




# GM-PHD Filter Based Sensor Data Fusion for Automotive Frontal Perception System

László Lindenmaier , Szilárd Aradi , *Member, IEEE*, Tamás Bécsi , *Member, IEEE*, Olivér Törő, and Péter Gáspár, *Member, IEEE*

**Abstract**—Advanced driver assistance systems and highly automated driving functions require an enhanced frontal perception system. The requirements of a frontal environment perception system cannot be satisfied by either of the existing automotive sensors. A commonly used sensor cluster for these functions consists of a mono-vision smart camera and automotive radar. The sensor fusion is intended to combine the data of these sensors to perform a robust environment perception. Multi-object tracking algorithms have a suitable software architecture for sensor data fusion. Several multi-object tracking algorithms, such as JPDAF or MHT, have good tracking performance; however, the computational requirements of these algorithms are significant according to their combinatorial complexity. The GM-PHD filter is a straightforward algorithm with favorable runtime characteristics that can track an unknown and time-varying number of objects. However, the conventional GM-PHD filter has a poor performance in object cardinality estimation. This paper proposes a method that extends the GM-PHD filter with an object birth model that relies on the sensor detections and a robust object extraction module, including Bayesian estimation of objects' existence probability to compensate for drawbacks of the conventional algorithm.

**Index Terms**—Advanced driver assistance, Gaussian mixture model, multi-object tracking, object detection, PHD filter, radar detection, sensor fusion, smart cameras.

## I. INTRODUCTION

THE need for autonomous vehicles results in a gradual increase in the number of automated functions [1]–[3]. The role of the environment perception becomes more significant with the higher automation level of vehicles. Most advanced driver assistance functions, such as Autonomous Emergency Braking (AEB) or Adaptive Cruise Control (ACC), rely on an

Manuscript received June 22, 2021; revised November 22, 2021 and February 23, 2022; accepted April 26, 2022. Date of publication April 28, 2022; date of current version July 18, 2022. This work is a part of Project Number BME-NVA-02, supported by the Ministry of Innovation and Technology of Hungary from the National Research, Development, and Innovation Office Fund under Grant TKP2021, in part by the BME has been supported by the NRDI Fund under Grant TKP2020 NC, Grant BME-NC based on the charter of bolster issued by the NRDI Office under the auspices of the Ministry for Innovation and Technology. The review of this article was coordinated by Prof. Arafat J Al-Dweik. (*Corresponding author: Tamás Bécsi.*)

László Lindenmaier, Szilárd Aradi, Tamás Bécsi, and Olivér Törő are with the Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics, 1111 Budapest, Hungary (e-mail: lindenmaier.laszlo@kjk.bme.hu; aradi.szilard@kjk.bme.hu; becsi.tamas@kjk.bme.hu; toro.oliver@kjk.bme.hu).

Péter Gáspár is with the Systems and Control Laboratory and Institute for Computer Science and Control, Budapest University of Technology and Economics, 1111 Budapest, Hungary (e-mail: gaspar.peter@sztaki.mta.hu).

Digital Object Identifier 10.1109/TVT.2022.3171040

enhanced frontal perception system [4], [5], hereinafter referred to as EFPS. The EFPS consists of three main modules: lane detection, object detection, and multi-object tracking subsystems. Object detection and tracking have several difficulties: the number of objects within the surveillance area is time-varying and unknown, missed detections, and false alarms may occur due to sensor uncertainty. Furthermore, the sensors cannot observe all the relevant state variables of the true objects, and the measurements are noisy. This paper focuses on multi-object tracking that aims to accurately estimate the cardinality and the state of present objects in front of the ego vehicle. However, none of the available sensors can satisfy alone the requirements of the EFPS. For example, radars can measure the spatial distance of the targets precisely, but they have a poor performance in azimuth estimation. Meanwhile, mono-vision cameras have a good lateral position resolution but inaccurately estimate the objects' longitudinal distances. Thus, sensor data fusion is needed to obtain a robust and precise representation of the environment [6]–[8].

## A. Related Work

Multi-object tracking algorithms have a suitable architecture for sensor fusion purposes. The commonly-used Probabilistic Data Association (PDA) [9] and Joint Probabilistic Data Association (JPDA) [10] multi-object tracking algorithms estimate the state of a known number of objects, often fusing the data of multiple sensors [11]–[13]. Since, in real applications, the number of objects is time-varying and unknown, an estimation of the object cardinality is needed. The Integrated Probabilistic Data Association (IPDA) [14] and Joint Integrated Probabilistic Data Association (JIPDA) [15] filters extend the previous approaches to estimate the existence probability of the objects as the basis of object management. However, the performance and runtime of these algorithms depend highly on the gating size applied around the objects. In addition, the complexity of JPDA and JIPDA is combinatorial, which makes real-time applications complicated. The Multiple Hypothesis Tracker (MHT) [16], [17] considers the dynamics of data association by propagating it in time. This multi-object tracking algorithm is also often used as a sensor fusion module of object perception as in [18], [19]; however, the computational cost of this approach is more significant compared to IPDA and JIPDA. Random Finite Sets (RFS) provide an alternative way to model the object list consisting of object states and possibly labels in an arbitrary order. RFS-based multi-object tracking can handle object birth and death, clutter measurements, and missed detections by inserting the multi-object PDF of RFS

into the recursive Bayesian framework. The implementation of optimal and exact RFS-based recursive filtering is complex and not tractable due to multiple set integrals. A suitable alternative approach, the Probability Hypothesis Density (PHD) filter, propagates and updates the first-order moment of the RFS [20]. Although the PHD filter does not have a closed-form solution, the Gaussian Mixture (GM-PHD) [21] and particle [22] approximations can provide tractable recursions.

The GM-PHD filter is a straightforward algorithm with a low computational cost that recursively estimates the first-order moment of the multi-object RFS. However, it requires an additional module to extract the potential objects from the updated PHD function consisting of Gaussian components. The GM-PHD filter provides an estimation of the object cardinality, but this variable follows a Poisson distribution. Since the standard deviation of a Poisson distribution is proportional to the mean value, this estimation leads to noisy cardinality estimates in the case of many targets. Another way for object extraction is to apply a threshold on the existence probability of Gaussian components computed by the GM-PHD filter. However, this approach underestimates the number of present objects. Numerous papers aim to overcome the limitations of the PHD filter. An approach is to introduce state dependent detection probabilities as reported in [23], [24], [25], [26]. In [27] state-dependent survival probability is considered in a scenario with low detectability. For visual based tracking, the GM-PHD filter, proposed in [28], creates a cost function for object assignment and creates trajectories using an auction algorithm. A multi-frame scheme is introduced in [29] to deal with estimates of undetected targets. To preserve the standard measurement model, a penalization scheme [30] and a competitive algorithm [31] is introduced to renormalize weights along the objects and measurements.

### B. Contributions of the Paper

This paper extends the conventional GM-PHD filter with a robust object extraction module that estimates the existence probability of the objects in a Bayesian manner. The proposed method efficiently tackles the missed detections and clutter measurements. Therefore, it provides a better estimation of the object cardinality, obtaining a robust object perception for ADAS and highly automated functions' EFPS. The object extraction module first performs a data association between the Gaussian components and the sensor detections by two different approaches based on the component weights considering them as association probabilities instead of existence probabilities. Then, the updated PHD function components and the detections are clustered by the k-Nearest Neighbor algorithm [32] avoiding exponential complexity. The elements of a cluster are fused, relying on the IPDA algorithm. The IPDA weights and the objects' PoE (Probability of Existence) are computed by the updated weights of the Gaussian components. This approach results in a non-parametric IPDA algorithm in terms of gating size. Although the Bayesian IPDA algorithm can be derived based on RFS theory [33]; however, this leads one to the already mentioned implementation problems. A connection between the JIPDA tracking algorithm and the GM-PHD filter has been pointed out in [34], but the object's existence probability is not computed based on the Bayesian rule. Furthermore, we

introduce a method for modeling the birth probability of measurements that efficiently identifies new components in the PHD function. The proposed robust GM-PHD filter significantly improves the conventional GM-PHD filter, and it achieves the performance of the state-of-the-art JIPDA algorithm, preserving the beneficial complexity of the conventional GM-PHD filter.

## II. THEORETICAL BACKGROUND

This section briefly summarizes the theoretical background of the proposed method. The GM-PHD filter is based on finite set statistics theory (FISST) [35] which provides mathematical apparatus for computations with random finite sets. After introducing RFS, the section discusses how the PHD workflow simplifies the filtering of the multi-object PDF and presents a practical variety, the Gaussian Mixture PHD filter.

### A. Random Finite Sets

A random finite set  $\mathbf{X}$  consists of  $n$  random vector variables  $\mathbf{X} = \{x_1, \dots, x_n\}$ , where the cardinality  $|\mathbf{X}| = n$  is also a random variable with  $\varrho(n)$  probability mass function. The PDF of a multi-object RFS is given by

$$p(\mathbf{X}) = n! \varrho(n) p_n(x_1, \dots, x_n). \quad (1)$$

The multiplier  $n!$  corresponds to the fact that the elements of an RFS are in arbitrary order, thus the permutation of the elements results in equivalent RFS. Suppose  $X = [x_1, \dots, x_n]$  is an ordered array, then the connection between two multi-object PDFs is written as

$$p(X = [x_1, \dots, x_n]) = \frac{1}{n!} p(\mathbf{X} = \{x_1, \dots, x_n\}). \quad (2)$$

Although the multi-object PDF given in (1) can be inserted into a recursive Bayes filter, the exact solution has a high computational cost because of its combinatorial complexity. The PHD filter provides a good approximation to compute the posterior RFS in real-time.

### B. Gaussian Mixture PHD Filter

The PHD function is the first-order moment of the multi-object PDF, which is defined as

$$D(x) = \int p(\mathbf{X}) \sum_{x' \in \mathbf{X}} \delta(x - x') \delta \mathbf{X}. \quad (3)$$

The higher the value of the PHD function, the more likely the corresponding state represents a real object. The expected number of objects (cardinality of RFS) over a given region  $\Omega$  of the state-space can be computed as the integral of the PHD function over this region:

$$E(|\mathbf{X} \cap \Omega|) = \int_{\Omega} D(x) dx. \quad (4)$$

In the PHD filter the object model is based on Poisson Point Process (PPP) in that the cardinality of the RFS follows a Poisson distribution [36]. Thus, the PDF  $p(\mathbf{X})$  of a multi-object RFS and its PHD (or intensity) function  $D(x)$  can be written as

$$p(\mathbf{X}) = e^{-\lambda} \prod_{x \in \mathbf{X}} \lambda p(x) \quad (5)$$

$$D(x) = \lambda p(x), \quad (6)$$

where  $p(x)$  denotes the PDF in the single-object state space and  $\lambda$  is the Poisson rate of object cardinality that equals the number of expected objects.

The PHD function can be inserted into the Bayesian framework to obtain an estimation of the multi-object RFS. The prediction for the PHD function, given the posterior  $D_{k-1|k-1}(x)$  from timestep  $k-1$  is performed according to

$$\begin{aligned} D_{k|k-1}(x) &= b_{k|k-1}(x) \\ &+ p_S(x) \int \pi_{k|k-1}(x|x_{k-1}) \\ &D_{k-1|k-1}(x_{k-1}) dx_{k-1}. \end{aligned} \quad (7)$$

In equation (7)  $b_{k|k-1}(x)$  denotes the birth PHD that corresponds to the appearing objects while the second part of the equation is the Chapman-Kolmogorov prediction of the prior, where  $p_S(x)$  denotes the survival probability and  $\pi_{k|k-1}(x_k|x_{k-1})$  is the transition density from timestep  $k-1$  to  $k$ .

The PHD function is updated using the measurement set  $\mathbf{Z}_k = \{z_1, \dots, z_{m_k}\}$  as

$$\begin{aligned} D_{k|k}(x) &= (1 - p_d(x)) D_{k|k-1}(x) \\ &+ \sum_{z \in \mathbf{Z}_k} \frac{p_d(x) g_k(z|x) D_{k|k-1}(x)}{\kappa(z) + \int p_d(x) g_k(z|x) D_{k|k-1}(x) dx}. \end{aligned} \quad (8)$$

where  $p_d(x)$  denotes the detection probability of an object with state  $x$ ,  $\kappa(z)$  is the clutter density and  $g_k(z|x)$  is the measurement likelihood. It should be noted that the measurement set does not only contain the object originated elements  $\mathbf{O}_k \subseteq \mathbf{Z}_k$  but the  $\mathbf{C}_k \subseteq \mathbf{Z}_k$  clutter measurements as well. In equation (8), the first term corresponds to miss-detected objects, and the second part of the expression refers to measurement update. This equation does not have a closed-form solution; the PHD function is usually approximated by an assumed density.

