree coverage. In order to correctly implement the existence estimation it is important to provide an appropriate detection model which considers the true detection and false alarm rate of each particular sensor. Generally speaking, the detection model gives a probability which specifies how likely it is that a particular object hypothesis causes an observation in the considered measurement domain. The detection model used to describe the behavior of connected vehicles is different to conventional perception sensors: for example, a classical FOV does not exists. At best, an approximation can be provided by an assumed maximum communication range $d_{\text{range}}$ which describes the radius of a circle centered at the position of the communication antenna at the ego vehicle. In section VI-A, an empirical detection model which considers the observed packet error rates in the conducted test drives is presented.

Finally it is worth mentioning that with this approach two types of inconsistencies can be detected: First, an artificially injected ghost vehicle which is not really present on the road as shown in [5] can be detected. Second, temporary sensor failures can be detected. For example, the low-cost GPS devices usually used to provide the absolute position information for the CAMs might be subject to multipath as present in urban areas. This temporary disturbance may significantly offset the position inside of the CAMs.

V. Implementation

The paper at hand follows a similar approach as recently presented in [18] where a system design for multi-sensor data fusion with a flexible configuration is proposed. However, the aforementioned system does not consider sensor input from CAMs and thus has not to deal with the challenge of integrating CAMs with absolute position measurements.

A. State Space Quantities

A fundamental fist step is the selection and definition of the state space quantities, that is the target and support variables, which should be estimated for each object by the MOT algorithm. We propose to use a Constant Velocity (CV) state space [19] which is augmented with the dimensions (width $w$ and length $l$) of the object, thus describing a planar box:

$$
\mathbf{x} = (x \ y \ \vartheta \ v \ \mid \ w \ l)^T,
$$

where $x$ and $y$ denote the relative position of the tracked object with respect to the ego vehicle frame (which originates at the middle of the rear axle) and $\vartheta$ represents the heading deviation of the track with respect to the ego heading. Moreover, the state space contains the absolute magnitude of the track’s velocity indicated by $v$. For the sake of simplicity the distance between the vehicle origin and the rear bumper is assumed a fixed value of -0.5 m. After defining the state space quantities, a dynamic model which describes the evolution of a track over time needs to be selected. For the dynamic parameters the CV vehicular motion model as described in [19] can be directly used while the dimensions are assumed to be constant. Finally, each track is hybridized with an additional binary random variable $\exists x = [\text{3, 3}]$ which represents the existence of the track given its associated noisy sensor observations. In this work, $\exists x$ is
estimated based on \textit{sequential probability ratio testing} (SPRT) as explained in [17]. Basically, this represents an independent discrete Bayes filter which estimates the existence based on associated and missed sensor observations.

\section{MobilEye Sensor}

\subsection{Observation Model:}
The MobilEye camera is an aftermarket system which implements an acoustic forward collision warning. Therefore, the detected objects are described in a particular coordinate frame which is appropriate to fulfill the anticipated function. The system is calibrated in such a way, that independent of the particular mounting position of the sensor, each vehicle detection \( m_{\mathbf{x}} \) describes the distance \( r \) and angle \( \varphi \) from the rear bumper of the detected object to the front bumper of the ego vehicle. Moreover, a range rate \( r \) is included. It is worth mentioning that the provided width observation was neglected, thus, the implemented observation model for the MobilEye sensor is rather similar to a classical radar observation model:

\begin{equation}
\mathbf{z} = \mathbf{h}(\mathbf{x}) = \begin{pmatrix} r \\ \varphi \\ r \dot{r} \end{pmatrix} = \begin{pmatrix} \sqrt{(x - x_r)^2 + (y - y_r)^2} \\ \arctan \frac{y - y_r}{x - x_r} \\ v \cdot \cos(\vartheta - \varphi) - v \cdot \cos(\varphi) \end{pmatrix},
\end{equation}

where \( x_r \) and \( y_r \) are the known offsets between the front bumper and the vehicle origin of the ego vehicle. It should be noted that the variable \( v \) which denotes the velocity of the ego vehicle is explicitly available inside of the measurement model and not taken from the state vector.

\subsection{Detection Model:}
According to [10] the MobilEye system has a sensor FOV of 47° with a detection probability of \( P_D = 0.995 \). However, for the used test vehicle this angular range was not symmetric. By empirical tests the minimum and maximum angle have been set to -30° and 17°, respectively. Moreover, the minimum and maximum lateral detection range was set to 10 m and 50 m. As the MobilEye system typically only emits one detection per real object, the detection model was implemented for \( n_z = 1 \) only (observation inside FOV) and \( n_z = 0 \) (outside of FOV or occluded).

