

A Plausibility-Based Fault Detection Method for High-Level Fusion Perception Systems

FLORIAN GEISSLER¹ (Member, IEEE), ALEXANDER UNNERVIK^{1,2}, AND
MICHAEL PAULITSCH¹ (Senior Member, IEEE)

¹Intel Labs, Intel Corporation, 85579 Neubiberg, Germany

²Department of Biometrics Security and Privacy, Idiap Research Institute, 1920 Martigny, Switzerland

CORRESPONDING AUTHOR: F. GEISSLER (e-mail: florian.geissler@intel.com)

This work was supported by the German Federal Ministry of Transport, Building and Urban Development (BMVI) within the projects KoRA9 under Grant 16AVF1032A and Providentia++ (P++) under Grant 01MM19008.

ABSTRACT Trustworthy environment perception is the fundamental basis for the safe deployment of automated agents such as self-driving vehicles or intelligent robots. The problem remains that such trust is notoriously difficult to guarantee in the presence of systematic faults, e.g., non-traceable errors caused by machine learning functions. One way to tackle this issue without making rather specific assumptions about the perception process is plausibility checking. Similar to the reasoning of human intuition, the final outcome of a complex black-box procedure is verified against given expectations of an object's behavior. In this article, we apply and evaluate collaborative, sensor-generic plausibility checking as a mean to detect empirical perception faults from their statistical fingerprints. Our real use case is next-generation automated driving that uses a roadside sensor infrastructure for perception augmentation, represented here by test scenarios at a German highway and a city intersection. The plausibilization analysis is integrated naturally in the object fusion process, and helps to diagnose known and possibly yet unknown faults in distributed sensing systems.

INDEX TERMS Intelligent transportation systems, plausibility checks, automated driving, smart infrastructure, dependability, sensor data fusion.

I. INTRODUCTION

THE VERIFICATION of a complex environment model without explicit knowledge of – or access to – the full processing pipeline is one of the key challenges of modern perception systems. Causes for incorrect observations can involve electronic system malfunctions of sensors or unintended system behavior, i.e., misperception in the absence of malfunctions such as misuse or incomplete information [1]. Notably, object detection further relies more and more on machine learning techniques such as deep neural networks, which are known to be sensitive to particular faults that defy a classical safety assurance (e.g., erroneous training data, ontology gaps, adversarial noise, etc.) [2]. We here refer to a

wider class of root causes for the incorrect environment perception of an intelligent sensing device as *perception faults*. The detection and evaluation of such perception faults is the prerequisite for a safety analysis concerning the user of the environment model, e.g., an automated vehicle (AV), as regulated by standardization bodies such as the ISO 26262 [3] or ISO/PAS 21448 (SOTIF) [4].

Among quantitative model-based verification approaches [5], [6], we see a characteristic trade-off between the complexity of the verification process and the coverage of potential faults: traditional fault diagnosis techniques, based, e.g., on a direct hardware or software redundancy, provide variable sets of high completeness, leading to a potentially high fault coverage. However, a replication of components is neither cost-efficient nor feasible in many scenarios. In contrast to that, analytical redundancies inferred from known functional dependencies

The review of this article was arranged by Associate Editor Abdulla Hussein Al-Kaff.

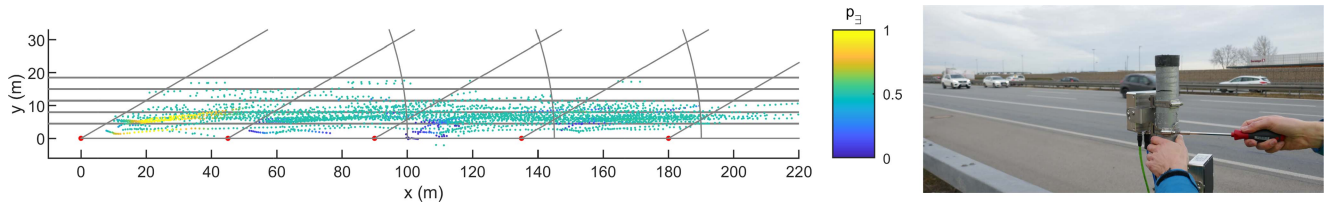


FIGURE 1. An example experiment performed in the KoRA9 project, see also [7]. (Left) Object tracks accumulated over time at a highway test segment. Plausibilization can be applied to identify false positive tracks, and reveal inconsistencies in regions of sensor overlaps. (Right) One elementary perception fault is the misorientation of individual sensors at the roadside.

TABLE 1. Typical safety-critical perception errors and faults, as identified from real smart infrastructure experiments. Faults with similar local effects in the system are combined to fault classes. The selected highlighted faults are injected in simulation in Section V to test plausibility-based fault evaluation.

Class	Local effect (Error)	Potential fault cause
1	Sensor reports readings with incorrect angles	<ul style="list-style-type: none"> • Misoriented sensor pose (e.g. poor calibration, air blasts of passing vehicles, accidents, malicious intervention) • Incorrect signal processing
2	Sensor has a higher false positive rate than expected	<ul style="list-style-type: none"> • Defect component causing increased thermal noise • Incorrect parametrization of algorithmic component (e.g. tracking or detection threshold) • Incorrect signal processing (e.g. ground plane filter, side lobe artefacts) • (Local) spoofing • Increased signal multipath propagation (e.g. highly reflecting object in sensor field of view)
3	Sensor has a higher false negative rate than expected	<ul style="list-style-type: none"> • Sensor surface polluted • Incorrect parametrization of algorithmic component • Incorrect signal processing

on the input variables, or a priori rules, allow to check output variables with greatly reduced complexity. For that reason, *plausibility checks* that verify complex environment variables against a range of expected outcomes are promising candidates for the verification of black-box perception systems [2]. On the other hand, functional relationships are often unknown or ambiguous, resulting in a lack of diagnostic residuals which can diminish the fault coverage. It is therefore of great interest to numerous safety-critical applications if and how such an approach is able to diagnose realistic perception faults.

In this article, we investigate the suitability of plausibility checking for the evaluation of perception faults in the context of a distributed sensor network (DSN) for collaborative environment perception. Our use case is a roadside sensor network that supports automated vehicles with complementary high-quality data [8], a concept also known as *smart infrastructure*, that represents one of the driving forces in the fast-growing market of intelligent transportation systems and smart cities [9]. Verification methods as the one studied in this article allow to quantify trust in the integrity of the perception data, and will be required to achieve certifiably safe infrastructure systems in the future (see, e.g., [10]).

DSNs typically adopt a *decentralized* tracking architecture, where each sensing device performs a series of processing steps and eventually provides with lists of tracked objects. The data from multiple sensors are then combined in a central node in the course of high-level sensor data fusion (SDF), or

object fusion. Starting from the architecture of [11] and the track-to-track (T2T) fusion implementation of [12], we here present a plausibility checking framework that integrates naturally into the object fusion process. Uncertainties are handled within the Dempster-Shafer (DS) theory of evidence [13], [14]. Besides an instantaneous plausibility assessment, we provide with statistical metrics in order to identify systematic faults. While additional customized plausibility checks can be readily incorporated due to the modular structure of the framework, we here focus on sensor-generic checks that are not limited to a specific modality of the underlying sensor network (most commonly camera, Radar, Lidar), in order to keep the methodology widely applicable.

