Object Tracking Based on Evidential Dynamic Occupancy Grids in Urban Environments

Sascha Steyer, Georg Tanzmeister, and Dirk Wollherr

Abstract— Occupancy grid mapping approaches, especially those that additionally estimate the dynamics, enable a robust and consistent modeling of the local environment in a cell-level representation. But a scene understanding of surrounding traffic participants requires a generalized object-level representation. This work presents an object tracking approach based on dynamic occupancy grids. The association of occupied grid cells with existing object tracks is solved individually on the cell-level without clustering or forming object hypotheses. New object tracks are extracted using a clustering strategy and a velocity variance analysis of neighboring occupied cells to reduce false positives. In order to improve the estimates of the position and size, an object boundary extraction is presented that takes the surrounding free space of the selected box representation into account. Experimental results with real sensor data show the effectiveness of the proposed object tracking approach in challenging urban scenarios with dense traffic.

I. INTRODUCTION

Estimating the dynamic environment by processing measurement data of multiple heterogeneous sensors is an essential part of autonomous driving. Object tracking has been a research topic for decades, but, since the scenarios to be mastered are becoming more complex and the requirements of the accuracy are rising, it remains a challenge. In addition to high-level fusion approaches of sensor-individual object tracking algorithms, grid-based methods are becoming more popular for modeling the local environment. Occupancy grid mapping [1] has been successfully adapted to deal with dynamic environments [2]–[5], typically referred to as dynamic occupancy grids. As a result, a consistent model of the environment in a grid representation including static and dynamic obstacles as well as free space is generated. Due to the spatial cell discretization and assumed independence of grid cells, no specific geometrical object models or associations are required.

A scene understanding of surrounding traffic participants requires a generalized spatial representation of objects. Based on a dynamic occupancy grid, the key challenges to obtain a robust tracking on the object-level are the determination of the set of measured occupied grid cells that describe one physical object and the correct association of these cells to the same object track over time.

Vu et al. [6], [7] classify sensor measurements as dynamic by considering inconsistencies of the local occupancy grid. Object hypotheses using a box model are formed by clustering dynamic measurements with a defined maximum

S. Steyer and G. Tanzmeister are with BMW Group, Munich, Germany. E-mail: sascha.steyer@bmw.de, georg.tanzmeister@bmw.de
D. Wollherr is with the Institute of Automatic Control Engineering (LSR), Technische Universität München, Munich, Germany. E-mail: dw@tum.de

distance. Object tracks are filtered using multiple hypotheses tracking (MHT), considering the m-best association hypotheses of the object hypotheses and tracks. Object hypotheses are extracted similarly in [8] and associated to the tracks using global nearest neighbor (GNN). In addition, the object tracking and the occupancy grid mapping are linked to improve each other by extending the object states by a list of associated and accumulated grid cells. In [9] cell clusters are extracted from dynamic grid cells using density-based clustering. Object tracking is performed using particle filters. Each particle consists, among others, of a set of clustered cells that allow free form object models. The object tracks are associated to the extracted clusters using joint probability data association (JPDA).

More generally, the problem of extended object tracking [10] is often solved by selecting features, forming object hypotheses, and performing the association on the object-level. But the information processing of the measurements before the association with the existing object tracks is performed can be critical in dense urban scenarios. For example, the preselection of dynamic cells would cause to lose track of moving vehicles that temporarily stop at traffic lights as their velocities get zero and they would be classified as static. Likewise, a cell clustering may not be able to distinguish multiple objects close to each other with similar velocities without previous knowledge. Conversely, a single object that is partly occluded may result in multiple clusters. The transformation to a geometrical model, like an oriented rectangle box, may also be critical, e.g., the orientation estimation in terms of the direction of Cartesian velocities is unreliable for slow objects, which directly affects the estimates of the length and width. Moreover, it may be beneficial to differ whether an object hypothesis or measurement is expected due to an already tracked object or if that hypothesis would form a new track to reduce false positives.

In a different way, Mekhnacha et al. [11] use information of predicted object tracks as a feedback to define regions of interest on the cell-level, but their main goal is a more efficient clustering and association. Each region of interest is used to restrict possible starting points of a clustering process on the cell-level, implying an association of the clustered cells to the corresponding track. An additional re-clustering strategy using k-means clustering is proposed since the clustering may still result in ambiguous associations.

In this work, the association of occupied cells and object tracks is solved on the cell-level to overcome the described problems of forming object hypotheses. In contrast to [11], the association is performed for each individual cell without
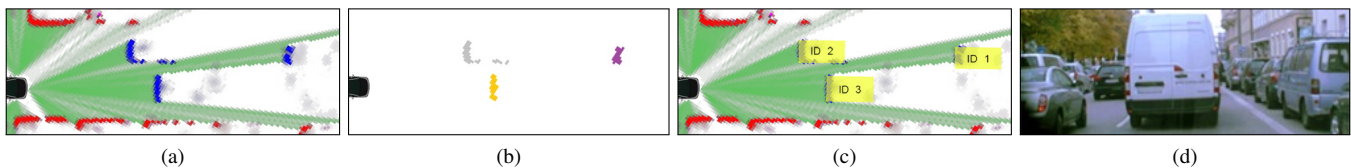
Fig. 1. Visualization of the proposed grid-based object tracking approach. Static obstacles, including parking vehicles, are not extracted. (a) Evidential dynamic occupancy grid. Color coding of evidence masses: red $m(S)$, blue $m(D)$, pink $m(\{S, D\})$, green $m(F)$. (b) Extracted and associated occupied grid cells ($\mathcal{C}_1$: purple, $\mathcal{C}_2$: gray, $\mathcal{C}_3$: orange). (c) Filtered object tracks using a box model representation. (d) Camera image of the scene.

clustering, hence, ambiguous associations are handled implicitly. New object tracks are formed by a clustering strategy that combines a density-based clustering of dynamic grid cells and a velocity consistency analysis of neighboring occupied cells to reduce false positives.

