This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

Grid-Based Object Tracking with Nonlinear Dynamic State and Shape Estimation

Sascha Steyer, Christian Lenk, Dominik Kellner, Georg Tanzmeister, and Dirk Wollherr

***Abstract*—Object tracking is crucial for planning safe maneuvers of mobile robots in dynamic environments, in particular for autonomous driving with surrounding traffic participants. Multi-stage processing of sensor measurement data is thereby required to obtain abstracted high-level objects, such as vehicles. This also includes sensor fusion, data association, and temporal filtering. Often, an early-stage object abstraction is performed, which, however, is critical, as it results in information loss regarding the subsequent processing steps. We present a new grid-based object tracking approach that, in contrast, is based on already fused measurement data. The input is thereby pre-processed, without abstracting objects, by the spatial grid cell discretization of a dynamic occupancy grid, which enables a generic multi-sensor detection of moving objects. On the basis of already associated occupied cells, presented in our previous work, this paper investigates the subsequent object state estimation. The object pose and shape estimation thereby benefit from the freespace information contained in the input grid, which is evaluated to determine the current visibility of extracted object parts. An integrated object classification concept further enhances the assumed object size. For a precise dynamic motion state estimation, radar Doppler velocity measurements are integrated into the input data and processed directly on the object-level. Our approach is evaluated with real sensor data in the context of autonomous driving in challenging urban scenarios.**

***Index Terms*—Autonomous vehicles, dynamic occupancy grids, environment perception, object detection, object tracking, radar Doppler measurements, shape estimation, state estimation.**

I. Introduction

OBJECT tracking is an essential task of environment perception with the aim of detecting and temporally filtering surrounding objects based on sensor measurements. Mobile robots, especially autonomous vehicles, require a robust object estimation to plan interactive maneuvers and avoid collisions with other traffic participants or obstacles. Therefore, measurement data of multiple sensors have to be processed in different ways. This includes several tasks, such as object extraction by detecting or abstracting features, sensor data fusion, data association of measurements and predicted objects, shape estimation of the spatial extent of an object, and state estimation with temporal filtering in general.

Objects are often detected and tracked for each individual sensor type by sensor-specific features. Data fusion is then realized afterwards based on those filtered high-level objects of each sensor, i.e., a high-level fusion, e.g. [1]–[3]. However, this requires an early-stage abstraction of measurement data to generalized objects by specific object assumptions, resulting in information loss and thus a more error-prone data fusion. With greater computing resources and higher sensor resolutions now available, low-level fusion approaches are increasingly used, as they reduce that information loss by fusing measurement data before object assumptions are made.

We present a new grid-based object tracking approach, where the input is based on pre-processed measurement data in the form of a dynamic occupancy grid. The basic concept of this uniform input representation is to use the spatial grid cell discretization for a generic cell-wise sensor data fusion and dynamic estimation, all processed without requiring specific object assumptions. The low-level dynamic estimation thereby results in a static/dynamic occupancy classification including velocity estimates of each grid cell. This enhances and also simplifies the detection and tracking of moving objects, since only measurement data classified as dynamic have to be considered. Parts of our overall grid-based approach have been presented in our previous work on the dynamic grid estimation [4], the extraction of new objects by clustering dynamic cells [5], and the association of dynamic cells with predicted objects [6]. This paper follows on from those concepts and focuses on the subsequent object state estimation. The contribution is thereby divided into three parts.

First, we show how the dynamic state of an object, estimated by an unscented Kalman filter (UKF), is generally updated by the associated dynamic occupied cells of the current measurement. A common box model is used for the object form and measurement abstraction, but in contrast to approaches that directly process the raw sensor data, we also evaluate the freespace information that is contained in the grid representation. Thereby, a generic orientation estimation based on minimizing the included freespace is proposed, and the position is updated by selecting the most visible reference point regarding the surrounding freespace of the box edges.

Second, the UKF-based dynamic state estimation is additionally improved by processing radar Doppler velocity measurements. A generic radar velocity-based UKF motion estimation is proposed that directly uses the radar velocity measurement space for updating the object state by projecting the UKF sigma points to the expected radial velocities, which is also applicable to other measured velocity components. The paper thus also discusses how radar measurements are represented and associated within our grid-based framework.

Third, the object shape estimation is addressed, i.e., the size of the selected object box model, which is estimated

Manuscript received February 06, 2019; revised May 20, 2019; accepted May 31, 2019. *(Corresponding author: Sascha Steyer.)*

S. Steyer, C. Lenk, D. Kellner, and G. Tanzmeister are with the BMW Group, 80788 Munich, Germany (e-mail: {sascha.steyer, christian.ch.lenk, dominik.m.kellner, georg.tanzmeister}@bmw.de).

D. Wollherr is with the Institute of Automatic Control Engineering, Technical University of Munich, 80333 Munich, Germany (e-mail: dw@tum.de).

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

differently with an assumed static state. A new combination of a histogram filter geometry distribution estimation and an object classification concept is proposed. This enables us to model non-Gaussian distributions of the length and width by distinguishing lower and upper bounds of the measurement, also evaluated by the freespace, and prior class knowledge of the assumed length or width if either has not been fully observed yet.

Overall, this grid-based object tracking has several advantages compared to a high-level fusion with a sensor-individual object tracking on the raw measurement data, in particular:

- *Static/dynamic occupancy classification*, simplifying the object detection and data association problem, since only measurement data classified as dynamic have to be considered for estimating moving objects.
- *Low-level dynamic estimation*, resulting in a filtered cell velocity estimation and a track-before-detect concept, enabling the detection of arbitrarily shaped moving objects, an object velocity and orientation initialization, and an improved data association.
- *Freespace information*, derived by sensor measurement models, improving the estimation of the object position by determining the most visible reference point, the object shape by distinguishing lower and upper bounds, and the object orientation by minimizing inside freespace.
- *Uniform occupancy grid representation*, enabling a generic multi-sensor object tracking with an implicit low-level sensor data fusion by the grid cell discretization, but also extendable with separate layers for sensor-specific measurements such as radar Doppler velocities.

In combination with the dynamic grid map, this approach results in a consistent multi-sensor estimation of moving objects, static obstacles, and freespace. In the context of autonomous driving, this eventually improves the scene understanding of the current surroundings and thus the safety and comfort.

This article is organized as follows. Section II discusses related work, while the specific grid-based estimation of our previous work that forms the input of this work is summarized in Section III. The dynamic state estimation of the object tracking is presented in Section IV, which is extended in Section V by the additional radar Doppler measurement integration. The histogram filter-based estimation of the spatial extent, combined with the object classification, is discussed in Section VI. Our approach is finally evaluated in Section VII with real sensor data in various challenging scenarios; additional experimental results are demonstrated in the attached video.

II. Related Work

Object tracking is a broad field of research with various approaches; a general overview is presented in [7]–[9], for example. This section focuses on similar grid-based approaches, as they enable a generic multi-sensor object tracking, with some additional relevant work on radar-based motion estimation.

Occupancy grid maps are typically accumulated in a fixed coordinate system given the odometry of the egomotion. Sensor data can thus be classified into static (stationary obstacles) and dynamic (moving objects) by evaluating the occurring measurement position in the accumulated grid map, while dynamic parts are characterized by inconsistencies of previously derived freespace and currently measured occupancy or vice versa [10]–[12]. By clustering measurements classified as dynamic, moving objects are detected in [12], [13], which are then filtered using multiple hypotheses tracking (MHT). Object hypotheses are extracted similarly in [14], [15] with a global nearest neighbor (GNN) association and a Kalman filter state estimation. In [16], clusters of dynamic grid cells are directly tracked as a free-form object model with a particle filter and a joint probability data association (JPDA). However, a moving object detection directly based on such occupied/free inconsistencies is error-prone to measurement inaccuracies of ranging or odometry sensors.

A more robust strategy requires a recursive dynamic state estimation of the occupancy grid. First approaches have estimated discrete velocities of each grid cell by neighborhood cell transition histograms [17], [18], whereas recent approaches efficiently estimate continuous velocity distributions of each grid cell using a grid-based particle filter as originally proposed in [19] and further extended in [20]–[23]. This robust particle-based dynamic estimation concept also forms the basis of our work as presented in [4], with a more detailed discussion on such dynamic occupancy grids. Overall, this temporally filtered estimation results in an accumulated occupancy grid map including accurate estimates of the grid cell velocities as well as a static/dynamic occupancy classification. Those cell velocity estimates thereby enable a more sensitive cell clustering of neighboring dynamic occupied cells by considering their velocity difference, which improves the object detection with dense traffic and also reduces false positives, as proposed in our previous work [5].

In [5], we also used the mean of the cell velocity estimates of all associated cells for a direct measurement update of the object velocity and orientation with a subsequent UKF-based object tracking. However, this results in correlated input data of the UKF and thus in a multi-filtering of the velocity and orientation, which can cause higher filtering latencies in nonlinear scenarios. In [24], a de-autocorrelation scheme is proposed to whiten the correlated velocity input of a Kalman filter object tracking based on such dynamic occupancy grids, which, however, uses a simplified linear approximation that is prone to deviations of that model, in particular in scenarios with critical nonlinear movements.

Other recent approaches [25]–[28] use deep learning techniques for the object detection based on dynamic occupancy grids. But, in contrast to the generic cell clustering extraction of arbitrarily shaped moving objects, those machine learning-based approaches require large labeled training datasets. Furthermore, those approaches describe only a single-frame object detection without considering the temporal filtering. This is extended in [29] by a tracking with a recurrent neural network (RNN), which, however, requires a stationary ego vehicle in that study and also does not investigate the object state estimation of the velocity, acceleration, or turn rate. Similarly, several recent vision-based deep learning approaches, e.g. [30]–[34], achieve promising results in the object detection and tracking, especially in the object classification as well, but usually

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

only in the image space, without a detailed dynamic state or shape estimation of the object tracks in the Cartesian world or odometry space. That, however, is necessary for a robust multi-sensor environment representation and applications such as maneuver planning of autonomously moving mobile robots. In addition, those approaches also require large training datasets with a wide variety of possibly occurring objects.

A different approach is proposed in [35] with virtual rays of a lidar sensor that are similar to a polar grid representation. The virtual rays are used to compute the measurement likelihood with a pre-defined cost function of a box model object shape, which also considers the expected freespace around an object. The object state is then estimated by a Rao-Blackwellized particle filter (RBPF) with a sampling of the motion state and Gaussian estimates of the geometry. Hence, that approach introduces using the freespace information for the object state estimation, but the polar grid structure of the virtual rays is not directly extendable to a generic multi-sensor approach and a static/dynamic classification is not used.

The object shape can also be estimated using an object local grid map [36], [37], enabling a detailed free-form shape estimation by accumulating occupancy probabilities with an individual grid for each object, also considering the freespace information of the grid. However, for most applications, an abstracted box model is sufficient and more robust, since such an object local grid map requires complex and error-prone re-alignment with the movement of the corresponding object. Furthermore, the memory-intensive representation is unfavorable for large trucks or in heavy traffic urban scenarios.

An accurate motion estimation requires Doppler radars that enable directly measuring velocity components. A multi-radar object tracking is proposed in [38], [39] by deriving the velocity profile that describes the varying Doppler velocities of an extended object over the azimuth angle. Depending on the number of sensors and measurements, a least-squares regression is used to resolve up to three degrees of freedom of the object state, i.e., the velocity, orientation, and turn rate. Similarly, in [40], the velocity profile is used for a single-radar tracking, with a separate orientation estimation based on the contour of all radar detections in the case of a nonlinear motion. In [41], a Gaussian process is used for the radar-based shape estimation, combined with an extended Kalman filter (EKF) for the motion estimation, which, however, results in complex partial derivatives that are prone to linearization errors. But since radars typically have a lower spatial accuracy than lidar sensors, all of those radar-only approaches have difficulties in estimating the object pose and size, which in turn affects the motion estimation due to the inaccurate center of rotation.

Overall, there are several promising object tracking approaches that already partly use benefits of the static/dynamic occupancy classification, the low-level dynamic grid estimation, the freespace information contained in the grid, or radar Doppler measurements. However, none of these approaches fully combines the different benefits, which is addressed in the following, with the aim of achieving a generic and robust multi-sensor object tracking for challenging urban scenarios.

III. Grid-Based Environment Estimation

This section gives a brief overview of our overall grid-based environment estimation approach [4]–[6] in order to understand the input of this work and the pre-processing steps of it. These processing steps are exemplarily illustrated in Fig. 1.

A. Evidential Grid Representation

The environment is modeled in an evidential occupancy grid representation that is based on the Dempster-Shafer theory of evidence (DST) [42], [43]. The DST framework enables modeling separate hypotheses rather than a single Bayesian occupancy probability. We use the frame of discernment

$$\Theta = \{F, S, D\} \tag{1}$$

with the individual hypotheses freespace (F), static occupancy (S), and dynamic occupancy (D), as initially proposed in [20]. In this framework, not only are these individual hypotheses taken into account, but also all possible combinations of them, i.e., the power set 2^Θ of all subsets of Θ. This allows us in particular to model unclassified occupancy $O = \{S, D\}$, without further specifying whether it is static or dynamic as used to represent an occupancy measurement. Moreover, passable area $\{F, D\}$ that may be currently free or dynamically occupied, primarily used in the temporal mapping [4], and the remaining unknown state Θ that models all possible hypotheses are thus also included in that representation.

In the DST framework, each hypothesis of the power set 2^Θ is estimated by an evidential basic belief mass $m(\cdot) \in [0, 1]$, while the sum of all basic belief masses equals 1. The belief

$$\mathrm{bel}(\theta) = \sum_{\tilde{\theta} \subseteq \theta} m(\tilde{\theta}) \;\leq\; p(\theta) \tag{2}$$

describes the sum of all subsets $\tilde{\theta} \subseteq \theta$ of a hypothesis $\theta \subseteq \Theta$, which is a lower bound of the Bayesian probability $p(\theta)$. Overall, this evidential grid structure represents a generic, sensor-independent environment model.