We test our plausibility-based fault diagnosis in simulation, using the injection of typical perception faults as identified from real-world infrastructure experiments. In particular, we use insights from the German publicly funded projects KoRA9 [15] and P++ [16]. In KoRA9, environment perception with a Radar-only roadside sensor network is studied at a German highway, see also Fig. 1. The P++ project on the other hand is set up to monitor a cohesive test area of highway, rural, and urban roads with a variety of sensor modalities including camera and Radar. The Tab. 1 gives examples of empirical, potentially safety-critical faults that challenge roadside perception systems due to their systematic, erratic, and unsignaled nature. Selected fault instances that are representative of broader classes with similar local effects will be used in this article to identify characteristic signatures.

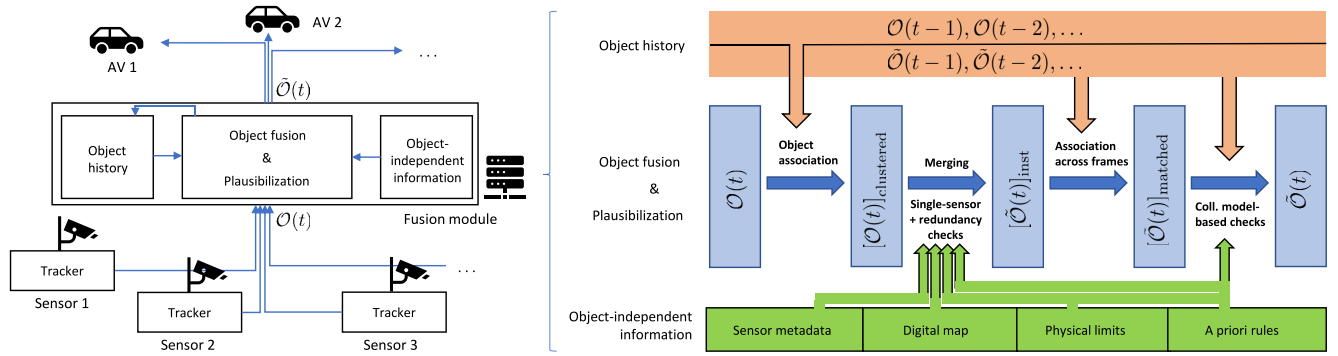


FIGURE 2. (Left) System sketch of the decentralized tracking architecture for collaborative perception. (Right) Schematic of the fusion module, performing combined object fusion and plausibilization.

In summary, the novel contribution of our work is twofold:

- While in [11] separate modules for T2T and existence fusion are envisioned, we here describe a more efficient combined approach, meaning that plausibility checks and existence estimation are naturally integrated into the object fusion process. We further introduce additional checks that correlate the *track status* of an object with other observed attributes. This allows for a more differentiated and realistic situational assessment.
- We demonstrate that plausibility-based metrics represent a valid strategy to detect realistic perception faults in the given setup. The signatures of three empirical faults are studied in simulation, and characteristic fingerprints are identified. We provide an interpretation of the plausibility metrics with respect to the underlying source of implausibility.

The article is structured as follows. In Section II we review related literature, before we present our model in Section III, including the process integration and the various plausibility checks. Next, Section IV describes the method for the detection of perception faults. The respective plausibility signatures are analyzed in Section V for two different scenarios, before we conclude in Section VI.

II. RELATED WORK

First, it is key to note the difference between measurement uncertainty, plausibility, and risk. The former refers to the statistical inaccuracy of a sensor observation, and its propagation throughout the system. Plausibility represents a notion of trust in a hypothesis (e.g., “object exists”), that is established by checking a piece of information against expectations based on a priori knowledge or consistency with other sources. As given in Section III, the plausibility of an existence hypothesis can be conveniently mapped onto a *probability of existence* measure. Risk on the other hand refers to an event, which can occur with a certain likelihood and a given impact on a user. Probabilities of existence and uncertainties are valuable inputs to a risk analysis, but can not provide with any context or user impact. In this article, we are not concerned with the explicit evaluation of measurement uncertainty or risk, but with plausibilization.

Different hierarchies of plausibility checking are to be distinguished, the authors of [17] define for example three categories of increasing complexity: i) Single signal monitoring, ii) redundancy-based checking, and iii) model-based checking. Interesting examples for the latter category are, e.g., the plausibilization of trajectory curvature against possible lateral acceleration [18], or unphysical overlaps of object shapes [19]. Among the established mathematical methods to quantify confidence, in particular Bayesian belief networks and the DS theory have proven to be appropriate tools [20].

In the domain of safety for automated driving, plausibility checking methods have been applied to verify the environment perception of (overlapping) in-vehicle sensors – including camera, Lidar, and Radar modalities – and to infer a probability of existence metric [11], [21], [22]. Most of the applied checks there are single-signal and redundancy checks. This assessment can take place at different stages of the SDF, and the checks take different forms depending on the available features. High-level object fusion architectures with DS plausibilization have for example been studied in [11], [21]. Here, the position of an object is for example checked against the position of other objects to identify implausible occlusions, and a non-uniform belief is assigned across a sensor’s field of view (FoV). Another body of literature harnesses plausibility checking to address secure vehicle-to-everything (V2X) communication, see, e.g., [19], [23], [24]. Information that is transmitted to a vehicle from external sources, such as the infrastructure or other vehicles, is plausibilized to detect misbehavior or spoofing. A basis for verification are the parameters contained, e.g., in the format of a basic safety message (BSM), including position, speed, acceleration or object dimensions [18], but also the sensing and communication range.

III. MODEL

A. OBJECT REPRESENTATION

The Fig. 2 gives a schematic overview of a high-level fusion system. Objects tracked by the individual sensors are passed to a central *fusion module*. In the highlighted infrastructure use case, this component runs for example on an edge cloud server at the roadside. We here focus on an integrated model

of object fusion and plausibilization as the core functionality of this module.

Let us define a tracked object reported by an individual sensor as a structured list O , while \mathcal{O} is the *local object list*, i.e., the set of all individually registered objects at a common time step,

$$O = (i, n, t, \mathbf{x}, \mathbf{P}, \mathbf{d}, \mathbf{tr}_{\text{status}}, \dots) \equiv O_n^i(t), \quad (1)$$

$$\mathcal{O}(t) = \left\{ \left\{ O_n^i(t) \mid n = 1, 2, \dots \right\} \mid i = 1, 2, \dots \right\}. \quad (2)$$

The minimal required object attributes in our model are: i sensor index, n track identifier (ID), t time stamp, \mathbf{x} object state, \mathbf{P} state covariance matrix, \mathbf{d} object dimensions, and $\mathbf{tr}_{\text{status}}$ tracking status information. Any object can be uniquely identified by (i, n, t) . For notational ease, we only give selected attributes if clear from the context (e.g., O^i refers to a given object reported by a sensor i , $\mathbf{x}_{O^i} \equiv \mathbf{x}^i$, etc.). We specify

$$\mathbf{x} = (x, y, z, v_x, v_y, v_z),$$

$$\mathbf{d} = (L, W, H, h),$$

$$\mathbf{tr}_{\text{status}} = (sc, cnf, cst, \dots). \quad (3)$$

Object position and velocity are represented by x, y, z and v_x, v_y, v_z , respectively. The dimension entails length (L), width (W), height (H), and heading (h) of the perceived object. In our simulation, we adopt a sequential multi-object tracking (MOT) scheme – different implementations are possible and do not affect the general model – where the track score sc is a log-likelihood ratio [25] that increases (decreases) with each incoming (missing) update. The binary value cnf indicates whether the object track is a confirmed ($cnf = 1$), or a tentative ($cnf = 0$) track. Furthermore, the binary cst incorporates information about the coasting status, i.e., whether the object track is updated at the current time step ($cst = 0$) or not ($cst = 1$).