The geometry of an object track is modeled by an oriented box as a bounding box is a suitable conservative approximation of dynamic objects. The measured free space evidence adjacent to each object box is taken into account in addition to the associated occupied grid cells. This allows for a better estimation of the size, the reference position, and the occupancy probability of a track on the cell-level.

The paper is structured as follows: A short overview of the used evidential dynamic occupancy grid and the proposed derivation of object tracks are given in Section II. Section III focuses on the cell-to-track association including a clustering strategy for potential new object tracks. A description of the object state filtering and the feedback of the object tracks to the cell-level for performing the association are given in Section IV. Results are presented in Section V and a conclusion is given in Section VI.

II. OVERVIEW OF GRID-BASED OBJECT TRACKING

In this section a definition of the used evidential dynamic occupancy grid representation and a brief overview of the proposed derivation of filtered object tracks are given.

A. Evidential Dynamic Occupancy Grid

A dynamic occupancy grid denotes an extension of the classical occupancy grid mapping [1] towards dynamic environments. In an evidential occupancy grid, based on the Dempster-Shafer theory [12], [13], the Bayesian occupancy probability $p(O)$ of each grid cell is generalized to an occupancy belief $\mathrm{bel}(O) \leq p(O)$ and a free space belief $\mathrm{bel}(F) \leq 1 - p(O)$. There are many approaches for both dynamic and evidential occupancy grids – the following definition refers to the approach by Tanzmeister et al. [3], [14].

The frame of discernment $\Theta = \{F, S, D\}$ with the hypotheses free space $\{F\}$, statically occupied $\{S\}$, and dynamically occupied $\{D\}$ is introduced in [3]. Since every element of the power set 2^Θ is considered in the Dempster-Shafer theory, the occupancy belief $\mathrm{bel}(O) = \mathrm{bel}(\{S, D\})$ can be subdivided into basic belief masses $m(\cdot)$ for static $m(S)$, dynamic $m(D)$, and unclassified occupied $m(\{S, D\})$:

$$\mathrm{bel}(O) = m(S) + m(D) + m(\{S, D\}). \quad (1)$$

The dynamic state of the occupied environment is estimated using a grid-based particle filter. Each particle has a

continuous 2D position and velocity and can move freely in the entire grid. Input to the particle filter are scan grids, i. e., an evidential grid representation of fused sensor measurements. The particles allow a filtering of both the velocity and occupancy. But in our work, the particle filtering is only used to estimate velocities and subdivide the measured occupancy beliefs. The proposed object tracking uses the evidential dynamic occupancy grids as input measurements and assumes uncorrelated input data. Occupancy filtering is therefore omitted, instead, unaccumulated measured sensor data is used. The velocity estimation is assumed to be uncorrelated.

The following particle evidence derivation is a slight modification of [14] and [3]. The particle population $\mathcal{X}_c$ located in a cell $c \in \mathcal{G}$ of the discrete grid structure $\mathcal{G}$ is divided into subsets of static $\mathcal{X}_c^S$ and dynamic $\mathcal{X}_c^D$ particles by considering the individual particle velocities. However, only particles that fulfill a minimum required age are classified ($\mathcal{X}_c^S \cup \mathcal{X}_c^D \subseteq \mathcal{X}_c$). Based on measured evidences $\mathrm{bel}(O)$ and $\mathrm{bel}(F) = m(F)$ of a scan grid, the occupancy belief masses can be derived for each grid cell $c \in \mathcal{G}$:

$$m(S_c) = \frac{|\mathcal{X}_c^S|}{|\mathcal{X}_c|} \mathrm{bel}(O_c), \quad (2)$$

$$m(D_c) = \left(1 - \frac{|\sigma_\phi|}{\sigma_{\max}}\right) \frac{|\mathcal{X}_c^D|}{|\mathcal{X}_c|} \mathrm{bel}(O_c), \quad (3)$$

$$m(\{S, D\}_c) = \mathrm{bel}(O_c) - m(S_c) - m(D_c). \quad (4)$$

The dynamic mass $m(D_c)$ is reduced by the standard deviation σ_ϕ of the individual particle velocity orientations compared to a maximum $\sigma_{\max}$ [14]. The velocity distribution of a cell c is extracted as a 2D Cartesian mean velocity vector $v_c \in \mathbb{R}^2$ of all particles in $\mathcal{X}_c^S \cup \mathcal{X}_c^D$.

In sum, the evidential dynamic occupancy grid forms for each grid cell $c \in \mathcal{G}$ a measurement

$$z_c = (\, m(S_c),\; m(D_c),\; m(\{S, D\}_c),\; m(F_c),\; v_c\,). \quad (5)$$

Fig. 1a shows an example of the belief masses of an evidential dynamic occupancy grid.

