Abstract—Autonomously driving vehicles require a complete and robust perception of the local environment. A main challenge is to perceive any other road users, where multi-object tracking or occupancy grid maps are commonly used. The presented approach combines both methods to compensate false positives and receive a complementary environment perception. Therefore, an environment perception framework is introduced that defines a common representation, extracts objects from a dynamic occupancy grid map and fuses them with tracks of a Labeled Multi-Bernoulli filter. Finally, a confidence value is developed, that validates object estimates using different constraints regarding physical possibilities, method specific characteristics and contextual information from a digital map. Experimental results with real world data highlight the robustness and significance of the presented fusing approach, utilizing the confidence value in rural and urban scenarios.

I. INTRODUCTION

A major challenge for autonomously driving vehicles is a complete and robust perception of the surroundings. For this reason, the vehicle is equipped with a large variety of sensors that generate a dense and precise depiction of the local environment. The main objective of these sensor measurements is to detect and track any other road users. Algorithms using temporal filtering of the sensor measurements are commonly used and are a well studied topic [1]. These object tracking approaches apply object-model-based prior information and use Bayesian filtering techniques to suppress uncertainties and clutter. Further, multi-object tracking approaches [2] are able to detect and estimate the state of multiple objects at once. Here, measurements and tracks have to be associated and assigned. To overcome explicit associations, the usage of a Random Finite Set (RFS) for probabilistic estimations of the objects states and the sets cardinality, showed impressive results [2], [3]. Despite the great success, multi-object tracking in urban scenarios with a large variety and amount of traffic participants is still a tough challenge. A second approach for an environment perception are occupancy grid maps [4], [5]. For this reason, the local environment is separated in single independent grid cells, where the occupancy probability of each cell is estimated using the sensor measurements. Due to the object-model-free representation detecting any object type, e.g. pedestrians, cars and trucks, is possible. The Dynamic Occupancy Grid Map (DOGMa) [6], [7] is an extension to the classical grid map and is able to estimate velocities for dynamic areas. These grid map implementations showed significant results for a generic environment perception. However, a huge disadvantage for an object detection is a missing association of grid cells to the corresponding object. Steyer et al. [8] recently published promising results to combine a grid map and a multi-object tracking approach. Here, object hypotheses are generated by clustering cells of an evidential DOGMa and tracked with a subsequent unscented Kalman filter. Additionally, Asvadi et al. [9] presented a similar approach using a 2.5D motion grid. Furthermore, different approaches [10], [11] showed the effectiveness of density-based clustering, e.g. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) to cluster cells as object hypotheses from an occupancy grid map.

Certainly, fully automated vehicles depend on an accurate estimation of the objects and their states. Especially, planning and control modules, like [12] presented, malfunction of an error propagation of incorrect data. In the presented approach an environment perception framework is introduced, where the advantages of a multi-object tracking and a DOGMa are combined to receive a complete and robust estimation of other road users. A common representation of any object state is defined using an ego-stationary coordinate system [13]. Since, the work of [14], [15] and [16] showed the significance of contextual information, a digital map with roads and buildings is incorporated into the framework. Additionally, a DOGMa object extraction is presented using a clustering of cells to estimate a full object state. Finally, extracted tracks of a Labeled Multi-Bernoulli (LMB) filter [17] and the extracted DOGMa objects are transmitted to a high-level fusion module. Here, the object estimates are validated to ensure a complete and accurate object list. For that reason, a confidence value is developed that determines the quality of an object estimate regarding physical, module and digital map constraints.

The remaining paper is organized as follows: Section II reviews the basic properties of an LMB filter and the DOGMa. In Section III the presented environment perception framework including all functional modules is described. Subsequently, in Section IV the fusion of the object estimates considering a confidence value is explained, followed in Section V by experimental results with real world data. This work is closed in Section VI with a summary and outlook.

II. BACKGROUND

This section gives a short summary of the characteristics and parameters of a multi-object tracking using the Labeled...
Multi-Bernoulli (LMB) filter. Furthermore, the occupancy grid map and especially its extension, the Dynamic Occupancy Grid Map (DOGMa) is reviewed.