The GM-PHD filter approximates the PHD function by a weighted sum of  $\mathcal{H}$  Gaussian components

$$D(x) \approx \sum_{h=1}^{\mathcal{H}} w^h \mathcal{N}(x; \hat{x}^h, \hat{P}^h), \quad (9)$$

where  $\mathcal{N}(x; \hat{x}^h, \hat{P}^h)$  denotes the normal distribution of the component with mean  $\hat{x}^h$  and covariance  $\hat{P}^h$ , while  $w^h$  is the weight of the corresponding component.

Given the posterior PHD function  $D_{k-1|k-1}(x)$  in the above form, the predicted GM-PHD function can be written as

$$D_{k|k-1}(x) = D_{k|k-1}^S(x) + b_{k|k-1}(x), \quad (10)$$

where  $D_{k|k-1}^S(x)$  denotes the predicted PHD of the previous components representing the surviving objects:

$$D_{k|k-1}^S(x) = \sum_{h=1}^{\mathcal{H}_{k-1|k-1}} w_{k|k-1}^h \mathcal{N}(x, \hat{x}_{k|k-1}^h, \hat{P}_{k|k-1}^h), \quad (11)$$

and  $b_{k|k-1}(x)$  is the birth PHD:

$$b_{k|k-1}(x) = \sum_{b=1}^{\mathcal{H}_k^b} w_{k|k-1}^b \mathcal{N}(x, \hat{x}_{k|k-1}^b, \hat{P}_{k|k-1}^b). \quad (12)$$

The birth PHD should capture the locations in the state-space where objects are assumed to appear. The birth model of our work is detailed in Section III-C.

The prediction of both surviving and birth components can be separated into two steps. In the first step, the spatial distribution of the components is predicted based on a Kalman prediction [37]. In the second step, the weights are scaled by the survival probability:

$$w_{k|k-1}(x) = p_S(x) w_{k|k-1}(x). \quad (13)$$

The number of the predicted components is computed as

$$\mathcal{H}_{k|k-1} = \mathcal{H}_{k-1|k-1} + \mathcal{H}_k^b. \quad (14)$$

Then the updated PHD consists of two parts: the PPP part representing the miss-detected objects and the multi-Bernoulli (MB) RFS, which corresponds to the detected objects:

$$D_{k|k}(x) = D_{k|k}^0(x) + D_{k|k}^Z(x). \quad (15)$$

The PPP intensity consists of  $\mathcal{H}_{k|k-1}$  components:

$$D_{k|k}^0(x) = \sum_{h=1}^{\mathcal{H}_{k|k-1}} (1 - p_d(x)) w_{k|k-1}^h \mathcal{N}(x; \hat{x}_{k|k-1}^h, \hat{P}_{k|k-1}^h). \quad (16)$$

The MB part of the posterior PHD contains  $m_k \mathcal{H}_{k|k-1}$  components capturing all the possible component-measurement pairs:

$$D_{k|k}^Z(x) = \sum_{i=1}^{m_k} \sum_{h=1}^{\mathcal{H}_{k|k-1}} w_{k|k}^{h,i} \mathcal{N}(x, \hat{x}_{k|k}^{h,i}, \hat{P}_{k|k}^{h,i}), \quad (17)$$

where the  $\mathcal{N}(x; \hat{x}_{k|k}^{h,i}, \hat{P}_{k|k}^{h,i})$  spatial PDF of the updated components is proportional to the product of the  $\mathcal{N}(z_k^i; H_k x, R_k^i)$  measurement likelihood and the  $\mathcal{N}(x; \hat{x}_{k|k-1}^h, \hat{P}_{k|k-1}^h)$  predicted PDF. Thus, the posteriori spatial PDF is computed by the Kalman update [37] based on the  $H_k$  measurement model and  $R_k^i$  measurement covariance matrix. The weight  $w_{k|k}^{h,i}$  representing the existence probability of the component is updated as:

$$\bar{w}_{k|k}^{h,i} = p_d(x) \mathcal{N}(z_k^i, H_k x_{k|k-1}^h, S_k^h) w_{k|k-1}^h, \quad (18)$$

where

$$w_{k|k}^{h,i} = \frac{\bar{w}_{k|k}^{h,i}}{\kappa(z_k^i) + \sum_{h=1}^{\mathcal{H}_{k|k-1}} \bar{w}_{k|k}^{h,i}}, \quad (19)$$

and  $S_k^h$  denotes the innovation covariance of measurement  $z_k^i$  and component state  $x_{k|k-1}^h$ .

The most beneficial property of the GM-PHD filter is that the complexity of the algorithm depends linearly on the  $\mathcal{H}_{k|k-1}$  number of predicted components. However, the posterior PHD function consists of  $\mathcal{H}_{k|k} = \mathcal{H}_{k|k-1}(m_k + 1)$  components which means that the complexity of the algorithm would diverge. In practical applications, a mixture reduction is needed to avoid increasing computational costs.

The reduction of the number of components is performed by merging and pruning. A simple manner of mixture reduction is to keep the first  $\hat{n}_{k|k}$  components of the updated PHD function after sorting the  $w_{k|k}^h$  updated weights, where  $\hat{n}_{k|k}$  can be computed as

$$\hat{n}_{k|k} = \text{round} \left( \sum_{h=1}^{\mathcal{H}_{k|k}} w_{k|k}^h \right). \quad (20)$$

In particle implementations of PHD filter, the resampled particles can also be given based on  $\hat{n}_{k|k}$  by Sequential Monte Carlo (SMC) simulation as in [22]. However, since the multi-object RFS is modeled as a PPP, this estimation of the expected number of objects follows a Poisson distribution. It means that the higher the number of objects, the more uncertain the cardinality estimation is. A more robust way of object extraction from the updated PHD is to compare the  $w_{k|k}^h$  weights to a defined  $\gamma_{\exists}$  threshold. If the weight of a component is greater than the threshold, the component initiates an object in the multi-object RFS as in [21], [23].

### III. METHODOLOGY

This section details the proposed robust GM-PHD filter fusing heterogeneous data from multiple sensors for EFPS of ADAS functions. In the evaluation, besides a simulation setup, we consider real measurement data to investigate our method in practical applications. The most commonly applied sensor cluster of frontal perception systems consists of a forward-looking radar and camera [38], [39]. The proposed method was evaluated on a similar sensor cluster with characteristics detailed in Section IV-B. The system architecture depicted in Fig. 1 consists of two main layers and an additional step. The fusion layer is separated from the sensor layer resulting in a modular fusion architecture. In the sensor layer, only the acquisition of the sensor data is performed. The sensor and fusion layers are connected by the data synchronization that handles asynchronous sensor data and identifies the processable measurements at a given time. The fusion layer involves the GM-PHD filter and the object extraction module. If the data synchronization assumes a sensor data processable at a given time, it updates the GM-PHD filter, and the object extraction updates the object list. Since the fusion layer works on a joint object list, it utilizes the data of both sensors fusing the measurements in the order defined by the data synchronization. Finally, the posterior delay of the fused data is compensated by predicting the fused objects to the current time. The state prediction is performed based on a straightforward constant acceleration (CA) model detailed in Section IV-C.

#### A. Data Synchronization

Data synchronization is a fundamental part of every sensor fusion algorithm. This block is needed because:

- Sensors provide data with different frequency
- Sensors work asynchronously
- Out-of-sequence measurements (OOSM)

A measurement is assumed to be out of the sequence if its timestamp is older than the timestamp of the latest measurement. This phenomenon can occur because of the sensor data latency. If the timestamp of measurement were older than the latest update time of the filter, the prediction would be performed to the past,

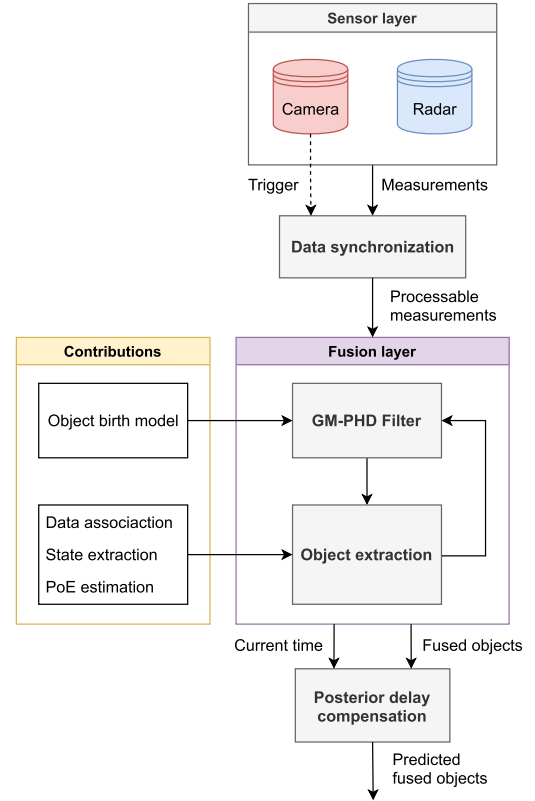


Fig. 1. The system architecture.

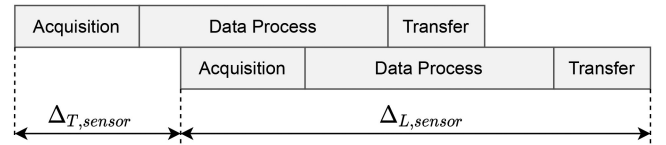


Fig. 2. The sensor data latency and frequency;  $\Delta_{L,sensor}$  denotes the latency of sensor data,  $\Delta_{T,sensor}$  is the time period of the sensor.

referred to as “negative-time measurement update”. There are different approaches to handle OOSM in multi-object tracking. One way is to approximate the solution by neglecting the process noise in the prediction step as in [40]. An exact solution is also provided in [41]. Other works deal with OOSM based on data buffering techniques as in [42], [43]. According to [42] the sensor data latency consists of three main parts: the data acquisition, the pre-process, and the transfer, as shown in Fig. 2. Deterministic data buffering means that the frequency and the latency information of the sensors are provided. In [42], when a new measurement is provided, the timestamp of the following data is assumed to be known. Although this information is not provided in our sensor configuration, based on measurement analysis, the maximum latency of each sensor can be accurately estimated. A measurement updates the object list at a given time if none of the sensors may send data with an older timestamp; otherwise, it is buffered. A buffered data is assumed to be processable if:

- the timestamp of the measurement is lower than the difference of the current time and the maximum latency of each sensor and
- it is the latest measurement of the sensor.

The other role of the data synchronization module is scheduling. According to [43], the two types of fusion algorithms are asynchronous and synchronous. Asynchronous fusion means that each sensor measurement triggers the algorithm. However, in the synchronous case, the fusion is called by one of the sensors or an independent scheduler with constant frequency. The deterministic data buffer allows asynchronous running of fusion; however, since the frequency of the smart camera is significantly lower compared to the radar, the camera triggers the fusion. Therefore, at least one camera measurement updates the fusion in almost every cycle.

### B. Implementation of GM-PHD Filter

The GM-PHD filter described in Section II-B is implemented as in Alg. 1, where  $F_k$ ,  $Q_k$ ,  $H_k$  and  $R_k$  denotes the model parameters of the Kalman filter detailed in Section IV-C: the transition matrix, the process noise covariance, the observation matrix, and the measurement covariance, respectively. While  $S_k$ ,  $e_k$ , and  $K_k$  are the computed variables, namely the innovation covariance, the innovation, and the Kalman gain.