Next, a *system object* \tilde{O} is defined to be a global tracked object, that can be observed by multiple sensors at the same time, and therefore can have multiple varying *local representations*. Similar to Eq. (2) we define the *system object list* $\tilde{\mathcal{O}}$ at a common time t ,

$$\tilde{O} = (n', t, \mathbf{x}, \mathbf{P}, \mathbf{d}, \mathbf{tr}_{\text{status}}, p_{\exists}, s_{\exists}, \dots) \equiv \tilde{O}_{n'}(t), \quad (4)$$

$$\tilde{\mathcal{O}}(t) = \left\{ \tilde{O}_{n'}(t) \mid n' = 1, 2, \dots \right\}. \quad (5)$$

Here, p_{\exists}, s_{\exists} are the respective probability existence and existence uncertainty metrics derived in Section III-D, using the concepts of belief and plausibility. The goal of the object fusion with integrated plausibility checking is the transformation $\mathcal{O}(t) \rightarrow \tilde{\mathcal{O}}(t)$, under consideration of temporal continuity. In the process, the metrics p_{\exists} and s_{\exists} are assigned to give a quantifiable estimate for trust in the observation.

B. FUSION MODULE

The integral part of this process is track-to-track fusion [25], which combines the dynamic state variables of multiple tracks. Furthermore, object fusion merges features like

dimensions, headings etc. We encounter the following steps (see Fig. 2):

- **Object association** Synchronous and confirmed local representations in \mathcal{O} are grouped to *track clusters*, $[\mathcal{O}(t)]_{\text{clustered}}$. We adopt the T2T distance calculation and the global nearest neighbor (GNN) association of [12], which can include the history of local representations for better association robustness.
- **Merging** All local object representations within one cluster are merged to a single system object [12]. The result of this fusion process is a system object list $[\tilde{\mathcal{O}}(t)]_{\text{inst}}$ (for *instantaneous*), with arbitrary IDs.
- **Association across frames** We associate the unique object labels from the previous and the current time frames via another GNN association, so that persistent objects keep the same unique ID in the matched list $[\tilde{\mathcal{O}}(t)]_{\text{matched}}$. Dimensions and existence estimations can be averaged across recent frames for improved robustness.

Our model now integrates various plausibility checks in the object fusion process to evaluate the quality of the sensor data. Extending the paradigm of [17], we have three different types of checks, see Fig. 2 and Sections III-C–III-E:

- **Single-sensor checks** are based on the attributes of local representations from a *single* sensor and result in the basic belief assignments of the DS framework. An attribute is here checked against predefined thresholds (signal-based) or known relations to other attributes (model-based).
- **Redundancy checks** compare local object representations from *different* sensors that refer to the same system object, as identified by object association. Such redundancy checks are applied after the clustering and before the merging step. Here, they are represented by the DS combination, together with the registration of misses and unexpected observations.
- **Collaborative model-based checks** aim to verify the final outcome of the fusion process against predefined rules about the expected behavior of system objects. This additional processing step transforms the output of the frame-association process $[\tilde{\mathcal{O}}(t)]_{\text{matched}}$ into the desired output $\tilde{\mathcal{O}}(t)$. Such checks may be based on instantaneous interrelations among merged object attributes, or on the evolution of merged attributes over time. In this work, we present two possible checks that can be used to tackle false positives or spoofed detections (see Section III-E). Various further checks of increasing complexity can be envisioned at this point, including, e.g., traffic-rule-compliant trajectory checking, or self-learned motion patterns.

The process diagram of Fig. 2 further illustrates which information beyond the sensor object data is required to perform the plausibility checks described in our model. Single-sensor (and partially the redundancy) checks build upon the knowledge of sensor metadata such as the sensor

field of view and mounting locations, or the expected road boundaries of a digital map. Physical limits of variables and predefined rules help to establish additional plausibility for all three types of checks.

C. SINGLE-SENSOR CHECKS

Compared to a Bayesian-based object existence estimation, the well-studied Dempster Shafer theory of evidence [13], [14], [26] postulates an extended probability space of $\Theta_{DS} = \{\emptyset, \exists, \nexists, \{\exists, \nexists\}\}$, where \emptyset is the null set, \exists and \nexists are the propositions that a given object exists or does not exist, respectively, and the element $\{\exists, \nexists\}$ represents the possibility that the existence of the object is unknown. Basic belief assignments (BBA) give rise to belief masses that indicate the degree of belief in a proposition according to a given sensor. Explicitly, $m_i(X|\tilde{O})$ denotes the belief mass for the proposition $X \in \Theta_{DS}$ of \tilde{O} based on the local observation O^i of sensor i .

The following sensor-generic plausibility checks define our BBA,

$$\begin{aligned} m_i(\exists|\tilde{O}) &= p_{\text{trust}}^i p_{\text{FoV}}(O^i) p_{\text{occ}}(O^i) \\ &\quad \times p_{\text{ex}}(O^i) p_{\text{dm}}(O^i) p_{\text{val}}(O^i), \\ m_i(\nexists|\tilde{O}) &= p_{\text{trust}}^i p_{\text{FoV}}(O^i) p_{\text{occ}}(O^i) \\ &\quad \times [1 - p_{\text{ex}}(O^i) p_{\text{dm}}(O^i) p_{\text{val}}(O^i)], \\ m_i(\{\exists, \nexists\}|\tilde{O}) &= 1 - [m_i(\exists) + m_i(\nexists)]. \end{aligned} \quad (6)$$

Note that despite the formal equivalence to some of the models of [21], [22], [27], [28], the BBA factors in Eq. (6) take a different form due to the fact that we study tracked objects instead of detections, which leads to important additional interrelations of the attributes of O^i . The plausibility factors p can be grouped into two categories: a reduced value of p_{trust} , p_{FoV} , p_{occ} will support ignorance of the system object, while a low weight on p_{ex} , p_{dm} , p_{val} favors the proposition of non-existence. The individual factors defining the BBA are explained in more detail in the following.

First, p_{trust}^i is an object-independent trust factor for the sensor i , which can be interpreted as its typical object detection capability [22].

The factor p_{FoV} checks whether the reported object is in the three-dimensional FoV, here defined as $\text{fov}^i = (r^i, \omega^i, \psi^i)$, where r is the range, ω and ψ the horizontal and vertical view angle, respectively. Given that at least one point across the body of an extended object (here we check the eight corners of the bounding box, object center, and two additional points in the middle front and rear of the bounding box) is found to be within the expected FoV, we denote $(x^i, d^i) \in \text{fov}$. Otherwise, a distance metric D is applied to measure the closest distance from the object center x^i to the FoV. For example, $D(x^i, r^i)$ is the difference of the radial distance of x^i and the sensor range, and analogous

for ω, ψ . We then define

$$p_{\text{FoV}}(O^i) = \begin{cases} 1 & \text{if } (x^i, d^i) \in \text{fov}^i, \\ e^{-\sum_{n=1}^3 D(x^i, \text{fov}_n^i)/(\text{fov}_n^i/2)} & \text{otherwise.} \end{cases} \quad (7)$$

This choice of weighting is motivated by two additional considerations: i) Tracked objects can leave the FoV while coasting, in which case it is legitimate to have $p_{\text{FoV}} > 0$; ii) Tracks far away from the original FoV should be suppressed, whether coasting or not (otherwise unrealistic, e.g., spoofed tracks that pop up at an arbitrary remote location and start coasting would be supported).