B. Measurement Grid Fusion

The measurements of each sensor, measured at time t, are abstracted in such a uniform evidential occupancy grid representation. Each individual cell $c \in \mathcal{G}$ of the 2-D grid structure $\mathcal{G}$ thereby contains a cell measurement z_t^c with evidential beliefs of unclassified occupancy $\mathrm{bel}(O) = \mathrm{bel}(SD)$ and freespace $\mathrm{bel}(F) = m(F)$, as well as the remaining unknown mass $m(\Theta)$ that is implicitly included. The measurement grid is thus described by

$$Z_t = \{z_t^c \mid c \in \mathcal{G}\}, \quad z_t^c = \left[\mathrm{bel}(O_{z,t}^c),\, \mathrm{bel}(F_{z,t}^c)\right]^\mathsf{T}. \tag{3}$$

All sensor-individual measurement grids are fused to one fused measurement grid by combining the corresponding occupancy and freespace beliefs for each grid cell with Dempster's rule of combination [43]. Thereby, the individual cells of the different grids exactly overlap with regard to the spatial cell discretization by avoiding rotations and performing only integer translations of the cell resolution using the accumulated

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

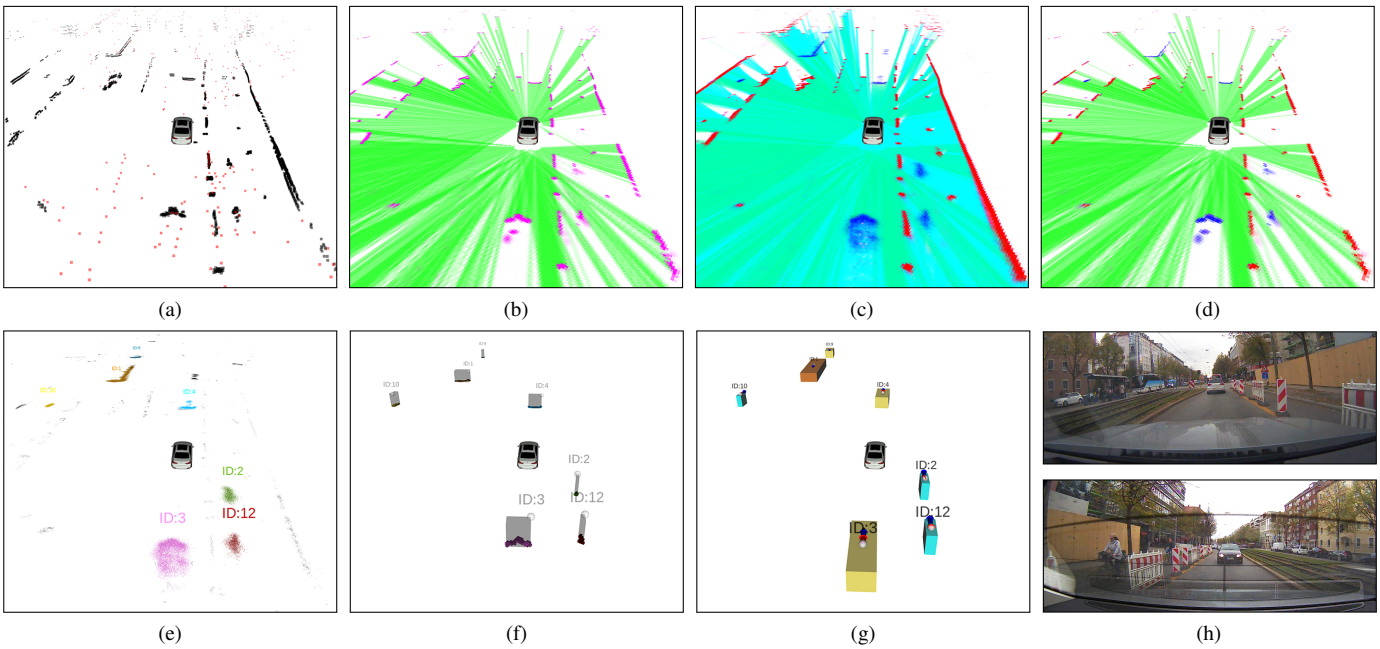


Fig. 1. Overview of the different processing steps of the grid-based environment estimation and the subsequent object tracking. (a) Raw measurement data (black squares: lidar, red squares: radar). (b) Fused 2-D measurement grid. (c) Filtered dynamic grid map. (d) Classified measurement grid. (e) Labeled particle population. (f) Associated occupied cells and measurement bounding boxes. (g) Filtered object tracks. (h) Camera images (front and rear) of the scene.

odometry as the uniform reference frame, whereas the pose of the ego vehicle inside the grid varies over time [4], [44], [45].

In this work, different lidar and radar sensors are used, which are described in more detail in Section VII. To take the different latencies and asynchronous measurement time instances of the individual sensors into account, all measurement data are buffered and then fused using an interval alignment based on a trigger sensor as proposed in [46].

The radar sensors also include a velocity measurement

$$z_{v,t}^{c} = \left[v_z^r,\, \theta_z^r, x_s\right]^{\mathsf{T}} \tag{4}$$

of the corresponding occupied cells, which contains the measured scalar radial Doppler velocity v_z^r, the azimuth angle θ_z^r, and the position x_s of the current sensor origin. In the case of multiple Doppler measurements in a cell, the most certain one with regard to the highest corresponding occupancy belief is selected in the measurement grid fusion.

C. Dynamic Grid Mapping and Low-Level Particle Tracking

The measurement data of different sensors, all measured at time t, have been abstracted to the evidential grid representation and fused cell-wise to one measurement grid Z_t. The fused measurement grids of different time instances are temporally accumulated in a dynamic grid map

$$\mathcal{M}_t = \left[m(S_t),\, m(D_t),\, m(SD_t),\, m(F_t),\, m(FD_t)\right]^{\mathsf{T}} \tag{5}$$

by an adapted evidential filtering combined with a low-level particle tracking as discussed in detail in [4]. Hence, this temporally filtered estimation enables us to distinguish static and dynamic occupancy and thus subdivide the unclassified occupancy belief of the current measurement

$$\mathrm{bel}(O_{z,t}^{c}) = m(S_{\tilde{z},t}^{c}) + m(D_{\tilde{z},t}^{c}) + m(SD_{\tilde{z},t}^{c}) \tag{6}$$

into individual basic belief masses for $\{S\}$, $\{D\}$, and $\{S, D\}$, i.e., resulting in a pseudo-measurement $\tilde{z}$ with a static/dynamic occupancy classification. Therefore, the object tracking of this work is mostly reduced and thus simplified to the set

$$\mathcal{G}_{D,t} = \{c \in \mathcal{G} \mid m(D_{\tilde{z},t}^{c}) \geq \Gamma_D\} \tag{7}$$

of currently occupied cells with a dynamic occupancy mass greater or equal than a defined threshold $\Gamma_D > 0$.

Dynamic occupancy of the filtered map $\mathcal{M}_t$ is initialized and predicted using a grid-based particle filter as initially presented in [19] and adapted in other recent approaches [20]–[23]. Each particle $\chi \in \mathcal{X}_t$ of the population $\mathcal{X}_t$ represents a hypothesis of dynamic occupancy at a particular position $x_\chi \in \mathbb{R}^2$ with velocity $\nu_\chi \in \mathbb{R}^2$ and an occupancy value o_χ. Hence, in addition to the dynamic occupancy prediction, the filtered particle population also estimates velocity distributions. The particle velocity initialization and weighting are thereby enhanced by the radar velocity measurements $z_{v,t}^{c}$ if available as proposed in [20], [47]. By evaluating all particles $\chi \in \mathcal{X}_t^c \subseteq \mathcal{X}_t$ in a cell $c \in \mathcal{G}$ with regard to the current particle positions x_χ, the 2-D weighted mean cell velocity results in

$$\nu_t^c = \left(\sum_{\chi \in \mathcal{X}_t^c} o_\chi\right)^{-1} \cdot \sum_{\chi \in \mathcal{X}_t^c} o_\chi \nu_\chi \,. \tag{8}$$

Overall, this particle-based dynamic estimation, as part of the dynamic grid mapping, enables a robust estimation of cell velocities and thus the detection of moving parts of the environment without requiring specific object assumptions.

D. Extraction of New Objects

In order to detect new occurring objects in the current measurement grid, the static/dynamic occupancy classification

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

and the particle-based cell velocity estimates ν_t^c are evaluated. As presented in more detail in [5], a combination of a density-based and connectivity-based clustering is used to extract new objects. The basic idea is that clusters of dynamic cells with similar velocities are compared with the local neighborhood regarding the velocity variance of the cells, which is less error-prone to areas wrongly classified as dynamic and thus reduces false positives.

This object extraction based on clusters of dynamic cells enables a robust detection of arbitrarily shaped moving objects without requiring specific features such as L-shapes or large training datasets for machine learning algorithms, e.g. [25]–[28]. Furthermore, the particle tracking thereby serves as a track-before-detect concept, i.e., movements are tracked by low-level particle velocity hypotheses before high-level objects are detected. Moreover, only new objects are extracted this way, whereas existing object tracks that have been extracted before are directly associated with occupied cells, which is described in the following. This means that the object extraction is not applied in areas of predicted tracks, i.e., occupied cells are associated to those existing tracks before possible new objects are analyzed in the remaining set of unassociated cells.

E. Particle Labeling Association

The object tracking is updated by a set of associated occupied cells of the current measurement grid. The individual cells are thereby directly associated to predicted object tracks, i.e., information about the existing tracks is considered before those cells are abstracted by a clustering. We use our particle labeling association approach as presented in [6]. The basic idea is to link particles $\chi \in \mathcal{X}_t$ of the underlying low-level particle tracking with the high-level objects by attaching an object label to each particle. Dynamic occupied cells are thus associated to objects by evaluating the particle label distribution in each cell. This association concept further contains a subsequent clustering, in which multiple clusters of an object are extracted and finally checked for plausibility to further increase the robustness of the association.

Overall, both the extraction of new objects and the association with existing ones result in a set of dynamic cells

$$\mathcal{C}_{\tau,t} = \{c \in \mathcal{G}_{D,t} \mid f_a(c) = \tau\} \tag{9}$$

that are associated to the corresponding track $\tau \in \mathcal{T}_t$, with $\mathcal{T}_t$ describing the set of all currently tracked objects including the newly extracted ones. This mapping of dynamic occupied cells to the object tracks given the cell measurements and predicted object track states is thereby formally described by the surjective function

$$f_a : \mathcal{G}_{D,t} \to \mathcal{T}_t \cup \{\zeta_\emptyset\} \;:\; c \mapsto \tau\,, \tag{10}$$

while cells that are not associated to a track are mapped to an auxiliary variable $\zeta_\emptyset$. Up to this point, the measurement data, including the different processing steps of the dynamic estimation, the association, and the extraction of new objects, are still retained in the low-level grid representation. Based on that pre-processing, the focus of this paper is the object state estimation, i.e., deriving abstracted high-level information of the individual objects, including the temporal state filtering.

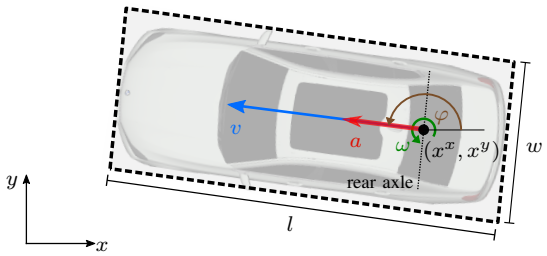


Fig. 2. Illustration of the object state representation used in this work.

IV. Dynamic State Estimation of Object Tracks

The grid-based environment estimation, as described in the previous section, results in a robust pre-processing of measurement data in a uniform evidential grid representation. This also includes the extraction of new object tracks and the association of measurement data with existing object tracks, both still retaining the grid cell representation. As discussed before, those steps of our grid-based object tracking approach already benefit directly from the dynamic grid estimation. This section focuses on the high-level object state estimation based on already associated occupied cells of the current measurement grid, which further benefits from the evidential grid representation of the input and the low-level particle tracking.

A. Dynamic State Representation

The dynamic state of each object track $\tau \in \mathcal{T}_t$ of all currently tracked objects $\mathcal{T}_t$ at time t is defined by

$$s_{\tau,t} = \left[x_t^x,\, x_t^y,\, v_t,\, a_t,\, \varphi_t,\, \omega_t\right]^\mathsf{T} . \tag{11}$$

This vector is composed of the 2-D position $x = \left[x^x,\, x^y\right]^\mathsf{T}$, the longitudinal velocity v with the corresponding acceleration $a = \dot{v}$, and the orientation φ with the turn rate $\omega = \dot{\varphi}$. As with the grid representation, the accumulated odometry is used as the uniform reference frame of the position of all objects. The position x of an object is fixed at the center of its rear axle as the assumed center of rotation in normal driving conditions. It is approximated between the center point and the middle rear point of the selected box representation for all object types. Accordingly, the velocity v, the acceleration a, and the turn rate ω are also referred to that position x. The selected object state representation is illustrated in Fig. 2.

This dynamic state s_τ is filtered by an unscented Kalman filter (UKF) [48], which is discussed in the following. The object shape, i.e., the length l and width w of the selected bounding box representation, is assumed to be static and estimated differently by a histogram filter combined with an object classification concept that is presented separately in Section VI.

Regarding the notation, since only a single track τ is considered in the following state estimation, variables of the state $s_{\tau,t}$ as well as the corresponding current measurement are denoted without reference to the track τ for better readability.

B. Prediction

The recursive UKF filtering requires a prediction to the time t of the current measurement, which in turn requires a motion model of the mean state $s_{\tau,t}$ and a process noise

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

model of the corresponding covariance $\Sigma_{\tau,t}$. The models used for this procedure are briefly described in the following for the sake of completeness.

1) Motion Model: The prediction of the state $s_{\tau,t-1}$ is denoted by $\hat{s}_{\tau,t}$, i.e., predicted variables are characterized by a hat symbol. The state transition is based on a nonlinear constant turn rate and acceleration (CTRA) motion model [49] that enables clothoid trajectories. To increase the overall robustness, we modify the CTRA motion model slightly

$$\hat{\omega}_t = (1-\varepsilon_\omega)\,\omega_{t-1} \tag{12}$$

$$\hat{a}_t = \underset{a_1,a_2}{\arg\min}\left(|a_1|,\,|a_2|\right),\; a_1 = (1-\varepsilon_a)a_{t-1},\; a_2 = \tfrac{-v_{t-1}}{t_{\text{horizon}}} \tag{13}$$

by ensuring slow convergence of the turn rate ω and acceleration a toward zero, which is modeled by the two reduction factors $\varepsilon_\omega, \varepsilon_a \in (0,1)$. The acceleration is additionally limited by the term $-v_{t-1}/t_{\text{horizon}}$ to prevent a sign change of the velocity without the object stopping or slowing down. For example, a fast-moving object with $v_{t-1} \gg 0$ that performs a full braking with $a \ll 0$ would otherwise, due to a too slowly decreasing absolute acceleration, eventually overshoot and result in a predicted negative velocity. The parameter $t_{\text{horizon}} \geq \Delta t$ defines the time horizon and hence the smoothness of the limitation. The remaining state is updated by the default CTRA motion model [49]

$$\hat{\varphi}_t = \varphi_{t-1} + \omega\Delta t \tag{14}$$

$$\hat{v}_t = v_{t-1} + a\Delta t \tag{15}$$

$$\hat{x}_t^x = x_{t-1}^x + \tfrac{1}{\omega^2}\big(\omega\hat{v}_t\sin(\hat{\varphi}_t) + a\cos(\hat{\varphi}_t) - \omega v_{t-1}\sin(\varphi_{t-1}) - a\cos(\varphi_{t-1})\big) \tag{16}$$

$$\hat{x}_t^y = x_{t-1}^y + \tfrac{1}{\omega^2}\big(-\omega\hat{v}_t\cos(\hat{\varphi}_t) + a\sin(\hat{\varphi}_t) + \omega v_{t-1}\cos(\varphi_{t-1}) - a\sin(\varphi_{t-1})\big) \tag{17}$$

using $\omega = \hat{\omega}_t$ and $a = \hat{a}_t$. For $\omega \approx 0$, the position update in (16)-(17) is replaced by a constant acceleration motion model to avoid singularities.