The miss-detected components are scaled based on the  $p_d(x)$  detection probability; therefore, it is crucial to obtain a proper detection model. Some papers focus on the state-dependent detection model in poor detectability circumstances as in [23], [24]. We also propose a state-dependent detection probability; however, this model considers the field of view (FoV) of the sensor. The missed detections are resolved by the robust object extraction detailed in Section III-D. Different detection probability is given if the object is within the sensor FoV and if it is outside the FoV. Therefore, the detection probability is given as:

$$p_d(x) = \begin{cases} \hat{p}_\emptyset, & \text{if } d_{xy,k}^h \cap \Omega_{FOV} = \emptyset \\ \hat{p}_d, & \text{otherwise} \end{cases}, \quad (21)$$

where  $\hat{p}_\emptyset$  and  $\hat{p}_d$  denotes the detection probability constant outside and within the sensor FoV. In (21),  $d_{xy,k}^h = \begin{bmatrix} d_x & d_y \end{bmatrix}_k^h$  refers the subspace of  $\hat{x}_k^h$  including the estimated position of the object. The expression  $d_{xy,k}^h \cap \Omega_{FOV}$  corresponds to the event that the  $\hat{x}_k^h$  object is located outside the  $\Omega_{FOV}$  space of the FoV. Wang *et al.* proposed a method that considers a state-dependent survive probability [27]; however, in our implementation the survive probability  $p_S(x) = p_S$  is given by a constant.

Most ADAS and HAD functions refer to the objects by their unique ID after the relevant target selection. Therefore, the  $\{w_{k|k}, \hat{x}_{k|k}, \hat{P}_{k|k}\}$  parameters of the GM-PHD components are extended with the required  $l_{k|k}$  label consisting of the unique ID of the objects. The labels of the updated components are inherited from the corresponding previous components. Hence, the labels of the updated components are not unique. In the proposed robust GM-PHD filter, the unique ID of the present objects is established by the object extraction module.

---

### Algorithm 1: Gaussian Mixture PHD Filter.

---

```

1: given  $\{w_{k-1|k-1}^h, \hat{x}_{k-1|k-1}^h, \hat{P}_{k-1|k-1}^h, l_{k-1|k-1}^h\}_{h=1}^{\mathcal{H}_{k-1|k-1}}$ ,
 $\{w_{k-1|k-1}^b, \hat{x}_{k-1|k-1}^b, \hat{P}_{k-1|k-1}^b, l_{k-1|k-1}^b\}_{b=1}^{\mathcal{H}_{k-1|k-1}^b}$ ,  $\mathbf{Z}_k$ 
Append birth components:
2:  $\mathcal{H}_{k|k-1} = \mathcal{H}_{k-1|k-1} + \mathcal{H}_{k-1|k-1}^b$ 
Prediction of previous and birth components:
3: for  $h = 1$  to  $\mathcal{H}_{k|k-1}$  do
4:  $\hat{x}_{k|k-1}^h = F_k \hat{x}_{k-1|k-1}^h$ 
5:  $\hat{P}_{k|k-1}^h = F_k \hat{P}_{k-1|k-1}^h F_k^\top + Q_k$ 
6:  $w_{k|k-1}^h = p_S(x) w_{k-1|k-1}^h$ 
7:  $l_{k|k-1}^h = l_{k-1|k-1}^h$ 
8: end for
Update missed detected components:
9: for  $h = 1$  to  $\mathcal{H}_{k|k-1}$  do
10:  $\hat{x}_{k|k}^h = \hat{x}_{k|k-1}^h$ 
11:  $\hat{P}_{k|k}^h = \hat{P}_{k|k-1}^h$ 
12:  $w_{k|k}^h = (1 - p_d(x)) w_{k|k-1}^h$ 
13:  $l_{k|k}^h = l_{k|k-1}^h$ 
14: end for
Update detected components:
15: for  $i = 1$  to  $m_k$  do
16: for  $h = 1$  to  $\mathcal{H}_{k|k-1}$  do
17:  $\hat{z}_{k|k-1}^h = H_k \hat{x}_{k|k-1}^h$ 
18:  $S_k^{h,i} = R_k^i + H_k \hat{P}_{k|k-1}^h H_k^\top$ 
19:  $K_k^{h,i} = \hat{P}_{k|k-1}^h H_k^\top (S_k^{h,i})^{-1}$ 
20:  $e_{k|k}^{h,i} = (z_k^i - \hat{z}_{k|k-1}^h)$ 
21:  $\hat{x}_{k|k}^{i\mathcal{H}_{k|k-1}+h} = \hat{x}_{k|k-1}^h + K_k^{h,i} e_{k|k}^{h,i}$ 
22:  $\hat{P}_{k|k}^{i\mathcal{H}_{k|k-1}+h} = (I - K_k^{h,i} H_k) \hat{P}_{k|k-1}^h$ 
23:  $g_k(z_k^i | x_k^h) = \mathcal{N}(z_k^i, \hat{z}_{k|k-1}^h, S_k^{h,i})$ 
24:  $\bar{w}_{k|k}^h = p_d(x) w_{k|k-1}^h g_k(z_k^i | x_k^h)$ 
25:  $l_{k|k}^{i\mathcal{H}_{k|k-1}+h} = l_{k|k-1}^h$ 
26: end for
27:  $w_{k|k}^{i\mathcal{H}_{k|k-1}+h} = \frac{\bar{w}_{k|k}^{h,i}}{\kappa(z_k^i) + \sum_{h=1}^{\mathcal{H}_{k|k-1}} \bar{w}_{k|k}^{h,i}}$ 
28: end for
29:  $\mathcal{H}_{k|k} = \mathcal{H}_{k|k-1} + m_k \mathcal{H}_{k|k-1}$ 
30: go to: Birth model
output:  $\{w_{k|k}^h, \hat{x}_{k|k}^h, \hat{P}_{k|k}^h, l_{k|k}^h\}_{h=1}^{\mathcal{H}_{k|k}}$ 

```

---

### C. Birth Model

The birth model is intended to capture the newborn objects considering the locations where new ones can appear (e.g., on the edge of the sensor FoV). According to [23], [44], it is beneficial to compose the birth components based on the  $\mathbf{Z}_{k-1}$  measurement set at time  $k - 1$ . Some works extend the IPDA and JIPDA algorithms with a birth model that computes the birth probability of the detections based on the present objects [45], [46]. This paper proposes an extension of the conventional GM-PHD filter with a similar birth model; therefore, only the measurements

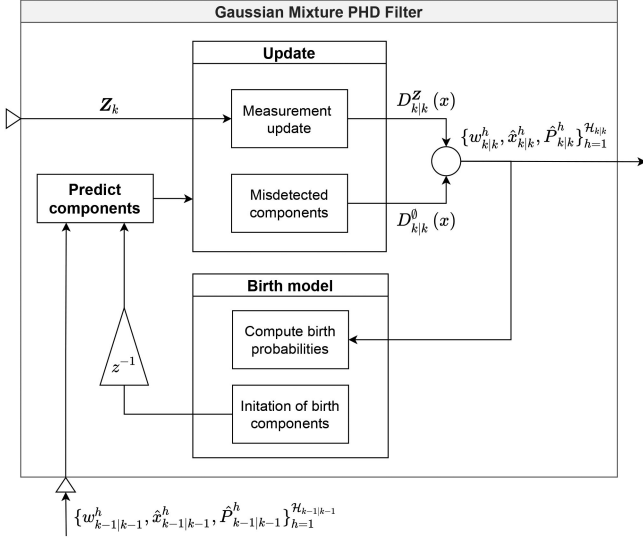


Fig. 3. The flowchart of the extended GM-PHD filter.

representing new objects are drawn to the birth pool. This model reduces the  $\mathcal{H}_{k|k-1}$  number of predicted components such as the runtime and avoids the duplicated measurements that can distort the estimated states  $\hat{x}_{k|k}$  of objects. The conditional birth probability  $p_b(z_k^i | D_{k|k}(x))$  considers the PHD function to compute the probability of the event that a measurement  $z_k^i$  represents a new object. The birth probability is computed based on the  $\{w_{k|k}^h\}_{h=1}^{\mathcal{H}_{k|k}}$  updated component weights assuming their association probabilities as:

$$p_b(z_k^i | D_{k|k}(x)) = 1 - \sum_{h=1}^{\mathcal{H}_{k|k-1}} w_{k|k}^{i\mathcal{H}_{k|k-1}+h} \quad (22)$$

Each measurement with a birth probability above a given  $p_{b,t}$  threshold initiates a new component in the prior PHD based on the inverse measurement model and it will be updated in the next cycle. The pseudo code of the birth model is given in Alg. 2, where  $\hat{P}_{0,n_z+}$  denotes the initial covariance containing the initial uncertainties of the state variables unobserved by the sensor providing the measurement. The weights  $w_{k|k}^b$  of the birth components equal to the initial existence probability  $p_{\text{init}}(\exists x_k^b)$  of the corresponding object are computed based on the confidence  $p_{TP}(z_k^i)$  of detection  $z_k^i$ , birth probability  $p_b(z_k^i | D_{k|k}(x))$ , clutter density  $\kappa(z_k^i)$ , and birth density  $b(z_k^i)$ . The  $l_{k|k}^b$  labels of the birth components are defined based on the  $l_{k|k}^h$  labels of the updated components and the user-defined  $n_{max}$  maximum number of objects by identifying the first unoccupied object ID. The GM-PHD filter extended with the proposed birth model is represented in Fig. 3.

#### D. Object Extraction

The GM-PHD filter detailed in Alg. 1 updates the PHD function by  $(m_k + 1)\mathcal{H}_{k|k-1}$  number of components. Without a mixture reduction, the number of components would diverge, and so does the algorithm's runtime. The purpose of the mixture reduction is not just to prune the unlikely components but also

#### Algorithm 2: Birth Model.

---

```

1: given  $\{w_{k|k}^h, l_{k|k}^h\}_{h=1}^{\mathcal{H}_{k|k}}, \mathcal{Z}_k, n_{max}$ 
2:  $b = 0, i_b = \{1, \dots, n_{max}\}$ 
3: for  $i = 1$  to  $m_k$  do
  Compute birth probability:
4:  $p_b(z_k^i | D_{k|k}(x)) = 1$ 
5: for  $h = 1$  to  $\mathcal{H}_{k|k-1}$  do
6:    $p_b(z_k^i | D_{k|k}(x)) = p_b(z_k^i | D_{k|k}(x)) - w_{k|k}^{i\mathcal{H}_{k|k-1}+h}$ 
7: end for
  Initiate component:
8: if  $p_b(z_k^i | D_{k|k}(x)) \geq p_{b,t}$  then
9:    $b = b + 1$ 
10:   $\hat{x}_{k|k}^b = H_k^{-1} z_k^i$ 
11:   $\hat{P}_{k|k}^b = H_k^{-1} R_k^i (H_k^{-1})^\top + \hat{P}_{0,n_z+}$ 
12:   $w_{k|k}^b = p_{TP}(z_k^i) p_b(z_k^i | D_{k|k}(x)) \frac{b(z_k^i)}{b(z_k^i) + \kappa(z_k^i)}$ 
13:   $l_{k|k}^b = \min(i_b \notin \{l_{k|k}^h\}_{h=1}^{\mathcal{H}_{k|k}})$ 
14: end if
15: end for
16:  $\mathcal{H}_{k|k}^b = b$ 
17: go to: Object extraction
output:  $\{w_{k|k}^b, \hat{x}_{k|k}^b, \hat{P}_{k|k}^b, l_{k|k}^b\}_{b=1}^{\mathcal{H}_{k|k}^b}$ 

```

---

to extract the states of present objects. Therefore, the mixture reduction and the state extraction greatly impact the performance of the multi-object tracking algorithm. Erdinc *et al.* investigated the effect of missed detections on the existence probability of an object in a single object scenario [47]. In case of single object tracking, the  $\hat{n}_{k|k}$  number of expected objects at timestep  $k$  equals to the  $p_{k|k}(\exists x_i)$  existence probability of object  $x_i$ , that is computed as:

$$\hat{n}_{k|k} = p(\exists x_i) |_{n_k=1} = (1 - p_d(x_k)) \hat{n}_{k|k-1}, \quad (23)$$

where  $\hat{n}_{k|k-1} = p_S(x_i) \hat{n}_{k-1|k-1}$  denotes the predicted number of objects (existence probability). However, this expression underestimates the existence probability (the expected number of objects) that can result false negative objects in the environment representation. In Bayesian manner the right estimation of the existence probability would be:

$$p_{k|k}(\exists x_i) = \frac{(1 - p_d(x_i)) p_{k|k-1}(\exists x_i)}{1 - p_{k|k-1}(\exists x_i) p_d(x_i)}. \quad (24)$$

Some papers also focus on robust object extraction to overcome the underestimation of the object cardinality [48], [49]. Choi *et al.* proposed a robust GM-PHD filter that considers component merging, duplication check, and the tracking score of the objects [49]. However, their method requires a gating parameter, such as IPDA or JIPDA, that may affect the tracking algorithm's performance and runtime. The proposed method associates the  $\{w_{k|k-1}^h, \hat{x}_{k|k-1}^h, \hat{P}_{k|k-1}^h\}_{h=1}^{\mathcal{H}_{k|k-1}}$  predicted components and the  $\mathcal{Z}_k$  measurements based on the  $p(\hat{x}_{k|k}^h \leftrightarrow z_k^i)$  association probabilities defined by the  $w_{k|k}^h$  updated component weights. Therefore, an adaptive gating parameter is realized. Furthermore,

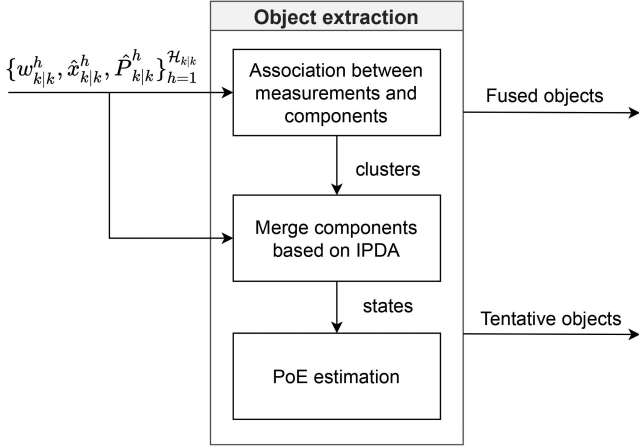


Fig. 4. The flowchart of the Object extraction module.