Next, p_{occ} verifies if an object is rightfully observable in relation to other objects perceived by the same sensor. If at least one of the checked surface points of O^i is in the line of sight of sensor i , given all other confirmed observations of i , we denote $(x^i, d^i) \in \text{LoS}^i(\{O^i\})$. If that this is *not* the case, and at the same time the track is not coasting, the object is occluded and the belief in this observation is forfeited. A track that is coasting in the shadow of another object will not be affected.

$$p_{\text{occ}}(O^i) = \begin{cases} 0 & \text{if } (x^i, d^i) \notin \text{LoS}^i(\{O^i\}) \wedge \text{cst}^i = 0, \\ 1 & \text{otherwise.} \end{cases} \quad (8)$$

Furthermore, $p_{\text{ex}}(O^i)$ is the existence probability of the object as estimated by sensor i . Here, we infer p_{ex} directly from the track score [25], [29],

$$p_{\text{ex}}(O^i) = 1 / (1 + e^{-\alpha \text{score}^i + \beta}), \quad (9)$$

where α and β are tunable coefficients to control the slope of the sigmoid function. We here chose α and β such that a newly initialized, tentative track starts with $p_{\text{ex}} = 0.9$, and increases to $p_{\text{ex}} = 0.99$ if the score reaches the confirmation threshold. A detection missed by sensor i will be handled as $p_{\text{ex}}(O^i) = 0$, leading with Eq. (6) to $m_i(\exists|\tilde{O}) = 0$, $m_i(\nexists|\tilde{O}) = p_{\text{trust}}^i$, and $m_i(\{\exists, \nexists\}|\tilde{O}) = 1 - p_{\text{trust}}^i$.

The factor $p_{\text{dm}}(O^i)$ compares the position of an object to the expected road boundaries. Let dm be a digital grid map with binary values corresponding to road (1), and non-road (0) squares. For each object, the metric $D(x^i, dm)$ then calculates the Euclidean distance from the measured position x^i to the closest road square. As a normalization factor we use a typical lane width [30] of $W_{\text{lane}} = 3.5m$,

$$p_{\text{dm}}(O^i) = e^{-D(x^i, dm)/W_{\text{lane}}}. \quad (10)$$

Finally, we have the signal value factor p_{val} which compares object attributes against physically possible limits. In particular, we check the vertical position relative to the road, dimensions, and absolute velocity,

$$p_{\text{val}}(O^i) = \exp\left(-\sum_{A \in \{z, W, L, H, |v\}} [A^i, A_{\text{max}}]_{+} / A_{\text{max}}\right). \quad (11)$$

Here we denote $[A, B]_{+} = \theta(A - B)(A - B)$ as the positive real difference, where $\theta(\dots)$ is the Heaviside function. The

predefined maximum limit of an attribute A is represented by A_{\max} , and we choose an empirically parameter set of $z_{\max} = 3m$, $W_{\max} = 5m$, $L_{\max} = 25m$, $H_{\max} = 5m$, and $|v|_{\max} = 80m/s$.

D. REDUNDANCY CHECKS

To obtain an overall notion of plausibility of \tilde{O} , the belief masses of the individual sensors referring to the same system object are fused – note \oplus – with the help of the DS combination rule [26]. Assuming a total number of N sensors, we arrive at the fused belief masses m_F ,

$$m_F(\tilde{X}) \equiv [m_1 \oplus m_2 \oplus \dots \oplus m_N](\tilde{X}). \quad (12)$$

Iterating through all sources, any sensor that makes a local observation of \tilde{O} can possibly generate non-zero belief masses for all propositions $X \in \Theta_{\text{DS}} \setminus \emptyset$. We register *unexpected* observations when a reported object is outside the non-occluded FoV of the reporting sensor, and at the same time is non-coasting. These situations will produce a high belief in the proposition $m_i(\{\exists, \# \} \tilde{O}) \simeq 1$. If a given sensor i' has *no* local representation in the respective cluster, it matters whether it failed or was unable to perceive the object. Therefore, we check if \tilde{O} (as given by the available reports of other sensors) is in the non-occluded field of view of sensor i' , in which case we register a *miss* for i' , along with the mass $m_{i'}(\tilde{X}) = 0$ (see Section III-E). Further, a miss is as well registered (without change of the masses) for a sensor that fails to update an object in its non-occluded FoV, as can be inferred from a positive coasting status. Finally, if i' has no representation but \tilde{O} is outside its FoV or occluded, i' is an irrelevant contributor, and we have $m_{i'}(\{\exists, \# \} \tilde{O}) = 1$. Such sensors do not have any impact under the DS combination rule. Unexpected observations and misses will be used for a statistical fault analysis (see Section IV).

E. COLLABORATIVE MODEL-BASED CHECKS

We describe two additional, sensor-generic checks that help to verify the integrity of high-level perception data based on collaborative knowledge. More complex or sensor-specific model-based checks – e.g., related to object classes or trajectories – can be incorporated the same way, and are envisioned for further research.

Plausible observation history: We check whether an *increase* of belief in the object existence across two subsequent time steps is justified. This is not the case if there is *no* current observation update obtained from *any* sensor (i.e., if a system object has global coasting status $cst = 1$). The existence probability and uncertainty of a coasting object then have to be corrected appropriately, with the belief increment Δ_{hist} being transferred to the ignorance hypothesis,

$$\begin{aligned} \Delta_{\text{hist}, \exists}(\tilde{O}(t)) &= -\left[m_F(\tilde{X}) \right]_+ \\ &\quad \times cst(t), \\ \Delta_{\text{hist}, \#}(\tilde{O}(t)) &= 0, \end{aligned}$$

$$\Delta_{\text{hist}, \{\exists, \# \}}(\tilde{O}(t)) = -\Delta_{\text{hist}, \exists}(\tilde{O}(t)). \quad (13)$$

This check will prevent objects that have already left the field of view, or spoofed objects with an alleged coasting status to become more relevant.

Dimension – velocity check: The type of an object is related to its physical speed limitations. Let us here distinguish the two classes of vulnerable road users (VRU, i.e., pedestrians, bicycles, etc.), and other objects (vehicles, trucks, etc.). Since not all sensor modalities can reliably identify object classes, we use the object dimensions for a basic type estimate: VRUs are characterized by the combination of small width and length extensions, and low speed, as parametrized here by the empirical upper thresholds $l_{\text{vru}} = 2m$ and $v_{\text{vru}} = 20m/s$. Small and fast objects will be considered implausible as a consequence, and full ignorance is assumed (θ is the multi-dimensional Heaviside function),

$$\begin{aligned} \Delta_{\text{dim-vel}, \exists}(\tilde{O}) &= -m_F(\tilde{X}) \\ &\quad \times \theta(l_{\text{vru}} - W, l_{\text{vru}} - L, |v| - v_{\text{vru}}), \\ \Delta_{\text{dim-vel}, \#}(\tilde{O}) &= 0, \\ \Delta_{\text{dim-vel}, \{\exists, \# \}}(\tilde{O}) &= -\Delta_{\text{dim-vel}, \exists}(\tilde{O}). \end{aligned} \quad (14)$$

Note that object occlusions can temporarily reduce the effectively observed object dimensions, and lead to a low-plausibility assessment in Eq. (14). However, we expect this effect to be short-lived and statistically uniform across all sensors of the network, such that no individual fault patterns will emerge due to this check.