B. Derivation of Spatial Object Tracks

Based on these evidential dynamic occupancy grids, the aim of this work is to develop a robust tracking on the object-level including a description of the spatial extent. This means estimating the object states s_τ of each object track $\tau \in \mathcal{T}$, where the set $\mathcal{T}$ describes all detected objects of the current local environment. The problem of recursively updating the tracks by the cell measurements z_c, assuming the tracks are already predicted to the measurement time stamp, is subdivided into two stages:

1) Cell-to-Track Association: First, the set of cell indices $\mathcal{C}_\tau \subseteq \mathcal{G}$ that belong to each track τ have to be determined, i.e., the direct association of a set of measurement cells $\mathcal{Z}_\tau = \{z_c \,|\, c \in \mathcal{C}_\tau\}$ to a track τ, as shown in Fig. 1b. This also includes creating new object tracks with an initial cluster of measurement cells.

2) Object State Update: Secondly, the state s_τ of each track τ has to be updated. Depending on the appropriate object model representation, i.e., the selected state space, the set of measurement cells $\mathcal{Z}_\tau$ has to be mapped to a corresponding measurement vector. Fig. 1c illustrates filtered object tracks using a box model representation.

These two steps are discussed separately in the following two sections.

III. CELL-TO-TRACK ASSOCIATION

The association of cell measurements z_c with individual object tracks τ can be performed on different levels of abstraction, in particular on the object-level or the cell-level, as illustrated in Fig. 2.

The association on the object-level requires extracting object hypotheses from the cell measurements. As shown in the bottom path of Fig. 2, this in turn typically requires an extraction of dynamically occupied cells, a clustering, and a transformation to the selected object representation. But these information processing steps may fail in urban environments with dense traffic, occlusion, or stopping objects, as indicated in the introduction.

In contrast, the association on the cell-level is performed for each individual cell measurement. This way, the drawbacks of forming object hypotheses for the association are omitted. In this work, this is achieved by an inverse transformation of the spatial object tracks to the cell-level representation. Alternatively, a direct distance criterion between a measurement cell and a track can be defined. However, this merely changes the calculation order of analyzing each cell determined to be occupied and checking if any track is a possible source of this measurement, i.e., whether a cell is inside the gating area of a track.

In this section, the proposed association of measurement cells to existing object tracks, i.e., already tracked objects, and a clustering strategy to form new tracks are presented.

A. Association of Cells to Existing Object Tracks

We associate individual cell measurements directly to existing tracks by a feedback of these spatial object tracks on the cell-level. This basic idea is similar to [11], but our approach associates each individual cell to a track without clustering. An association probability $\alpha(z_c, \tau)$ of a cell measurement z_c and a track τ is calculated which considers the object pose and size as well as the difference in velocity. These are discussed separately in the following subsections. Finally, a cell is associated with at most one track by evaluating the maximum association probability. Hence, in contrast to [11], no additional re-clustering strategy for ambiguous associations is required.
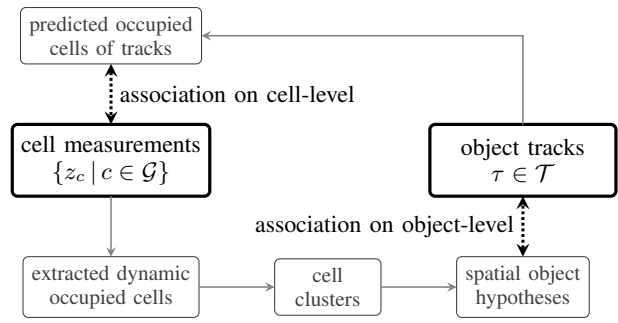
Fig. 2.   Possible associations of cell measurements with object tracks on the object-level or the cell-level.

1) Cell Occupancy Probabilities of a Track: First, each spatial object track is transformed to the cell-level representation. This yields an occupancy association probability $\alpha_o(c, \tau)$ of a cell c and a track τ, that is the probability that c is occupied by τ, similarly described as a track cover probability in [15]. The probability α_o is an approximation of the occupancy probability $p(O_c|s_\tau)$ given the object state s_τ at the measurement time stamp, i.e., the occupancy measurement likelihood.

In sum, cells that can be occupied by the tracked objects regarding their poses and geometrical dimensions are extracted. This allows tracking of stopping objects without considering all occupied measurement cells of the grid. Since the source of an occupied cell may be ambiguous due to multiple objects close to each other or clutter measurements, the corresponding association probability to each possible track is calculated.

Note that $\alpha_o(c, \tau)$ does not depend on the cell measurement z_c, but only on the modeled transformation of the selected object-level representation to the position of the cell c. For example, this can be the occupancy probability of an object local grid map representation [16], a probabilistic model of the pose and geometrical shape, or, in the simplest case, a binary validation gate of the track. A derivation of $\alpha_o(c, \tau)$ for an oriented box model is described in Section IV-C.

2) Velocity Difference of Cell and Track: Next, an association probability $\alpha_v(z_c, \tau)$ regarding the velocities of z_c and τ is calculated. The similarity of the measured cell velocity v_c and estimated track velocity v_τ is analyzed by an unnormalized Gaussian distribution

$$\alpha_v(z_c, \tau) \propto \mathcal{N}(v_c; v_\tau, \Sigma_{\tau,v}), \ \alpha_v \in [0, 1] \qquad (6)$$

with mean v_τ and the covariance matrix $\Sigma_{\tau,v}$ of the object state s_τ evaluated at v_c.