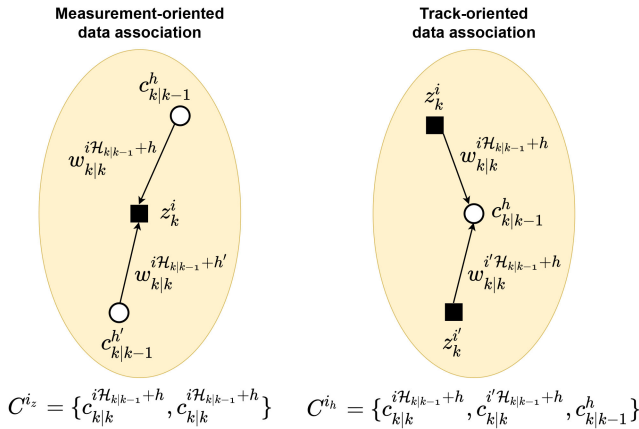


Fig. 5. The two types of data association.

the existence probability of the objects is computed based on the Bayesian rule given in (24), hence the underestimation of the number of objects is resolved without computing tracking scores. The flowchart of the object extraction is shown in Fig. 4.

The first step of the module is the data association that obtains the clusters whose members will be fused. This step is performed based on a locally adaptive k-nearest neighbor (Local kNN) algorithm [50]. The  $k_{max}$  parameter is applied as a saturation value for the number of cluster members, and the neighbors are identified based on the Greedy algorithm. Two data association approaches are proposed: the measurement-oriented and the track-oriented, compared in Fig. 5 and described in Alg. 3 and Alg. 4. The measurement-oriented approach assigns the  $\{c_{k|k-1}^h\}_{h=1}^{\mathcal{H}_{k|k-1}} = \{w_{k|k-1}^h, \hat{x}_{k|k-1}^h, \hat{P}_{k|k-1}^h, l_{k|k-1}^h\}_{h=1}^{\mathcal{H}_{k|k-1}}$  predicted components to the  $\{z_k^i\}_{i=1}^{m_k}$  current measurements; in contrast, the detections are assigned to the predicted components in the track-oriented. The measurement-oriented algorithm creates two cluster groups: the  $C_Z$  measurement-updated and the  $C_\emptyset$  misdetection cluster group. The measurement-updated group consists of clusters that contain one measurement, while the misdetection group involves those components that are not assigned to either of the measurements. The drawback of this approach is that the clusters in the measurement-update group, in most cases, represent one Bernoulli component of the updated

---

**Algorithm 3: Object Extraction - Measurement-Oriented Data Association.**


---

```

1: given  $\{w_{k|k}^h, \hat{x}_{k|k}^h, \hat{P}_{k|k}^h, l_{k|k}^h\}_{h=1}^{\mathcal{H}_{k|k}}, k_{max}$ 
2:  $C_Z = \emptyset, i_{C_Z} = 0$ 
3:  $C_\emptyset = \emptyset, i_{C_\emptyset} = 0$ 
4:  $\forall i_z \in \{1, \dots, m_k\} \quad k_{i_z} = 0$ 
Association probability:
5:  $p(\hat{x}_{k|k}^h \leftrightarrow z_k^i) := w_{k|k}^{i\mathcal{H}_{k|k-1}+h}$ 
6: for  $h = 1$  to  $\mathcal{H}_{k|k-1}$  do
7:    $i_z = \operatorname{argmax}_{i \in \{0, \dots, m_k\}} p(\hat{x}_{k|k}^h \leftrightarrow z_k^i)$ 
8:    $c = \{w_{k|k}^{i_z, h}, \hat{x}_{k|k}^{i_z, h}, \hat{P}_{k|k}^{i_z, h}, l_{k|k}^{i_z, h}\}$ 
9:   if  $i_z = 0$  then
10:     $p(\hat{x}_{k|k}^h \leftrightarrow z_k^{i_z}) = 0$ 
11:     $i_{C_\emptyset} = i_{C_\emptyset} + 1$ 
12:     $C_\emptyset^{i_{C_\emptyset}} = c := c_\emptyset$ 
13:   else if  $i_z > 0 \wedge k_{i_z} < k_{max}$  then
14:     $p(\hat{x}_{k|k}^h \leftrightarrow z_k^{i_z}) = 0$ 
15:     $k_{i_z} = k_{i_z} + 1$ 
16:    if  $k_{i_z} = 1$  then
17:      $i_{C_Z} = i_{C_Z} + 1$ 
18:    end if
19:     $C_Z^{i_{C_Z}} = \{C_Z^{i_{C_Z}}, c\}$ 
20:  end if
21: end for
22:  $C = \{C_Z, C_\emptyset\}$ 
23: go to: Object extraction - Merging and PoE estimation
output:  $\{C^{i_C}\}_{i_C=1}^{i_{C_Z}+i_{C_\emptyset}}$ 

```

---

PHD, neglecting the misdetection probability that is resolved in the merging process. The track-oriented approach assigns the current measurements to the predicted components, meaning multiple measurements can be associated with a component. Furthermore, since the bases of the clusters are the predicted components, each cluster is initialized by its  $c_\emptyset$  missed detected component. The track-oriented data association creates one  $\{C^h\}_{h=1}^{\mathcal{H}_{k|k-1}}$  cluster group with clusters centralized to the predicted components. In both approaches, the cluster members correspond to a component of the updated PHD function: a measurement-previous component pair represents a Bernoulli component, and a singleton component refers to the PPP part of the updated PHD.

The second step of the object extraction is the merging and PoE estimation described in Alg. 5, that takes the  $\{C^{i_C}\}_{i_C=1}^{n_C}$  clusters, the  $\{w_{k|k-1}, l_{k|k-1}\}_{h=1}^{\mathcal{H}_{k|k-1}}$  weights and labels of the predicted components, and the  $\mathcal{O}_{k-1} = \{\hat{x}_{k-1}^o, \hat{P}_{k-1}^o, l_{k-1}^o\}_{o=1}^{\hat{n}_{k-1|k-1}}$  previous objects defined by their estimated state, covariance and unique ID as inputs. First, the corresponding predicted component of a cluster must be identified by its  $l_{i_C}$  unique label for the recursive Bayesian PoE estimation. This process is slightly different for measurement-oriented and track-oriented clusters. Since the measurement-oriented clusters may contain more than one predicted component, the corresponding predicted component is identified by fusing the  $l_{k|k}^c \in C^{i_C}$  labels of the cluster members. Furthermore, suppose a measurement-oriented cluster

---

**Algorithm 4:** Object Extraction - Track-Oriented Data Association.
 

---

```

1: given  $\{w_{k|k}^h, \hat{x}_{k|k}^h, \hat{P}_{k|k}^h, l_{k|k}^h\}_{h=1}^{\mathcal{H}_{k|k}}, k_{max}$ 
2:  $\forall h \in \{1, \dots, \mathcal{H}_{k|k-1}\}, \mathbf{C}^h = \emptyset$ 
3:  $\forall i_h \in \{1, \dots, \mathcal{H}_{k|k-1}\}, k_{i_h} = 0$ 
4:  $h = \{1, \dots, \mathcal{H}_{k|k-1}\}$ 
Association probability:
5:  $p(\hat{x}_{k|k}^h \leftrightarrow z_k^i) := w_{k|k}^{i\mathcal{H}_{k|k-1}+h}$ 
Assign missed detected component to every cluster:
6: for  $h = 1$  to  $\mathcal{H}_{k|k-1}$  do
7:  $c_\emptyset = \{w_{k|k}^h, \hat{x}_{k|k}^h, \hat{P}_{k|k}^h, l_{k|k}^h\}$ 
8:  $\mathbf{C}^h = \{\mathbf{C}^h, c_\emptyset\}$ 
9: end for
Assign measurements to clusters:
10: for  $i = 1$  to  $m_k$  do
11:  $i_h = \operatorname{argmax}_{h \in \{1, \dots, \mathcal{H}_{k|k-1}\}} p(\hat{x}_{k|k}^h \leftrightarrow z_k^i)$ 
12:  $c_Z = \{w_{k|k}^{i,i_h}, \hat{x}_{k|k}^{i,i_h}, \hat{P}_{k|k}^{i,i_h}, l_{k|k}^{i,i_h}\}$ 
13: if  $k_{i_h} < k_{max}$  then
14:  $\mathbf{C}^{i_h} = \{\mathbf{C}^{i_h}, c_Z\}$ 
15:  $k_{i_h} = k_{i_h} + 1$ 
16:  $p(\hat{x}_{k|k}^{i_h} \leftrightarrow z_k^i) = 0$ 
17: end if
18: end for
19: go to: Object extraction - Merging and PoE estimation
output:  $\{\mathbf{C}^{i_C}\}_{i_C=1}^{\mathcal{H}_{k|k-1}}$ 

```

---

is the member of the  $\mathbf{C}_Z$  measurement-updated cluster group. In that case, the cluster does not consider the misdetection of the corresponding predicted component. Therefore, after identifying the  $l_{i_C}$  ID of the corresponding predicted component, the  $c \in \mathbf{C}_\emptyset | l_{c|k}^c = l_{i_C}$  miss-detected component generated by it is appended to the cluster and removed from the  $\mathbf{C}_\emptyset$  misdetection cluster group. In the track-oriented case, the corresponding predicted component's ID equals the cluster's index because the clusters are organized around the predicted components. In both cases, the label of the updated cluster will be inherited from the corresponding predicted component. The state of the merged cluster is extracted based on the  $\beta_c$  IPDA weights of the cluster members. The  $p_{k|k}(\exists x_{i_C})$  existence probability of the objects represented by the  $x_{i_C}$  fused clusters is computed based on the Bayes rule as in the conventional IPDA algorithm. The Gaussian components of the reduced posterior are given by the clusters that contain at least one measurement and the miss-detected components with  $w_{k|k}^{c_0}$  weight above a given  $\gamma_h$  component existence threshold. If a component meets one of these criteria, it is drawn to the reduced PHD function, but with  $w_{k|k}^{i_h} = p_{k|k}(\exists x_{i_C})$  weight updated in Bayesian manner. The present objects are extracted similarly; however, two existence thresholds are applied. If the  $l_{i_C}$  cluster label is represented in the  $\mathcal{O}_{k-1}$  previous object list, the  $p_{k|k}(\exists x_{i_C})$  existence probability of the cluster is compared to a  $\gamma_{\exists,low}$  lower existence threshold. If it is tentative yet, the existence probability must reach a  $\gamma_{\exists,upp}$  higher threshold for the confirmation. The clusters that meet the existence criteria are called confirmed objects, and the other

clusters are referred to as tentative objects. The output of the object extraction is the  $\mathcal{O}_k$  list of the confirmed (present) objects and the reduced posterior PHD considering both the confirmed and tentative objects.

