VDI

REIHE 08
MESS-,
STEUERUNGS-
UND REGELUNGS-
TECHNIK

Fortschritt-Berichte VDI

Sascha Steyer, M.Sc.,
München

NR. 1272

Grid-Based Object Tracking

BAND
1|1

VOLUME
1|1

Lehrstuhl für Steuerungs- und Regelungstechnik
Technische Universität München
Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Grid-Based Object Tracking

Sascha Jannik Steyer

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

Vorsitzender: Prof. Dr.-Ing. Hans-Georg Herzog

Prüfer der Dissertation:

1. Priv.-Doz. Dr.-Ing. habil. Dirk Wollherr
2. Prof. Dr.-Ing. Marcus Baum

Die Dissertation wurde am 02.11.2020 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 12.04.2021 angenommen.

VDI

REIHE 08
MESS-,
STEUERUNGS-
UND REGELUNGS-
TECHNIK

Fortschritt-Berichte VDI

Sascha Steyer, M.Sc.,
München

NR. 1272

Grid-Based Object Tracking

BAND
1|1

VOLUME
1|1

VDI verlag

VDI verlag

Steyer, Sascha

Grid-Based Object Tracking

Fortschritt-Berichte VDI, Reihe 08, Nr. 1272. Düsseldorf: VDI Verlag 2021.

196 Seiten, 66 Bilder, 2 Tabellen.

ISBN 978-3-18-527208-0, ISSN 0178-9546

71,00 EUR/VDI-Mitgliederpreis: 61,30 EUR

Für die Dokumentation: Autonomes Fahren – Objekterkennung – Objektverfolgung – Occupancy Grid Mapping – Sensordatenassoziation – Sensordatenfusion – Umfelderfassung – Umgebungswahrnehmung – Zustandsschätzung

Keywords: Autonomous Vehicles – Data Association – Environment Perception – Moving Object Detection – Object State Estimation – Object Tracking – Occupancy Grid Mapping – Sensor Data Fusion

Mobile robots require an accurate environment perception to plan intelligent maneuvers and avoid collisions. This thesis presents a novel multi-sensor environment estimation strategy that fully combines tracking moving objects and mapping the static environment. The basic idea is to fuse and accumulate measurement data by a dynamic occupancy grid model, whereas moving objects are extracted subsequently based on that generic low-level grid representation. Overall, this work results in a robust and consistent estimation of arbitrary objects and obstacles, which is demonstrated in the context of autonomous driving in complex unstructured environments.

Bibliographische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet unter www.dnb.de abrufbar.

Bibliographic information published by the Deutsche Bibliothek (German National Library)
The Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie (German National Bibliography); detailed bibliographic data is available via Internet at www.dnb.de.

ISBN 978-3-18-527208-0, ISSN 0178-9546

Foreword

This thesis is the result of my research within the last years as a doctoral student at the Technical University of Munich in conjunction with the Department of Automated Driving of the BMW Group in Munich. I am very grateful to all the people who supported me during this time and made this work even possible.

First of all, I would like to thank PD Dr. habil. Dirk Wollherr, my doctoral advisor at the Institute of Automatic Control Engineering (LSR) of the Technical University of Munich, for his great support and guidance as well as the scientific freedom during my entire doctoral time.

I would also like to express my special thanks to Dr. Georg Tanzmeister for his outstanding supervision and mentoring on site at the BMW Group. He also supervised me before during my internship at the BMW Group Research and Technology, while he was a doctoral student himself, and my subsequent master thesis there, which all eventually originated this work.

I am also thankful to all the other great colleagues at the BMW Group that I have worked with over the past years. Thank you all for the supportive and friendly atmosphere, various interesting discussions, and the overall very exciting working time. Special thanks to the other PhD students in our group, in particular Jens Schulz, Constantin Hubmann, Christian Pek, and Kai Stiens, but also to Dr. Dominik Kellner and the other team members of the feature team I have been part of.

This work has also been supported by student projects, for which I would like to thank my former students Vinzenz Dallabetta, Christian Lenk, David Raba, and Max Pittner.

Lastly, I want to express my deepest gratitude to my family and Britta for their endless support, encouragement, and love.

As a final remark, I sincerely hope that the work of this thesis will make a small contribution to making driving safer and more automated in the future, with the ultimate goal of minimizing the number of road fatalities – in memory of M. B.