3) Joint Association Probability: The association probability $\alpha(z_c, \tau)$, considering both the occupancy probability of c by τ and the velocity difference, is combined to

$$\alpha(z_c, \tau) = \alpha_o(c, \tau) \ (\lambda_v \alpha_v(z_c, \tau) + 1 - \lambda_v). \qquad (7)$$

An additional weighting coefficient $\lambda_v \in [0, 1]$ is introduced, which is a design parameter that regulates the impact of the velocity association probability α_v to the total association probability α. In that way, incorrect velocity estimates or measurements can be handled, i.e., when an occupied cell

must be of that track as to the spatial extent and small prediction time since the last measurement. Only the maximum association probability $\alpha(z_c, \tau^*)$ of z_c to the best fitting track

$$\tau^* = \arg\max_{\tau \in \mathcal{T}} \alpha(z_c, \tau) \qquad (8)$$

is considered, meaning that a nearest neighbor approach for each individual cell is performed. To finally associate z_c to τ^*, the association probability has to fulfill the inequation

$$\alpha(z_c, \tau^*)\, \mathrm{bel}(O_c) \geq \alpha_{\min} \in [0,1] \qquad (9)$$

with a minimum required association probability $\alpha_{\min}$. The minimum required measurement occupancy belief and minimum required association probability depend on each other, so that a lower association probability requires a higher certainty that a cell is occupied in order to be extracted and vice versa. In sum, the set $\mathcal{Z}_\tau = \{z_c \,|\, c \in \mathcal{C}_\tau\}$ with

$$\mathcal{C}_\tau = \Big\{ c \in \mathcal{G} \;\big|\; \alpha(z_c, \tau)\, \mathrm{bel}(O_c) \geq \alpha_{\min} \;\wedge$$
$$\alpha(z_c, \tau) > \max_{\tau' \in \mathcal{T}\setminus\{\tau\}} \alpha(z_c, \tau') \Big\} \qquad (10)$$

is associated to τ. However, only cells that fulfill

$$\mathrm{bel}(O_c) \geq \alpha_{\min} \;\wedge\; \exists \tau \in \mathcal{T} : \alpha_o(c, \tau) \geq \alpha_{\min} \qquad (11)$$

are able to be associated and are considered.

B. Cell Clustering of Potential New Object Tracks

The unassociated measurement cells of the remaining grid set

$$\mathcal{G}' = \mathcal{G} \setminus \bigcup_{\tau \in \mathcal{T}} \mathcal{C}_\tau \qquad (12)$$

have to be analyzed for potential new object tracks. The proposed approach is divided into four steps: An extraction of dynamic cells, two consecutive cell clustering steps, and finally a velocity analysis to decide whether a cluster forms a valid new track or not.

1) Extraction of Dynamic Cells: Only measurement cells of $\mathcal{G}'$ where the measured dynamic occupancy mass $m(D_c)$ is above a defined minimum threshold $m_{D,\min}$ are extracted, i. e., the set $\{z_c | c \in \mathcal{G}'_D\}$ with

$$\mathcal{G}'_D = \{c \in \mathcal{G}' \mid m(D_c) \geq m_{D,\min}\} . \qquad (13)$$

As the evidence derivation of $m(D)$ in turn depends on a defined velocity threshold, new objects are only extracted when they are moving with a velocity above that threshold.

2) Clustering of Dynamic Cells: Cell clusters $\mathcal{C}_k \subseteq \mathcal{G}'_D$, where each index k describes a potential new object, of the extracted dynamic cells are formed using a density-based clustering. In this work, the DBSCAN algorithm [17] is used with the ε-neighborhood

$$N_c^\varepsilon = \Big\{ c' \in \mathcal{G}'_D \;\big|\; \|x_c - x_{c'}\|_2 \leq \varepsilon_d \;\wedge$$
$$\|v_c - v_{c'}\|_2 \leq \varepsilon_v \;\wedge \sum_{\tilde{c} \in \mathcal{D}(c,c')} m(F_{\tilde{c}}) \leq \varepsilon_f \Big\} . \qquad (14)$$

of a cell c. It describes the set of cells c' that are neighbors to the cell c regarding a defined ε-distance which is divided into three conditions here. In addition to a maximum position
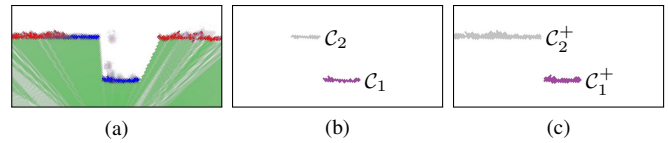

Fig. 3. Incorrect cell velocity estimation at a guardrail. Both clusters $\mathcal{C}_1$ (purple) and $\mathcal{C}_2$ (gray) of dynamic occupied cells represent potential objects moving right. But only $\mathcal{C}_1$ describes a real vehicle, $\mathcal{C}_2$ is a false positive. (a) Evidential dynamic occupancy grid. (b) Density-based clustering of dynamic occupied cells. (c) Region growing of neighboring occupied cells.

distance ε_d of the 2D cell centers $x_c \in \mathbb{R}^2$ and a maximum velocity difference ε_v, the sum of the in-between free space evidence masses $m(F)$ is taken into account, where $\mathcal{D}(c, c')$ defines the set of in-between cells of c and c'. A simple and conservative choice for $\mathcal{D}$ is the rectangular area between two cells, which can be implemented as a summed area table [18]. Alternatively, cells of the connected line can be extracted by Bresenham's line algorithm [19]. By considering the available in-between free space and the estimated velocities, objects close to each other can be better separated in comparison to a direct clustering of sensor measurement data. Further ε_d can be enlarged if measurements occur on different parts of one object.