#### IV. EVALUATION

In the evaluation, the proposed GM-PHD based sensor fusion extended with the robust object extraction methods is compared with two conventional GM-PHD filters, namely the Poisson and threshold approaches and the state-of-the-art JIPDA algorithm. In the Poisson approach, the  $\hat{n}_{k|k}$  expected number of objects is computed according to (20) and after sorting the components by their weights, the first  $\hat{n}_{k|k}$  components of the PHD function initiates an object in the output. In the other conventional approach, the states of the components with weight greater than  $\gamma_{\exists}$  threshold are extracted as present objects. The  $\gamma_{\exists}$  existence threshold is set equal to  $\gamma_{\exists,upp}$  the object initiation threshold of the proposed object extractions. Both of the conventional GM-PHD filters are extended with the proposed birth model described in Section III-C. The JIPDA filter is parameterized in the same manner as GM-PHD filters but according to [15] it expects  $z$  sensor detections in a  $V_k^t$  validation gate around track  $t$  at time  $k$  neglecting unlikely data associations. We applied 7 different  $V = V_k^t$  constant validation gate parameters but only the results of the optimal  $V = 100$  is explained in Section V.

The performance of the five sensor fusion algorithms is evaluated in a simulation environment and based on real-world measurement data as well. In the simulation environment, the performance of the algorithms is investigated based on GOSPA (Generalized Optimal Sub-pattern Assignment) metric [51], that considers the localization error of the objects and the cardinality errors. The cardinality error involves both the false-negative (missed detections) and false-positive objects. The GOSPA is parametrized by  $d_c = 10$  cut-off distance threshold,  $p = 2$  order and  $\alpha$  parameter. In many cases, the performance of the environment perception is given by  $Pr$  precision, and  $Rc$  recall metrics as in [4]. If the alpha parameter is set as  $\alpha = 2$ , the localization error, the missed detection, and the false positive detection components of the total GOSPA metric can be separated. Therefore, the precision and recall can be computed as:

$$Pr = \frac{TP}{TP + FP}, \quad (25)$$

$$Rc = \frac{TP}{TP + FN}, \quad (26)$$

where  $TP$ ,  $FP$ , and  $FN$  denote the total number of true-positive, false-positive, and missed detections.

The ground truth of the real measurement data is performed by video annotation. Thus it is unsuitable for computing an accurate localization error. Therefore, in the evaluation of the real measurement data, the F1 Score and MOTA (Multi-Object Tracking Accuracy) metric of the algorithms and sensors are computed as in [52], considering the precision, recall, and track ID mismatches. The annotated objects are transformed to the vehicle coordinate system based on the homography matrix of the mono-vision camera. The output tracks are associated with the ground truth based on GNN (Global Nearest Neighbor)



**Algorithm 5:** Object Extraction - Merging and PoE Estimation.

---

```

1: given  $\{C^{i_C}\}_{i_C=1}^{n_C}, \{w_{k|k-1}, \hat{x}_{k|k-1}, l_{k|k-1}\}_{h=1}^{\mathcal{H}_{k|k-1}}$ ,
 $\mathcal{O}_{k-1} = \{\hat{x}_{k-1}^o, \hat{P}_{k-1}^o, l_{k-1}^o\}_{o=1}^{\hat{n}_{k-1}}$ 
2:  $i_h = 0, i_o = 0$ 
3: for  $i_C = 1$  to:  $n_C$  do
4:  $C^{i_C} = \{w_{k|k}^c, \hat{x}_{k|k}^c, \hat{P}_{k|k}^c, l_{k|k}^c\}_{c=1}^{n(C^{i_C})}$ 
Find corresponding predicted component:
5: if  $C^{i_C}$  measurement-oriented then
6:  $\hat{c} = \text{round}(\sum_c \beta_c c)$ 
7:  $l_{i_C} = l_{k|k}^{\hat{c}} \in C^{i_C}$ 
Append missed detected component:
8: if  $\exists c \in C^{i_C} \neq c_\emptyset$  then
9:  $C^{i_C} = \{C^{i_C}, c \in C_\emptyset | l_{k|k}^c = l_{i_C}\}$ 
10:  $C_\emptyset = C_\emptyset \setminus \{c \in C_\emptyset | l_{k|k}^c = l_{i_C}\}$ 
11: end if
12: else
13:  $l_{i_C} = i_C$ 
14: end if
Merging components:
15: for each  $\{w_{k|k}^c, \hat{x}_{k|k}^c, \hat{P}_{k|k}^c, l_{k|k}^c\}_{c=1}^{n(C^{i_C})} \in C^{i_C}$  do
16:  $\beta_c = \frac{w_{k|k}^c}{\sum_c w_{k|k}^c}$ 
17: end for
18:  $\hat{x}_{k|k}^{i_C} = \sum_c \beta_c \hat{x}_{k|k}^c$ 
19:  $\epsilon^c = (\hat{x}_{k|k}^{i_C} - \hat{x}_{k|k}^c)(\hat{x}_{k|k}^{i_C} - \hat{x}_{k|k}^c)^\top$ 
20:  $\hat{P}_{k|k}^{i_C} = \sum_c \beta_c (\hat{P}_{k|k}^c + \epsilon^c)$ 
Bayesian PoE estimation:
21:  $p_{k|k-1}(\exists x_{i_C}) = \frac{w_{k|k-1}^{i_C} | l_{k|k-1}^{i_C} = l_{i_C}}{\sum_c w_{k|k-1}^c}$ 
22:  $p_{k|k}(\exists x_{i_C}) = \frac{\sum_c w_{k|k}^c | l_{k|k}^c = l_{i_C}}{\sum_c w_{k|k}^c + 1 - p_{k|k-1}(\exists x_{i_C})}$ 
Create reduced posterior PHD components
23: if  $\exists c \in C^{i_C} \neq c_\emptyset \vee w_{k|k}^{c_\emptyset} > \gamma_h$  then
24:  $i_h = i_h + 1$ 
25:  $w_{k|k}^{i_h} = p_{k|k}(\exists x_{i_C})$ 
26:  $\hat{x}_{k|k}^{i_h} = \hat{x}_{k|k}^{i_C}, \hat{P}_{k|k}^{i_h} = \hat{P}_{k|k}^{i_C}$ 
27:  $l_{k|k}^{i_h} = l_{i_C}$ 
28: end if
Extract present objects
29: if  $\exists l_{k-1}^o \in \mathcal{O}_{k-1} : l_{i_C} = l_{k-1}^o$  then
30:  $\gamma_\exists = \gamma_{\exists, \text{low}}$ 
31: else
32:  $\gamma_\exists = \gamma_{\exists, \text{upp}}$ 
33: end if
34: if  $p_{k|k}(\exists x_{i_C}) > \gamma_\exists$  then
35:  $i_o = i_o + 1$ 
36:  $\hat{x}_k^{i_o} = \hat{x}_{k|k}^{i_C}, \hat{P}_k^{i_o} = \hat{P}_{k|k}^{i_C}$ 
37:  $l_k^{i_o} = l_{i_C}$ 
38: end if
39: end for
40:  $k = k + 1$ 
output:  $\{w_{k|k}^{i_h}, \hat{x}_{k|k}^{i_h}, \hat{P}_{k|k}^{i_h}, l_{k|k}^{i_h}\}_{i_h=1}^{\mathcal{H}_{k|k}}$  PHD components,
 $\mathcal{O}_k = \{\hat{x}_k^{i_o}, \hat{P}_k^{i_o}, l_k^{i_o}\}_{i_o=1}^{\hat{n}_{k|k}}$  confirmed objects

```

---

algorithm with a pre-defined  $d_c$  cut-off distance threshold as in GOSPA; however, the localization error is neglected.

Besides the performance metrics, the proposed robust GM-PHD filters are compared to the state-of-the-art JIPDA filter in terms of computational effort. Since the real-world measurement detailed in Section IV-B includes different sections with a varying number of objects, the runtime comparison is detailed only for the measurement evaluation in Section V-B. Since the  $O(\mathcal{H}_{k|k-1}, m_k)$  algorithm complexity depends on the  $\mathcal{H}_{k|k-1}$  number of tracked components corresponding to potential objects and the  $m_k$  number of detections, the runtime characteristics of the fusion algorithms according to them by computing the average runtime of a  $\mathcal{H}_{k|k-1}, m_k$  pair.

### A. Simulation Environment

The simulation environment performs a high-level simulation of radar and smart camera detections. The real sensor data is pre-evaluated to obtain the clutter and the detection model of the sensors performing a realistic sensor simulation. The detection model is intended to describe the detection probability obtaining the parameters given in (21) based on the sensor FoVs. The clutter model is more complex since the false alarms are not uniformly distributed according to real sensor data. For example, the radar frequently detects the highway guardrail's poles and tracks these irrelevant detections of actual objects. Therefore, the spatial distribution of these false alarms is assumed to be a Gaussian mixture instead of uniform distribution. Therefore, the clutter model of the sensor is described by the sum of two independent parts, the uniform distribution, and the Gaussian mixture, as in:

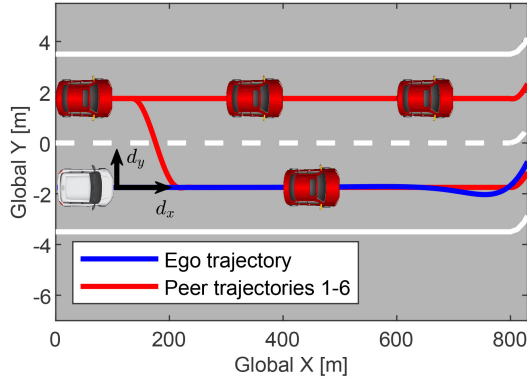
$$\hat{\kappa}(z) = \bar{\lambda}_{\mathcal{U}} \mathcal{U}(V_z) + \bar{\lambda}_{\mathcal{N}} \sum_{i=1}^N \mathcal{N}(z, \hat{c}_i, \hat{P}_i), \quad (27)$$

where  $\mathcal{U}(V_z)$  and  $\mathcal{N}(z, \hat{c}_i, \hat{P}_i)$  denote the uniform distribution in  $V_z$  measurement space volume and the  $i$ -th component of the Gaussian mixture with  $\hat{c}_i$  mean value and  $\hat{P}_i$  covariance. The cardinality of the false detections is assumed to be distributed according to Poisson distribution with different  $\bar{\lambda}_{\mathcal{U}}$  and  $\bar{\lambda}_{\mathcal{N}}$  Poisson rates for the uniformly distributed false alarms and the Gaussian mixture. The states of the guardrail poles are identified based on the dynamic state of the ego vehicle and the lane information. Since the radar tracks the poles, the parameters of the Gaussian components are updated by a Kalman filter detailed in Section IV-C. The measurement delay is simulated by a random value between the minimum and maximum latency of the sensors. The parameters of the sensors in the simulation environment are summarized by Table I.

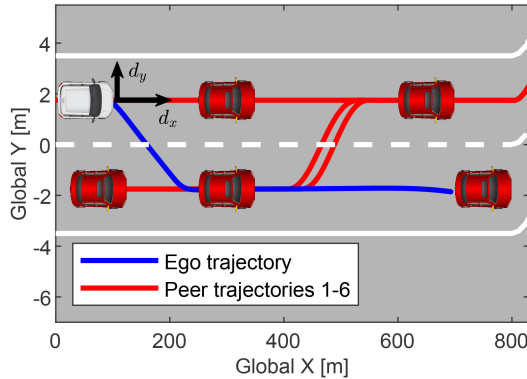
Two common ADAS scenarios, an Adaptive Cruise Control (ACC) and Autonomous Emergency Brake (AEB), are tested in the simulation environment. The scenarios generated with IPG CarMaker are demonstrated by Fig. 6. The track consists of a 900 meters long straight section and an arc with a radius of 750 meters and a total turn angle of  $90^\circ$ . In both scenarios, six traffic vehicles (peer objects) are involved, lasting 40 seconds. A peer vehicle cuts between the ego vehicle and ACC target vehicles during the ACC scenario and forces the ego vehicle to decelerate. Four other objects also participate in this scenario to simulate more realistic traffic. The AEB scenario is more complex and dynamic, including lane change performed by the

TABLE I  
PARAMETERS OF THE SIMULATED SENSORS

Parameter	Notation	Radar	Camera
Detection probability outside the FoV	$\hat{p}_\emptyset$	0.15	0.05
Detection probability within the FoV	$\hat{p}_d$	0.85	0.95
Uniform distributed Poisson rate	$\bar{\lambda}_U$	0.1	0.01
Gaussian mixture Poisson rate	$\bar{\lambda}_N$	4	0
Minimum latency [s]	$\Delta L, min$	0.007	0.01
Maximum latency [s]	$\Delta L, max$	0.004	0.01
Data frequency [Hz]	$f$	13	9



(a)



(b)

Fig. 6. Simulated ADAS scenarios. (a) ADAS scenario I: Adaptive Cruise Control (ACC). (b) ADAS scenario II: Autonomous Emergency Brake (AEB).

ego vehicle. In the beginning, a peer vehicle is driving in front of the ego vehicle inside. Then, the ego vehicle changes into the outside lane behind two other cars and follows them. The two cars initiate a lane change into the inside lane after  $\sim 400$  meters revealing a standstill object. The ego vehicle starts to decelerate and finally stops behind the standstill object while two other objects are passing by the ego vehicle in the left lane, simulating a real traffic flow.