A. Labeled Multi-Bernoulli Filter

The LMB filter [17] is an accurate and fast multi-object tracking filter using Random Finite Sets (RFSs) [2]. The multi-object state is represented by the RFS \( X = \{x^{(1)}, \ldots, x^{(N)}\} \subset \mathbb{X} \), where \( x^{(i)} \in \mathbb{X} \) are the single-target state vectors and \( \mathbb{X} \) the state space. If not empty, a multi-Bernoulli RFS comprises multiple independent Bernoulli RFSs that represent the spatial distribution \( p(x) \) with a probability \( r \). In a scenario with multiple objects, a major challenge is to estimate the current state as well as the identity of an object. Therefore, Vo et al. [18] introduced the class of labeled RFSs. Here, a distinct label \( \ell \) identity of an object. Therefore, V o et al. [18] introduced

So, any tracked object comprise an unique label \( \ell \) represented with a parameter set \( \pi \) finite label space. The multi-object posterior LMB RFS is is appended to each state vector \( x \).

B. Occupancy Grid Mapping

An occupancy grid map is a discrete representation of the environment, where the local surroundings of a vehicle is separated into single grid cells \( c \). In classical approaches [4], [5] an occupancy probability of each cell is estimated using sensor measurements, that are updated with the Bayesian rule. For modeling inconsistencies of free and occupied cells the Dempster-Shafer theory of evidence [19] is implemented, where each cell holds a mass for free \( M_F \in (0,1) \) and occupied \( M_O \in (0,1) \). The Bayesian occupancy probability can be calculated with \( p_O(o) = 0.5\cdot M_O + 0.5\cdot (1.0 - M_F) \in (0,1) \). These classical grid maps assume, that the environment contain only static objects and each grid cell is independent of all other grid cell states. An extension of the classic occupancy grid map is the DOGMa, that additionally enables an estimation of the dynamic environment [7], [6]. In the presented approach, a particle filter estimates a velocity in \( x \) and \( y \) direction of each occupied cell, what corresponds to the implementation of [6]. The DOGMa provides cells in \( \mathbb{R}^{W \times H} \) with width \( W \) and height \( H \) around the vehicle pointing to stationary coordinates \( x \) and \( y \), respectively. The spatial resolution of grid maps depend on the quadratic grid cell size \( a_c \). The higher the resolution the more precisely an object can be detected, given a sensors uncertainty. Finally, each grid cell holds a state

\[ s_c = \{M_O, M_F, p, v, P\} \],

with a mass for occupied \( M_O \) and free space \( M_F \). Further, a two dimensional cell position \( p = [x, y]^T \), with velocity

\( v = [v_x, v_y]^T \) and the corresponding covariance matrix

\[ P = \begin{bmatrix}
\sigma_{v_x}^2 & \sigma_{v_x v_y} \\
\sigma_{v_y v_x} & \sigma_{v_y}^2
\end{bmatrix} \]

is given. For a more detailed mathematical description of the DOGMa, refer to [6].

III. Environment Perception

In this section, the system architecture with all necessary components is presented. Here, all functional modules are described with their main characteristics and attributes.

A. System Architecture

The environment perception framework is shown in Figure 1. The system architecture comprises out of a sensor layer with multiple sensor types and a perception layer with various functional modules resulting in a high-level fusion module. In the sensor layer a GPS timestamp is applied to every measurement, that will be used to queue and sort input data. Primarily, the presented perception layer focuses on perceiving other road users like vehicles, cyclist, pedestrians or trucks. For this reason, a module for estimating the ego motion, a multi-object tracking and a DOGMa is implemented. The results of these functional modules are passed to a high-level fusion module. A main idea is the generic structure of the functional modules, where each can be replaced with any other methods generating equal, worse or even better results. In the following, the functional modules with their input and output interfaces are described.

B. Ego Motion and Localization

As a common representation for any external objects, an ego-stationary coordinate system is defined. For that purpose, the measurements of an inertial measurement unit (IMU) sensor are applied to a dead reckoning motion model [13]. The ego vehicle’s state

\[ s_{ec} = [x, y, v, a, \varphi, \omega]^T \]

is given with a two dimensional position \( x, y \), velocity \( v \), acceleration \( a \), orientation \( \varphi \) and a turn rate \( \omega \). Starting from the initial state, where \( x, y \) and \( \varphi \) are equal zero, the IMU measurements are used to calculate the vehicle’s motion with a constant turn rate and acceleration (CTRA) model [20]. Since, the measurements of the IMU can be noisy, a standard
linear kalman filter estimates the dynamic states $v$, $a$ and $\omega$. Additionally, to utilize information from a global digital map a localization of the ego vehicle in the world is necessary. Accordingly, the ego vehicle’s global state