\section{Incorporation of CAM Detections}

\subsection{Transformation of CAMs to Relative Measurements:}
Due to the fact that the MOT is estimating the objects in relative coordinates with respect to the coordinate frame of the ego vehicle as proposed in [18], the absolute position measurements contained in the CAMs cannot be directly incorporate by the tracker with a classical observation model. Therefore, an additional transformation step is introduced which transforms the CAMs to relative measurements which are compatible with the previously defined state space of the tracker. In order to perform this transformation the position and heading of the ego vehicle needs to be known in absolute coordinates (e.g. UTM). In our work we are taking these values \( \mathbf{x}_{ego} = (x_{ego} \ y_{ego} \ \vartheta_{ego})^T \) at time \( k \) from a high-frequency GNSS/IMU-based ground truth sensor which is installed at the ego vehicle. The spatial alignment of CAMs at time \( k \), that is, the transformation of the absolute position from the remote vehicle’s CAM to a relative distance vector \( \mathbf{d}_k = (\Delta x \ \Delta y \ \Delta \vartheta)^T \), is based on the idea proposed in [7]. Basically, this is an adoption of the \textit{absolute position differencing} (APD) as explained in [20] that accounts for the measurement uncertainties. Therefore, a temporary joint state space \( \mathbf{x}_k \) which comprises \( \mathbf{x}_{ego} \) of the ego vehicle and the position and heading from the received CAM \( \mathbf{x}_{cam} = (x_{cam} \ y_{cam} \ \vartheta_{cam})^T \) in Cartesian coordinates (e.g. UTM) is constructed by concatenation:

\begin{equation}
\mathbf{x}_k = \begin{pmatrix} \mathbf{x}_{ego} \\ \mathbf{x}_{cam} \end{pmatrix}, \quad \mathbf{P}_k = \begin{pmatrix} \mathbf{P}_{ego} & \mathbf{C}_{ego, cam} \\ \mathbf{C}_{ego, cam}^\top & \mathbf{P}_{cam} \end{pmatrix}.
\end{equation}

Apparently, the mean and covariance parameters inside of (4) can be directly taken from the individual densities of the ego and remote vehicle. However, the correlation \( \mathbf{C}_{ego, cam} \) between both estimates, which are typically based on the GNSS observations at each vehicle, is generally not known and is therefore considered as a process parameter. In [7], a rather pessimistic choice with \( \mathbf{C}_{ego, cam} = 0 \) is made. Based on this concatenated state \( \mathbf{x}_k \), a transformation function \( f(\cdot) \) which projects the absolute CAM position and heading to the distance vector \( \mathbf{d}_k \sim \mathcal{N}(\mu_d, \mathbf{P}_d) \) is given by \( f : \mathbf{x}_k \mapsto \mathbf{d}_k \):

\begin{equation}
\mathbf{d}_k = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta \vartheta \end{pmatrix} = f(\mathbf{x}_k)
\end{equation}

\begin{equation}
= \begin{pmatrix} \sin \vartheta_{ego}(x_{cam} - x_{ego}) + \cos \vartheta_{ego}(y_{cam} - y_{ego}) \\ -\cos \vartheta_{ego}(x_{cam} - x_{ego}) - \sin \vartheta_{ego}(y_{cam} - y_{ego}) \\ \vartheta_{ego} - \vartheta_{cam} \end{pmatrix}
\end{equation}

Although, both states are represented by Normal distributions, the transformation function (6) is non-linear. Hence, the computation of the PDF for \( \mathbf{d}_k \) can be performed by an \textit{unscented transformation} in a consistent way which correctly considers the uncertainties of the 2D positions and the heading [21]. In [22], typical accuracies or these relative distance vectors, which are calculated from two absolute GPS fixes, are given.

\subsection{Observation Model:}
After the absolute position and heading measurements from the CAM have been transformed into the local coordinate frame of the ego vehicle as described in the previous section, the observation model of the relative CAM can be formulated in a straightforward manner:

\begin{equation}
\mathbf{z} = \mathbf{h}(\mathbf{x}) = \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta \vartheta \\ v \\ w \end{pmatrix} = \begin{pmatrix} x \\ y \\ \vartheta \\ v \\ w \end{pmatrix}.
\end{equation}

Apparently, this one-to-one relation of the state and observation space is only true as long as the position inside of the CAMs is related to the same vehicle origin (middle of rear axle). For this work, this was ensured by configuration of the utilized ITS-G5 devices.