3) Consideration of Neighboring Occupied Cells: Occupied grid cells next to the density-based dynamic clusters $\mathcal{C}_k$ are also considered, because the cell velocities and thus the dynamic masses of the evidential dynamic occupancy grid may be incorrect. For example, straight guardrails are prone for incorrect velocity estimations on a cell-level, especially when parts are dynamically occluded. However, apart from boundary regions of newly detected occupancy, the evidential dynamic occupancy grid still converges correctly, as shown in Fig. 3a. This behavior is used to reduce false positives of the dynamic clusters $\mathcal{C}_k$.

As to that, each $\mathcal{C}_k$ is expanded to a cluster $\mathcal{C}_k^+ \supseteq \mathcal{C}_k$ by iteratively adding all neighboring unassociated and unclustered occupied cells in a seeded region growing [20] with $\mathcal{C}_k$ as the starting seed points, cf. Fig. 3b and Fig. 3c. The maximum number of iterations can be limited in terms of efficient implementation, since the main motivation is to reduce false positives. The connectivity criterion is defined as the 8-connected neighborhood with a defined minimum occupancy belief $\mathrm{bel}_{\min}(O)$. Velocity differences are ignored in this connectivity-based clustering. Instead, the velocity variance of the new cluster is analyzed.

4) Velocity Variance Analysis: If any cell has been added to the initial cluster ($\mathcal{C}_k^+ \neq \mathcal{C}_k$), the velocity variance

$$\sigma_{v_k}^2 = \frac{1}{|\mathcal{C}_k^+|} \sum_{c \in \mathcal{C}_k^+} \|v_c - \bar{v}_k\|_2^2 \qquad (15)$$

of $\mathcal{C}_k^+$ with the weighted mean velocity

$$\bar{v}_k = \frac{\sum\limits_{c \in \mathcal{C}_k} (m(D_c) + m(S_c))\, v_c}{\sum\limits_{c \in \mathcal{C}_k} (m(D_c) + m(S_c))} \qquad (16)$$

is calculated. Note that $\bar{v}_k$ is formed only by using the cells of $\mathcal{C}_k$, since the variance is related to the initial dynamic

cluster. Valid clusters that form new tracks have to fulfill the condition $\sigma_{v_k}^2 \leq \sigma_{v,\max}^2$. Alternatively, an initial existence probability could be derived from the velocity variance.

In sum, new tracks are extracted, when there are dense clusters of unassociated dynamic occupied cells with similar velocities. Furthermore, these clusters must not be directly next to static obstacles with a large velocity variance, assuming that the grid resolution and measurement certainty is generally high enough to model gaps between occupied cells of different sources. However, this does not affect already tracked objects, as the association of measurement cells is handled without forming cell clusters.

IV. Object Tracking Using a Box Model

This section gives details about the object state model and describes the object state update process given the set of associated cell measurements. Furthermore, the transformation of a track to the cell-level representation as required by the proposed cell-to-track association is explained.

A. Object State Update and Measurement Model

Each object track $\tau \in \mathcal{T}$ is described by its state

$$s_\tau = (\, x_\tau,\, y_\tau,\, v_\tau,\, \phi_\tau,\, a_\tau,\, \omega_\tau,\, l_\tau,\, w_\tau\,) \in \mathbb{R}^8, \quad (17)$$

containing the 2D position (x, y), scalar velocity v, orientation ϕ, acceleration a, and turn rate ω, and a corresponding covariance matrix $\Sigma_\tau \in \mathbb{R}^{8 \times 8}$. A constant turn rate and acceleration (CTRA) motion model is used as the state transition model to represent clothoid trajectories [21]. The object size is described by an oriented box model with length l and width w. Since the selected motion model is nonlinear, the full track state s_τ is filtered by an unscented Kalman filter (UKF) [22].

The set of measurement cells $\mathcal{Z}_\tau = \{z_c \,|\, c \in \mathcal{C}_\tau\}$ associated to τ has to be mapped to a measurement vector

$$z_\tau = (\, x_\tau^z,\, y_\tau^z,\, v_\tau^z,\, \phi_\tau^z,\, l_\tau^z,\, w_\tau^z\,) \in \mathbb{R}^6 \,. \quad (18)$$

The velocity magnitude v_τ^z and orientation ϕ_τ^z are determined from the weighted mean of the associated 2D Cartesian cell velocities v_c as calculated in (16). Since ϕ_τ^z is unreliable for slow or stopped objects, its measurement uncertainty is, in addition to the velocity variance, inversely proportional to v_τ^z. The orientation ϕ also reflects the length axis of the object. Hence, it directly affects l_τ^z and w_τ^z, as they are estimated regarding the minimum and maximum extracted cell position along the corresponding axes. To prevent an uncertain orientation estimate ϕ_τ^z from distorting l_τ^z and w_τ^z, the orientation estimate of the track is updated first and then the filtered track orientation ϕ_τ is used for extracting l_τ^z and w_τ^z of the current measurement.