Final combined metric: The corrections from all model-based checks are combined as

$$m_F(\tilde{O}) \rightarrow \left[m_F(\tilde{O}) + \Delta_{\text{hist}}(\tilde{O}) + \Delta_{\text{dim-vel}}(\tilde{O}) \right]_0^1, \quad (15)$$

with $m_F(\tilde{O}) = (m_F(\tilde{X}), m_F(\tilde{\#}), m_F(\{\exists, \# \} \tilde{O}))$ and $\Delta(\tilde{O}) = (\Delta_{\exists}(\tilde{O}), \Delta_{\#}(\tilde{O}), \Delta_{\{\exists, \# \}}(\tilde{O}))$. In case the accumulated increments exceed the correct belief mass interval bounds, we reestablish them by the mapping $[\dots]_0^1$ applying subsequent shift and renormalization operations. The DS framework defines the quantities of overall belief and plausibility [26], however, we here work instead with the related object existence probability p_{\exists} , and belief interval s_{\exists} , which can be retrieved from the fused belief masses by the pignistic transformation [28]

$$\begin{pmatrix} p_{\exists}(\tilde{O}) \\ s_{\exists}(\tilde{O}) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} m_F(\tilde{X}) \\ m_F(\{\exists, \# \} \tilde{O}) \end{pmatrix}. \quad (16)$$

With p_{\exists} , a Bayesian-like probability of the existence of \tilde{O} from all observed evidence is recovered. On the other hand, s_{\exists} can be interpreted as an epistemic uncertainty measure, since it is directly proportional to the weight of the ignorance proposition $m_F(\{\exists, \# \} \tilde{O})$.

IV. FAULT ANALYSIS

A. ADDRESSED FAULTS

For the roadside sensor network at hand, a service failure occurs if the system does not deliver data of sufficient information quality to safely support AVs on the road. This can happen in the form of various failure modes, commonly categorizing failures from the view points of domain, detectability, consistency, and consequences on the environment [6]. We here focus on system failures that can not be detected easily with component monitoring, but leave statistical signatures to be studied in the plausibility metrics. Therefore, our addressed failure mode is erratic (data delivery does not halt but is silently corrupted, or mistimed), un signaled, systematic, and catastrophic (data corruption can be safety-critical). Importantly, in this article, we adopt a single fault hypothesis. This is justified by the assumption that our analysis, possibly in combination with additional sensor diagnostics, allows for a quick fault detection, that renders the occurrence of multiple faults in that very time interval unlikely. Perception faults that were found to be relevant in roadside sensing experiments and testbeds, and that all manifest in the above failure mode, were listed in Tab. 1. The highlighted items there are chosen for further simulation studies in the following section, to investigate their respective fingerprints in the plausibility metrics. Our analysis will however be representative of a wider spectrum of faults that belong to the same class, and thus manifest themselves in a similar form.

B. STATISTICAL EVALUATION

Systematic faults as the ones given in Tab. 1 alter sensor readings over the course of many sensing cycles, such that we expect statistically relevant deviations from a baseline scenario. The goal of the fault evaluation process is to identify: i) which of the sensors is faulty (assuming single failures), and ii) what fault or fault class is present. To that end, we inject the selected faults in separate setups, and consider three characteristics for a statistical interpretation: The miss ratio (MR) per sensor, the unexpected observation rate (UOR) per sensor, and most importantly the probability of existence of an object (involving all sensors, see Eq. (16)), here in relation to the observed position in the monitored region. For the former two, we define for a given time and a given sensor,

$$\begin{aligned} \text{MR} &= \frac{\text{misses}}{\text{misses} + \text{all observations}}, \\ \text{UOR} &= \frac{\text{unexpected observations}}{\text{all observations}}. \end{aligned} \quad (17)$$

The above metrics are evaluated at each time step, however, it is important to note that the corresponding variables are not independent due to the nature of the underlying object tracking. For example, an object with a low track score causing a low plausibility at a given time step will likely have a low plausibility in the next time step as well, since the track score only changes gradually. To make the quantities

quasi-iid variables, either of the two following steps can be applied: i) The setup is run various times with independent random distributions of the objects, where the respective signatures are averaged out across one individual scenario. ii) One long scenario run is split into quasi-independent time intervals, where each interval is long enough such that the scene changes significantly from one interval to another. The plausibility metrics within one time interval are averaged out, and the statistics is performed over the resulting iid interval means, assuming a normal distribution. The data basis for our fault analysis is therefore the averages and errors of the various interval means. Using standard methods [31], every data point is presented with its 95% confidence interval (CI) in the figures of Section V, and a fault is diagnosed if the CI of a interval mean does not overlap with the CI of the baseline scenario (no injected fault). For MR and UOR, a statistical baseline without faults is established by performing an additional cross-sensor weighted-least-squares average.

In the following section, we analyze three selected fault mechanisms in two different smart infrastructure scenarios motivated by real experiments. The first setup is a highway segment monitored by a network of overlapping Radar sensors, where various vehicle types move on very structured trajectories on multiple lanes. Objects moving in parallel on different lanes occlude each other frequently. In the second setting, we explore an urban intersection monitored by two Lidar sensors with high accuracy. This setup includes not only various vehicle types but also cyclists and pedestrians, that all accelerate, stop, and take turns. Each temporal interval in our simulations has a sample rate of 0.1s and a duration of 5s, which we found to be a sufficient duration for the highly dynamic scene to change significantly.

V. RESULTS

A. SCENARIO 1: RADAR HIGHWAY INFRASTRUCTURE

This scenario of highway surveillance by a network of Radar sensors is guided by long-term observations within the KoRA9 project [7], [15]. In our setup, a four-lane road segment of about 600m is monitored by 12 Radars at a separation of 50m, where we discard the first and last sensor for the statistics to minimize finite size effects, see Fig. 3. The simulation is done in MATLAB [32] using the *radarDetectionGenerator* class and decentralized MOT with extended Kalman filters (EKF). Key characteristics are $\text{fov} = (90m, 30^\circ, 8^\circ)$, detection probability $p_d = 0.9$, false alarm rate of $p_{fa} = 10^{-6}$, and a track confirmation threshold [25] of $\Lambda_{\text{cnf}} = 1.5 \log(p_d/p_{fa})$. To reflect experiences with realistic Radar signal behavior, we implement an additional signal generation up to a range of 100m, with a reduced likelihood, which results in unexpected observations. The adopted highway traffic model features various vehicle classes such as cars, trucks, buses, and generates speed, lane, and type distributions based on realistic highway statistics [33]. Vehicle maneuvers such as overtaking are neglected here since they do not impact the observation quality significantly, see also [7].

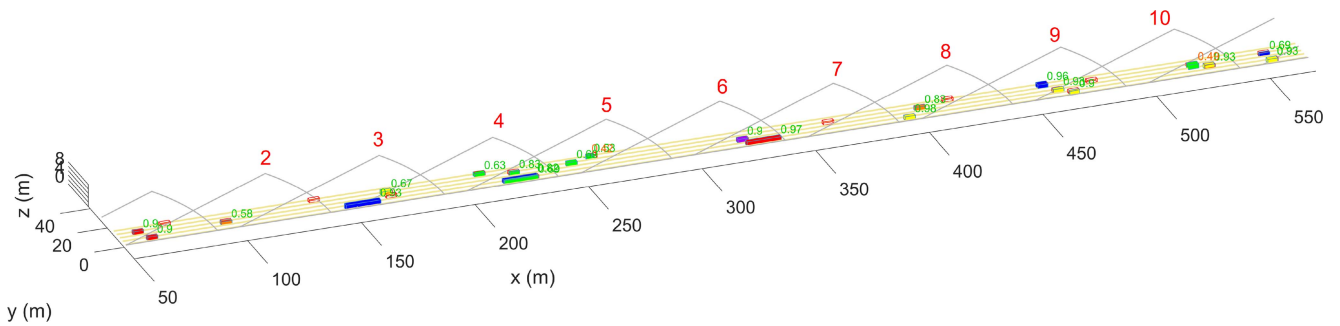


FIGURE 3. Scenario of a four-lane highway segment monitored by a network of Radar sensors. An existence estimate is assigned to each observed object. Sensor indices are shown in red color.