Munich, 2020 *Sascha Steyer*

Contents

Notations

The following overview of the notations of this thesis is limited to the most common abbreviations and symbols regarding the overall context of this work. Some of the symbols are also used as part of subscripts and superscripts.

Abbreviations

CTRA	Constant Turn Rate and Acceleration (motion model)
DST	Dempster–Shafer Theory of evidence [34, 35, 111]
FOV	Field Of View
GPU	Graphics Processing Unit
GTAM	Grid-based Tracking And Mapping [118]
ID	Identifier
OCS	Object Coordinate System
RMSE	Root Mean Square Error
UKF	Unscented Kalman Filter [55]
VRU	Vulnerable Road User (pedestrians, cyclists, etc.)
Cell	Individual grid cell c of the discretized grid representation
Map	Accumulated dynamic grid map $\mathcal{M}_t$ of the occupancy grid mapping
Particle	Individual particle filter hypothesis χ of the low-level particle tracking
Track	Individual object track instance τ of the high-level object tracking

Symbols

General

$t, \Delta t$	Time instance t and time difference Δt between two time instances
v	Velocity component
x, y	Position components, partly summarized as 2-D position vector $x = [x_x, x_y]$
z	Measurement
η	Normalization coefficient
Γ	Specified threshold
$\mathcal{N}(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2 (evaluation: $\mathcal{N}(\,\cdot\,; \mu, \sigma^2)$)
$\mathcal{U}(\cdot\,,\,\cdot)$	Uniform distribution with specified lower and upper bounds of the interval
$p(\cdot), p(\cdot\,|\,\cdot)$	Probability density function p, with conditional probability $p(\cdot\,|\,\cdot)$
$E_{(\cdot)}[\cdot]$	Expected value of a random variable
$[\cdot]^\mathsf{T}$	Transpose of a vector or matrix
R_φ	Rotation matrix with angle φ

Evidential Occupancy Grid Representation

c	Cell of the discretized grid structure
d_c	Grid cell size (square with $d_c \times d_c$)
x_c	Position of the center of cell c
$\mathcal{A}(c)$	Spatial area (2-D interval) of the cell c with center position x_c
$\mathcal{G}$	Grid structure, describing the set of all individual cells c, i.e., $c \in \mathcal{G}$
F, S, D	Hypotheses of freespace (F), static occupancy (S), dynamic occupancy (D)
O	Hypothesis of occupancy, abbreviation for $\{S, D\}$, i.e., $O = \{S, D\}$
$\{F, D\}$	Hypothesis of passable area (freespace or dynamic occupancy)
$\{S, D\}$	Hypothesis of occupancy (static or dynamic occupied)
Θ, Θ_z	Frame of discernment (full set), with $\Theta = \{F, S, D\}$, $\Theta_z = \{F, O\}$
2^Θ	Hypotheses power set of all combinations of the different hypotheses of Θ
θ	Arbitrary hypothesis subset of Θ, with $\theta \in 2^\Theta$
$m(\cdot)$	Basic belief assignment (evidence mass) of a specified hypothesis set
$m(\cdot \mid \cdot \oplus \cdot)$	Evidence mass based on the evidential combination ($\oplus$) of two sources
$\mathrm{bel}(\cdot)$	Belief of a specified hypothesis set
$\mathrm{pl}(\cdot)$	Plausibility of a specified hypothesis set
$\zeta(\cdot\,,\,\cdot)$	Conflict mass between two sources

Dynamic Grid Mapping and Low-Level Particle Tracking

$\mathcal{M}_t$	Accumulated dynamic grid map with the individual evidence masses $m(\cdot)$
$\overline{\mathcal{M}}_t$	Predicted dynamic grid map
$\widehat{\mathcal{M}}_t$	Particle-based map prediction of the dynamic part
$\mathcal{M}'_t$	Adapted map prediction of the non-dynamic part
$\mathcal{M}_{z,t}$	Fused measurement grid
$\mathcal{M}_{z^*,t}$	Fused measurement grid enhanced by additional occupancy classification
f_D	Function for the particle-based convergence toward dynamic occupancy (D)
β_S, β_D	Sensor-based occupancy classification coefficients for S and D, respectively
γ_D	Assignment uncertainty parameter of $\{D\}$ based on $\{F, D\}$ and $\{S, D\}$
$\lambda_{(\cdot)}^{(\cdot)}$	Individual terms of the adapted evidential occupancy filtering
χ	Individual particle hypothesis (position x_χ, velocity v_χ, occupancy value o_χ)
$\mathcal{X}, \hat{\mathcal{X}}$	Total population of all particles $\chi \in \mathcal{X}$ and corresponding prediction $\hat{\mathcal{X}}$
$\lvert\mathcal{X}^c\rvert$	Number of particles in a cell c
$n_{\max}$	Maximum number of particles per grid cell
Δn_χ	Difference between the number of existing particles and its target in a cell
ν_t^c	Mean particle-based 2-D velocity estimate of a grid cell c
v_R, v_T	Radial and tangential velocity component, respectively
w_χ	Particle (velocity) weight based on radar and camera measurement data