2) Process Noise Model: The process noise model used is based on a Wiener-sequence acceleration model [50], [51]

$$Q_w = \sigma_{w,\ddot{\rho}}^2 \begin{bmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^3}{2} & \Delta t^2 & \Delta t \\ \frac{\Delta t^2}{2} & \Delta t & 1 \end{bmatrix}, \tag{18}$$

where, for a state $[\rho, \dot{\rho}, \ddot{\rho}]^\mathsf{T}$, a change of the acceleration $\ddot{\rho}$ is assumed to be an independent white noise process with variance $\sigma_{w,\ddot{\rho}}^2$. Applying (18) independently to the tangential acceleration a and the radial acceleration $\dot{\omega}$ of the state as defined in (11), the UKF process noise matrix finally results in

$$Q = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix} \tag{19}$$

with

$$Q_1 = \sigma_{w,a}^2 \begin{bmatrix} \frac{\Delta t^4}{4}\cos^2\varphi & \frac{\Delta t^4}{4}\sin\varphi\cos\varphi & \frac{\Delta t^3}{2}\cos\varphi & \frac{\Delta t^2}{2}\cos\varphi \\ \cdot & \frac{\Delta t^4}{4}\sin^2\varphi & \frac{\Delta t^3}{2}\sin\varphi & \frac{\Delta t^2}{2}\sin\varphi \\ \cdot & \cdot & \Delta t^2 & \Delta t \\ \cdot & \cdot & \cdot & 1 \end{bmatrix} \tag{20}$$

$$Q_2 = \sigma_{w,\dot{\omega}}^2 \begin{bmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & \Delta t^2 \end{bmatrix} \tag{21}$$

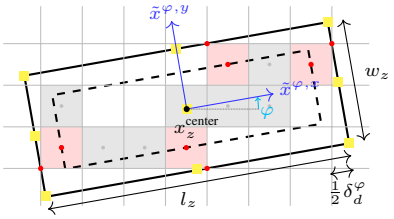


Fig. 3. Transformation of associated grid cells (gray, with crucial cells highlighted in red) to a box model. The inner dotted rectangle fits the cell centers, the solid rectangle also considers the cell extent by adding δ_d^φ. All possible reference points are visualized by yellow squares.

where dots $\cdot$ in (20) represent symmetric entries such that $Q_1 = Q_1^\mathsf{T}$. Overall, these motion and process noise models are used for the prediction of the dynamic state and the corresponding covariance of the UKF, which represents the first step of the recursive state estimation. For the second step, the measurement update, the measurement data of the grid structure have to be abstracted to the representation of the object state, which is addressed in the rest of the section.

C. Transformation of Associated Cells to Box Representation

The set $\mathcal{C}_{\tau,t}$ of occupied cells of the current measurement grid that have been associated to the track τ, see (9), is transformed to an oriented minimum bounding box, i.e., occupied grid cells are abstracted to the same box representation as the filtered object tracks in this processing step. The length axis of an object is thereby defined along the orientation φ and the width axis is defined orthogonal to it. Hence, given the orientation φ, the length and width of the measurement box

$$l_z = \max_{c\in\mathcal{C}_{\tau,t}}(\tilde{x}_c^{\varphi,x}) - \min_{c\in\mathcal{C}_{\tau,t}}(\tilde{x}_c^{\varphi,x}) + \delta_d^\varphi \tag{22}$$

$$w_z = \max_{c\in\mathcal{C}_{\tau,t}}(\tilde{x}_c^{\varphi,y}) - \min_{c\in\mathcal{C}_{\tau,t}}(\tilde{x}_c^{\varphi,y}) + \delta_d^\varphi \tag{23}$$

are computed as the distance of the maximum and minimum of the positions x_c of the associated cells $c \in \mathcal{C}_{\tau,t}$, with

$$\left[\tilde{x}_c^{\varphi,x},\, \tilde{x}_c^{\varphi,y}\right]^\mathsf{T} = R_\varphi \left[x_c^x,\, x_c^y\right]^\mathsf{T} \tag{24}$$

describing the cell position x_c of the odometry reference frame mapped to the object coordinate system $(\tilde{x}^{\varphi,x}, \tilde{x}^{\varphi,y})$ using the rotation matrix R_φ around the angle φ. As x_c represents only the center of a cell, similar to a dilation, an additional extent

$$\delta_d^\varphi = d_c\left(|\sin(\varphi)| + |\cos(\varphi)|\right) \in [d_c, \sqrt{2}\,d_c] \tag{25}$$

is added to the length and width to cover the complete quadratic cell with a size of $d_c \times d_c$ by the extracted object size along the orientation φ. The orientation φ of the bounding box is determined by analyzing the freespace of the measurement grid for various box hypotheses with different orientations, which is described separately in Section IV-F. A geometric illustration of the transformation to the box model is shown in Fig. 3.

D. Position Measurements with Reference Point Selection

The associated occupied grid cells have been abstracted to an oriented measurement box as described above, which

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

enables a position update. The position of the center of the measurement box in the odometry frame is calculated as

$$x_z^{\text{center}} = R_\varphi^\mathsf{T} \begin{bmatrix} \min\limits_{c\in\mathcal{C}_{\tau,t}} (\tilde{x}_c^{\varphi,x}) + \frac{1}{2}(l_z - \delta_d^\varphi) \\ \min\limits_{c\in\mathcal{C}_{\tau,t}} (\tilde{x}_c^{\varphi,y}) + \frac{1}{2}(w_z - \delta_d^\varphi) \end{bmatrix} . \quad (26)$$

The position of the filtered track is referred to the assumed center of rotation, approximated as the position between the center and the middle rear point of the box, see (11). Hence, a direct position update of $\hat{x}_t$ is achieved using the extracted position with $\frac{1}{4}l_z$ rather than $\frac{1}{2}l_z$ in (26). The size of the measurement box, however, may change through occlusion or a changing sensor field of view, as it is only a minimum bounding box. Consequently, the center position x_z^{center} may vary even if the object does not move, with noise-free measurement data as well.

A robust estimate of the position therefore requires using reference points that describe the position in reference to a certain fixed point of an object. Thereby, the position update of the UKF is achieved by transforming the position of the internal state $\hat{s}_\tau$ to the selected reference point of the measurement space, i.e., a predicted measurement is calculated using the UKF sigma points. In this work, the reference point can be either at a corner, an edge center, or the center of the box, see Fig. 3, as also used in [15], [52]. The position in terms of a reference point is thus defined as

$$x_z^{\text{ref}} = x_z^{\text{center}} + R_\varphi^\mathsf{T} \begin{bmatrix} \delta_{z,l}^{\text{ref}}\, l_z \\ \delta_{z,w}^{\text{ref}}\, w_z \end{bmatrix} , \ \delta_{z,l}^{\text{ref}}, \delta_{z,w}^{\text{ref}} \in \left\{0, \tfrac{1}{2}, -\tfrac{1}{2}\right\}, \quad (27)$$

where $\delta_{z,l}^{\text{ref}}$ and $\delta_{z,w}^{\text{ref}}$ have to be selected with respect to the most robust position estimate.

Rather than using the reference point with the shortest distance to a certain sensor origin, we consider the surrounding freespace evidence of the measurement grid to determine the best reference point. Hence, this concept is suitable for a multi-sensor setup and takes all current sensor observabilities, modeled in the evidential grid representation, into account. Each edge $e \in \{\text{front}, \text{rear}, \text{left}, \text{right}\}$ of the measurement box is analyzed with regard to its current visibility $\vartheta_z^e \in [0, 1]$, while $\vartheta_z^e = 1$ means that the edge e is currently a fully visible boundary and thus denotes a true object boundary. Thereto, the ratio

$$r_F(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \sum_{c\in\mathcal{A}} m(F_{z,t}^c) , \quad \mathcal{A} \subseteq \mathcal{G} \quad (28)$$

defines the sum of the current freespace evidence $m(F_{z,t}^c)$ of all cells $c \in \mathcal{A} \subseteq \mathcal{G}$ in an area $\mathcal{A}$ compared to the number of cells $|\mathcal{A}|$ of that area, i.e., its cardinality. Hence, the estimated edge visibility is approximated by

$$\vartheta_z^e \approx r_F(\mathcal{A}_e) , \quad (29)$$

i.e., the surrounding freespace ratio $r_F(\mathcal{A}_e)$ in an area $\mathcal{A}_e$ around the edge e of the measurement box that is expected to be freespace in order to represent a visible boundary. In this work, an outside rectangle area is used, as shown in Fig. 4.

The position along the length axis is finally selected as

$$\delta_{z,l}^{\text{ref}} = \begin{cases} +\frac{1}{2} , & \text{if } \vartheta_z^{\text{front}} \geq \vartheta_{\min} \ \wedge \ \vartheta_z^{\text{rear}} < \vartheta_{\min} \\ -\frac{1}{2} , & \text{if } \vartheta_z^{\text{front}} < \vartheta_{\min} \ \wedge \ \vartheta_z^{\text{rear}} \geq \vartheta_{\min} \\ 0 , & \text{else} \end{cases} , \quad (30)$$

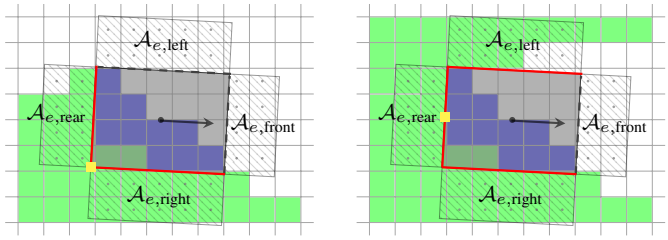


Fig. 4. Analysis of track edge visibility considering the freespace evidence in the surrounding striped rectangle areas $\mathcal{A}_{e,\tau}$. Visible edges are highlighted red, a yellow square denotes the selected reference point. (a) Scenario where the rear and right edge are visible. (b) Scenario with additional freespace at the left edge such that the extracted width represents also an upper bound.

meaning that the position is at an edge if the corresponding visibility is above a defined threshold $\vartheta_{\min} > 0$. If both the front and the rear edge are visible, the center depicts the most robust position along the length axis. The position $\delta_{z,w}^{\text{ref}}$ along the width axis is determined equivalently with the visibilities of the left edge $\vartheta_z^{\text{left}}$ and the right edge $\vartheta_z^{\text{right}}$. In sum, the corresponding corner point is selected if exactly two orthogonal edges are visible, whereas the in-between edge center is selected when multiple corner points are available. This edge visibility ϑ_z^e is also considered for the length and width estimation, which is discussed separately in Section VI. Overall, this reference point selection with the freespace evaluation enables a robust position update of the track.

E. Velocity and Orientation Estimation of Particle Tracking

The object position is updated by abstracting a bounding box of the associated occupied cells. In addition to the occupancy and freespace evidence of the measurement grid, the input data also contain 2-D cell velocities ν_t^c, estimated by the low-level particle tracking with the particle population $\mathcal{X}_t$ as defined in (8). Hence, the 2-D weighted mean velocity vector

$$\overline{\nu} = \left(\sum_{c\in\mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \right)^{-1} \cdot \sum_{c\in\mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c)\, \nu_t^c \quad (31)$$

of all associated cells of the set $\mathcal{C}_{\tau,t}$, weighted by the corresponding dynamic evidence mass $m(D_{\tilde{z},t}^c)$, represents an estimate of the velocity and orientation of the object track in terms of the velocity magnitude and movement direction, i.e.,

$$v_{\tilde{z}} = \| \overline{\nu} \| , \quad (32)$$

$$\varphi_{\tilde{z}} = \arctan(\overline{\nu}) . \quad (33)$$

The corresponding weighted variances are calculated as

$$\sigma_{v,\tilde{z}}^2 = \eta_\sigma \sum_{c\in\mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \left(\|\nu_t^c\| - v_{\tilde{z}} \right)^2 \quad (34)$$

$$\sigma_{\varphi,\tilde{z}}^2 = \eta_\sigma \sum_{c\in\mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \left((\arctan(\nu_t^c) - \varphi_{\tilde{z}}) \bmod 2\pi \right)^2 \quad (35)$$

with an unbiased weight normalization [53]

$$\eta_\sigma = \frac{\sum\limits_{c\in\mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c)}{\left(\sum\limits_{c\in\mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \right)^2 - \sum\limits_{c\in\mathcal{C}_{\tau,t}} (m(D_{\tilde{z},t}^c))^2} . \quad (36)$$

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

That 2-D velocity $\overline{v}$ represents a rough but robust estimate since it is temporally filtered by the particle tracking using a simple constant velocity motion model without specific object assumptions. As described before, the particle tracking thereby serves as a track-before-detect concept. Hence, the velocity and orientation, in terms of the movement direction, of a newly extracted high-level object are directly initialized with that estimate even without directly measuring those states.

Since the particle tracking is performed permanently, i.e., even after objects are extracted, this particle-based estimation can be extracted in each update step. Hence, $v_{\tilde{z}}$ and $\varphi_{\tilde{z}}$ can be interpreted as pseudo measurements and directly used for the UKF update as performed in our previous work [5]. However, these estimates are already filtered by the particle tracking, resulting in a correlation of those pseudo measurements and thus violating the required uncorrelated input noise of a Kalman filter. This multi-filtering can increase the filtering latency, i.e., the system becomes slower if there are any deviations from the motion model, which is critical for scenarios with a fast-changing turn rate or acceleration. Therefore, in this work, the particle-based velocity and orientation estimates are only used directly for the object state initialization.

Nonetheless, since the particle-based estimation does not rely on specific object shape assumptions, the estimated mean and variance are used to form a confidence interval

$$\mathcal{I}_{\varphi}^{\mathcal{X}} = \left\{ \varphi \in [-\pi, \pi] \;\middle|\; |\varphi - \varphi_{\tilde{z}}| \bmod \pi \leq \gamma \sigma_{\varphi,\tilde{z}} + \sigma_{\varphi,\min}^{\mathcal{X}} \right\} \tag{37}$$

of the assumed object orientation based on the movement of the point-mass particles. The parameter γ scales the respective variance, with defined added uncertainty $\sigma_{\varphi,\min}^{\mathcal{X}}$ of the particle filtering in general. Hereby, since the UKF state estimation is not used for determining this orientation interval, a feedback of the object state and a drift off toward a wrong local convergence is avoided.

Overall, the low-level particle tracking enables a robust velocity and orientation initialization of the object state, while a subsequent direct measurement update is avoided as this would result in multi-filtering. Instead, the orientation is further estimated by a local optimization of the filtered state regarding the currently measured freespace, which is described in the following, whereas the velocity is only updated with radar Doppler measurements, as presented in the next section.

F. Orientation Estimation Based on Freespace Evidence

The freespace evidence $m(F_{z,t}^{c})$ of the measurement grid is evaluated in terms of the surrounding areas of the box edges to determine the current visibility of each edge as used for the reference point selection of the position update. In the following, this freespace information is also used for estimating the object orientation by analyzing the inside area of the object box. For this, the freespace ratio $r_F(\mathcal{A}_i^{\varphi})$ as defined in (28) is evaluated with the area $\mathcal{A}_i^{\varphi}$ enclosing all cells inside the measurement minimum bounding box with length l_z and width w_z given the set of associated cells $\mathcal{C}_{\tau,t}$ and the respective evaluated orientation φ, see (22)-(23). This concept is based on the assumption that objects do not contain freespace inside the object shape.

In other words, this means that a low freespace ratio $r_F(\mathcal{A}_i^{\varphi})$ implies that the corresponding measurement box fits well with the data of the measurement grid, whereas an inaccurate orientation estimate results in a higher freespace ratio. Hence, this is described by the optimization problem

$$\varphi_z^* = \underset{\varphi \in \mathcal{I}_{\varphi}^{\mathcal{X}}}{\arg\min}\ \kappa(\varphi), \quad \kappa(\varphi) = r_F(\mathcal{A}_i^{\varphi}), \tag{38}$$

minimizing the cost function $\kappa(\varphi)$ that equals the ratio $r_F(\mathcal{A}_i^{\varphi})$ of the included freespace. The possible range of the orientation φ is thereby limited by the particle confidence interval $\mathcal{I}_{\varphi}^{\mathcal{X}}$ as defined in (37).