$$s_{gc} = [\text{UTM}_E, \text{UTM}_N, v, a, \varphi_{gc}, \omega]^T$$  \hspace{1cm} (4)

is estimated using a differential global positioning system (DGPS) and represented in Universal Transverse Mercator (UTM) coordinates. Thus, each point $p_{gc} = [\text{UTM}_E, \text{UTM}_N, \varphi_{gc}]^T$ in global coordinates can be transformed to the corresponding ego-stationary coordinates $p_{es} = [x, y, \varphi]^T$ and vice versa. The ego vehicle’s state is directly transmitted to the remaining modules.

C. Digital Map Processing

The digital map contains road courses and buildings from two different sources. Firstly, the Openstreetmap (OSM) [21] is used, where especially the contour of buildings is a valuable information. Even though, OSM contains inaccurate information it can be used as prior. Here, the global corner points of OSM buildings $B$ are transformed to the ego-stationary coordinates.

Additionally, a set of lanes $L$ were recorded with the experimental vehicle of Ulm University [22] using a high precise DGPS. Each lane represents the reference line of a single road using a vector of two dimensional equidistant points $p_{gc}$ in global coordinates and an unique line ID. Further, to employ the lanes in an appropriate way, an iterative end point-fit algorithm [23] approximates the lanes. As a result, every lane is approximated with a set of rectangles $R$ each consisting of a center point with orientation, width, length and an identifier. The number of required rectangles and their corresponding length results from the algorithm and depend on the road curvature. Since, the rectangles depict parts of a lane, the calculated orientation is interpreted as the course of a road. The width is specified with a fixed value for the offline map, but could be calculated using an online road detection. A benefit of the rectangles is an easy association to any road user. Finally, the lanes and rectangles can be incorporated to the ego-stationary coordinates.

The complete digital map $\mathcal{M} = \{B, L, R\}$ is shown in Figure 2 where the OSM buildings (cyan), the lanes (black) and the rectangles (grey) are visualized in an urban area.

D. Multi-Object Tracking

An object-model-based tracking of other road user is an essential task for the environment perception. Consequently, every object should be detected and their state estimated over subsequent time steps. For a temporal filtering of sensor measurements and associating these to the dynamic state of multiple road users at once, a multi-object tracking is used. In the presented perception framework, a generic functional module architecture is realized, so that the multi-object tracking can be replaced with similar implementations.

In this work an LMB filter is implemented to track vehicles using radar and camera detections, and is noted with $T$. For detailed information of the filter processing, see Section II-A. The implemented LMB filter uses a CTRA motion model to predict the tracked objects and consequently, the orientation and velocity of tracks can be estimated directly.

To extract tracks of the posterior LMB distribution the existence probability $r^\ell$ of a track with label $\ell$ has to exceed a minimum threshold $\theta_r$. Finally, the state of extracted tracks

$$\hat{s}_{T} = \{p_{rp}, v, a, \varphi, \omega, B, P, r, c, \ell\}$$  \hspace{1cm} (5)

consists of a two dimensional reference point position $p_{rp} = [x, y]^T$ with the set

$$rp = \{b, bl, l, fl, f, fr, r, br\},$$  \hspace{1cm} (6)

of labels back (b), back left (bl), left (l), front left (fl), front (f), front right (fr), right (r) or back right (br), the absolute velocity $v$, acceleration $a$, orientation $\varphi$ and a turn rate $\omega$. In addition, the length and width of an object is estimated resulting in an orientated bounding box

$$B = [p_{bl}, p_{fr}, p_{fr}, p_{br}].$$  \hspace{1cm} (7)

The uncertainty of the states is estimated with the covariance matrix $P$ accordingly. Finally, the existence probability $r$, the class type $c$ and an unique label $\ell$ is stated. After the filter processing cycle is finished, the set $S_T$ includes all extracted tracks $\hat{s}_T \in S_T$ and is transmitted to the high-level fusion module at every sample time $t_T$.

E. Dynamic Occupancy Grid Map

In the presented framework, the DOGMa separates the local environment into single grid cells $c$ and uses lidar measurements to estimate the occupancy, free space and velocity per cell. The implemented DOGMa is further described in Section II-B and in the following labeled with $\mathcal{G}$. Here, the essential characteristics and parameters of using a DOGMa in the environment perception framework are highlighted.