Object Extraction and Association

$\mathcal{C}_{\tau,t}$	Set of dynamic occupied cells associated to a track τ
$\mathcal{C}^k_{\tau,t}$	Sub-cluster of the set $\mathcal{C}_{\tau,t}$
$\mathcal{G}_{D,t}$	Set of currently dynamic occupied grid cells above a threshold $\Gamma^D_{\min}$
$\mathcal{G}^{\zeta_\emptyset}_{D,t}$	Set of unassociated dynamic occupied grid cells, with $\mathcal{G}^{\zeta_\emptyset}_{D,t} \subseteq \mathcal{G}_{D,t}$
$\mathcal{G}_{O,t}$	Set of currently occupied grid cells above a threshold $\Gamma^O_{\min}$
τ	Individual object track, $\tau \in \mathcal{T}_t$
$\mathcal{T}_t$	Set of currently tracked objects
$f_a(c)$	Association function for each grid cell c
l_χ	Particle label that connects a particle χ to a track τ
$\mathcal{X}^\tau_t$	Particle population of all particles linked to a track τ
$\zeta_\emptyset$	Symbol for denoting no association
N^c_ε	ε-neighborhood of a cell c for density-based clustering

Object State Estimation

a	Acceleration state component, $a = \dot{v}$
g_τ	Geometric state of a track τ (bounding box size)
l	Length of the bounding box
$s_{\tau,t}$	Dynamic state of a track τ at time t
w	Width of the bounding box
φ	Orientation state component
ω	Turn rate state component, $\omega = \dot{\varphi}$
$\mathcal{I}^{\mathcal{X}}_\varphi$	Particle-based confidence interval of the assumed object orientation
k	Specific object class instance, $k \in \mathcal{K}$
$\mathcal{K}$	Set of possible classes (car, truck, pedestrian, cyclist, motorcycle, other)
$r_F(\mathcal{A})$	Ratio of the freespace evidence within a defined grid area $\mathcal{A}$
ϑ^e_z	Edge visibility of the measurement box, $e \in \{\text{front, rear, left, right}\}$
$\overline{v}$	Weighted mean cell velocity vector of all associated cells

Abstract

Mobile robots require an accurate environment perception to plan intelligent maneuvers and avoid collisions with traffic participants or obstacles. To obtain a robust model of the current surroundings, measurement data of multiple sensors have to be processed in different ways. This includes several tasks, such as abstracting object instances, data association, temporal filtering, and sensor data fusion. Popular approaches are commonly based on a sensor-individual object tracking with an early-stage object abstraction and a late-stage sensor fusion. However, that procedure causes significant information loss of the raw measurement data for the subsequent processing steps, potentially resulting in object ambiguities and thus in an error-prone high-level object fusion. This is particularly critical for autonomous driving in complex urban scenarios with densely moving traffic, partial occlusions, and unstructured parts of the static environment.

This thesis proposes a novel multi-sensor environment estimation strategy – the *Grid-Based Object Tracking*. The fundamental idea of this work is to fuse and temporally filter measurement data by a low-level environment model based on the generic concept of dynamic occupancy grids, whereas the object estimation is performed subsequently based on that grid representation. That way, objects are extracted robustly and consistently by using the full information of the fused and temporally accumulated data of all sensors, without requiring any early-stage object abstraction. Moreover, this approach fully comprises and combines tracking moving objects and mapping the static environment, taking into account all measurement detections, which overall ensures that all occurring objects and obstacles are reliably contained in the resulting environment representation.