If the track τ is not newly extracted and its predicted orientation is within the particle confidence interval, then the optimization starts with that predicted track orientation ($\varphi_0 = \hat{\varphi}_t$), otherwise the mean particle orientation is used ($\varphi_0 = \varphi_{\tilde{z},\tau}$). A hill-climbing optimization with a discrete orientation delta δ_{φ}, e.g., $\delta_{\varphi} = 2°$, is then performed, resulting in the discrete optimum φ_{i^*} with $\varphi_i = \varphi_0 + i \cdot \delta_{\varphi} \in \mathcal{I}_{\varphi}^{\mathcal{X}}$ and $i \in \mathbb{Z}$. Finally, a continuous optimum is approximated by the weighted mean with the two neighbors $\varphi_{i^*} \pm \delta_{\varphi}$, i.e.,

$$\varphi_z^* \approx \left(\sum_{\varphi'} \frac{1}{\kappa(\varphi')} \right)^{-1} \cdot \sum_{\varphi'} \frac{1}{\kappa(\varphi')} \cdot \varphi', \tag{39}$$

$$\varphi' \in \{\varphi_{i^*},\, \varphi_{i^*} + \delta_{\varphi},\, \varphi_{i^*} - \delta_{\varphi}\}. \tag{40}$$

The corresponding variance

$$\sigma_{\varphi,z}^2 = (\nabla_r - \nabla_l)^{-2} \tag{41}$$

is determined by the right-sided ($i > i^*$) and left-sided ($i < i^*$) gradients with the maxima $\varphi_i^{\max,r}$ and $\varphi_i^{\max,l}$, respectively, i.e.,

$$\nabla_r = \frac{\kappa(\varphi_i^{\max,r}) - \kappa(\varphi_{i^*})}{\varphi_i^{\max,r} - \varphi_{i^*}}, \quad \varphi_i^{\max,r} = \max_{i > i^*} \varphi_i\,. \tag{42}$$

In contrast to an L-shape fitting, this freespace evaluation concept does not require a feature-specific extraction. Furthermore, occupancy can occur not only on the outer edge, but also anywhere inside that object. Moreover, only a local optimization is performed based on the orientation of the predicted track and the estimated movement direction of the particle tracking, thus resulting in an approach that is also robust against deviations from the box model shape.

In summary, this section has presented a UKF-based dynamic state estimation of an object with measurement data modeled by the dynamic grid representation. The object position and orientation are updated by an abstracted measurement box of the associated occupied cells and an evaluation of the freespace of the grid. For the object state initialization, the mean of the 2-D cell velocities of the underlying low-level particle tracking is used as the initial estimate of the object velocity and orientation in terms of the movement direction.

V. Additional Radar Velocity Measurements

The dynamic state of each track, in particular the object pose, is robustly estimated based on the occupancy and freespace evidence of the current measurement grid as described in the previous section. This section extends that concept by evaluating radar Doppler velocity measurements that

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

may occur at a track, which significantly improves the motion state estimation for highly dynamic movements. A novel radar velocity-based UKF motion estimation is proposed, which directly uses the measurement space of the Doppler velocity measurements for updating the object state by a projection of the UKF sigma points to expected radial velocity components. The radar velocity measurements are thereby integrated into the overall grid-based framework to utilize the advantages of the fused grid-based input representation, including the spatially accurate lidar measurements.

A. Association of Radar Doppler Velocity Measurements

The radar velocity measurements $z^c_{v,t}$ are represented in the measurement grid structure as separate layers, as defined in (4), containing the radial velocity v^r_z, azimuth angle θ^r_z, and the position x_s of the corresponding sensor origin at the current time t. The radar measurements thereby typically have a higher spatial uncertainty than those of the lidar sensors, thus resulting in a lower occupancy evidence and multiple cells containing the same radial velocity measurement, which in turn is useful for the velocity weighting of the low-level particle tracking. The particle labeling association [6] used in this work, however, discards cells with a low dynamic occupancy evidence, as only cells with a higher occupancy evidence should be considered to form the measurement minimum bounding box of the object tracking.

Hence, all occupied cells of the extended set of associated cells $\mathcal{C}^+_{\tau,t} \supseteq \mathcal{C}_{\tau,t}$, representing all cells associated to the track τ by the particle label distributions but without discarding cells with a low occupancy evidence, are considered in the following, while only cells that contain a velocity measurement are relevant, i.e.,

$$\mathcal{C}^r_\tau = \{\, c \in \mathcal{C}^+_{\tau,t} \mid \exists\, z^c_{v,t} \,\}\,. \tag{43}$$

This extended association is illustrated in Fig. 5, which forms the basis for updating the tracks by the associated radar Doppler measurements.

B. Geometric Relations of the Radial Velocity Component

The velocity of a track τ is represented as a scalar velocity v_τ along the orientation φ_τ with the turn rate ω_τ, which is referenced to the center of the rear axle, see Fig. 2. The 2-D Cartesian velocity v_i in the odometry reference frame at an arbitrary point x_i of the object is calculated as

$$\begin{bmatrix} v^x_i \\ v^y_i \end{bmatrix} = v_\tau \begin{bmatrix} \cos(\varphi_\tau) \\ \sin(\varphi_\tau) \end{bmatrix} + \omega_\tau \begin{bmatrix} -(x^y_i - x^y_\tau) \\ x^x_i - x^x_\tau \end{bmatrix}, \tag{44}$$

as described in [39] using the instant center of rotation (ICR) and the assumption of the Ackermann steering geometry with a drift-free driving state.

The radial velocity regarding the angle θ^r_i results in

$$\begin{aligned} v^r_i &= \begin{bmatrix} \cos(\theta^r_i),\, \sin(\theta^r_i) \end{bmatrix} \begin{bmatrix} v^x_i \\ v^y_i \end{bmatrix} \\ &= v_\tau \left(\cos(\theta^r_i)\cos(\varphi_\tau) + \sin(\theta^r_i)\sin(\varphi_\tau) \right) \\ &\quad + \omega_\tau \begin{bmatrix} \cos(\theta^r_i),\, \sin(\theta^r_i) \end{bmatrix} \begin{bmatrix} -(x^y_i - x^y_\tau) \\ x^x_i - x^x_\tau \end{bmatrix}, \end{aligned} \tag{45}$$

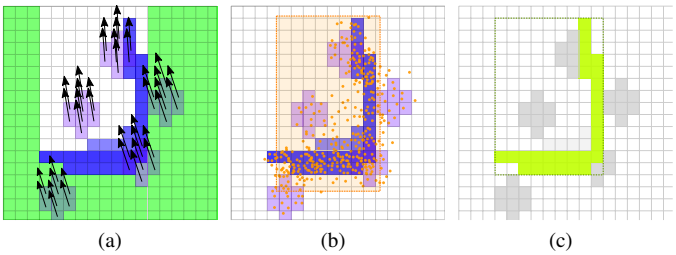


Fig. 5. Extended particle labeling association with radar measurements. (a) Classified measurement grid with radar measurements illustrated by arrows. (b) Dynamic cells with labeled particles and predicted track. (c) Extracted minimum bounding box formed by cells with a high dynamic mass. The remaining cells with radar velocities are considered for the dynamic state estimation, but not utilized for the extraction of the bounding box.

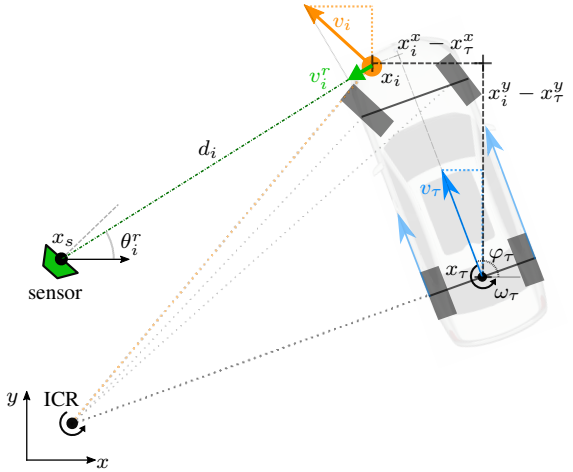


Fig. 6. Geometric relations of the velocity components of a turning object using the instant center of rotation (ICR) as the stationary rotation point of the object and the assumption of the Ackermann steering geometry with a drift-free driving state. The radial velocity v^r_i depends on the angle θ^r_i of the radial component, the dynamic object state s_τ, and the position difference between the track x_τ and the evaluated position x_i or the sensor position x_s.

which depends on the evaluated position x_i if $\omega_\tau \neq 0$. Using the cell centers x_c of $z^c_{v,t}$ as the position, however, introduces a bias error due to the grid discretization and the additional spatial uncertainty over multiple grid cells as shown in Fig. 5. The position of the radar detection can also be described

$$\begin{bmatrix} x^x_i \\ x^y_i \end{bmatrix} = d_i \begin{bmatrix} \cos(\theta^r_i) \\ \sin(\theta^r_i) \end{bmatrix} + \begin{bmatrix} x^x_s \\ x^y_s \end{bmatrix} \tag{46}$$

relative to the corresponding position x_s of the sensor origin at the time of the measurement with the distance $d_i = \|x_i - x_s\|$ between the evaluated position x_i and the sensor origin x_s and the azimuth angle θ^r_i of the measurement. Since (46) also fulfills the relation

$$\begin{bmatrix} \cos(\theta^r_i),\, \sin(\theta^r_i) \end{bmatrix} \begin{bmatrix} -x^y_i \\ x^x_i \end{bmatrix} = \begin{bmatrix} \cos(\theta^r_i),\, \sin(\theta^r_i) \end{bmatrix} \begin{bmatrix} -x^y_s \\ x^x_s \end{bmatrix}, \tag{47}$$

the radial velocity of (45) can also be evaluated without any additional bias error, which finally results in

$$\begin{aligned} v^r_i =& v_\tau \cos(\theta^r_i - \varphi_\tau) \\ &+ \omega_\tau \big(\sin(\theta^r_i)(x^x_s - x^x_\tau) - \cos(\theta^r_i)(x^y_s - x^y_\tau) \big) \end{aligned} \tag{48}$$

using the trigonometric addition formula. The different geometric relations of the velocity parts are summarized in Fig. 6.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

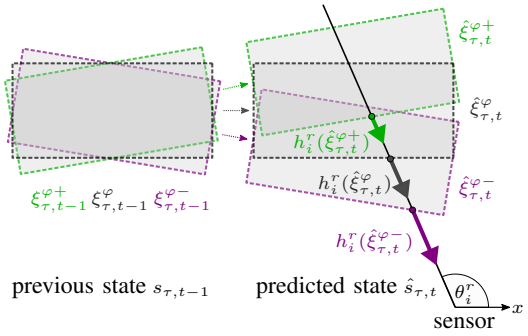


Fig. 7. Illustration of different state orientations φ of the UKF sigma points ξ, their prediction $\hat{\xi}$, and the different expected radial velocity components $h_i^r(\hat{\xi}_{\tau,t})$ given the observation angle θ_i^r of the radar sensor.

C. Radar Velocity-Based UKF Motion Estimation

Based on the relation between the object state s_τ and the radial velocity v_i^r at a particular position x_i as derived in (48), expected radial velocities can be determined by the predicted state $\hat{s}_\tau$. This forms the basic idea of the measurement update with radar Doppler measurements, since the innovation between the predicted state and the measurement is evaluated in the measurement space for a Kalman filter, while the UKF sigma points $\xi \in \Xi$ enable the nonlinear projection $h^r(\xi)$. Each predicted sigma point $\hat{\xi}_{\tau,t}$ of the track τ is projected

$$\begin{aligned} h_i^r(\hat{\xi}_{\tau,t}) =& \hat{v}_{\tau,t,\xi} \cos(\theta_{z,i}^r - \hat{\varphi}_{\tau,t,\xi}) \\ &+ \hat{\omega}_{\tau,t,\xi} \big(\sin(\theta_{z,i}^r)(x_{s,i}^x - \hat{x}_{\tau,t,\xi}^x) \\ &\quad - \cos(\theta_{z,i}^r)(x_{s,i}^y - \hat{x}_{\tau,t,\xi}^y) \big) \end{aligned} \tag{49}$$

to the measurement space of the radial velocity for each measured velocity $v_{z,i}^r$ with azimuth angle $\theta_{z,i}^r$ and the corresponding sensor position $x_{s,i}$. The predicted radial velocity is calculated by the weighted sum of all sigma points, i.e.,

$$\hat{v}_i^r = \sum_{\xi \in \Xi} w_\xi h_i^r(\hat{\xi}_{\tau,t}) \,, \tag{50}$$

with a total of $|\Xi| = 2 \cdot |s_{\tau,t}| + 1 = 13$ sigma points for the dynamic state $s_{\tau,t}$, finally resulting in the innovation $v_{z,i}^r - \hat{v}_i^r$.

The measured azimuth angle $\theta_{z,i}^r$ is assumed to be noise-free in this work, i.e., errors of this parameter are not considered in (49). However, implausible measurements, e.g. caused by micro-Doppler, Doppler ambiguities, or a wrong association, are discarded by evaluating the innovation with a 3σ-range of the predicted measurement variance as the valid gating area.

Overall, this UKF-based measurement update by the radar Doppler velocities enables a generic dynamic state estimation, which is also applicable to other measured velocity components. Not only is the velocity estimated that way, but all variables of the dynamic state $s_{\tau,t}$ are estimated implicitly, as they all influence the predicted state $\hat{s}_\tau$ and the derived predicted radial velocity $\hat{v}_i^r$. For example, a varying track orientation also results in a different angle difference to the observation angle, leading to a different radial velocity as illustrated in Fig. 7. In contrast, an explicit determination of the velocity, orientation, and turn rate by a least-squares fitting of the equation set (48) for all measured radial velocities requires at least three detections of two sensors at the same measurement time and with varying azimuth angles.

VI. Shape Estimation and Object Classification

This section focuses on the shape estimation in terms of the length and width of the bounding box of an object track. Furthermore, an object classification is performed based on the geometry distribution estimation and the maximum filtered velocity of the track. Finally, a specific object track size is extracted using both the geometry distribution and the object classification. Overall, this combined estimation enables modeling of prior class knowledge of the assumed length and width if either has not been fully observed yet.

A. Histogram Filter Based Geometry Distribution Estimation

The length l_τ and width w_τ of a track τ, with random variables L_τ and W_τ, respectively, form the geometric shape

$$g_\tau = \left[l_\tau, \, w_\tau \right]^\mathsf{T} \tag{51}$$

of the selected box model representation. This state is assumed to be static, i.e., it does not change over time, thus the time index of g_τ is omitted. The box shape is updated iteratively by the geometry measurement, which primarily contains the extracted length l_z and width w_z of the measurement box as defined in (22)–(23). However, the real length and width are usually not fully observable, which means that this measurement box generally describes only a minimum bounding box, i.e., a lower bound of the geometric shape. To determine whether l_z and w_z also represent upper bounds and thus the real size of that track, the visibility $\vartheta_z^e \in [0, 1]$ of each edge e of the current measurement box is considered, which was calculated in (29) as part of the reference point selection. For example, an upper bound of the length l_τ is measured if both edges at the front and the rear are currently fully visible, i.e., $\vartheta_z^{\text{front}} = \vartheta_z^{\text{rear}} = 1$. Hence, with separation of the length and width components, the geometry measurement of the current time instance t is described by the two vectors

$$z_t^{l,\vartheta} = \left[l_z, \, \vartheta_z^{\text{front}}, \, \vartheta_z^{\text{rear}} \right]^\mathsf{T} , \tag{52}$$

$$z_t^{w,\vartheta} = \left[w_z, \, \vartheta_z^{\text{left}}, \, \vartheta_z^{\text{right}} \right]^\mathsf{T} . \tag{53}$$

Overall, the goal is to estimate the geometry of a track τ given all measurements up to time t, denoted as $1:t$, i.e.,

$$p(l_\tau, w_\tau \mid z_{1:t}^{l,\vartheta}, z_{1:t}^{w,\vartheta}) = p(l_\tau \mid z_{1:t}^{l,\vartheta}) \, p(w_\tau \mid z_{1:t}^{w,\vartheta}) \,. \tag{54}$$

The length and width are estimated separately, assuming independence of both. In the following, only the estimation of the length distribution $p(l_\tau \mid z_{1:t}^{l,\vartheta})$ is described, since the width distribution $p(w_\tau \mid z_{1:t}^{w,\vartheta})$ is determined equivalently.