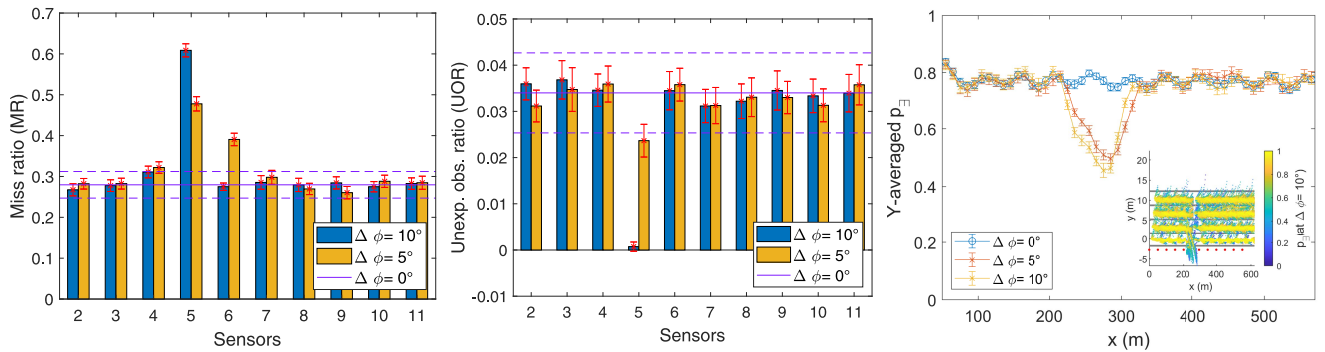


FIGURE 4. Plausibility signatures of the highway scenario of Fig. 3 with injected azimuthal misorientation fault (at sensor 5). Due to the rotated track headings, already small variations $\Delta\phi$ of a few degrees lead to a characteristic fingerprint in all considered metrics MR, UOR, and p_{Ξ} .

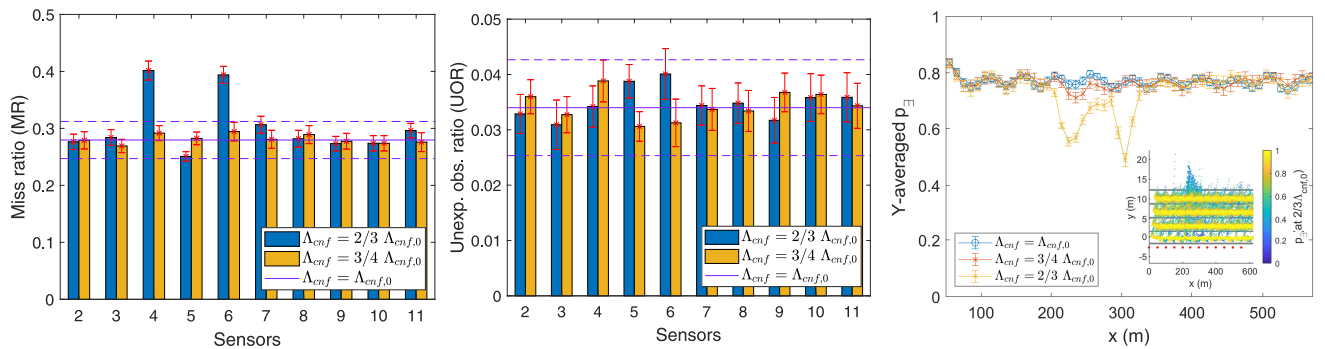


FIGURE 5. Plausibility signatures of the highway scenario with injected tracking parametrization fault (reduced track confirmation threshold Λ_{cnf} at sensor 5). This fault can be identified from the locally reduced average p_{Ξ} as well as the increased MR of the neighboring sensors.

Fig. 4 illustrates the plausibility signatures of an injected miscalibration fault at one of the central sensors (here sensor number 5 located at $x = 200m$, see Fig. 3), representing fault class 1 in Tab. 1. The affected Radar is misoriented counter-clockwise by an azimuthal disturbance angle $\Delta\phi$. All objects observed by the affected sensor then appear rotated (in the opposite direction) to the fusion module, which is not aware of the altered sensor pose. As a result, object tracks do not run parallel to the lanes anymore, and coast out of the road boundaries. We find that the lane-averaged existence probability p_{Ξ} is significantly reduced in a locally confined region of $200m \lesssim x \lesssim 300m$, which clearly indicates a dependability issue of nearby sensors (see Fig. 4). The deviation appears on top of undulations in p_{Ξ} caused by the

varying distance of a passing objects to the respective closest sensor, which can be observed also in the absence of faults. The inconsistencies of observations made by the faulty sensor with its neighbor sensors leaves a trace in the misses and unexpected object statistics as well: the MR of sensor 5 is significantly increased, as this sensor will no longer observe all targets close to the lower road boundary, in contradiction to its neighbor sensor 6. At the same time the UOR is reduced, which can be explained the following way: the dominant contribution of unexpected observations in our model comes from objects that are in the sensing cone but beyond the regular sensing range. As sensor 5 is misoriented counter-clockwise, this sensor field segment is no longer covering the road, and thus no objects are to appear there.

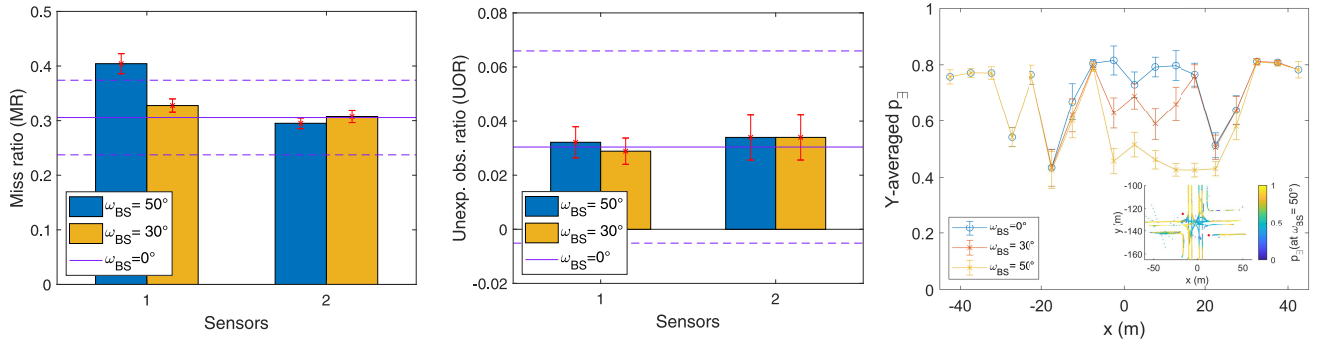


FIGURE 6. Plausibility signatures derived from the two Lidar sensors covering the intersection shown in Fig. 7. The perception of sensor 1 is impaired by a blind spot of azimuthal extension ω_{BS} , inducing a high false negative rate.

In the next scenario analyzed in Fig. 5 we simulate a fault of class 2 (see Tab. 1) by altering the tracking parameters of the test sensor 5. While an appropriately parametrized tracking system is quite robust against noise, the tuning of track management parameters has a significant impact on the false object observation rate. A poor parametrization can be caused for example by human intervention, software updates, or – if machine learning is involved – by the deployment of non-optimal training sets etc. Such faults seem a valid concern given that the use of deep learning methods for MOT is a popular trend not only for video applications [34]. To demonstrate tracking errors we here modify the track confirmation threshold Λ_{cnf} . When reduced with respect to its original value $\Lambda_{cnf,0}$, tentative tracks are accepted after fewer matching measurement updates, such that more false positive confirmed tracks are generated across the sensor’s FoV. As a consequence, our simulation showcases that the MR of the *neighboring* sensors 4 and 6 are significantly increased, since those overlapping sensors can not confirm the large number of false positive tracks generated by sensor 5. On the other hand, this mismatch is effectively compensated by the likewise increased number of overall observations for sensor 5 itself. In the UOR statistics, no significant fingerprint can be found since the false positive tracks are mostly born in the regular FoV and thus will not be registered as unexpected. Finally, the lane-averaged p_E again gives a clear indication of the locally confined dependability issue. The dip in the projected existence probability here follows a mexican hat potential, which reflects the geometry of the sensor field overlaps.