A main advantage of using a DOGMa for the object detection is the object-model-free environment perception. Thus, any road users like cars, pedestrian, bicycle or trucks can be detected. To estimate the dynamic areas particles are predicted with a constant velocity (CV) model and updated with the next sample of a measurement grid. In the end, every cell holds a state $[\ell]$, and is updated with sample time $t_G$. Since, no turn rate is depicted in a CV model, the cell orientation is assumed constant and calculated
The set of cell states in the search mask shows the remaining cells with white pixels. This threshold ensures to remove unknown cells. Figure 3(b) includes any cells with a minimum occupancy mass $\varepsilon_C$ passing each other in front of the ego vehicle. Hereafter, the scene is recorded in a urban area, where two vehicles occupy cells and white pixels for free space. Unknown cells in sharp curves can have an uncertain estimation for the orientation. Furthermore, any cell state of a DOGMa is independent of all other cell states, thus there is no association between cells and the corresponding objects, that caused the measurements. For that reason, in the presented approach a grid map object extraction is developed to assign each cell to an object and finally receive a full state description of any road user perceived with the DOGMa.

F. Grid Map Object Extraction

In the following, every single step is described to extract objects from the grid map cell states, what is visualized in Figure 3. An exemplary scene shows Figure 3(a) where the occupancy probability $p_c(O)$ is indicated with dark pixel for occupied cells and white pixels for free space. Unknown cells with $p_c(O) = 0.5$, or $M_F = 0$ and $M_O = 0$ are marked gray. The scene is recorded in a urban area, where two vehicles passing each other in front of the ego vehicle. Hereafter, the set $C_G$ of all cell states of the entire DOGMa $G$ is evaluated:

$$s_c = \{M_O, M_F, p, v, P\} \in C_G.$$  \hspace{1cm} (8)

1) Search Mask: The search mask is the basis for assigning cells to a corresponding object. Here, the search mask identifies cells, where any measurement was incorporated. The set of cell states in the search mask

$$C_s = \{s_c \in C_G | M_O > \varepsilon_{M_O} \}$$  \hspace{1cm} (9)

includes any cells with a minimum occupancy mass $\varepsilon_{M_O}$. This threshold ensures to remove unknown cells. Figure 3(b) shows the remaining cells with white pixels.

2) Cluster Cells: In this step a cluster is formed using the DBSCAN [24], as density based clustering technique. The presented approach performs a clustering on the search mask $C_s$ and uses the $\varepsilon$-neighborhood for an euclidean position and velocity distance between the cell $c$ and every other cell $c^*$. The resulting clusters $n$ are a set of cells

$$C_c^{(n)} = \{s_c \in C_s | \|p_c - p_c^*\|_2 \leq \varepsilon_p \land \|v_c - v_c^*\|_2 \leq \varepsilon_v \},$$  \hspace{1cm} (10)

that represent potential detected objects. Firstly, using a maximum two dimensional position distance $\varepsilon_p$ ensures clustering of cells next to each other and secondly, the maximum velocity deviation $\varepsilon_v$, implicitly considers the cells orientation using the velocity in $x$- and $y$-direction. Figure 3(d) visualizes the DBSCAN clustering applied on the search mask, where cells with same color indicate a cluster.

3) Validation Mask: Obviously, since the search mask contains almost every occupied cell it is not possible to separate between background and other road users. Thus, a validation mask is used where each cell has to fulfill strict criteria. Additional, besides the cell state (8), for every cell a Mahalanobis distance

$$d_0 = \sqrt{v^2_x \cdot P^{-1} \cdot v}$$  \hspace{1cm} (11)

relative to zero velocity is calculated [6]. A high value of $d_0$ indicates a great certainty that there is any movement. In addition, the absolute cell velocity $v_c = \sqrt{v_x^2 + v_y^2}$ is calculated. The validation mask

$$C_v = \{s_c \in C_s | p_c(O) \geq \varepsilon_p(O) \land v_c \geq \varepsilon_v \land \sigma_{v_x,v_y}^2 \leq \varepsilon_{\sigma_{v_x,v_y}^2} \land \sigma_{v_x,v_y}^2 \leq \varepsilon_{\sigma_{v_x,v_y}^2} \land d_0 \geq \varepsilon_{d_0} \}$$  \hspace{1cm} (12)

examines the cells occupancy probability, absolute velocity and Mahalanobis distance to reach a minimum threshold. In addition, the velocity variances must not exceed a maximum threshold. Figure 3(d) visualizes the few remaining cells of the validation mask.