The inverse sensor model of the length is modeled as a piecewise function

$$\begin{aligned} p(l \mid z_t^{l,\vartheta}) &\propto \exp\left(-\frac{(l - l_z)^2}{2 \varsigma_{l,\vartheta} \, \sigma_l^2} \right) , \\ \varsigma_{l,\vartheta} &= \begin{cases} \left(\vartheta_z^{\text{front}} \, \vartheta_z^{\text{rear}} \right)^{-1} , & \text{if } \; l \geq l_z \\ 1 \,, & \text{else} \end{cases} , \end{aligned} \tag{55}$$

i.e., a normal distribution in which the variance below and above the measured length l_z differs by a scaling variable $\varsigma_{l,\vartheta}$. This scaling variable is used to model the impact of the edge

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

visibility. If both edges are fully visible, then the probability $p(l_\tau \mid z_t^{l,\vartheta})$ corresponds to an unmodified normal distribution with l_z representing an estimate of the real length. If at least one edge is not visible, then l_z is only a lower bound of the object track length l_τ, in which (55) is modeled similar to a sigmoid function. This sensor model is illustrated in Fig. 8 for varying edge visibilities.

The probability distribution of the length $p(l_\tau \mid z_t^{l,\vartheta})$ is estimated by a 1-D histogram filter, requiring a decomposition of the continuous state space $l_\tau \in \mathbb{R}^+$ into discrete values $\{l_i\}_{i=1}^{I_l} = \{l_1, l_2, \dots, l_{I_l}\}$. The probability density function of the random variable L_τ is thus approximated by the discrete probability mass function $\{p^i_{l_\tau,t}\}$ in that interval, i.e.,

$$p(l_\tau \mid z_{1:t}^{l,\vartheta}) \approx \frac{p^i_{l_\tau,t}}{\delta^i}\,, \quad l_\tau \in \left(l_i - \tfrac{\delta_i}{2}, l_i + \tfrac{\delta_i}{2}\right], \tag{56}$$

with δ_i defining the interval size of the corresponding bin i. The individual interval size δ_i can be adapted over time based on the current distribution using dynamic decomposition [54] to increase the histogram approximation. However, we use a constant interval size here, i.e., static decomposition, since the number of required intervals I_l is rather low due to the accuracy limitation of the input grid resolution and the limited range of the possible values of the modeled classes.

The posterior probability distribution of the discrete length intervals $\{l_i\}_{i=1}^{I_l}$ is then recursively estimated

$$p^i_{l_\tau,t} = \frac{p(l_i \mid z_t^{l,\vartheta})\, p^i_{l_\tau,t-1}}{\sum_{j=1}^{I_l} p(l_j \mid z_t^{l,\vartheta})\, p^j_{l_\tau,t-1}} \tag{57}$$

by applying Bayes rule and using the inverse sensor model as defined in (55). The initial length distribution $\{p^i_{l_\tau,0}\}_{i=1}^{I_l}$ is set to a uniform distribution with $p^i_{l_\tau,0} = \frac{1}{I_l}\ \forall i$. To ensure convergence toward new measurements and avoid singularities, the minimum value of each discrete probability $p^i_{l_\tau,t-1}$ of the previous time instance is limited to a value $\epsilon_{\min} > 0$ in (57).

Overall, this histogram filter-based length and width estimation enables modeling non-Gaussian distributions, which is beneficial as lower and upper bounds of the measurement box are distinguished by the freespace information.

B. Classification Based on Geometry and Velocity Information

In the following, the object track is classified with regard to the modeled classes $k \in \mathcal{K}$ of the set

$$\mathcal{K} = \{\text{car}, \text{truck}, \text{pedestrian}, \text{cyclist}, \text{motorcycle}, \text{other}\}\,. \tag{58}$$

This classification is based on the features length l, width w, and velocity v, resulting in the estimation problem

$$p(k \mid l, w, v) = \frac{p(k)\, p(l, w, v \mid k)}{p(l, w, v)}\,. \tag{59}$$

Those features are assumed to be conditionally independent of each other, i.e.,

$$p(l, w, v \mid k) = p(l \mid k)\, p(w \mid k)\, p(v \mid k)\,, \tag{60}$$

which simplifies (59) to a naive Bayes classifier with

$$p(k \mid l, w, v) \propto p(k)\, p(l \mid k)\, p(w \mid k)\, p(v \mid k)\,. \tag{61}$$

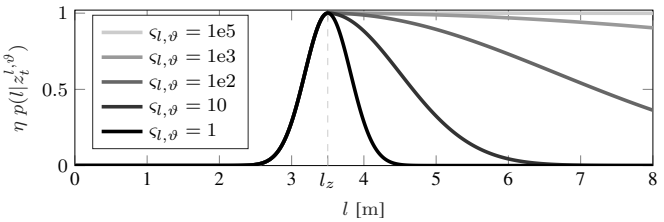


Fig. 8. Inverse sensor model with boundary visibility consideration. Example shows length distribution with measured length $l_z = 3.5\,\text{m}$ (dashed line).

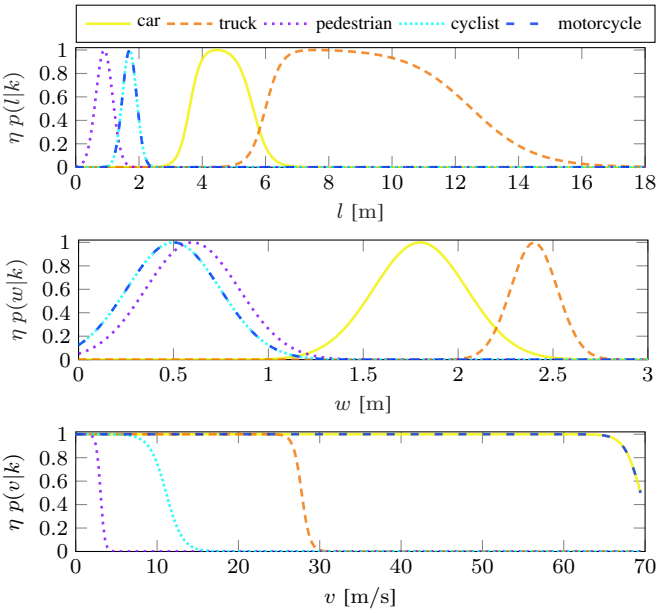


Fig. 9. Modeled likelihood $p(\cdot|k)$ of the features length l, width w, and velocity v given the individual classes $k \in \mathcal{K}$. The distributions are illustrated unnormalized with η indicating different normalization factors.

The modeled likelihood functions of the length $p(l \mid k)$, the width $p(w \mid k)$, and the velocity $p(v \mid k)$ are shown in Fig. 9 for all classes $k \in \mathcal{K}$ except the class $k = \{\text{other}\}$, which is modeled uniformly for all features. The length is modeled as a normal distribution $\mathcal{N}(l;\, \mu_{l,k},\, \sigma^2_{l,k})$ with mean $\mu_{l,k}$ and variance $\sigma^2_{l,k}$ for the classes $k \in \{\text{pedestrian}, \text{cyclist}, \text{motorcycle}\}$. The car and truck classes are modeled differently due to their greater variety in the length. Hence, both classes are represented using lower and upper boundaries of the possible length, which is modeled

$$p(l \mid k) \propto S(l, l^k_{\min}, \alpha^k_{l,\min})\left(1 - S(l, l^k_{\max}, \alpha^k_{l,\max})\right)\,, \quad k \in \{\text{car}, \text{truck}\} \tag{62}$$

using a combination of two sigmoid (logistic) functions with

$$S(x, x_0, \alpha) = \frac{1}{1 + e^{-\alpha(x - x_0)}}\,. \tag{63}$$

Similarly, the likelihood of the velocity v given a class k is also modeled as a negated sigmoid function

$$p(v \mid k) \propto 1 - S(v, v^k_{\max}, \alpha^k_v)\,. \tag{64}$$

This means that a class k is unlikely if the observed velocity v exceeds the maximum velocity $v^k_{\max}$ of this class. This feature is only used as an exclusion criterion, and thus the likelihood in (64) is not normalized over the individual classes here. For

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

example, a velocity of $v = 1\,\mathrm{km/h}$ does not represent a gain of information and should result in equal class probabilities of all classes $k \in \mathcal{K}$ regarding the velocity feature. This is equivalent to normalizing (64) and adjusting the individual class priors $p(k)$ in (61) accordingly.

The best class regarding the maximum a posteriori classification given the features (l, w, v) then results in

$$\begin{aligned} k^* &= \arg\max_{k \in \mathcal{K}} \; p(k \mid l, w, v) \\ &= \arg\max_{k \in \mathcal{K}} \; p(k)\, p(l \mid k)\, p(w \mid k)\, p(v \mid k)\,. \end{aligned} \tag{65}$$

However, rather than classifying the individual measurements and recursively filtering that classification, we perform the classification directly on the filtered track τ. This means that the classifier is based on the probability distributions of the length L_τ and width W_τ, which are recursively estimated by the histogram filter. Therefore, instead of evaluating a single length l of the likelihood $p(l \mid k)$, the expectation

$$E_L[\,p(L_\tau \mid k)\,] = \sum_{i=1}^{I_l} p(l_i \mid k)\, p^i_{l_\tau,t} \tag{66}$$

regarding the likelihood of L_τ that considers all possible discretized lengths $\{l_i\}_{i=1}^{I_l}$ weighted by their corresponding probability $p^i_{l_\tau,t}$ has to be taken into account. Accordingly, the expectation $E_W[\,p(W_\tau \mid k)\,]$ of the likelihood for the width distribution W_τ has to be determined. For the velocity feature that is evaluated by the classifier, the maximum of the filtered velocity over all time instances of the track is used, i.e.,

$$v_\tau^{\max} = \max_{t'=1,\dots,t} (v_{\tau,t'})\,. \tag{67}$$

Overall, based on (65), the best fitting class k^*_τ of the track τ regarding the filtered geometry state and the maximum of the filtered velocity results in

$$\begin{aligned} k^*_\tau &= \arg\max_{k \in \mathcal{K}} \; E_{L,W}[\,p(k \mid L_\tau, W_\tau, v_\tau^{\max})\,] \\ &= \arg\max_{k \in \mathcal{K}} \; p(k)\, p(v_\tau^{\max} \mid k)\, E_{L,W}[\,p(L_\tau \mid k)\, p(W_\tau \mid k)\,] \\ &= \arg\max_{k \in \mathcal{K}} \; \Bigg(\; p(k)\; p(v_\tau^{\max} \mid k) \sum_{i=1}^{I_l} p(l_i \mid k)\, p^i_{l_\tau,t} \\ &\qquad \sum_{i=1}^{I_w} p(w_i \mid k)\, p^i_{w_\tau,t} \Bigg)\,. \end{aligned} \tag{68}$$

This simple classification concept can further be combined with other extracted features, especially direct camera-based classification information, which is addressed in future work.

C. Extraction of Estimated Length and Width of Box Model

The histogram filter-based geometry estimation results in probabilities $\{p^i_{l_\tau,0}\}_{i=1}^{I_l}$ and $\{p^i_{w_\tau,0}\}_{i=1}^{I_w}$ of the discrete intervals of the length and width. Overall, however, a specific value of the assumed length l_τ and width w_τ is required for the box model representation of the track. For this purpose, the estimated length and width distributions are combined with the best fitting class k^*_τ as defined in (68). Thus, an unobservable full length or width, e.g., a preceding vehicle for which only a minimum length is observable, is then estimated by evaluating the corresponding likelihood of the length or width given the class k^*_τ as depicted in Fig. 9 rather than extracting that observed minimum bounding box. The finally extracted length

$$l^*_\tau = l_{i^*_{l_\tau}}\,, \tag{69}$$

and equivalently the width, is determined by the length of the histogram bin

$$i^*_{l_\tau} = \arg\max_{i=1,\dots,I_l} \left\{ p^{i,l}_{\text{comb}} : p^{i,l}_{\text{comb}} > \left((1+\epsilon) \max_{j=1,\dots,i} p^{j,l}_{\text{comb}}\right) \right\} \tag{70}$$

with the highest combined probability

$$p^{i,l}_{\text{comb}} \propto p(l_i \mid k^*_\tau) \cdot p^i_{l_\tau,t} \tag{71}$$

regarding the length likelihood $p(l_i \mid k^*_\tau)$ of the best class k^*_τ and the length histogram distribution $p^i_{l_\tau,t}$. In the case of multiple lengths with similarly high probabilities, the smallest length is chosen by (70) in order to extract the minimum bounding box. The factor $(1+\epsilon)$ with a small $\epsilon > 0$ increases the robustness by ensuring that larger values of the length are only extracted if the corresponding probability is significantly higher than the previous maximum.

By considering only the best class k^*_τ in (71) rather than all classes weighted by their individual class probability as determined in (68), it is ensured that the extracted length l^*_τ also fits to the assumed length of the best class k^*_τ of that track. The length can also be calculated by the mean length weighted by the probability $p^{i,l}_{\text{comb}}$. However, that weighted mean increases when only lower bounds of the length are observed, since then the normalized probabilities of the unobserved larger length values are higher, resulting in a larger extracted length, which is unfavorable in that application and thus avoided by (69).

The priors $p(k)$ of the classes $k \in \{\text{pedestrian}, \text{cyclist}\}$ are selected higher than the prior of the class car such that slow-moving small objects – even when no upper boundary of the length or width has been observed – tend to be classified as such smaller objects. This means that if none of the evaluated features contradicts with the class pedestrian, that class should dominate the others. This results in a more robust association, since the extracted size is also used for the gating area of the association [6], which is selected more conservatively that way.

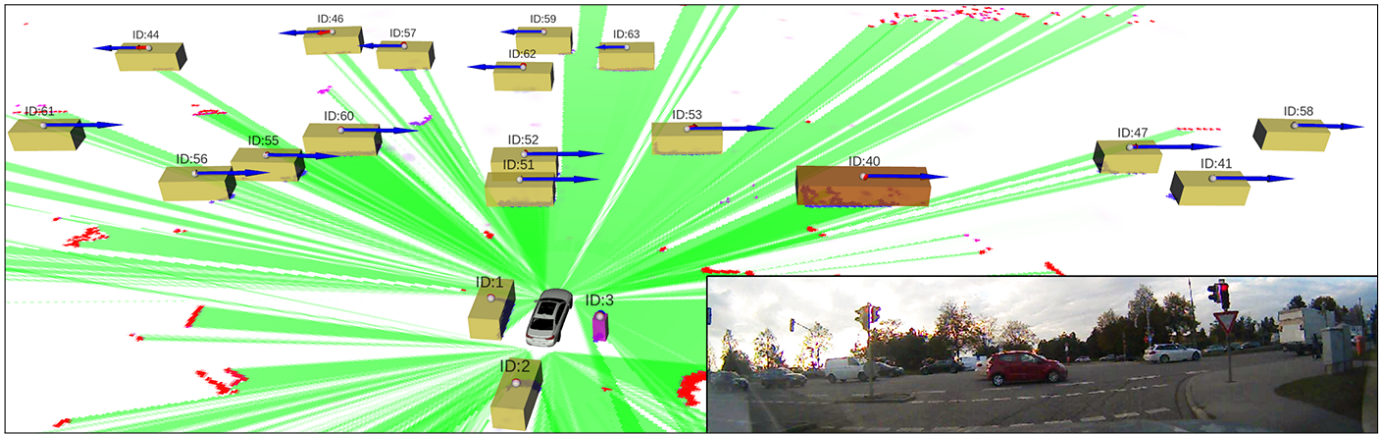
Overall, this approach combines the temporal filtering of the measured box size, including distinguishing lower and upper bounds by evaluating the freespace information, with the corresponding likelihood of the modeled classes, which results in a robust and reasonable object shape estimation.