B. SCENARIO 2: URBAN INTERSECTION WITH LIDAR

According to recent surveys [35], about 37% of fatal vehicle crashes in the USA are intersection-related. Junctions as critical hotspots will thus benefit in particular from infrastructure assistance, and represent an interesting use case for plausibility-related fault evaluation. In this section, we study an urban intersection scenario that is guided by the real infrastructure testbed in [36] (with the difference that Lidars are deployed instead of cameras for ease of the simulation pipeline). The recordings were performed in the open-source

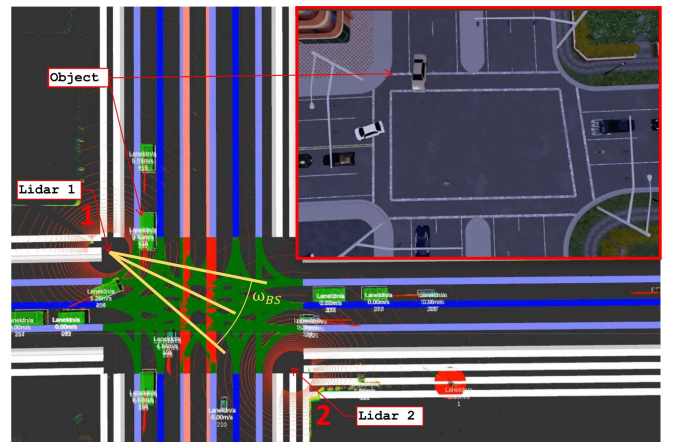


FIGURE 7. Overview of the CARLA multi-lane intersection used for the urban scenario. Two Lidar sensors are located on two opposite corners of the junction, concentric circles indicate the point clouds generated by each Lidar. A blind spot fault with tunable occlusion angle ω_{BS} is injected at sensor 1.

simulator for autonomous driving research CARLA [37]. Based on the Unreal Engine for the rendering, it provides with a set of towns and maps in addition to various sensor models such as RGB cameras, depth sensors, and Lidar sensors. For this scenario, a four-way intersection of the *Town03* map that resembles the setup in [36] was chosen, see Fig. 7. Multiple lanes at each intersection leg, and a large variety of randomly spawned vehicles allow for a diverse urban traffic scenario featuring cars, trucks, cyclists as well as pedestrians roaming on the sidewalks. The recorded scenario has an overall length of 370s, which is split into quasi-independent intervals of 5s each, and on average 18 tracks are detected per frame.

At this intersection, two Lidar sensors are placed at two opposing corners, with a $fov = (80m, 360^\circ, 28^\circ)$ that allows to monitor all events taking place at the junction. This layout represents a minimal example setup for accurate object detection, under mitigation of occlusion events. Each 16-channel Lidar sensor provides with a cloud of 230400 points per second, at an update rate of 20Hz. The point cloud feeds into a deep neural network performing 3D object detection with the PointPillars algorithm, see [38], [39] for performance

TABLE 2. Overview of the simulation results obtained in Section V, confer also with Tab. 1. The three selected empirical perception faults analyzed in the two smart infrastructure scenarios lead to distinct fingerprints in the plausibility metrics.

Fault	Characteristic signatures
Misoriented sensor pose	Reduced local p_{\exists} , increased MR for affected sensor, reduced UOR for affected sensor
Incorrect parametrization of algorithmic component (tracker threshold)	Reduced local p_{\exists} , increased MR for other overlapping sensors
Sensor polluted	Reduced local p_{\exists} , increased MR for affected sensor

and accuracy benchmarks on the KITTI data set. A separate study with synthetic CARLA data was conducted to verify that PointPillars achieves a mean average precision similar to KITTI on the classes car, cyclist and pedestrian, both from a bird-eye-view and a 3D perspective [40]. The generated object list includes information about 3D dimension, position, and heading. Subsequently, the extracted bounding boxes are tracked with an EKF. CARLA further provides with an autonomous agent behavior. Vehicles stay in their lanes, follow traffic lights and take random turns at intersections (whenever possible), drive at a predefined speed, and do not perform any maneuvers such as overtaking. Another limitation of the CARLA simulator is the fact that the traffic lights at an intersection follow a round-robin schedule, where only one road has a green light at a time. Even though the traffic model in the simulator is simplistic, there are enough actors to produce occlusions.

In Fig. 6, we see that the established baseline of the averaged p_{\exists} shows an unsteady behavior at the intersection interfaces, even in the absence of injected faults. This can partially be attributed to the fact that vehicles are frequently turning and accelerating in this zones, which poses severe challenges for the trackers. A class-3-fault (see Tab. 1) is now injected into the simulation, in the form of a blind spot for sensor 1 that represents, e.g., a significant mud splash on the Lidar cover. Sensor 1 is now blind to detections under certain azimuthal angles facing the intersection center, where the extension of the blind spot is controlled by the parameter ω_{BS} , see Fig. 7. The simulation results in Fig. 6 demonstrate that the observed probability of existence is statistically reduced in the inner junction, given a sufficiently drastic sensor pollution. At the same time, the miss ratio of the affected sensor 1 is increased due to the inconsistent observations with respect to sensor 2. Note that this gives us a handle to identify the faulty Lidar within the described plausibilization scheme, in a busy urban environment, even though in this experiment only two overlapping sensors were used.

VI. SUMMARY AND CONCLUSION

This article investigated whether specific plausibility metrics – in our model represented by sensor miss ratios, unexpected

observation ratios, and the collaborative existence probability derived from various plausibility checks across the monitored region – can act as fault indicators in complex decentralized perception systems. For two realistic roadside infrastructure scenarios, we have tested selected perception faults in simulation, and demonstrated that the resulting statistical fingerprints in the plausibility metrics allow for a conclusive fault detection and evaluation. The injected faults were hereby identified from experience with the real-world smart infrastructure systems KoRA9 and P++. Tab. 2 summarizes the results of this analysis.

Since the deployed plausibility checks are independent of the input sensor modality, or the underlying perception algorithms, our analysis can be used in particular to verify the integrity of systems with black-box behavior. An extension with checks focused on machine learning-based perception faults is envisioned for further research. The plausibility signatures are the more pronounced the more sensor overlap can be generated. With this work, we therefore emphasize as well the importance of redundant, independent sources of information for fault tolerant automated driving.

ACKNOWLEDGMENT

The authors further thank Soeren Kohnert, Hochschule of Augsburg, for recording highway Radar data.