### B. Real-World Measurement Setup

The sensor cluster consists of an automotive Continental ARS408 radar and a Mobileye EyeQ2 smart camera providing the object detections as input data for the EFPS. Fig. 7 depicts

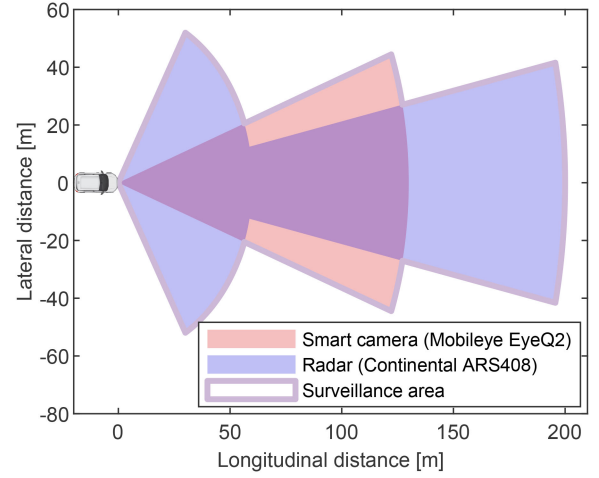


Fig. 7. The FoV of the sensor cluster and the surveillance area.

the FoV of the sensors and the surveillance area. The Mobileye camera has a 130 m view range and  $40^\circ$  FoV angle. The radar has multiple FoVs consisting of a near scan zone up to 60 m with  $120^\circ$  opening angle, and a far-looking one with 200 m view range and  $20^\circ$  angle. The frequency of the Continental radar is about 12 Hz, while the Mobileye camera has a lower  $\sim 9$  Hz frequency. Although the surveillance area extends up to 200 meters due to the radar; however, the measurement evaluation is performed on a smaller area since the ground truth objects can be annotated confidently up to 80 meters. The measurement setup is extended with a mono-vision camera supporting the annotation-based ground truth generation and visualization. The sensor data was logged through CAN communication by Vector CANape, and the video of the mono-vision camera was recorded in sync. The measurement was taken place on the beginning section of the Hungarian M1/M7 highway in Budapest towards Budaörs. The ground truth objects are created on a video section of 6000 frames ( $\sim 60$  FPS, 100 seconds). The annotated frames were selected to evaluate the multi-object tracking in the heaviest traffic and various dynamic scenarios such as lane-keeping in different lanes and lane changes performed by the surrounding objects and the ego vehicle as well.

### C. Model Description

The proposed sensor fusion tracks dynamic objects with  $x_k$  state at time  $k$  relevant to planning a safe trajectory, such as pedestrians, bicycles, and road vehicles including cars, trucks, motorbikes. In the GM-PHD filter, the states of the Gaussian components are estimated by a Kalman filter that requires a process and measurement model. It is beneficial to apply a multi-model estimation considering different vehicle maneuvers as in [53], [54]. Since this paper focuses on the object extraction module, therefore the  $x_k$  object state at timestep  $k$  is computed by a constant linear acceleration (CA) model based on the  $x_{k-1}$  object state at timestep  $k-1$  as in:

$$x_k = F_k x_{k-1} + \nu_k, \quad (28)$$

where  $F_k$  and  $\nu_k$  denote the the transition matrix and the process noise with covariance  $Q_k$ . The state vector is defined as:

$$x_k = \begin{bmatrix} d_x & v_x & a_x & d_y & v_y & a_y \end{bmatrix}^\top \quad (29)$$

capturing the  $x$  longitudinal and  $y$  lateral components of the  $d$  position,  $v$  velocity, and  $a$  acceleration concerning the ego vehicle coordinate system. The transition and process noise covariance matrices defined as

$$F_k = I_2 \otimes \begin{bmatrix} 1 & \Delta t & \Delta t^2/2 \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix}, \quad (30)$$

$$Q_k = \sigma_a^2 \begin{bmatrix} Q_{x,k} & \mathbf{0} \\ \mathbf{0} & Q_{y,k} \end{bmatrix}, \quad (31)$$

and

$$Q_{x,k} = Q_{y,k} = \begin{bmatrix} \Delta t^4/4 & \Delta t^3/2 & \Delta t^2/2 \\ \Delta t^3/2 & \Delta t^2 & \Delta t \\ \Delta t^2/2 & \Delta t & 1 \end{bmatrix}. \quad (32)$$

In (30) and (32)  $\Delta t$  denotes the elapsed time between timestep  $k$  and  $k - 1$ . In (31)  $\sigma_a$  refers to the acceleration scale.

The smart camera provides detections in a Cartesian coordinate system with longitudinal velocity and acceleration estimation. The radar also provides the position of targets in a Cartesian coordinate system, however, it estimates the velocity of objects with longitudinal and lateral components. Therefore observation matrix of smart camera  $H_{k,c}$  and the radar  $H_{k,r}$  are therefore:

$$H_{k,c} = H_c = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \quad (33)$$

$$H_{k,r} = H_r = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \quad (34)$$

The radar estimates the  $R_{k,r}$  measurement covariance of each detection, however, the  $R_{k,c}$  of the camera detections are estimated based on the  $\sigma_{d_x}$ ,  $\sigma_{d_y}$ ,  $\sigma_{v_x}$ ,  $\sigma_{a_x}$  position, velocity, and acceleration uncertainties given in the technical specification of the sensor as in:

$$R_{k,c} = R_c = \text{diag} \left( \sigma_{d_x}^2, \sigma_{d_y}^2, \sigma_{v_x}^2, \sigma_{a_x}^2 \right). \quad (35)$$

The parameters of the GM-PHD filter are obtained in accordance with the simulation environment since it simulates the real sensors. The  $p_d(x)$  state-dependent detection probability is given with the same parameters as in Table I. The  $p_S(x)$  survive probability is given by  $p_S(x) = p_S = 0.9999$  constant value. However, the  $\hat{\kappa}(z)$  clutter model of the simulation environment is not adapted to the tracking algorithm because the proposed method does not consider the lane information, and the parameters of that model depend highly on the environmental circumstances. Therefore, the  $\kappa(z) = \bar{\lambda}_c/V_z$  clutter density is given by a simple uniform distribution above the  $V_z$  measurement volume

TABLE II  
PRECISION, RECALL AND AVERAGE LOCALIZATION ERROR AND GOSPA IN ADAPTIVE CRUISE CONTROL (ACC) SCENARIO

Algorithm	Prec. [-]	Rec. [-]	Localization [m]	GOSPA [m]
Poisson	0.93	0.68	0.73	5.92
Threshold	0.95	0.68	0.71	5.84
JIPDA	0.98	0.96	1.07	2.05
<b>Robust MO</b>	0.99	0.90	0.90	2.56
<b>Robust TO</b>	0.99	0.96	0.96	1.65

with a  $\bar{\lambda}_c$  Poisson rate set to 10 for the radar and 0.01 for the smart camera. Similar to the clutter density, the  $b(z) = \bar{\lambda}_b/V_z$  birth density is defined by a uniform distribution, and the  $\bar{\lambda}_b$  number of expected newborn objects is given as  $\bar{\lambda}_b = 1$  for both sensors. The  $\gamma_{\exists,low}$  and  $\gamma_{\exists,upp}$  probability existence thresholds are set as  $\gamma_{\exists,low} = 0.08$  and  $\gamma_{\exists,upp} = 0.65$ , while the  $\gamma_h$  component surviving threshold  $\gamma_h = 0.03$ .

## V. RESULTS

In this section, the evaluation results of the simulation environment and real-world measurement data are described. The performance of the simulation environment and real-world measurement are investigated in different aspects. In the simulation environment, the exact states of the ground truth objects are known. Therefore, the GOSPA metric of the different multi-object tracking approaches and their components (missed detections, false detections, localization error) are compared. In practical applications, the precision and recall of the environment perception are essential metrics. Hence, they are also included in the simulation results. However, the state of the objects annotated on the mono-vision images of real-world measurement is approximated with uncertainty, suitable only for the assignment between fused objects and ground truth objects. Thus, the sensor fusion algorithms are evaluated by the MOTA metric considering the precision, recall, and track ID switches neglecting the localization error. Furthermore, since the real sensors also provide tracked objects, their performance metrics are also obtained. The performance gained by the sensor fusion compared to the raw sensor performance is a significant aspect of the EFPS.

The complexity of the proposed robust GM-PHD filter is investigated according to the runtime map over the number of tracked objects and the number of sensor detections compared to the state-of-the-art JIPDA algorithm. The runtime of conventional GM-PHD filters is not detailed since the proposed sensor fusion relies on it; therefore, their complexity has the same characteristics. The runtime evaluation is performed only for the real-world measurement since it includes different sections, consisting of relatively few and many objects as well.

### A. Simulation Results

The results of the simulated Adaptive Cruise Control (ACC) and Autonomous Emergency Brake (AEB) scenarios are summarized in Table II and III, detailing the precision, recall, average localization error, and GOSPA metric of the algorithms, highlighting the proposed algorithms with bold fonts. According

TABLE III  
PRECISION, RECALL AND AVERAGE LOCALIZATION ERROR AND GOSPA IN  
AUTONOMOUS EMERGENCY BRAKE (AEB) SCENARIO

Algorithm	Prec. [-]	Rec. [-]	Localization [m]	GOSPA [m]
Poisson	0.86	0.69	0.80	6.82
Threshold	0.89	0.69	0.80	6.56
JIPDA	0.93	0.95	1.21	2.93
<b>Robust MO</b>	0.99	0.94	1.18	2.15
<b>Robust TO</b>	0.99	0.95	1.17	1.98

to Table II and III, the two conventional GM-PHD filters, the Poisson and the Threshold concept, have very similar results in each aspect of the performance metrics. Both of them filter the false detections quite efficiently, resulting in  $\sim 86 - 95\%$  precision values; however, scenario objects are often missed detected, causing a lower recall metric. The state-of-the-art JIPDA algorithm performs significantly better than the conventional GM-PHD filters, reducing the average GOSPA metric by  $\sim 55 - 66\%$ . The proposed robust proposed GM-PHD filters have a similarly powerful performance to the JIPDA. The measurement-oriented (MO) GM-PHD filter has a slightly lower recall in the ACC scenario. Still, the track-oriented approach has a better result than the JIPDA filter in every performance aspect. In the dynamic AEB scenario, both proposed robust GM-PHD filters reduce the GOSPA metric of the JIPDA by 27% and 32%. The evolution of results is shown in Fig. 8 over the simulation time, detailing the number of missed and false detections, the localization error, and the GOSPA metric.