VII. Experimental Results

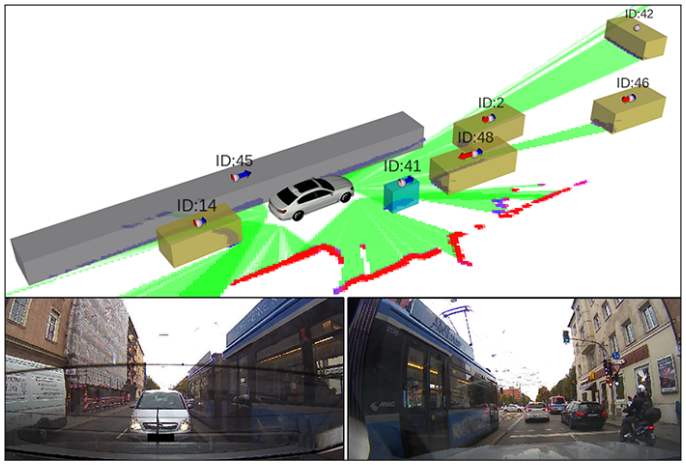
The proposed grid-based object tracking has been tested in various scenarios in the context of urban autonomous driving. The sensor setup of the test vehicle as well as the evaluation setup are briefly described first, followed by qualitative results in different urban environments. The dynamic state estimation of an object is then evaluated quantitatively with critical full braking and turning scenarios using real sensor data with a reference measurement system. Finally, the geometry estimation and object classification is evaluated by a simulated overtaking maneuver by another vehicle with different variants. Further results are demonstrated in the attached video.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

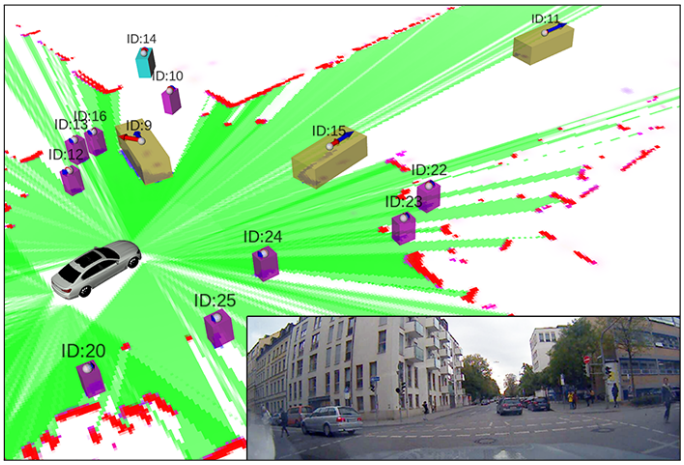
This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248



(a) Scenario at an intersection with a lot of cross traffic.



(b) Scenario with an unknown object (tram) and surrounding vehicles.



(c) Scenario with several surrounding pedestrians and a turning car.

Fig. 10. Different complex urban scenarios with real sensor data. The classified measurement grid (without occupancy accumulation) is shown on the ground. The proposed grid-based object tracking is shown by boxes with the following object classification color coding: car (yellow), truck (orange), pedestrian (purple), cyclist (cyan), motorcycle (blue), unknown (gray). The scalar object velocity is indicated by a blue arrow, the acceleration by a red arrow, which also illustrates the yaw rate transformed to the lateral acceleration $a_{\text{lat}} = v \cdot \omega$.

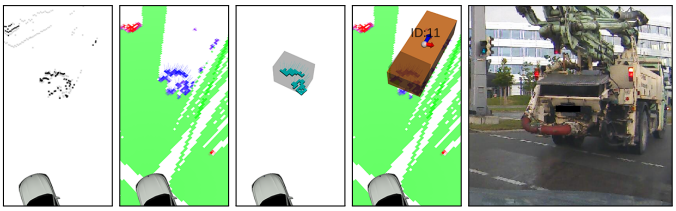
A. Sensor and Evaluation Setup

The sensor setup of the test vehicle is as follows:

- 5 lidar sensors (1 in front + each corner)
- 3 long-range radars (1 in front, 2 to the rear)
- 4 short-range radars (each corner)

This enables full 360° coverage with a multi-sensor fusion of lidar and radar sensors. The integration of camera information into the proposed grid-based approach will be addressed in future work. The test vehicle also consists of high-precision inertial measurement units (IMU) for the odometry estimation. Furthermore, for the quantitative evaluation, the observed target vehicle is equipped with a measurement reference system, resulting in ground truth data of the dynamic motion state.

The processing pipeline of the preceding grid-based estimation, including object extraction and cell association, was summarized in Section III. The grid resolution is 1024×1024 grid cells with a cell size of $0.15\,\text{m} \times 0.15\,\text{m}$, and a maximum of 100 particles per grid cell for the low-level particle tracking. The grid fusion is triggered by the front lidar sensor with a measurement rate of $25\,\text{Hz}$. The implementation is primarily parallelized for fast GPU computing, enabling real-time application of the overall system in our test vehicles with a total cycle runtime below $40\,\text{ms}$.



(a) (b) (c) (d) (e)

Fig. 11. Moving object with an unstructured shape; the lidar detection pattern at the sensing height does not fit a rectangular shape. (a) Lidar detections of front sensor (black: obstacles, gray: ground). (b) Classified measurement grid (blue lines indicate cell velocities). (c) Associated cells and measurement bounding box. (d) Grid-based object tracking. (e) Camera image.

B. Qualitative Results

Qualitative results of the proposed approach in three different urban scenarios are highlighted in Fig. 10. The upper scenario in Fig. 10a shows an intersection with a lot of cross traffic. All vehicles are correctly detected and their moving direction is robustly estimated, even though the left-crossing vehicles in the back are highly occluded by the right-crossing vehicles in the front. One truck is included in that sequence, which is correctly classified since the observed length is larger than those of cars. It can also be seen that stopped vehicles,

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

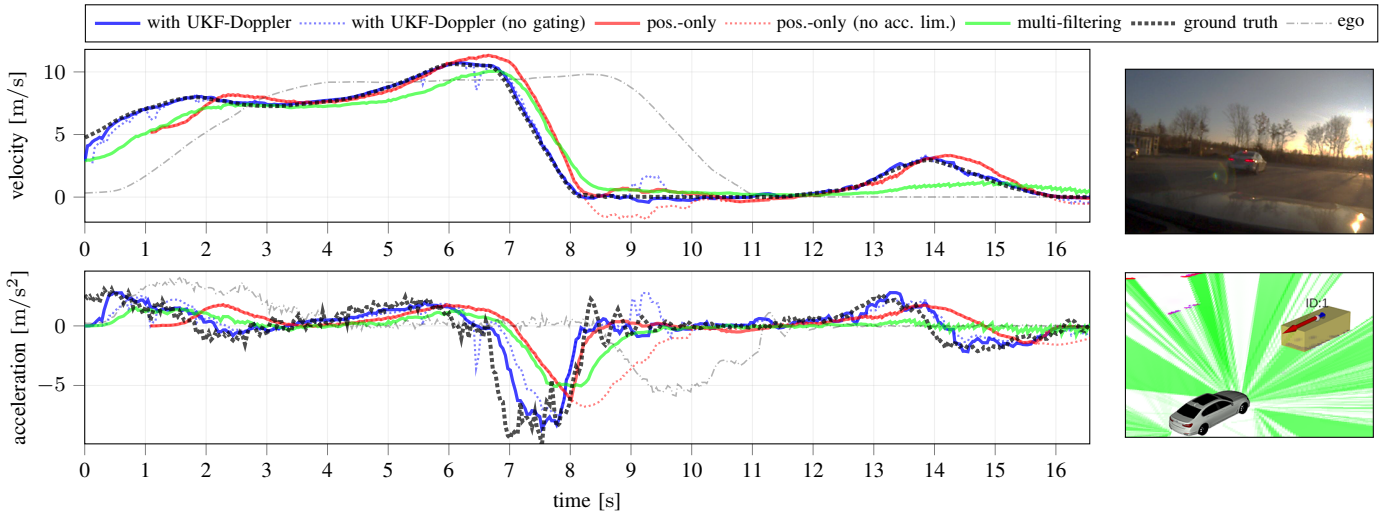


Fig. 12. Full braking scenario with real sensor data. The shown time of all data corresponds to the original measurement time to focus on the filtering behavior, i.e., sensor and system latencies as well as the processing time are excluded in the comparison with the ground truth data.

directly next to the ego vehicle, remain being tracked since they have been observed moving and thus extracted before. Furthermore, a pedestrian who stopped on the right side of the ego vehicle is also correctly extracted, which is outside the camera field of view at the shown time instance.

The second scenario in Fig. 10b shows a situation with a tram, which is classified as unknown as the large length and small width does not fit the modeled classes. A scooter is detected on the right side, which is classified as a cyclist, since the geometry classification of a cyclist and motorcycle is modeled equivalently with a higher prior of cyclists, i.e., an object is only classified as a motorcycle when the maximum estimated velocity is higher than the velocity likelihood of a cyclist as modeled in Fig. 9, which is not the case in the slow-moving scenario. Since the extracted object geometry is determined by the histogram filter distribution combined with the classification, unobserved parts, like the lengths of the partially occluded vehicles in the front (ID 46) or behind the ego vehicle (ID 14), are estimated by the classification likelihood, resulting in reasonable box dimensions of those vehicles.

The third scenario in Fig. 10c includes several slow-moving pedestrians crossing the road, which are all correctly extracted, tracked, and distinguished. The vehicle in front performs a left turn, the occurring turn rate is illustrated by the lateral acceleration component $a_{\text{lat}} = v \cdot \omega$.

Another challenging example is presented in Fig. 11, which contains an object with an unstructured shape of the raw lidar data, caused by the specific low mounted pipe construction on the rear of that truck. Since the proposed object tracking is based on the generic cell movement estimation of the dynamic grid, even such objects can be initially extracted by the density-based cell clustering of dynamic occupied cells [5] and in the following steps temporally filtered by the cell-individual association with the linked particle label distribution [6]. Hence, our approach is also suitable for lower resolution sensor data, whereas, in contrast, an L-shape feature detection or a box fitting without considering the particle-based movement direction estimation or the predicted object track are prone to such deviations of the box shape model.

C. Evaluation of Dynamic State Estimation

The dynamic state estimation is evaluated quantitatively for a test vehicle in two challenging scenarios:

1) Full Braking Scenario: The first scenario represents an emergency braking situation with an abrupt high negative acceleration, up to $-9\,\mathrm{m/s^2}$, which is presented in Fig. 12. The target vehicle first increases the velocity up to $10.6\,\mathrm{m/s}$, then stops by a full braking around $t \in [6.7\,\mathrm{s}, 8\,\mathrm{s}]$, and finally performs a short slow forward movement around $t \in [12\,\mathrm{s}, 16\,\mathrm{s}]$. The ego vehicle also starts moving with a similar velocity, with a position offset to the right of the target vehicle, and performs a slightly smoother braking action around $t \in [8.5\,\mathrm{s}, 11\,\mathrm{s}]$, and then remains stopped. Three different variants of the UKF measurement update are analyzed in the following.

The UKF state estimation using the position of the extracted minimum bounding box as the measurement update is shown in red. The velocity is implicitly estimated by position changes, i.e., first order derivative of the measurable position, whereas the acceleration is determined by the change of the estimated velocity, i.e., second order derivative of the measurable position. A robust balancing of position measurement noise and process noise therefore results in a sluggish state estimation with a higher filtering latency. Moreover, the red dotted line shows the filtering behavior without the additional limitation of the acceleration as defined in (13), i.e., without the time horizon t_{horizon}. Thereby the velocity overshoots to the negative (up to $-1.7\,\mathrm{m/s}$) without that limitation, since the acceleration converges very slowly toward zero. In contrast, the estimation with the limitation, with $t_{\text{horizon}} = 0.25\,\mathrm{s}$, shown by the red solid line, converges fast toward zero by that additional constraint.

The multi-filtering approach where the mean cell velocity $\overline{v}$, see (31), estimated by the low-level particle tracking, is additionally used as a direct velocity update of the object state is shown in green. As discussed before, this results in correlated input data of the UKF, even though the particles are partly weighted by radar Doppler measurements in the case shown. That approach is robust against wrong object assumptions and position updates, e.g., due to an incorrect association, but in

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

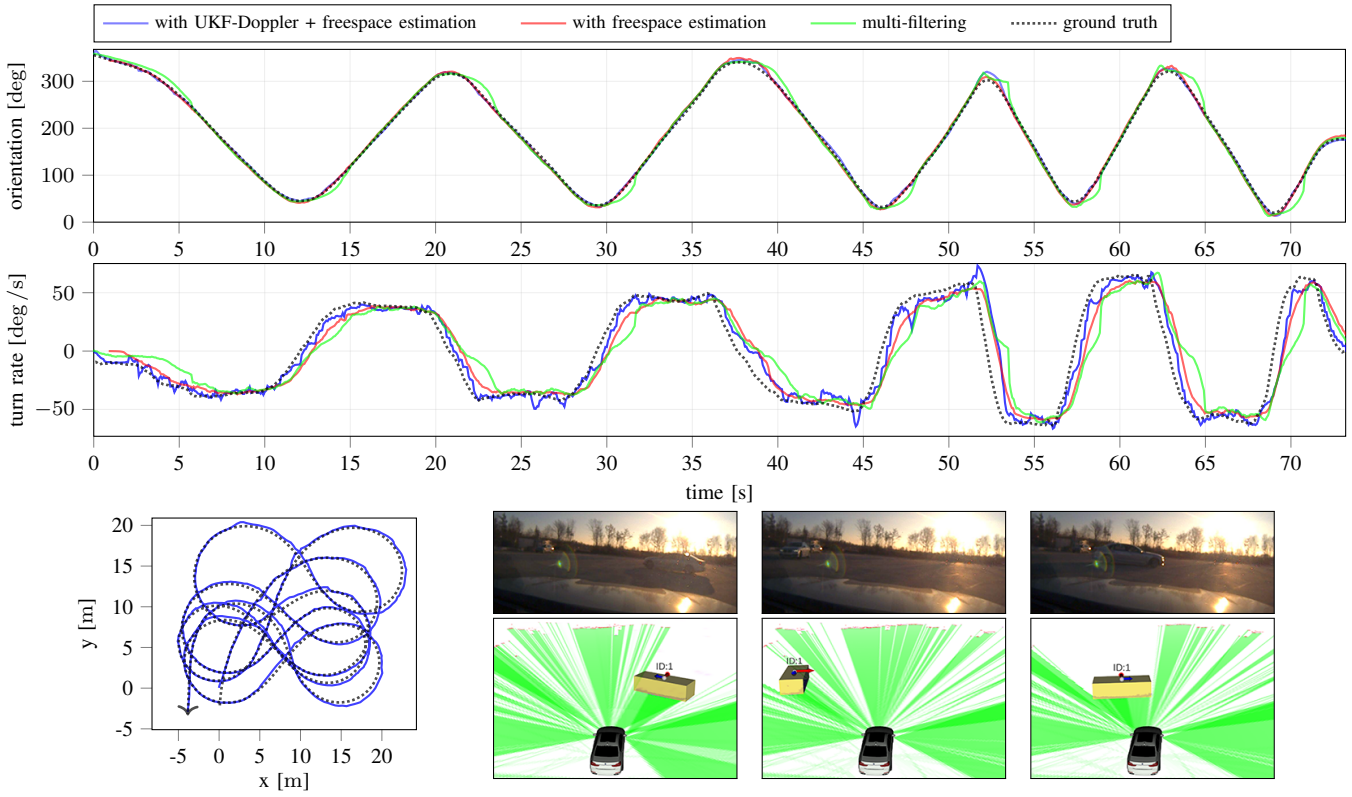


Fig. 13. Turning scenario with multiple "figure-eight" trajectories and real sensor data. As with Fig. 12, the shown time of all data corresponds to the original measurement time by excluding sensor and system latencies as well as the processing time in the comparison with the ground truth data.

that highly nonlinear case, it causes a high filtering latency, which is unfavorable in such critical situations. Furthermore, accurately estimating slow velocities, in the case of the slow forward movement after the full braking, is not possible by the particles, in particular when primarily the long side of the vehicle is observed, since that slow movement cannot be fully resolved for the individual cells of that part without considering neighboring cells, i.e., the overall object movement.