B. Object Boundary Extraction and Reference Point

Besides the associated occupied cells, the evidential dynamic occupancy grid also provides free space evidence masses $m(F)$, which can be used for the object tracking as well. In [23] the measurement likelihood of virtual rays of laser range scanners are computed by differentiating

geometric regions of occlusion, free space around the object, the object's visible surface, and rays that passed through the object. The basic idea is that an object has to be surrounded by free space, which is used in the following.

The extracted length l_τ^z and width w_τ^z of the measurement z_τ represents a minimum bounding box. For each edge of that box the adjacent free space evidences are used to analyze whether l_τ^z and w_τ^z form true object boundaries or if they are only lower bounds due to the sensor field of view or occlusion.

The belief that an edge of the box is a visible boundary $b \in \{\text{front}, \text{rear}, \text{left}, \text{right}\}$ is calculated as

$$\text{bel}(b) = \frac{1}{|\mathcal{B}_b|} \sum_{c \in \mathcal{B}_b} m(F_c)\,, \quad (19)$$

where $\mathcal{B}_b$ defines the set of underlying cells of the area around the visible boundary b that is expected to be free space. In this work, an outside rectangle area is used, as shown in Fig. 4a. This means, for example, w_τ^z does not only represent a lower bound, but an estimate of the actual object width w, when both boundaries $b_w \in \{\text{left}, \text{right}\}$ have a high belief, as shown by the bottom object in Fig. 4a.

The measurement estimates of an object's size are changing with occlusion or the current sensor field of view. This also means that the center of each minimum bounding box may vary even if the object does not move. Thus, a robust estimate of (x_τ, y_τ) requires a reference point which does not have to depend on l_τ^z or w_τ^z. In this work, the reference point can be either at a corner, an edge center, or the center of the box, as proposed in [24]. The object boundary extraction is used to select the most robust reference point in terms of the visible boundaries. For example, if two orthogonal boundaries are visible, the corresponding corner point is selected, whereas when multiple corner points are available, the in-between edge center is selected, as shown in Fig. 4a.

C. Predicted Cell Occupancy Association Probability

As introduced in Section III-A.1, an occupancy association probability $\alpha_o(c, \tau)$ of the object tracks and underlying cells is required to perform the proposed association on the cell-level. As mentioned before, α_o represents an approximation of the occupancy measurement likelihood. Thus, for each track τ predicted to the time stamp of the incoming measurement, the probability that a cell $c \in \mathcal{G}_\tau$, inside an allowed gating area $\mathcal{G}_\tau \subseteq \mathcal{G}$ of τ, is occupied by that track is calculated. This probability is modeled as the spatial distance of the center of a cell to a track. Hereto, the distances along the length and width axis of the box model are treated separately, assuming independence of both. In the following, the object coordinate system $(\tilde{x}, \tilde{y})$ is used, which represents the length and width axis. The origin is located at the center of the box, the coordinates $\tilde{x}$ and $\tilde{y}$ point to the front and left edge respectively.

The association probability regarding the position distance $d_c^{\tilde{x}}$ of a cell c to the object origin along the $\tilde{x}$-axis is approximated by an unnormalized zero-mean Gaussian distribution with variance $\sigma_{x,l}^2$, modeling the spatial extent
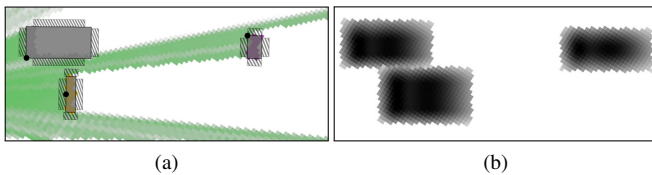
(a)   (b)

Fig. 4.   (a) Measurement minimum bounding boxes (solid rectangles), reference points (black dots), and analyzed free space regions $\mathcal{B}_b$ (around striped areas) of the box edge boundaries b. (b) Predicted cell occupancy association probabilities $\alpha_o(c, \tau)$. Both figures show the scenario of Fig. 1.

and uncertainty of the position, evaluated at $d_c^{\tilde{x}}$. In addition, the analyzed object boundaries b are taken into account to achieve a more accurate measurement likelihood of the currently observable object edges. This results in a Gaussian mixture of the distances to the center and the extracted boundaries, weighted by κ_c and κ_b respectively:

$$\alpha_o^{\tilde{x}}(c, \tau) = \kappa_c \exp\left(\frac{-(d_c^{\tilde{x}})^2}{2\sigma_{x,l}^2}\right) +$$
$$\kappa_b \sum_{b_l \in \{\text{front,rear}\}} \text{bel}(\bar{b}_l) \exp\left(\frac{-(d_c^{\tilde{x}} \mp \frac{1}{2}l_\tau)^2}{2\sigma_b^2}\right). \quad (20)$$

The impact of a boundary b is scaled by a defined variance σ_b^2. Each boundary is weighted by the estimated boundary belief $\text{bel}(\bar{b})$. Here, a moving average of the boundary beliefs $\text{bel}(b)$ of the last n measurements is used, since the sensors that contributed to the currently fused measured free space belief $m(F)$ may be not consistently.