REFERENCES

- [1] P. Rau, C. Becker, and J. Brewer, "Approach for deriving scenarios for safety of the intended functionality," in *Proc. ESV*, 2019, pp. 1–15. [Online]. Available: <https://www.esv.nhtsa.dot.gov/Proceedings/26/26ESV-000258.pdf>
- [2] S. Burton, L. Gauerhof, and C. Heinzemann, "Making the case for safety of machine learning in highly automated driving," in *Computer Safety, Reliability, and Security (SAFECOMP)* (Lecture Notes in Computer Science), vol. 10489, S. Tonetta, E. Schoitsch, and F. Bitsch, Eds. Cham, Switzerland: Springer, 2017.
- [3] "Road vehicles—Functional safety part 2—Management of functional safety," Int. Org. Stand., Geneva, Switzerland, Rep. ISO 26262-2, 2018. [Online]. Available: <https://www.iso.org/standard/68384.html>
- [4] "Road vehicles—Safety of the intended functionality," Int. Org. Stand., Geneva, Switzerland, Rep. 70939:2019, 2019. [Online]. Available: <https://www.iso.org/obp/ui/#iso:std:iso:pas:21448:ed-1:v1:en>
- [5] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, "A review of process fault detection and diagnosis part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, 2003.
- [6] A. Avižienis, J. C. Laprie, B. Randell, and C. Landwehr, "Basic concepts and taxonomy of dependable and secure computing," *IEEE Trans. Depend. Secure Comput.*, vol. 1, no. 1, pp. 11–33, Oct. 2004.
- [7] F. Geissler, S. Kohnert, and R. Stolle, "Designing a roadside sensor infrastructure to support automated driving," in *Proc. IEEE Int. Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 703–708.
- [8] S. Seebacher *et al.*, "Infrastructure data fusion for validation and future enhancements of autonomous vehicles' perception on Austrian motorways," in *Proc. 8th IEEE Int. Conf. Connected Veh. Expo*, 2019, pp. 1–7.
- [9] S. Garg, "Intelligent transportation system market—Global forecast to 2023," MarketsandMarkets, Pune, India, Rep., 2018.
- [10] X. D. A. Carreras, J. Erhart, and S. Ruehrup, "Road infrastructure support levels for automated driving," in *Proc. 25th ITS World Congr.*, Sep. 2018, pp. 12–20. [Online]. Available: https://www.inframix.eu/wp-content/uploads/ITSWC2018-ASF-AAE-Final-paper_v4.pdf

- [11] M. Aeberhard and N. Kämpchen, "High-level sensor data fusion architecture for vehicle surround environment perception," in *Proc. 8th Int. Workshop Intell. Transp. Syst.*, 2011, pp. 729–734. [Online]. Available: https://www.researchgate.net/publication/267725657_High-Level_Sensor_Data_Fusion_Architecture_for_Vehicle_Surround_Environment_Perception
- [12] A. Houenou, P. Bonnifait, V. Cherfaoui, and J. F. Boissou, "A track-to-track association method for automotive perception systems," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 704–710.
- [13] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
- [14] S. Challa and D. O. Koks, "Bayesian and Dempster–Shafer fusion," *Sadhana Acad. Proc. Eng. Sci.*, vol. 29, no. 2, pp. 145–176, 2004.
- [15] Bundesministerium fuer Verkehr Und Digitale Infrastruktur (BMVI). (2017). *KoRA9*. [Online]. Available: <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/AVF-projekte/KoRA9>
- [16] Bundesministerium fuer Verkehr und digitale Infrastruktur (BMVI). (2020). *Providentia++*. [Online]. Available: <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/AVF-projekte/providentia-plusplus>
- [17] H. Versmold and M. Saeger, "Plausibility checking of sensor signals for vehicle dynamics control systems," in *Proc. Symp. Adv. Veh. Control*, 2006, p. 136.
- [18] C. Yavvari, Z. Duric, and D. Wijesekera, "Vehicular dynamics based plausibility checking," in *Proc. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2018, pp. 1–8.
- [19] N. Bißmeyer, C. Stresing, and K. M. Bayarou, "Intrusion detection in VANETs through verification of vehicle movement data," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, 2010, pp. 166–173.
- [20] P. J. Graydon and C. M. Holloway, "An investigation of proposed techniques for quantifying confidence in assurance arguments," *Safety Sci.*, vol. 92, pp. 53–65, May 2017.
- [21] M. Maehlich, W. Ritter, and K. Dietmayer, "De-cluttering with integrated probabilistic data association for multisensor multitarget ACC vehicle tracking," in *Proc. IEEE Intell. Veh. Symp.*, 2007, p. 178.
- [22] M. S. M. Aeberhard, "Object-level fusion for surround environment perception in automated driving applications," Ph.D. dissertation, Faculty Elect. Eng. Inf. Technol., TU Dortmund, Dortmund, Germany, 2017.
- [23] M. Obst, L. Hobert, and P. Reisdorf, "Multi-sensor data fusion for checking plausibility of V2V communications by vision-based multiple-object tracking," in *Proc. IEEE Veh. Netw. Conf. (VNC)*, 2015, pp. 143–150.
- [24] M. Ambrosin, L. L. Yang, X. Liu, M. R. Sastry, and I. J. Alvarez, "Design of a misbehavior detection system for objects based shared perception V2X applications," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2019, pp. 1165–1172.
- [25] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House, 1999.
- [26] J. Dezert, P. Wang, and A. Tchamova, "On the validity of Dempster–Shafer theory," in *Proc. Int. Conf. Inf. Fusion*, 2012, pp. 655–660.
- [27] M. Munz, K. C. Dietmayer, and M. Mählich, "A sensor independent probabilistic fusion system for driver assistance systems," in *Proc. IEEE Conf. Intell. Transp. Syst. (ITSC)*, 2009, pp. 1–6.
- [28] M. Aeberhard, S. Paul, N. Kaempchen, and T. Bertram, "Object existence probability fusion using Dempster–Shafer theory in a high-level sensor data fusion architecture," in *Proc. IEEE Intell. Veh. Symp.*, 2011, pp. 770–775.
- [29] D. Mušicki, "Track score and target existence," in *Proc. 9th Int. Conf. Inf. Fusion*, 2006, pp. 1–7.
- [30] "Richtlinien Für Die Anlage Von Autobahnen (RAA)," Forschungsgesellschaft Für Straßen- Und Verkehrswesen, Cologne, Germany, Rep., 2008. [Online]. Available: <https://www.forschungsinformationssystem.de/servlet/is/275112/>
- [31] K. S. Trivedi, *Probability and Statistics With Reliability, Queuing and Computer Science Applications*. New York, NY, USA: Wiley, 2016.
- [32] *MATLAB Automated Driving Systems Toolbox Release 2019b*, Mathworks Inc., Natick, MA, USA, 2019. [Online]. Available: <https://de.mathworks.com/products/automated-driving.html>
- [33] Traffic Authority of Southern Bavaria (ABDSB), unpublished raw data, 2018.
- [34] G. Ciaparrone *et al.*, "Deep learning in video multi-object tracking: A survey," *Neurocomputing*, vol. 381, pp. 61–88, Mar. 2020.
- [35] "Traffic safety facts 2017: A compilation of motor vehicle crash data," NHTSA, Washington, DC, USA, Rep. DOT HS 812 806, 2019. [Online]. Available: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812806>
- [36] T. Fleck *et al.*, "Towards large scale urban traffic reference data: Smart infrastructure in the test area autonomous driving Baden-Württemberg," *Adv. Intell. Syst. Comput.*, vol. 867, pp. 964–982, Mar. 2019.
- [37] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proc. 1st Conf. Robot Learn. (CoRL)*, 2017, pp. 1–16. [Online]. Available: <http://arxiv.org/abs/1711.03938>
- [38] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun. (2018). *KITTI Vision Benchmark Suite*. [Online]. Available: <http://www.cvlibs.net/datasets/kitti/>
- [39] A. H. Lang, S. Vora, H. Caesar, L. Zhou, J. Yang, and O. Beijbom, "Pointpillars: Fast encoders for object detection from point clouds," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 12697–12705.
- [40] *Unpublished Material*, Intel Labs, Santa Clara, CA, USA, 2020.