In contrast to multi-filtering based on the particle estimation, the proposed radar Doppler-based UKF measurement update directly uses those measured velocity components in the object state estimation, which is shown in blue. This results in an accurate estimation of the object velocity and a significant reduction of the filtering latency, in particular of the acceleration, which is important for the reaction speed and thus the safety of autonomous mobile robots. The 3σ-validation gating of Doppler measurements, i.e., ignoring all measurements that are implausible with regard to the standard deviation of the expected measurement, further increases the system robustness, as it is less prone to outliers. The blue dotted line illustrates the variant without gating, which in turn requires increasing the Doppler measurement variance and thus decreasing the influence of the radar sensors here.

This discussion of the different approaches is confirmed by the root mean square error (RMSE) of the velocity and acceleration in that scenario, which is summarized in Table I.

To demonstrate the influence of Doppler measurements on the object extraction, the position-only approach (red line) is performed here without using any radar sensors at all, not

TABLE I
ERROR (RMSE) OF DYNAMIC STATE ESTIMATION FOR HIGHLY NONLINEAR MOVEMENTS WITH REAL SENSOR DATA.

Full braking scenario	v-RMSE [m/s]	a-RMSE [m/s^2]
with UKF-Doppler	0.1902	1.0931 (0.9301)*
with UKF-Doppler (no gating)	0.4044	1.4730 (1.3499)*
position-only	0.8641	2.0248 (1.9264)*
position-only (no acc. lim.)	0.9580	2.4813 (2.3844)*
multi-filtering	1.1097	1.9468 (1.8341)*
Turning "figure-eight" scenario	**φ-RMSE [deg]**	**ω-RMSE [deg/s]**
UKF-Doppler + freespace estim.	3.6869	11.4381
freespace estimation	4.8958	15.0031
multi-filtering	9.7007	19.9929

* Calculated with moving-average smoothing of the measured ground truth data.

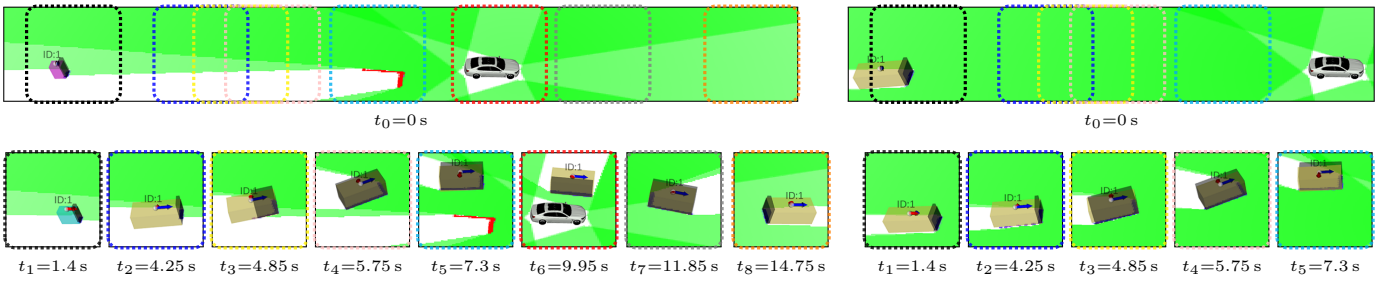
for the underlying grid-based estimation either. For the other two approaches, in contrast, the dynamic grid estimation is enhanced by radar measurements, resulting in a faster object extraction, about 1 s in that configuration. Nonetheless, all approaches result in a good initial object velocity estimate, as they all use the mean particle-based cell velocity for the initialization.

2) Turning Scenario with "Figure-Eight" Trajectories: The second scenario focuses on the orientation and turn rate estimation, which is tested with challenging multiple "figure-eight" trajectories as shown by the path in Fig. 13, resulting in fast changing turn rates with up to 65 deg/s.

The green line shows the multi-filtering approach, where a measurement update of the object orientation is performed by the mean movement direction of the particle-based 2D cell velocity estimation. Even though the vehicle is successfully

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248



(a) Scenario with a parking vehicle occluding parts of the approaching vehicle. (b) Scenario without occlusion.

Fig. 14. Simulated scenario of an approaching overtaking vehicle in two variants: with and without a parking vehicle causing partial occlusion of the approaching vehicle, shown in (a) and (b) respectively. The ego vehicle is stopped in that scenario, the approaching vehicle accelerates in the beginning before overtaking. In the lower row, only image parts are shown; not shown image parts in (b) are equivalent to those shown in (a).

tracked for the complete sequence, the estimation of the orientation and in particular the turn rate is rather inaccurate for that highly nonlinear movement.

The orientation estimation based on the currently measured freespace, see Section IV-F, is shown in red. That generic approach results in a robust orientation estimation, which is primarily based on the high spatial accuracy and thus the derived freespace information of lidar sensors.

Best results are achieved by combining that freespace-based orientation estimation with the UKF-Doppler state estimation, shown in blue, as those radial velocity components at different parts of the object also implicitly enable measuring the object orientation and turn rate. Table I also summarizes the RMSE analysis for the different approaches of that scenario.

D. Evaluation of Shape Estimation and Object Classification

The geometry estimation and object classification are primarily demonstrated by a simulated scenario, whereby the object classification is additionally evaluated for a real urban test drive, as included in the attached video.

1) Simulated Scenario: The simulated sequence depicts an overtaking vehicle with a stopped ego vehicle, which is shown in Fig. 14. Two different variants are analyzed: The first variant in Fig. 14a includes an additional parked vehicle that occludes parts of the overtaking vehicle in the beginning; the second variant in Fig. 14b is without occlusion. The trajectory of the overtaking vehicle is equal in both variants, the vehicle accelerates in the first $4\,\mathrm{s}$ of the sequence up to a velocity of $5\,\mathrm{m/s}$ and then moves with constant velocity. The simulated ground truth object box has a length $l = 3.9\,\mathrm{m}$ and a width $w = 1.8\,\mathrm{m}$. As the focus here is on the shape estimation, only lidar data are simulated, directly on the edge of the simulated box, with a small simulated distance error variance of $0.03\,\mathrm{m}^2$. A detailed evaluation of the measured and extracted length and width as well as the classification is shown in Fig. 15.

The first variant with occlusion, see Fig. 14a, illustrated by solid lines in Fig. 15, is simulated such that the measured width of the minimum bounding box in the initial acceleration phase is about $0.75\,\mathrm{m}$, and a noteworthy length is not observable, i.e., only the front left corner of that target is visible from the position of the ego vehicle. As denoted before, the priors of the pedestrian and cyclist classes are selected higher than the prior

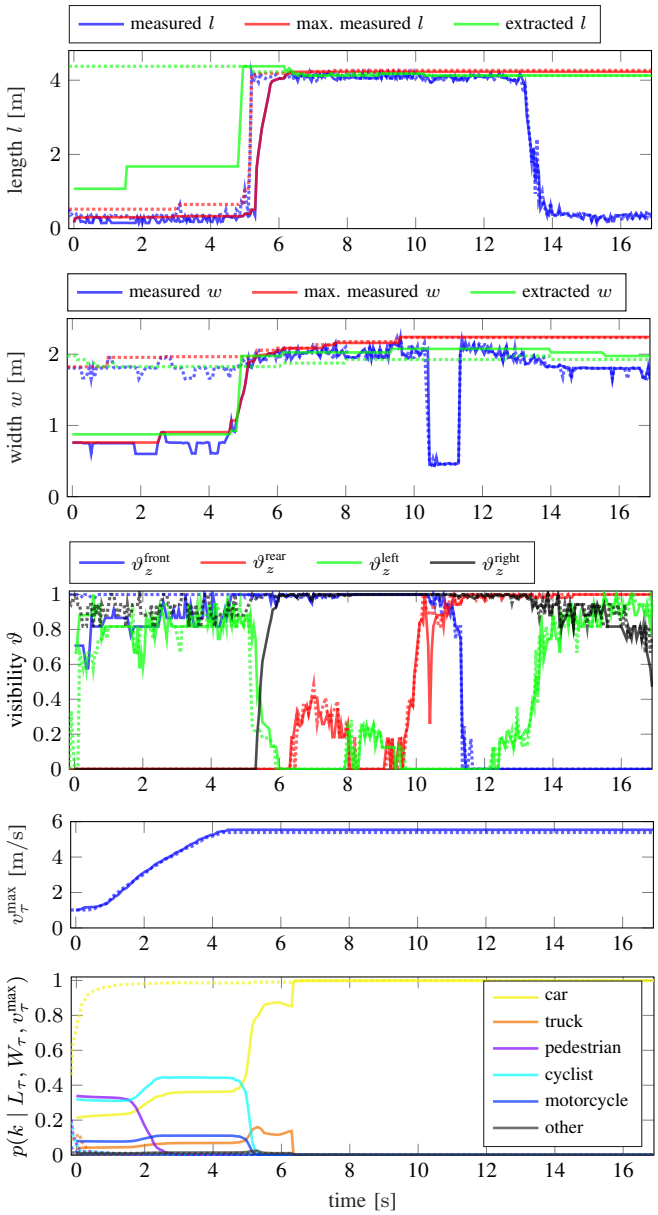


Fig. 15. Evaluation of the geometry estimation and classification of the overtaking vehicle scenario of Fig. 14 for both variants. Results for the scenario with occlusion (Fig. 14a) are shown by solid lines and for the scenario without occlusion (Fig. 14b) by dotted lines, respectively. The simulated ground truth object box has a length $l = 3.9\,\mathrm{m}$ and a width $w = 1.8\,\mathrm{m}$.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

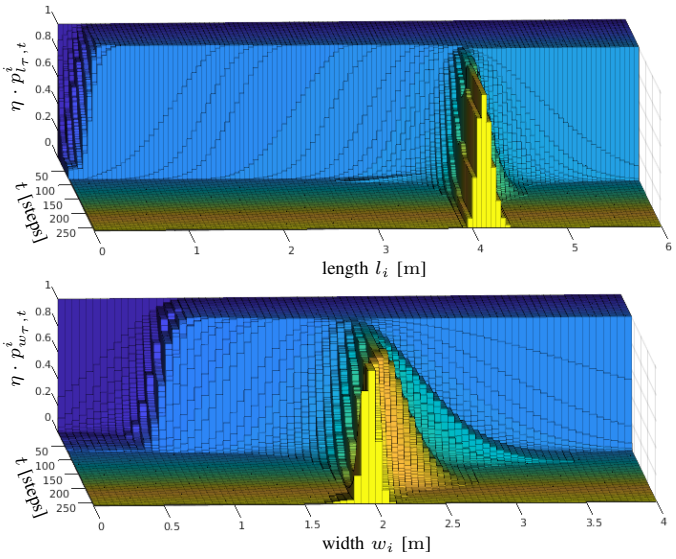


Fig. 16. Histogram distributions over time of the length and width estimation for the simulated overtaking scenario with occlusion as shown in Fig. 14a. All probabilities $p^i_{\cdot,t}$ are normalized to the highest probability of all bins i in each time step t for better representation. The interval size of each bin is $\delta_i = 0.05\,\mathrm{m}$; only a part of the estimated length and width range is shown.

of the class car for a more conservative object size estimation, since the assumed object size is also used for the gating area of the association. Hence, in the beginning, the pedestrian class results in the highest classification probability, since none of the observed features contradicts with that model.

Around $t \approx 2\,\mathrm{s}$, the accelerating simulated object has a velocity about $3\,\mathrm{m/s}$ and thus the maximum estimated object velocity $v_\tau^{\max}$ also increases to that value. Due to the increasing velocity, the pedestrian class becomes less probable; the cyclist class achieves the highest probability. Around $t \approx 5\,\mathrm{s}$, the vehicle starts overtaking and larger parts of the width are observable, whereas the cyclist class also becomes unlikely; the object is then classified as a car, with an increased extracted length and width given by the likelihood of the class.

While the target drives alongside the ego vehicle, the front and rear edge are both visible, thus the real length, i.e., not only a lower bound, is observed, which is selected smaller than the likelihood of the class car in this simulation. The measured length, and thus the estimated object size, is slightly higher than the ground truth box due to the grid cell and histogram discretization combined with the modeled spatial measurement uncertainty and the simple measurement simulation directly on the outer edge of the simulated rectangle. The full object width is observed near the end of the sequence when the target continues moving straight forward in front of the ego vehicle. The geometry estimation is clarified further in Fig. 16 by the corresponding histogram probability distributions of the length and width over time for that sequence. Thereby, the inverse sensor model with the boundary visibility consideration is also recognizable, as introduced in Fig. 8, which is used to distinguish whether the current length or width measurement is only a lower bound or also an upper bound.

The second variant without occlusion, shown in Fig. 14b and illustrated by dotted lines in Fig. 15, is simpler, since the full width of the target is visible directly from the beginning.

TABLE II
CONFUSION MATRIX OF THE OBJECT CLASSIFICATION IN A REAL TEST SEQUENCE. COLUMNS DEPICT THE ACTUAL CLASS, ROWS THE ESTIMATED CLASS. ALL EXTRACTED OBJECTS ARE EVALUATED EACH TIME THE OVERALL TRACKING IS UPDATED, ONLY THE BEST CLASS IS CONSIDERED.

	Car	Truck	Ped.	(Mot.-)Cycl.*1	Other	(FP)*2
Car	23781	39	945	1035	114	882
Truck	45	437	0	0	567	324
Ped.	76	0	9530	1792	0	167
(Mot.-)Cycl.*1	1800	0	834	9888	0	0
Other	294	0	0	0	403	0
# Occurr.*3	25996	476	11309	12715	1084	1373
Rate	0.915	0.918	0.842	0.778	0.372	0.026 (FPR)

*1 The cyclist and motorcycle classes are combined here due to slow object speeds.
*2 FP: false positives (falsely extracted objects).
*3 Total number of evaluated occurrences of that actual class (sum of column).

Due to the observed width, the object is directly classified as a car in the initialization, even though the velocity is very slow at that starting point. The assumed length is directly adjusted with regard to the likelihood of the class, resulting in an extracted length of approximately $4.4\,\mathrm{m}$ here, which is much more accurate than the maximum observed length that is only about $0.5\,\mathrm{m}$ up to $t \approx 5\,\mathrm{s}$.

Overall, as shown by the comparison of both variants with and without high occlusion in the initial phase, the proposed geometry estimation and object classification adapts as desired with regard to the measured box size, the observability in terms of the adjacent freespace, and the target velocity.