Finally, the occupancy association probability is given by

$$\alpha_o(c, \tau) = \alpha_o^{\tilde{x}}(c, \tau)\ \alpha_o^{\tilde{y}}(c, \tau), \quad (21)$$

while $\alpha_o^{\tilde{y}}$ is calculated along the $\tilde{y}$-axis equivalently to $\alpha_o^{\tilde{x}}$.

Fig. 4b exemplarily shows occupancy association probabilities of three objects predicted to the measurement time stamp as used in the scenario of Fig. 1.

V. EXPERIMENTAL RESULTS

The proposed grid-based object tracking approach was tested with real sensor data of a test vehicle in various urban and highway scenarios. The test vehicle was equipped with four multi-layer laser scanners, one at each side of the vehicle, and four short range radar sensors, one at each corner. The implementation, including the calculation of the dynamic evidential occupancy grid, is primarily based on parallel GPU computing and processes the fused sensor data with an input rate of about $25\,\text{Hz}$ in real time on a medium development computer. The cell size of the grid was set to $0.2\,\text{m} \times 0.2\,\text{m}$. In the following, the successful application of the proposed approach is illustrated qualitatively by means of two different challenging urban traffic scenarios.

The first analyzed scenario is shown in Fig. 5 where the focus lies on two cyclists that are approaching very close to each other until the measured occupied grid cells are directly adjacent. Consequently, a clustering on the cell-level cannot distinguish the two objects, neither by the position nor the velocity, without using prior knowledge of the objects. The approach, however, successfully associates the individual
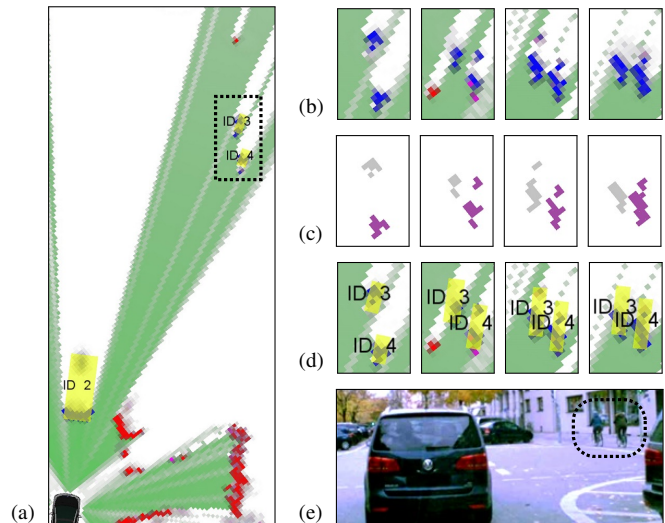

Fig. 5.   Scenario with two cyclists (dotted area) that are approaching very close to each other until the occupied grid cells are directly adjacent. (a) Scene at starting time. (b) Sequence of evidential dynamic occupancy grids of the two cyclists. (c) Extracted and associated occupied grid cells ($\mathcal{C}_3$: gray, $\mathcal{C}_4$: purple). (d) Resulting object tracks. (e) Camera image.

cell measurements to the corresponding cyclists. Thus, the proposed association allows to individually keep track of closely spaced objects when they were previously extracted as separate objects.

Another challenging scenario is shown in Fig. 6. In this scene, the ego vehicle tries to turn left, stops at an intersection, and waits for the oncoming traffic to pass. A vehicle following the ego vehicle also has to stop. This urban traffic situation involves many different challenges, including:

- Stop-and-go traffic for almost all vehicles, resulting in multiple evidence derivation changes of static and dynamic evidence masses as well as a mixture of both.
- Slow and stopped vehicles that do not allow an orientation estimation given 2D Cartesian cell velocities.
- Partly occluded objects due to a restricted sensor field of view, resulting in large gaps of the occupied cells belonging to one object or varying minimum bounding box estimates.

The proposed association on the cell-level considers both statically and dynamically occupied cells. Larger gaps between occupied cells of one object do not have a negative effect because no cell clustering is performed for tracked objects. Thus, it is kept track of all objects without losing stopped vehicles or splitting partly occluded objects, cf. Fig. 6b and Fig. 6c. The object size is estimated robustly, since the orientation is updated separately with regard to the velocity magnitude and, moreover, the adjacent free space evidences are taken into account.