A moving average filters the localization error and GOSPA metric with 2 seconds window size. Since the two conventional (Poisson and Threshold) and proposed robust (MO and TO) GM-PHD filters have similar results to their counterparts, only the better ones (Threshold and TO) are visualized in Fig. 8. The proposed robust object extraction realizes much fewer missed detections compared to the conventional GM-PHD filter in both scenarios according to Fig. 8(a) and Fig. 8(b). The estimation of the existence probability explains this: since the detection probability is high if an object is miss-detected, its PoE would be underestimated by the conventional GM-PHD filter, and it is pruned. Although the radar provides much clutter, the difference between the algorithms more negligible in false detections. The conventional tracker confirms an object provided by the radar if it detects it in two consecutive cycles; therefore, it can filter most false detections. However, according to the comparison between the conventional and TO object extraction approaches, it can be said that the object cardinality is estimated more robustly by the proposed methods in the corner cases of scenarios. The proposed robust GM-PHD filter has a similar performance to JIPDA in object cardinality estimation. Still, it filters the false detections more efficiently in both scenarios because JIPDA duplicates some objects. The localization error of the algorithms increases when the relative velocity between the surrounding objects and ego vehicle is high. Under these conditions, the conventional GM-PHD filter has slightly more accurate localization than the proposed method because the fused state of the objects relies more on the measurement. However, the robust GM-PHD filter estimates the object state more precisely than

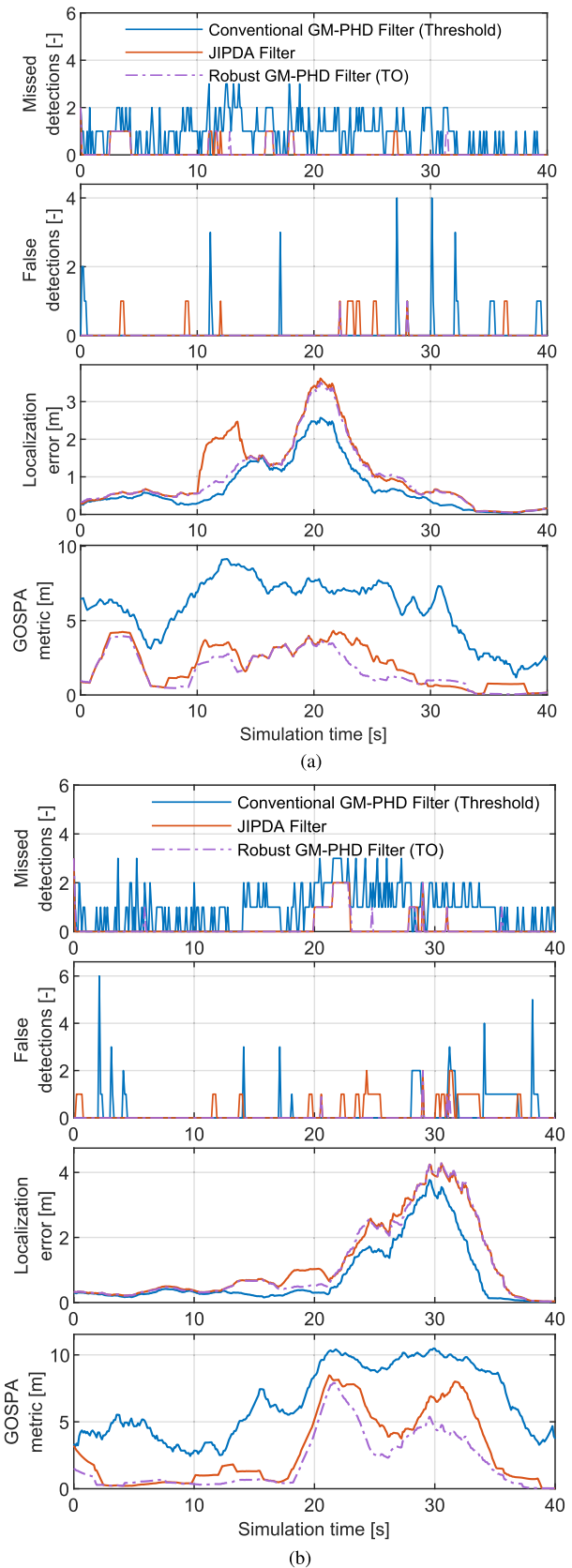


Fig. 8. Missed detections, false detections, Localization error and total GOSPA metric of simulated ADAS scenarios. (a) Adaptive Cruise Control (ACC) scenario. (b) Autonomous Emergency Brake (AEB) scenario.

TABLE IV  
PERFORMANCE METRICS OF RAW SENSOR DATA AND FUSION ALGORITHMS

Sensor/Fusion	Precision [-]	Recall [-]	F1 Score [-]	MOTA [-]
Radar	0.48	0.89	0.63	-0.08
Camera	0.98	0.72	0.83	0.71
Poisson	0.77	0.72	0.74	0.50
Threshold	0.95	0.68	0.79	0.62
JIPDA	0.95	0.86	0.90	0.81
<b>Robust MO</b>	0.96	0.87	0.91	0.83
<b>Robust TO</b>	0.96	0.86	0.90	0.82

the state-of-the-art JIPDA. According to the simulation results, the proposed GM-PHD filters provide a robust object extraction that slightly exceeds even the JIPDA filter's performance thanks to the existence probability computed in a Bayesian manner. Furthermore, the data association method does not influence its performance because the results of the MO and TO approaches do not differ.

### B. Real-World Measurement Results

The results of the real-world measurement are summarized in Table IV, including the precision, recall, F1 score, and the MOTA metric of the raw sensor data and the five sensor fusion algorithms. As it was described in Section IV-A, the radar frequently detects and tracks the poles of the highway guardrail and other irrelevant objects that are considered as false positive objects. Hence the precision of the radar is low compared to the smart camera. The MOTA metric is lower than zero because the number of false detections is larger than true positives. Although the smart camera has high precision, the recall is lower than the radar despite its high detection probability because it can provide four object detections maximally. The result of the five fusion algorithms has similar characteristics as in the simulation environment. However, in the case of Poisson distributed object cardinality, the real objects are often duplicated, resulting in more false detections and lower precision. The object extraction performed based on the Threshold method filters most of the false detections reported by the radar, but it frequently miss-detects real objects. The proposed robust object extraction approach tackles the false detections even more efficiently, exceeding the 95% precision of the JIPDA based fusion too. Although the 86 – 87% recall is a bit lower than precision, the proposed GM-PHD filters reduce the number of missed detections of the conventional object extraction methods by 15 – 19%. The JIPDA and proposed robust GM-PHD filters have similar F1 scores, increasing the resultant performance of the conventional GM-PHD filters by 12 – 17%. Still, the proposed fusion algorithm with MO data association exceeds the F1 score of the JIPDA, with 1%. Furthermore, the fusion algorithm performed by the proposed methods increases the F1 score of the best sensor (smart camera) by 7 – 8%. The difference is more significant in terms of the MOTA metric because the corresponding track ID of an object is not switched since the multi-object tracking is more stable.

The runtime evaluation was performed in Matlab R2020b environment on notebook with Intel Core i7-3520 M (2.9 GHz)

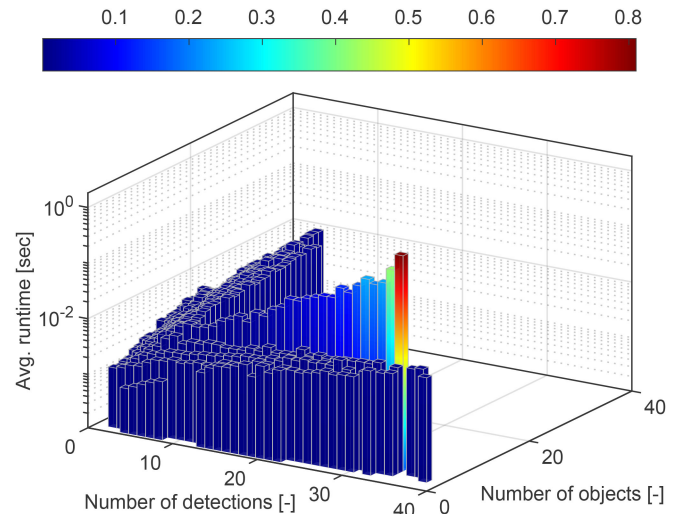


Fig. 9. The average runtime of the JIPDA based sensor fusion depending on the number of tracked objects and detections.

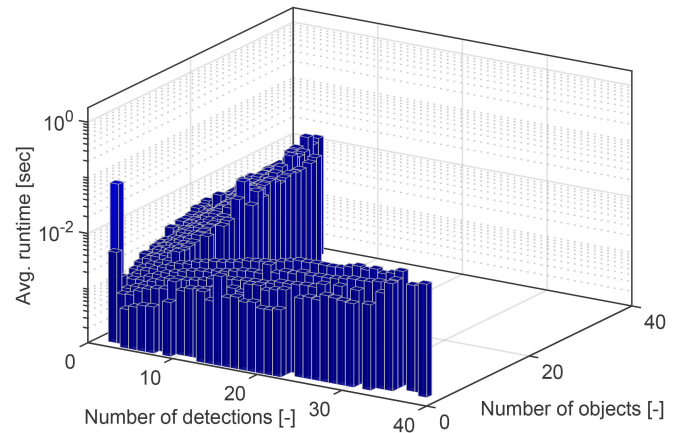


Fig. 10. The average runtime of the robust GM-PHD based sensor fusion depending on the number of tracked objects and detections.

processor and 16 GB memory. The resulting characteristics of the state-of-the-art JIPDA algorithm and proposed robust GM-PHD filter over the number of tracked objects and the number of input detections are depicted with logarithmic scale in Fig. 9 and Fig. 10 by the average runtime. Because of the properties of the different sensors, both Fig. 9 and Fig. 10 form an L shape. The radar detects many more objects, including false positives; in contrast, the smart camera can provide four detections at most. Therefore, the radar initiates plenty of objects, and the camera confirms some of them and vice versa. Fig. 9 shows a >800 ms significant peak in average runtime. This drawback of the JIPDA occurs when some objects are relatively close to each other and share some detections within their validation gate, forming one cluster. In this case, the runtime increases exponentially with the number of objects and detections within the cluster due to the combinatorial complexity clearly visible on the linear stage (logarithmic scale) in Fig. 9. However, Fig. 10 demonstrates that the complexity of the proposed robust GM-PHD based sensor fusion is scaled linearly with the number of tracked objects and

sensor detections over the entire measurement. The runtime has a 60 ms peak occurred at the initialization of the filter that is still lower than the period time of the sensors. Therefore this algorithm is assumed to be applicable in real traffic conditions considering that Matlab is an insufficient environment in terms of runtime compared to a C/C++ implementation.

## VI. CONCLUSION

Although the conventional GM-PHD filter can track multiple time-varying numbers of objects by inserting the first-order moment of the random finite set PDF into the Bayesian framework. However, due to this simplification, the object extraction results in a poor performance in the estimation of object cardinality because the existence probability estimation does not rely on Bayesian recursion. Furthermore, the conventional GM-PHD filter does not involve a sophisticated model that identifies newborn objects. The proposed methods extend the conventional GM-PHD filter with an object birth model that relies on the detections of the sensors considering their birth probability. This birth model reduces the number of duplicated objects and the runtime of the algorithm. The two robust object extraction modules significantly increase the performance of the conventional GM-PHD filters based on the Bayesian existence probability estimation. According to the KPI of the proposed sensor data fusion algorithm, besides the fused data results in a more reliable environment perception than using either of the sensors alone, it reaches and slightly exceeds the performance of the state-of-the-art JIPDA algorithm. Therefore, they provide a robust multi-object tracking fusing multiple sensor data that is a fundamental part of the enhanced frontal perception system of ADAS and other highly automated functions. Furthermore, our solution preserves the favorable complexity of the GM-PHD filter, the linear scaling with the number of objects and detections; hence, there is a powerful runtime reduction compared to the exponential characteristic of JIPDA based sensor fusion.