2) Object Classification in Real Urban Scenarios: The proposed object classification concept is further demonstrated by real traffic scenarios in an urban environment as shown in the attached video. The evaluated sequence has a duration of about $8\,\mathrm{min}$ with various vehicles, pedestrians, cyclists, etc. Quantitative results are highlighted by the confusion matrix in Table II. Thereby only extracted moving objects are evaluated, since non-moving obstacles or parked vehicles remain in the static grid representation in our approach. Hence, a simple ground truth labeling of the actual class of each of the extracted object tracks is performed, including falsely extracted objects that are labeled as false positives. All extracted objects are evaluated each time the overall tracking is updated, while only the best class k_τ^* as determined in (68) is considered.

These classification results are not directly comparable with other approaches, also since only moving objects are extracted here, but they indicate that this simple classification approach achieves promising results with real sensor data. Note that our approach is based on 2-D measurement occupancy grids, without additional camera information or height estimation, and that the lidar sensors used here only contain four vertical layers, i.e., in total, a limited density of the measurement data. Thereby the object shape and classification estimation are implicitly improved with a more detailed occupancy/freespace derivation of the measurement grid, in particular using lidar sensors with a high vertical resolution, and also when camera classification information is integrated into the measurement grid with a corresponding consideration in the object classification concept. Both aspects are addressed in future work.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

VIII. Conclusion

This paper presented a new grid-based object tracking approach. It is based on pre-processed measurement data in an occupancy grid representation, including low-level data fusion and dynamic estimation of grid cells, which simplifies the moving object detection and association.

This approach results in a robust multi-sensor detection and tracking of surrounding objects, even with many occurring objects, unstructured shapes, or partial occlusion. The object position and orientation are accurately updated by additionally evaluating the measured freespace, which is a generic concept for determining the most visible reference point and optimizing the box orientation without requiring an L-shape fitting. Similarly, the freespace information is used to distinguish lower and upper bounds of the measurement box, while a reasonable object size is extracted by further combining the estimated histogram filter geometry distributions with the likelihood of an object classification that also considers the maximum observed velocity.

The approach is also designed for an accurate dynamic state estimation in critical scenarios by additionally processing radar Doppler velocity measurements. Those measurements are thereby evaluated directly in the object tracking, while the proposed velocity-based UKF update robustly utilizes the measurement space of the radial velocities. In contrast to an indirect multi-filtering velocity update by the underlying dynamic grid, this significantly reduces the filtering latency and therefore further increases the estimation accuracy of the velocity, acceleration, orientation, and turn rate.

Experimental results were demonstrated in the context of autonomous driving with a real test vehicle equipped with multiple lidar and radar sensors. The robust object tracking and classification were presented by a test drive in a real urban environment, while the precise dynamic state estimation was evaluated quantitatively using a reference target vehicle performing full braking and turning maneuvers.

Overall, we thereby also combine the advantages of the spatially accurate lidar sensors for the object pose and shape estimation with the direct velocity measurements of radars for the motion estimation, since occupancy and freespace of the grid are primarily derived by the lidar data in our approach. Ongoing research focuses on integrating camera classification information, thus eventually resulting in a multi-sensor environment estimation that combines the individual benefits of lidar, radar, and camera within our generic grid-based framework.

References

[1] K. C. Chang, R. K. Saha, and Y. Bar-Shalom, "On optimal track-to-track fusion," *IEEE Trans. Aerosp. Electron. Syst*, vol. 33, no. 4, pp. 1271–1276, Oct. 1997.

[2] S. Matzka and R. Altendorfer, "A Comparison of Track-to-Track Fusion Algorithms for Automotive Sensor Fusion," in *Multisensor Fusion and Integration for Intelligent Systems*. Berlin, Germany: Springer, 2009, pp. 69–81.

[3] M. Aeberhard, S. Schlichtharle, N. Kaempchen, and T. Bertram, "Track-to-Track Fusion With Asynchronous Sensors Using Information Matrix Fusion for Surround Environment Perception," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1717–1726, Dec. 2012.

[4] S. Steyer, G. Tanzmeister, and D. Wollherr, "Grid-Based Environment Estimation Using Evidential Mapping and Particle Tracking," *IEEE Trans. Intell. Veh.*, vol. 3, no. 3, pp. 384–396, Sep. 2018.

[5] S. Steyer, G. Tanzmeister, and D. Wollherr, "Object Tracking Based on Evidential Dynamic Occupancy Grids in Urban Environments," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 1064–1070.

[6] S. Steyer, G. Tanzmeister, C. Lenk, V. Dallabetta, and D. Wollherr, "Data Association for Grid-Based Object Tracking Using Particle Labeling," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2018, pp. 3036–3043.

[7] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, pp. 13–es, Dec. 2006.

[8] Y. Bar-Shalom, P. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing, 2011.

[9] K. Granström, M. Baum, and S. Reuter, "Extended Object Tracking: Introduction, Overview, and Applications," *J. Adv. Inf. Fusion*, vol. 12, no. 2, 2017.

[10] C.-C. Wang and C. Thorpe, "Simultaneous localization and mapping with detection and tracking of moving objects," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 3, 2002, pp. 2918–2924.

[11] J. Moras, V. Cherfaoui, and P. Bonnifait, "Credibilist occupancy grids for vehicle perception in dynamic environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2011, pp. 84–89.

[12] T.-D. Vu, O. Aycard, and N. Appenrodt, "Online Localization and Mapping with Moving Object Tracking in Dynamic Outdoor Environments," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 190–195.

[13] T.-D. Vu, J. Burlet, and O. Aycard, "Grid-based localization and online mapping with moving objects detection and tracking: New results," in *Proc. IEEE Intell. Veh. Symp.*, 2008, pp. 684–689.

[14] M. Bouzouraa and U. Hofmann, "Fusion of Occupancy Grid Mapping and Model Based Object Tracking for Driver Assistance Systems using Laser and Radar Sensors," in *Proc. IEEE Intell. Veh. Symp.*, 2010, pp. 294–300.

[15] K. Schueler, T. Weiherer, E. Bouzouraa, and U. Hofmann, "360 Degree multi sensor fusion for static and dynamic obstacles," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 692–697.

[16] R. Jungnickel and F. Korf, "Object tracking and dynamic estimation on evidential grids," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2014, pp. 2310–2316.

[17] C. Coué, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere, "Bayesian Occupancy Filtering for Multitarget Tracking: An Automotive Application," *Int. J. Robot. Res.*, vol. 25, no. 1, pp. 19–30, Jan. 2006.

[18] M. K. Tay *et al.*, "The Bayesian occupation filter," in *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*. Springer, 2008, pp. 77–98.

[19] R. Danescu, F. Oniga, and S. Nedevschi, "Modeling and Tracking the Driving Environment With a Particle-Based Occupancy Grid," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1331–1342, Dec. 2011.

[20] G. Tanzmeister and D. Wollherr, "Evidential Grid-Based Tracking and Mapping," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 6, pp. 1454–1467, Jun. 2017.

[21] A. Nègre, L. Rummelhard, and C. Laugier, "Hybrid Sampling Bayesian Occupancy Filter," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 1307–1312.

[22] D. Nuss *et al.*, "A random finite set approach for dynamic occupancy grid maps with real-time application," *Int. J. Robot. Res.*, vol. 37, no. 8, pp. 841–866, Jul. 2018.

[23] A. Vatavu *et al.*, "Environment Estimation with Dynamic Grid Maps and Self-Localizing Tracklets," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2018, pp. 3370–3377.

[24] T. Yuan, J. Peukert, B. Duraisamy, M. Maile, and A. Gern, "Object tracking with de-autocorrelation scheme for a dynamic occupancy gridmap system," in *Proc. IEEE Int. Conf. Multisensor Fusion and Integration for Intell. Syst.*, 2016, pp. 603–608.

[25] F. Piewak, T. Rehfeld, M. Weber, and J. M. Zöllner, "Fully Convolutional Neural Networks for Dynamic Object Detection in Grid Maps," in *Proc. IEEE Intell. Veh. Symp.*, 2017, pp. 392–398.

[26] S. Hoermann, M. Bach, and K. Dietmayer, "Dynamic occupancy grid prediction for urban autonomous driving: A deep learning approach with fully automatic labeling," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2018, pp. 2056–2063.

[27] S. Wirges, T. Fischer, C. Stiller, and J. B. Frias, "Object Detection and Classification in Occupancy Grid Maps Using Deep Convolutional Networks," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2018, pp. 3530–3535.

[28] D. Stumper, F. Gies, S. Hoermann, and K. Dietmayer, "Offline Object Extraction from Dynamic Occupancy Grid Map Sequences," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2018, pp. 389–396.

[29] N. Engel, S. Hoermann, P. Henzler, and K. Dietmayer, "Deep Object Tracking on Dynamic Occupancy Grid Maps Using RNNs," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2018, pp. 3852–3858.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TITS.2019.2921248

[30] L. Wang, W. Ouyang, X. Wang, and H. Lu, "Visual Tracking with Fully Convolutional Networks," in *Proc. IEEE Int. Conf. Comput. Vision*, 2015, pp. 3119–3127.
[31] H. Nam and B. Han, "Learning Multi-domain Convolutional Neural Networks for Visual Tracking," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2016, pp. 4293–4302.
[32] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vision*, 2016, pp. 850–865.
[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[34] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[35] A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Auton. Robots*, vol. 26, no. 2-3, pp. 123–139, Apr. 2009.
[36] J. Aue, M. R. Schmid, T. Graf, and J. Effertz, "Improved object tracking from detailed shape estimation using object local grid maps with stereo," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2013, pp. 330–335.
[37] M. Schütz, N. Appenrodt, J. Dickmann, and K. Dietmayer, "Occupancy grid map-based extended object tracking," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 1205–1210.
[38] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Instantaneous full-motion estimation of arbitrary objects using dual Doppler radar," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 324–329.
[39] D. Kellner, M. Barjenbruch, J. Klappstein, J. Dickmann, and K. Dietmayer, "Tracking of Extended Objects with High-Resolution Doppler Radar," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 5, pp. 1341–1353, May 2016.
[40] J. Schlichenmaier, L. Yan, M. Stolz, and C. Waldschmidt, "Instantaneous Actual Motion Estimation with a Single High-Resolution Radar Sensor," in *Proc. IEEE Int. Conf. Microwaves for Intell. Mobility*, 2018, pp. 1–4.
[41] K. Thormann, J. Honer, and M. Baum, "Extended Target Tracking Using Gaussian Processes with High-Resolution Automotive Radar," in *Proc. Int. Conf. Inform. Fusion*, 2018, pp. 1764–1770.
[42] A. P. Dempster, "Upper and Lower Probabilities Induced by a Multi-valued Mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, Apr. 1967.
[43] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[44] K. C. J. Dietmayer, S. Reuter, and D. Nuss, "Representation of Fused Environment Data," in *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort*, H. Winner, S. Hakuli, F. Lotz, and C. Singer, Eds. Cham, Switzerland: Springer Int. Publishing, 2016, pp. 567–603.
[45] G. Tanzmeister, "Grid-Based Environment Estimation for Local Autonomous Vehicle Navigation," Ph.D. Dissertation, Dept. Elect. Comput. Eng., Technical University of Munich, Munich, Germany, 2016.
[46] G. Tanzmeister and S. Steyer, "Spatiotemporal alignment for low-level asynchronous data fusion with radar sensors in grid-based tracking and mapping," in *Proc. IEEE Int. Conf. Multisensor Fusion and Integration for Intell. Syst.*, 2016, pp. 231–237.
[47] D. Nuss *et al.*, "Fusion of laser and radar sensor data with a sequential Monte Carlo Bayesian occupancy filter," in *Proc. IEEE Intell. Veh. Symp.*, 2015, pp. 1074–1081.
[48] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
[49] R. Schubert, E. Richter, and G. Wanielik, "Comparison and evaluation of advanced motion models for vehicle tracking," in *Proc. Int. Conf. Inform. Fusion*, 2008, pp. 1–6.
[50] X. R. Li and V. Jilkov, "Survey of Maneuvering Target Tracking. Part I: Dynamic Models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1333–1364, Oct. 2003.
[51] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, 2004.
[52] M. Aeberhard, "Object-Level Fusion for Surround Environment Perception in Automated Driving Applications," Ph.D. dissertation, Technische Universität Dortmund, 2017.
[53] "GNU Scientific Library – Reference Manual: Weighted Samples," http://gnu.org/software/gsl/manual/html_node/Weighted-Samples.html, retrieved 2018-10-02.
[54] S. Thrun, W. Burgard, and D. Fox, *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. Cambridge, MA, USA: The MIT Press, 2005.

Sascha Steyer received the B.Sc. and M.Sc. degrees in electrical engineering and information technology from Karlsruhe Institute of Technology, Karlsruhe, Germany, in 2012 and 2015, respectively. He is currently pursuing his Dr.-Ing. degree at Technical University of Munich, Munich, Germany, in conjunction with the BMW Group. His research interests include autonomous vehicles, sensor data fusion, and tracking, with a focus on the grid-based estimation of dynamic environments.

Christian Lenk received the B.Sc. and M.Sc. degrees in electrical and information technology from Technical University of Munich, Munich, Germany, in 2015 and 2019, respectively. Since 2019, he has been a Research Engineer for automated driving with the BMW Group. His research interests include autonomous vehicles, sensor data fusion, and object tracking, with a focus on grid-based environment estimation and motion state estimation of extended objects.

Dominik Kellner received the Dipl.-Ing. degree in mechatronics and information technology from Technical University of Munich, Munich, Germany, in 2010, and the Dr.-Ing. degree in 2017 from Ulm University, Ulm, Germany. In 2015, he joined Autoliv in Dachau, Germany, as an algorithm developer in the field of high-level sensor data fusion and radar tracking. Since 2017, he has been a Research Engineer with the BMW Group. His research interests include environmental perception with a focus on motion state estimation of extended objects, low- and high-level sensor data fusion and automotive radar.

Georg Tanzmeister received the B.Sc. degree in computer science in 2009 from Technische Universität Wien, Vienna, Austria, the Dipl.-Ing. degree in computer graphics and computer vision in 2011 from Technische Universität Wien, and the Dr.-Ing. degree in electrical engineering in 2016 from Technische Universität München, Munich, Germany. Since 2012, he has been a Research Engineer for automated driving with the BMW Group. His research interests include autonomous vehicles, grid-based environment estimation, sensor data fusion, and motion planning. In 2016, he was the recipient of the dissertation award from VDI München and the dissertation award from VDE Südbayern.

Dirk Wollherr received the Dipl.-Ing. degree in electrical engineering in 2000, the Dr.-Ing. degree in electrical engineering in 2005, and the Habilitation degree in 2013 from Technische Universität München, Munich, Germany. From 2001 to 2004, he was a Research Assistant with the Control Systems Group, Technische Universität Berlin, Berlin, Germany. In 2004, he was a Research Fellow with the Japanese Society for the Promotion of Science, Yoshihiko–Nakamura-Lab, University of Tokyo, Tokyo, Japan. From 2006 to 2008, he was the General Manager with the Cluster of Excellence Cognition for Technical Systems (CoTeSys), where he has been a Principal Investigator, the Independent Junior Research Group Leader, and a Research Area Leader since 2005. He was a TUM Carl-von-Linde Junior Fellow with the Institute of Advanced Studies from 2010 to 2013. He has been active in dissemination, such as the General Chair of the German Robotik 2008 Conference, and the Finance Chair of RO-MAN 2008, the Program Co-Chair of the Workshop on Advanced Robotics and its Social Impacts 2013, and also in several EU projects, such as Robot@CWE, CyberWalk, Movement, and IURO. His research interests include automatic control, robotics, autonomous mobile robots, human robot interaction, and humanoid walking.

Copyright (c) 2019 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.