4) Validate Clusters: Certainly, creating clusters on the search mask causes many false object hypotheses. Thus, the resulting clusters are validated, using cells of the validation mask. For each cluster, a minimum number of corresponding validated cells are required. For that reason, the ratio between the amount of cells $N_n$ in a cluster and the remaining number of validated cells $N_n^*$ is calculated. If the ratio

$$r_N = \frac{N_n^*}{N_n}$$  \hspace{1cm} (13)

exceeds a minimum threshold $\varepsilon_{r_N}$, the cluster is valid. Finally, the resulting clusters

$$C_c^{(n')} = \{s_c \in C_c^{(n)} | r_N \geq \varepsilon_{r_N} \},$$  \hspace{1cm} (14)

are used to create objects. Figure 3(e) shows the validated and clustered cells.
5) Object Creation and Label Assignment: For each validated cluster \( n' \) an object is created. Here, the cell states are used to estimate the objects position \( p_{rp} = [x, y]^T \) with a corresponding reference point label \( \ell \), the mean absolute velocity \( v \), mean orientation \( \varphi \), a bounding box \( B \) as orientated rectangle of the corner points and an unique label \( \ell \). Additionally, in every time step, the last extracted grid map objects are predicted with a CV model to the current time. Afterwards, an association for the predicted and currently extracted objects is executed. This is performed, calculating an euclidean distance of the reference points and finding the optimal associations with the Hungarian Method [25]. If two objects are successful associated the label \( \ell \) of the predicted object is assigned to the currently extracted object. Figure 10 shows the resulting objects from the presented scene, including the corresponding cells transformed to the ego vehicle coordinates. Finally, the estimated object state
\[
\hat{\mathbf{s}}_G = \{p_{rp}, v, \varphi, B, \ell\} \in \mathbf{S}_G
\]
is passed to the high-level fusion module as set of extracted grid map object \( \mathbf{S}_G \) at every sample time \( t_G \).

IV. HIGH-LEVEL FUSION

In this section, the processing of the results of the functional modules in the perception layer is described. The main goal of the high-level fusion is to define a common representation for every object and fuse and validate the object estimates in a complete set of objects.

A. Fusion of Objects

To receive a complete and accurate set of objects in the ego vehicles local environment, the tracks of the LMB filter and the extracted objects from the DOGMa are aggregated and fused to a common representation in the high-level fusion module. Here, the ego-stationary coordinates are used as base representation of all necessary information, e.g. object states or contextual information from the digital map.

Since, the LMB filter and DOGMa are processed in vehicle coordinates, a simple transformation of the extracted objects to ego-stationary coordinates is performed. Hereafter, to simplify notations of the object states, \( \hat{\mathbf{s}}_T \) and \( \hat{\mathbf{s}}_G \) are in ego-stationary coordinates. Additionally, the digital map is transformed from UTM coordinates to ego-stationary coordinates. Furthermore, the LMB \( \mathbf{S}_T(t_T) \) and DOGMa \( \mathbf{S}_G(t_G) \) have different sample times, where each module transmits their result as soon the processing is finished. For this reason, the output of the modules is sequentially ordered in a waiting queue at the input of the high-level fusion module and processed in the correct order.

The high-level fusion aggregates every detected object of the functional modules and incorporates them in a complete set of objects \( \mathbf{S}_H \). This set is sequentially updated with every sample from the functional modules. Because the state representation of the extracted grid map objects \( \hat{\mathbf{s}}_G \) slightly differs from the extracted tracks \( \hat{\mathbf{s}}_T \) a new meta object state
\[
\hat{\mathbf{s}}_H = \{p_{rp}, v, \varphi, B, c, \ell, \eta, \hat{\mathbf{s}}_G, \hat{\mathbf{s}}_T, t_G, t_T\}
\]
is created. Here, \( p_{rp} = [x, y]^T \) is a two dimensional position at the reference point \( \ell \), with an absolute velocity \( v \) and an orientation \( \varphi \). In addition, the length and width of the object is represented with the bounding box \( B \) and the object type \( c \) is estimated. New meta objects attached with an unique label \( \ell \) and comprise the corresponding time and state of the DOGMa and LMB, that created or updated the meta object. Since, in complex scenarios, a multi-object tracking or the DOGMa object extraction can generate false alarms or uncertain states a new parameter \( \eta \), the confidence of an object is introduced.