\subsection{Detection Model:}
Compared to a typical on-board sensor, the communication range of ITS-G5 devices is rather large and can be approximately described by a circular area around the ego vehicle with the radius \( d_{\text{max}} \) [14]. However, in contrast to the MobilEye sensor, there are no exactly defined detection areas with a fixed detection probability \( P_D \). Instead, the detection probability is dependent on several parameters such as the communication distance \( d \) and environmental conditions. Consequently, the actual \( d_{\text{max}} \) might be smaller in high congested urban areas than on rural motorways. In this work the test drives where conducted under constant
sub-urban conditions. Thus, we will use a fixed PDF which describes the reception probability of a CAM dependent on the communication range. The particular statistics of this PDF describing \( P_D(d) \) for the conducted test drives are derived in section VI-A from empirical data.

### D. Time Synchronization & Fusion Schema

Due to the asymmetric sensor latencies between the MobilEye system and CAMs, the multi-sensor data fusion will receive the individual observation out of order regarding their time of validity—an OOSM scenario is present. In [7] the OOSM problem was mitigated by predicting the object state contained in the CAMs at reception time \( t_r \) to time of the most recent state estimate \( t_k \) inside of the tracker. The CAM prediction was implemented by using a constant turn rate and acceleration (CTRA) motion model and afterwards incorporated into the track list by track-to-track fusion. Unfortunately, such a temporal alignment strategy cannot be applied in this work as the CAMs of the used ITS-G5 devices do not contain the momentary acceleration and turn rate. Thus, a forward prediction of the CAM solely based on the position, the heading and the velocity for a time span of up to 800 ms will generate rather inaccurate results [23]. Therefore, we decided to use a buffering strategy as explained in [24] to generate first baseline results which can be used later on for further evaluations and comparisons. Due to the fact that the used ITS-G5 devices derive the vehicle dynamics from an attached low-cost GPS device and do not apply any filtering refinement, the CAMs are incorporated by measurement-to-track fusion.

### VI. Quantitative Results

In this section the results of the empirical evaluation are presented. The three utilized test vehicles are equipped with different sensors for vision, range, high-accurate localization and communication. In [25], a comprehensive description of these prototyping platform is available. In addition, all vehicles have been equipped with a communication device based on the Atheros AR5414A-B2B chipset with custom firmware to support communication on the ITS-G5 control channel. On top of this device we ran an ETSI compliant implementation of GeoNetworking, Basic Transport Protocol (BTP) and Cooperative Awareness Basic Service (CA BS) from Hitachi Solutions. This implementation was used as reference implementation for the ETSI conformance tests and was one of the best scoring implementations during all ETSI Plugtests™ to date.

#### A. Packet Delivery Ratio

The packet delivery ratio (PDR) is the ratio of received packets to the number of sent packets. Let \( \text{sent}_i \) be the number of sent packets by vehicle \( i \) and \( \text{rcvd}_i^j \) the number of received packets from vehicle \( i \) by vehicle \( j \). The packet delivery ratio at vehicle \( j \) for vehicle \( i \) is then calculated as \( \frac{\text{rcvd}_i^j}{\text{sent}_i} \). Figure 1 shows the PDR as a function of the inter-vehicle distance. The results show as expected a decreasing PDR for increasing distances. Our results are however limited to 50 m because the vehicles were driving too close in our tests to collect statistically significant results for any larger distance, even though we have seen transmissions and receptions at larger distances. Our results show a steeper decrease of the PDR compared to [12]. This result is likely caused by the used chipset which was originally not designed for high-speed scenarios and shadow fading caused by the vehicle roof bars. Finally, this data was used to derive the characteristics of the reliability of the ITS-G5 wireless channel by giving a detection probability \( P_D \) that is dependent on the communication range.

#### B. Assessment of Valid Situations

In a first step, the proposed MOT approach for plausibility checking is evaluated for a scenario which contains consistent measurement data of both sensor systems. The target vehicle, which is observable by both sensors at the beginning, is driving in front of the ego vehicle. After a period of a straight motion the target vehicle turns left and thus leaves the FOV of the MobilEye system. However, the CAMs are still received as the communication range is rather large compared to the on-board perception sensor. In Fig. 2, the scenario is depicted after the left turn maneuver of the leading vehicle. In order to compare the influence of the multi-sensor data fusion we first evaluated the tracking results which can be achieved with using CAMs only. Figure 3 illustrates the existence estimate of the V2V entity. Every time a new CAM is received (indicated by the red circles) and successfully assigned to an existing track the
existence estimate is updated and increased to a maximum value of app. 15 log-odds. Apparently, in between of the CAM existence, is lower compared to the middle part. This can be explained by the decreasing reliability of the wireless channel as already stated in section VI-A which directly depends on the distance between both vehicles. It can be summarized that for this use-case the V2V entity has a rather high existence value in the range of 15 log-odds but is still only confirmed as a neutral object as the leading vehicle is solely observed by the V2X sensor. In addition, Fig. 4, illustrates the same sequence but now with a multi-sensor data fusion among CAMs and the MobilEye system (vision detections are indicated by green crosses).Apparently, the MobilEye systems only perceives the leading vehicle for $t < 3.8$ s which is before the left turn maneuver. As both sensors deliver consistent observations the max. estimated existence now reaches a higher value of app. 16.5 log-odds. Thus, the CAMs of the V2V entity can be confirmed as valid. Starting from $t > 3.8$ s, the leading vehicle leaves the FOV of the MobilEye system and the existence estimate reverts to 15 log-odds again. In terms of plausibility checking this corresponds to a neutral classification result.