## REFERENCES

- [1] D. Földes and C. Csiszár, "Framework for planning the mobility service based on autonomous vehicles," in *Proc. Smart City Symp. Prague*, 2018, pp. 1–6.
- [2] C. Csiszár and A. Zarkeshev, "Demand-capacity coordination method in autonomous public transportation," *Transp. Res. Procedia*, vol. 27, pp. 784–790, 2017.
- [3] Q. Lu and T. Tettamanti, "Traffic control scheme for social optimum traffic assignment with dynamic route pricing for automated vehicles," *Periodica Polytechnica Transp. Eng.*, vol. 49, no. 3, pp. 301–307, 2021.
- [4] H. Zhu, K. Yuen, L. Mihaylova, and H. Leung, "Overview of environment perception for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 10, pp. 2584–2601, Oct. 2017.
- [5] J. Van Brummelen, M. O'Brien, D. Gruyer, and H. Najjaran, "Autonomous vehicle perception: The technology of today and tomorrow," *Transp. Res. Part C: Emerg. Technol.*, vol. 89, pp. 384–406, 2018.
- [6] H. Lee, J. Yang, and K. Moessner, "Implementation of a collision avoidance system to assist safe driving based on data fusion in vehicular networks," in *Proc. Int. Conf. Inf. Commun. Technol. Convergence*, 2020, pp. 795–800.
- [7] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Range-focused fusion of camera-IMU-UWB for accurate and drift-reduced localization," *IEEE Robot. Automat. Lett.*, vol. 6, no. 2, pp. 1678–1685, Apr. 2021.
- [8] R. Liu *et al.*, "Cost-effective mapping of mobile robot based on the fusion of UWB and short-range 2-D LiDAR," *IEEE/ASME Trans. Mechatronics*, to be published, doi: [10.1109/TMECH.2021.3087957](https://doi.org/10.1109/TMECH.2021.3087957).
- [9] Y. Bar-Shalom and E. Tse, "Tracking in a cluttered environment with probabilistic data association," *Automatica*, vol. 11, no. 5, pp. 451–460, 1975.
- [10] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, Y. Bar-Shalom and T. E. Fortmann Eds. Boston, MA, USA: Academic Press, 1988.
- [11] J. K. Tugnait, "Tracking of multiple maneuvering targets in clutter using multiple sensors, IMM, and JPDA coupled filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 1, pp. 320–330, Jan. 2004.
- [12] F. García, A. de la Escalera, and J. M. Armingol, "Joint probabilistic data association fusion approach for pedestrian detection," in *Proc. IEEE Intell. Veh. Symp. (IV)*, 2013, pp. 1344–1349.
- [13] F. García, D. Martín, A. de la Escalera, and J. M. Armingol, "Sensor fusion methodology for vehicle detection," *IEEE Intell. Transp. Syst. Mag.*, vol. 9, no. 1, pp. 123–133, Jan. 2017.
- [14] D. Musicki, R. Evans, and S. Stankovic, "Integrated probabilistic data association," *IEEE Trans. Autom. Control*, vol. 39, no. 6, pp. 1237–1241, Jun. 1994.
- [15] D. Musicki and R. Evans, "Joint integrated probabilistic data association: JIPDA," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 40, no. 3, pp. 1093–1099, Jul. 2004.
- [16] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. 24, no. 6, pp. 843–854, Dec. 1979.
- [17] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 19, no. 1, pp. 5–18, Jan. 2004.
- [18] R. O. Chavez-Garcia, J. Burlet, T. Vu, and O. Aycard, "Frontal object perception using radar and mono-vision," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 159–164.
- [19] Z. Zhang, K. Fu, X. Sun, and W. Ren, "Multiple target tracking based on multiple hypotheses tracking and modified ensemble Kalman filter in multi-sensor fusion," *Sensors*, vol. 19, no. 14, 2019, Art. no. 3118.
- [20] R. P. S. Mahler, "Multitarget bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.
- [21] B. Vo and W. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.
- [22] H. Sidenbladh, "Multi-target particle filtering for the probability hypothesis density," in *Proc. 6th Int. Conf. Inf. Fusion*, 2003, vol. 2, pp. 800–806.
- [23] O. Törő, T. Bécsi, and P. Gáspár, "PHD filter for object tracking in road traffic applications considering varying detectability," *Sensors*, vol. 21, no. 2, 2021, Art. no. 472.
- [24] X. Chen, Y. Li, Y. Li, and J. Yu, "PHD and CPHD algorithms based on a novel detection probability applied in an active sonar tracking system," *Appl. Sci.*, vol. 8, no. 1, 2018, Art. no. 36.
- [25] M. Yazdian-Dehkordi, O. Rojhani, and Z. Azimifar, "Visual target tracking in occlusion condition: A GM-PHD-based approach," in *Proc. 16th CSI Int. Symp. Artif. Intell. Signal Process.*, 2012, pp. 538–541.
- [26] C. Li, W. Wang, T. Kirubarajan, J. Sun, and P. Lei, "PHD and CPHD filtering with unknown detection probability," *IEEE Trans. Signal Process.*, vol. 66, no. 14, pp. 3784–3798, Jul. 2018.
- [27] S. Wang, Q. Bao, and Z. Chen, "Refined phd filter for multi-target tracking under low detection probability," *Sensors*, vol. 19, no. 13, 2019, Art. no. 2842.
- [28] J. Wu and S. Hu, "PHD filter for multi-target visual tracking with trajectory recognition," in *Proc. 13th Int. Conf. Inf. Fusion*, 2010, pp. 1–6.
- [29] L. Gao, H. Liu, and H. Liu, "Probability hypothesis density filter with imperfect detection probability for multi-target tracking," *Optik*, vol. 127, no. 22, pp. 10428–10436, 2016.
- [30] M. Yazdian-Dehkordi, Z. Azimifar, and M. A. Masnadi-Shirazi, "Penalized Gaussian mixture probability hypothesis density filter for multiple target tracking," *Signal Process.*, vol. 92, no. 5, pp. 1230–1242, 2012.
- [31] M. Yazdian-Dehkordi and Z. Azimifar, "An improvement on GM-PHD filter for target tracking in presence of subsequent miss-detection," in *Proc. 23rd Iranian Conf. Elect. Eng.*, 2015, pp. 765–769.
- [32] N. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, Aug. 1992.
- [33] S. Challa, B.-N. Vo, and X. Wang, "Bayesian approaches to track existence-IPDA and random sets," in *Proc. 5th Int. Conf. Inf. Fusion*, 2002, vol. 2, pp. 1228–1235.
- [34] R. Chakravorty and S. Challa, "Multitarget tracking algorithm - joint IPDA and Gaussian mixture PHD filter," in *Proc. 12th Int. Conf. Inf. Fusion*, 2009, pp. 316–323.
- [35] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA, USA: Artech House, 2007.

- [36] R. L. Streit, *The Poisson Point Process*. Boston, MA: Springer, 2010, pp. 11–55.
- [37] M. S. Grewal and A. P. Andrews, *Kalman Filtering: Theory and Practice With MATLAB*, 4th ed. Hoboken, NJ, USA: Wiley-IEEE Press, 2014.
- [38] R. Nabati and H. Qi, “Radar-camera sensor fusion for joint object detection and distance estimation in autonomous vehicles,” in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Oct. 25th, 2020, Las Vegas, USA.
- [39] F. Nobis, M. Geisslinger, M. Weber, J. Betz, and M. Lienkamp, “A deep learning-based radar and camera sensor fusion architecture for object detection,” in *Proc. Sensor Data Fusion: Trends, Solutions, Appl.*, 2019, pp. 1–7.
- [40] Y. Bar-Shalom and X.-R. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. vol. 19. Storrs, CT, USA: YBs, 1995.
- [41] Y. Bar-Shalom, “Update with out-of-sequence measurements in tracking: Exact solution,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 3, pp. 769–777, Jul. 2002.
- [42] N. Kaempchen and K. Dietmayer, “Data synchronization strategies for multi-sensor fusion,” in *Proc. IEEE Conf. Intell. Transp. Syst.*, 2003, vol. 85, no. 1, pp. 1–9.
- [43] A. Houenou, P. Bonnifait, V. Cherfaoui, and J. Boissou, “A track-to-track association method for automotive perception systems,” in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 704–710.
- [44] B. Ristic, D. Clark, and B.-N. Vo, “Improved SMC implementation of the PHD filter,” in *Proc. 13th Int. Conf. Inf. Fusion*, 2010, pp. 1–8.
- [45] C. Otto, “Fusion of data from heterogeneous sensors with distributed fields of view and situation evaluation for advanced driver assistance systems,” Ph.D. dissertation, Karlsruhe Inst., Karlsruhe, Germany, 2013.
- [46] L. Lindenmaier and V. Tihanyi, “Comparison of different tracking approaches on pre-fused data for automotive perception system,” in *Proc. IEEE 19th Int. Symp. Comput. Intell. Informat. 7th IEEE Int. Conf. Recent Achievements Mechatronics, Automat., Comput. Sci. Robot.*, 2019, pp. 000199–000204.
- [47] M. Liu, T. Huai, R. Zheng, and S. Zhang, “Multisensor multi-target tracking based on GM-PHD using out-of-sequence measurements,” *Sensors*, vol. 19, no. 19, 2019, Art. no. 4315.
- [48] X. Luo, C. Yang, R. Chen, and Z. Shi, “Improved SMC-PHD filter for multi-target track-before-detect,” in *Proc. IEEE 84th Veh. Technol. Conf.*, 2016, pp. 1–6.
- [49] M.-E. Choi and S.-W. Seo, “Robust multitarget tracking scheme based on Gaussian mixture probability hypothesis density filter,” *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4217–4229, Jun. 2016.
- [50] D. Wettschereck and T. G. Dietterich, “Locally adaptive nearest neighbor algorithms,” in *Proc. Adv. Neural Inf. Process. Syst.*, 1994, pp. 184–184.
- [51] A. S. Rahmathullah, Á. F. García-Fernández, and L. Svensson, “Generalized optimal sub-pattern assignment metric,” in *Proc. 20th Int. Conf. Inf. Fusion*, 2017, pp. 1–8.
- [52] L. Kovacs, L. Lindenmaier, H. Nemeth, V. Tihanyi, and A. Zarandy, “Performance evaluation of a track to track sensor fusion algorithm,” in *Proc. 16th Int. Workshop Cellular Nanoscale Netw. Their Appl.*, 2018, pp. 1–2.
- [53] O. Törő, T. Bécsi, S. Aradi, M. Kolat, and P. Gáspár, “Maneuver classification for road vehicles with constrained filtering techniques,” *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 15495–15500, 2020.
- [54] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with Markovian switching coefficients,” *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, Aug. 1988.



sensor data fusion.

**László Lindenmaier** received the B.Sc. and M.Sc. degrees in 2017 and 2020 from the Budapest University of Technology and Economics, Budapest, Hungary, where he is currently working toward the Ph.D. degree with the Department of Control for Transportation and Vehicle Systems. He was with Automotive Industry as a Development Engineer between 2017 and 2020 in field of environment perception and sensor data fusion. His research interests include vehicle control, automotive environment perception, object detection, multi-object tracking, and



control. His research interests include embedded systems, communication networks, vehicle mechatronics, and reinforcement learning. His research and industrial works include railway information systems, vehicle on-board networks, and vehicle control.



information systems and vehicle control.

**Tamás Bécsi** (Member, IEEE) received the M.Sc. and Ph.D. degrees from the Budapest University of Technology and Economics, Budapest, Hungary, in 2002 and 2008, respectively. Since 2005, he has been an Assistant Lecturer and since 2014, he has also been an Associate Professor with the Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics. His research interests include linear systems, embedded systems, traffic modeling, and simulation. His research and industrial works include railway information systems and vehicle control.



**Olivér Törő** received the M.Sc. degree from Eötvös Loránd University, Budapest, Hungary, in 2010. He is currently working toward the Ph.D. degree with the Budapest University of Technology and Economics, Budapest, Hungary. Since 2018, he has been an Assistant Research Fellow with the Department of Control for Transportation and Vehicle Systems, Budapest University of Technology and Economics. His research interests include object detection and tracking in road traffic applications, multi-object state estimation, and nonlinear filtering.



**Péter Gáspár** (Member, IEEE) received the M.Sc. and Ph.D. degrees from the Faculty of Transportation Engineering and Vehicle Engineering Közlekedésmérnöki és Járműmérnöki Kar (KJK), Budapest University of Technology and Economics Budapesti Műszaki és Gazdaságtudományi Egyetem (BME), Budapest, Hungary, in 1985 and 1997, respectively, and the D.Sc. degree in control from the Hungarian Academy of Sciences (MTA), Budapest, Hungary, in 2007. Since 1990, he has been a Senior Research Fellow with the Institute for Computer Science and Control Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI). Since 2016, he has also been a Research Professor. In 2004, he became the Head of the Vehicle Dynamics and Control Research Group and then in 2017, he became the Head of the Systems and Control Laboratory, SZTAKI. He was habilitated with the BME, in 2008, and he was appointed as the University Professor. Since 2013, he has also been the Head with the Department of Control for Transportation and Vehicle Systems Közlekedés- és Járműirányítási Tanszék (KIIT), BME KJK. His research interests include linear and nonlinear systems, robust control, multi-objective control, system identification, and identification for control and artificial methods. His research and industrial works include mechanical systems, vehicle structures, and vehicle dynamics and control. Since 2016, he has also been a Corresponding Member of MTA. He is also a Member of the IFAC Automotive Control and Transportation Systems Technical Committee, and the Chair of the International Federation of Automatic Control (IFAC) Hungary National Member Organization.