B. Object Confidence

To determine the quality of object estimates, received from either the DOGMa or LMB, a confidence value is calculated.

For that reason, highly uncertain object estimates or false positives from the functional modules should be detected and not included in an update for the meta object. For this purpose, the confidence is determined using physical, module and digital map constraints.

1) Physical Constraint: The physical constraints validate the state \( \hat{\mathbf{s}}_G \) or \( \hat{\mathbf{s}}_T \) relative to a last meta object state \( \hat{\mathbf{s}}_H \). For that reason, physical possible motions of the objects are examined using the relative movement between the last and current sample. The validity is represented with the confidence \( \eta(\hat{\mathbf{s}}|P) \).

2) Module Constraint: For the module constraint, the functional module specific characteristics are examined. Here, extracted tracks of the LMB filter estimate a covariance \( P \) of the object state and an existence probability \( r \). These parameters are analyzed to confirm the objects state. Furthermore, objects of the DOGMa are checked, if they are detected over subsequent time steps.

In addition, the confidence \( \eta(\hat{\mathbf{s}}|\mathcal{E}) \) of the functional modules \( \mathcal{E} = \{\mathcal{G}, \mathcal{T}\} \) increases, if an object is detected by both modules. This is evaluated in consideration of the modules specific sensors field of view.

3) Digital Map Constraint: The most significant weight generates the digital map constraints. Here, the map data, described in Section III-C, is used to validate every objects state. First, the confidence \( \eta(\hat{\mathbf{s}}|\mathcal{M}) \) reduces, when an objects position is clearly inside a building \( B \). In addition, the position and orientation of the vehicles are validated using the lanes \( L \) and corresponding rectangles \( R \). Currently, the uncertainty of the global ego position is not considered due to the high precise DGPS localization system, but can be integrated when using a different approach. Finally, all constraints are evaluated to update an object.

C. Object Update

To include a new object to the set of high-level fusion meta objects \( \mathbf{S}_H \) the label of the DOGMa or LMB object is analyzed. If a label is unknown and the estimated confidence value
\[
\eta = \eta(\hat{\mathbf{s}}|P) \cdot \eta(\hat{\mathbf{s}}|\mathcal{E}) \cdot \eta(\hat{\mathbf{s}}|\mathcal{M}) \in (0, 1)
\]
(17) exceeds a minimum, a new meta object will be created using the corresponding object state and a new unique
label. If the label $ℓ_σ$ or $ℓ_τ$ is already known, the object state is updated. Since, the high-level fusion receives sets of objects sequentially, an association between the already known objects from one module to the other module have to be identified. For that reason, the last sample is predicted with a simple CV model and a matrix of possible associations is created. Here, the optimal assignment is determined using the Hungarian method [25] regarding an euclidean distance of the reference points.

To update an object the confidence is examined. If $η$ exceeds a minimum, the position $p_{rp}$, velocity $v$ and orientation $ϕ$ of the meta object is updated. Additionally, the length $l$ and width $w$ is updated regarding the reference point $rp$. In particular, the following separations are considered:

\[
\begin{align*}
    w & \quad \text{when } rp \in \{b, f\} \\
    l & \quad \text{when } rp \in \{l, r\} \\
    w, l & \quad \text{when } rp \in \{bl, fl, fr, br\}.
\end{align*}
\]

Finally, the corresponding bounding box B is calculated using the updated width and length. If an object does not receive any samples over a short period, it will be removed from the set.