VI. CONCLUSION

Evidential dynamic occupancy grids provide a robust sensor data fusion and dynamic estimation on the cell-level. This paper has presented a novel object tracking approach based on that grid representation. Occupied grid cells are associated individually to object tracks using a feedback of
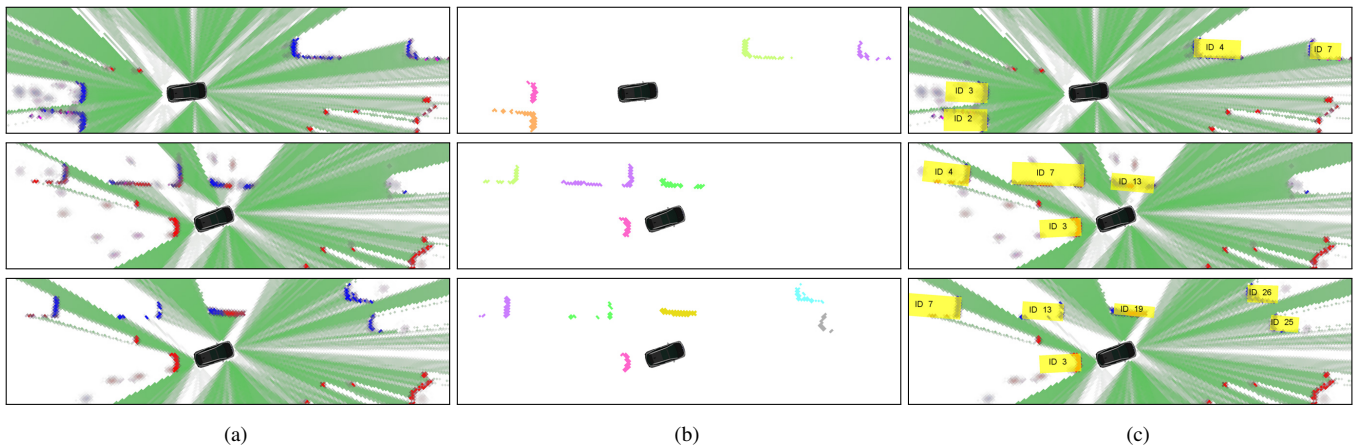
| (a) | (b) | (c) |

Fig. 6. Complex traffic situation at an intersection. Each row represents one time instance ($t_1 = 2.4\,\text{s}, t_2 = 9.6\,\text{s}, t_3 = 19.1\,\text{s}$). The ego vehicle is oriented to the north-east and tries to turn left into a road in the north. It stops at an intersection and waits for the oncoming traffic (*IDs 4, 7, 13, 19, 25, 26*) to pass, which also is in stop-and-go traffic. The vehicle following the ego vehicle (*ID 3*) has to stop as well. (a) Evidential dynamic occupancy grids. (b) Extracted and associated occupied cells. (c) Filtered object tracks.

these tracks to the cell-level representation. In contrast to an association on the object-level, no clustering or object hypothesis forming is required for already tracked objects. Results with real sensor data have shown that this approach is able to keep track of objects that may be close to each other, be partly occluded, or consist of occupied grid cells classified as static. The free space information modeled in the occupancy grid is taken into account to improve the estimation of the position and size of each object. The robustness of the extraction of new objects depends on the accuracy of the velocity estimation and thus the static/dynamic classification on the cell-level. However, false positives have been significantly reduced using a combination of a density- and a connectivity-based clustering with a velocity variance analysis of adjacent occupied cells. Future work will focus on improving the dynamic estimation on the cell-level, combining the object tracking approach with the underlying grid-based particle filter tracking, and a quantitative evaluation.

REFERENCES

[1] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *Computer*, vol. 22, no. 6, pp. 46–57, Jun. 1989.

[2] 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.

[3] G. Tanzmeister and D. Wollherr, "Evidential grid-based tracking and mapping," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–14, 2016, doi:10.1109/TITS.2016.2608919.

[4] D. Nuss, T. Yuan, G. Krehl, M. Stuebler, S. Reuter, and K. Dietmayer, "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.

[5] A. Nègre, L. Rummelhard, and C. Laugier, "Hybrid sampling bayesian occupancy filter," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 1307–1312.

[6] 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.

[7] 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.

[8] 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.

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

[10] K. Granström and M. Baum, "Extended object tracking: Introduction, overview and applications," *CoRR*, vol. abs/1604.00970, 2016. [Online]. Available: http://arxiv.org/abs/1604.00970

[11] K. Mekhnacha, Y. Mao, D. Raulo, and C. Laugier, "The 'fast clustering-tracking' algorithm in the bayesian occupancy filter framework," in *Proc. IEEE Int. Conf. Multisensor Fusion and Integration for Intell. Syst.*, 2008, pp. 238–245.

[12] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *Ann. Math. Statist.*, vol. 38, no. 2, pp. 325–339, April 1967.

[13] G. Shafer, *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[14] G. Tanzmeister, J. Thomas, D. Wollherr, and M. Buss, "Grid-based mapping and tracking in dynamic environments using a uniform evidential environment representation," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2014, pp. 6090–6095.

[15] D. Nuss, M. Stuebler, and K. Dietmayer, "Consistent environmental modeling by use of occupancy grid maps, digital road maps, and multi-object tracking," in *Proc. IEEE Intell. Veh. Symp.*, 2014, pp. 1371–1377.

[16] 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.

[17] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." AAAI Press, 1996, pp. 226–231.

[18] F. C. Crow, "Summed-area tables for texture mapping," in *Proc. Comput. Graph. and Interactive Techniques Conf.*, ser. SIGGRAPH '84, 1984, pp. 207–212.

[19] J. E. Bresenham, "Algorithm for computer control of a digital plotter," *IBM Syst. J.*, vol. 4, no. 1, pp. 25–30, Mar. 1965.

[20] R. Adams and L. Bischof, "Seeded region growing," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 6, pp. 641–647, Jun 1994.

[21] 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.

[22] S. J. Julier and J. K. Uhlmann, "Unscented filtering and nonlinear estimation," *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar 2004.

[23] A. Petrovskaya and S. Thrun, "Model based vehicle detection and tracking for autonomous urban driving," *Autonomous Robots*, vol. 26, no. 2, pp. 123–139, 2009.

[24] 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.