C. Assessment of Invalid Situations

In the evaluated scenario, which is similar as in [5], we assume an attacker which has tampered a system and is able to send CAMs with valid certificates and signatures while yet being able to control the message content. Without any plausibility check, a receiver could wrongly consider these messages as valid and as such, an attacker is able to generate a number of attacks, including the generation of ghost vehicles by changing the location data of broadcasted CAMs. Such situation is illustrated in figure 5. An attacker $A$ is positioned alongside the road and is broadcasting CAMs with forged position $A_{CAM}$. Vehicle $R$ is driving on the road and receives the broadcasted CAMs of attacker $A$. In addition, vehicle $R$ detects objects in front of the vehicle using its built-in camera. In this situation, the camera does not detect vehicle $A$ at position $A_{CAM}$ and for that reasons the plausibility check of Section V should classify the CAM data as implausible. In Fig. 6 the plausibility checking result for the spoofing CAM use-case is illustrated. At the beginning ($t < 3.8$ s) the faked CAM pretends that the injected vehicle is outside of the FOV of the on-board perception system. In fact, it is approaching from behind and starting to overtake the host vehicle. For this first part of this sequence the proposed system classifies the V2V entity as neutral similar to VI-B. However, starting from $t > 3.8$ s the injected vehicle performs a cut-in maneuver and thus enters the FOV of the MobilEye system. Instantly, the MOT starts to estimate a low existence value for this track as the object cannot be confirmed by the MobilEye system which does not detect a corresponding vehicle. Apparently, the track’s existence is still updated by the faked CAMs but never reaches similar values as observed in section VI-B—the V2V entity is classified to invalid. Consequently, the ghost-vehicle is reliably detected and can be marked as non trustworthy.
VII. CONCLUSION & OUTLOOK

In this paper, we evaluated a MOT-based approach for plausibility checking of CAMs as provided by ITS-G5. Therefore, we applied a multi-sensor data fusion among a MobiEye on-board perception sensor and V2V communication which is typically used to achieve an extended perception. Both sensors are considered off-the-shelf hardware and have been fused with a MOT algorithm as provided by BASELABS Create. We demonstrated that the existence estimate which is naturally included in MOT applications can be directly used to infer the plausibility of CAMs. Therefore, we evaluated three distinct use-cases which comprise plausible and implausible scenarios. It has been proven that the estimated existence values clearly correspond to the three used sensors states neutral, valid and invalid. Besides demonstrating that MOTs can leverage plausibility checking of sensors, we faced several challenges in multi-sensor data fusion: for example, the particular sensors’ behavior were characterized by dedicated sensor models which describe the FOV and the detection probability. For the V2V system an empirical detection probability that is dependent on the communication range has been deduced. One shortcoming of the presented system is the rather limited overlapping FOV of both sensors which leads to a rather narrow surveillance area where plausibility checking can be performed. Nevertheless, the proposed system correctly classifies V2V entities outside of that area as neutral. By adding more perception sensors this limitation can be removed without any need to modify the plausibility checking strategy. Moreover, the paper at hand aimed to evaluate the capabilities of MOT for plausibility checking in general. Therefore, we deliberately mitigated the issue of OOSM by using a buffering strategy. For a plausibility checking which has to operate under real-time conditions, another OOSM strategy—perhaps similar to [7]—is more appropriate. Finally, it is worth mentioning that the APD method used to transform the absolute position information of the CAMs to a relative observation as required by the MOT, depend to a big extend on the number of common satellites between both vehicles. In our work we took the observed statistics for the relative measurements from [22]. For further work, this statistics should be derived online.

ACKNOWLEDGMENT

This work was supported by the European Commission under AutoNet2030, a collaborative project part of the Seventh Framework Programme for research, technological development and demonstration (Grand Agreement no 610542). The test vehicles used to collect measurement data were provided by Technische Universität Chemnitz.

REFERENCES