V. EXPERIMENTS

Evaluating the presented approach, the experimental vehicle of Ulm University [22] is used. Here, the vehicle is equipped with lidar, camera and radar sensors, which are used to record real world data in many different scenarios. The environment perception framework was evaluated in various scenes at urban and rural areas. To highlight the advantage of using a subsequent high-level fusion with a confidence value, two challenging scenarios are chosen as experimental results. The multi-object tracking with an LMB filter detects and tracks vehicles with a wide-angle camera and a frontal long-range radar. Here, extracted tracks with an existence probability $r > 0.2$ are passed to the high-level fusion. The DOGMa incorporates a frontal lidar sensor with an opening angle of $100°$ and range of $100m$ and a $360°$ Velodyne PUK VLP-16 lidar mounted on the top of the vehicle with reflections up to $40m$. For the grid map object extraction the following thresholds are used: $ε_{MO} = 0.3, ε_{p(O)} = 0.8, ε_{v_c} = 0.3, ε_{p_{xy}} = 5, ε_{d_0} = 9, ε_{p} = 1.2m, ε_{v} = 1m$ and $ε_{τN} = 0.1$. The DOGMa cell size is defined to $0.15m × 0.15m$ to achieve a total width $W = 120m$ and length $L = 120m$.

Figure 4 visualizes the first scenario. Here, the ego vehicle (green) is exiting a roundabout following a leading vehicle on a rural road. After exiting the roundabout, an oncoming vehicle is detected and false radar measurements are generated at a traffic island. Due to that, the LMB filter initializes a false track $ℓ_T = 7$ (blue) and confirms it over subsequent frames. Here, the grid map objects (red) do not contain this object and consequently the module constraint decreases. In addition, because of a high orientation error relative to the lane, the digital map constraint further reduces the confidence value of the false track (dashed black). Finally, the false track is not validated and not included in the set of meta objects $S_H$ (magenta).

The second experiment is a complex inner city scenario, with a wide variety of different road users and stop-and-go traffic. Here, a lot of objects are occluded and correct associations between tracks and measurements are ambiguous. Figure 5 shows the scene with its progressive frames. Here, the tracks (blue) and extracted grid map objects (red) are fused to the corresponding meta objects (magenta). Additionally, the object (dashed black) with a low confidence value is not included in the set of meta objects.

First, a false grid map object with label 1103 (red) is removed from the meta objects (magenta) because it is clearly inside a building. This objects was created due to lidar reflections from a buildings glass front, what can occur in common inner city scenarios. Additionally, two bicyclists are crossing the road in front of the ego vehicle. Consequently, the tracks with label $ℓ_T = 61$ and $ℓ_T = 64$ (blue) are occluded and no measurements are generated. The LMB filters object estimation is now based on its prediction and the uncertainty increases, whereas the confidence value decreases in the last frame (dashed black). Finally, the environment perception comprises a complete set of meta objects including only confident object estimates.

VI. CONCLUSION

In this work, an environment perception framework is introduced using various functional modules. For this reason, a common representation is developed using the ego-stationary coordinates to prevent uncertainties in the fusion of perceived objects. Furthermore, a multi-object tracking with an LMB filter and a DOGMa are implemented to detect and track road users in the local surroundings. Since, any object estimation method can generate uncertain states or false positives in complex scenarios a validation of the objects state is realized using a high-level fusion. To determine the quality of an object estimate a confidence value is introduced, that incorporates physical, module and digital map constraints.
Obviously, the multi-object tracking and the DOGMa have strengths and weaknesses, that strongly depend on the current scenario, sensor setup and environment conditions. Especially, an adjustment of the parameters in the grid map object extracting algorithm has a strong impact on the performance. However, the main goal in the presented approach is, that a fusion of two independent functional modules for an object estimation results in a complete and accurate set of meta objects. Consequently, weaknesses of one functional module can be compensated utilizing the strength of the other module.

For future work, a full evaluation on a public data set is necessary to show the generalization of the presented approach. In addition, the digital map will be extended with further context information for the object validation.

ACKNOWLEDGMENT

The research leading to these results was conducted within the Tech Center a-drive. Responsibility for the information and views set out in this publication lies entirely with the authors.

REFERENCES

[1] Y. Bar-Shalom and T. Fortmann, Tracking and Data Association, ser. Mathematics in Science and Engineering Series. Academic Press, 1988.
[2] R. P. S. Mahler, Statistical Multisource-Multitarget Information Fusion. Norwood, MA, USA: Artech House, Inc., 2007.
[3] S. Reuter, “Multi-Object Tracking Using Random Finite Sets,” Ph.D. dissertation, Ulm University, 2014.
[4] A. Elfes, “Using Occupancy Grids for Mobile Robot Perception and navigation,” Computer, vol. 22, no. 6, pp. 46–57, June 1989.
[5] S. Thrun, D. Fox, and W. Burgard, Probabilistic Robotics. MIT Press, 2005.
[6] D. Nuss, “A Random Finite Set Approach for Dynamic Occupancy Grid Maps,” Ph.D. dissertation, Ulm University, Institute of Measurement, Control and Microtechnology, 2017.
[7] G. Tanzmeister and D. Wollherr, “Evidential Grid-Based Tracking and Mapping,” IEEE Transactions on Intelligent Transportation Systems, pp. 1–14, 2016.
[8] S. Steyer, G. Tanzmeister, and D. Wollherr, “Object Tracking Based on Evidential Dynamic Occupancy Grids in Urban Environments,” in IEEE Intelligent Vehicles Symposium (IV), June 2017, pp. 1064–1070.
[9] A. Asvadi, P. Peixoto, and U. Nunes, “Detection and Tracking of Moving Objects Using 2.5D Motion Grids,” in IEEE 18th International Conference on Intelligent Transportation Systems, Sept 2015, pp. 788–793.
[10] R. Jungnickel and F. Korf, “Object Tracking and Dynamic Estimation on Evidential Grids,” in 17th International IEEE Conference on Intelligent Transportation Systems (ITSC), Oct 2014, pp. 2310–2316.
[11] T. Yuan, K. Krishnan, B. Duraisamy, M. Maile, and T. Schwarz, “Extended Object Tracking using IMM Approach for a Real-World Vehicle Sensor Fusion System,” in IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Nov 2017, pp. 638–643.
[12] S. Ulbrich, A. Reschka, et al., “Towards a Functional System Architecture for Automated Vehicles,” ArXiv e-prints, Mar. 2017.
[13] T.-T. Weiss, “Hochgenaue Positionierung und Kartographie mit Laser-Scannern für Fahrerassistenzsystemen,” Ph.D. dissertation, Ulm University, 2011.
[14] T. Gindеле, S. Brechel, J. Schroder, and R. Dillmann, “Bayesian Occupancy Grid Filter for Dynamic Environments using Prior Map Knowledge,” in IEEE Intelligent Vehicles Symposium, June 2009, pp. 669–676.
[15] S. Hossainyulamardary, Y. Balazadegan, and C. Toth, “Tracking 3D Moving Objects Based on GPS/IMU Navigation Solution, Laser Scanner Point Cloud and GIS Data,” ISPRS International Journal of Geo-Information, vol. 4, no. 3, pp. 1301–1316, 2015.
[16] D. Nuss, M. Stuebler, and K. Dietmayer, “Consistent Environmental Modeling by use of Occupancy Grid Maps, Digital Road Maps, and Multi-Object Tracking,” in 2014 IEEE Intelligent Vehicles Symposium Proceedings, June 2014, pp. 1371–1377.
[17] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer, “The Labeled Multi-Bernoulli Filter,” IEEE Transactions on Signal Processing, vol. 62, no. 12, pp. 3246 – 3260, 2014.
[18] B.-T. Vo and B.-N. Vo, “Labeled Random Finite Sets and Multi-Object Conjugate Priors,” IEEE Transactions on Signal Processing, vol. 61, no. 13, pp. 3460–3475, 2013.
[19] G. Shafer, A Mathematical Theory of Evidence. Princeton university press, 1976, vol. 42.
[20] R. Schubert, E. Richter, and G. Wanke, “Comparison and Evaluation of Advanced Motion Models for Vehicle Tracking,” in 11th International Conference on Information Fusion, June 2008, pp. 1–6.
[21] (2018, Apr.). [Online]. Available: https://www.openstreetmap.org
[22] F. Kunz, D. Nuss, J. West, et al., “Autonomous Driving at Ulm University: A Modular, Robust, and Sensor-Independent Fusion Approach,” in Intelligent Vehicles Symposium, 2015, pp. 666–673.
[23] U. Ramer, “An Iterative Procedure for the Polygonal Approximation of Plane Curves,” Computer Graphics and Image Processing, vol. 1, no. 3, pp. 244 – 256, 1972.
[24] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., “A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise,” in Kdd, vol. 96, no. 34, 1996, pp. 226–231.
[25] H. W. Kuhn, “The Hungarian Method for the Assignment Problem,” Naval Research Logistics (NRL), vol. 2, no. 1-2, pp. 83–97, 1955.