In this work, measurement data of lidar, radar, and camera sensors are processed. They are modeled in a uniform occupancy grid representation with uncertainties using an evidential Dempster–Shafer model, which is further extended by separate velocity and classification grid layers. The sensor fusion is performed cell-wise based on the grid cell discretization. For the temporal accumulation, a new dynamic grid mapping approach combined with a low-level particle filter tracking is proposed, resulting in accurate cell velocity estimates and a differentiation of static and dynamic occupancy over time. Finally, moving objects are extracted based on the dynamic grid estimation, which thereby serves as a track-before-detect strategy that enables a generic detection of arbitrary-shaped moving objects primarily by identifying their cell-wise occupancy motion. To utilize the full potential of the overall approach, also new concepts for the object state estimation and association are presented that benefit from the grid-based representation as well, e.g., by evaluating the current object visibility based on the additional freespace information of the grid.

The proposed approach is evaluated with real sensor data and test vehicles, demonstrating that it is well suited for real-time multi-sensor environment perception applications, especially in the context of autonomous driving in challenging urban environments.

Zusammenfassung

Mobile Roboter benötigen eine genaue Umgebungswahrnehmung zur intelligenten Fahrmanöverplanung und Kollisionsvermeidung mit Verkehrsteilnehmern oder Hindernissen. Um ein robustes Modell der aktuellen Umgebung zu erhalten, müssen Messdaten mehrerer Sensoren verarbeitet werden. Dies beinhaltet verschiedene Schritte, wie etwa die Abstraktion von Objekten, die Datenassoziation, die zeitliche Filterung und die Sensordatenfusion. Bekannte Ansätze basieren üblicherweise auf einem sensorindividuellen Objekt-Tracking mit einer frühen Abstraktion der Messdaten zu Objektinstanzen und einer späten Fusion der Sensordaten. Allerdings führt diese Vorgehensweise zu einem deutlichen Informationsverlust der Messdaten für die darauffolgenden Verarbeitungsschritte, was potenziell Objekt-Mehrdeutigkeiten und somit eine fehleranfällige Sensorfusion auf Objektebene zur Folge haben kann. Dies ist insbesondere für autonomes Fahren in komplexen urbanen Szenarien mit dichtem Straßenverkehr, teilweisen Verdeckungen und einer unstrukturierten statischen Umgebung kritisch.

Diese Thesis stellt eine neuartige Strategie zur Multi-Sensor Umgebungsschätzung vor – genannt *Grid-Based Object Tracking*. Die Grundidee dieser Arbeit besteht darin, Messdaten in einem Low-Level Umgebungsmodell zu fusionieren und zeitlich zu filtern, basierend auf dem generischen Konzept von Dynamic Occupancy Grids. Die Objektschätzung erfolgt hingegen erst aufbauend auf dieser Grid-Repräsentation. Dies ermöglicht eine robuste und konsistente Extraktion von Objekten, da hierbei die vollständigen Informationen der fusionierten und zeitlich akkumulierten Messdaten aller Sensoren betrachtet werden, ohne dass diese zuvor zu Objektinstanzen abstrahiert werden müssen. Zudem kombiniert dieser Ansatz vollumfänglich die Schätzung von sich bewegenden Objekten mit der Schätzung der statischen Umgebung und berücksichtigt dabei alle Messdetektionen. Insgesamt wird somit sichergestellt, dass in der resultierenden Umgebungsrepräsentation alle auftretenden Objekte und Hindernisse zuverlässig enthalten sind.

In dieser Arbeit werden Messdaten von Lidar-, Radar-, und Kamerasensoren verarbeitet. Diese werden in einer einheitlichen Occupancy Grid Darstellung mit Unsicherheiten mithilfe eines Dempster–Shafer Evidenzmodells sowie weiterer Grid-Ebenen für Geschwindigkeits- und Klassifikationsinformationen modelliert. Die Sensorfusion erfolgt dabei zellweise basierend auf der Grid-Zelldiskretisierung. Für die zeitliche Akkumulation wird ein neuer Dynamic Grid Mapping Ansatz vorgestellt, der mit einer Partikelfilter-basierten Dynamikschätzung kombiniert ist. Hieraus resultiert eine genaue Schätzung von Geschwindigkeiten der Grid-Zellen sowie eine Unterscheidung von statischer und dynamischer Belegung über die Zeit. Bewegte Objekte werden dann aufbauend auf dieser Dynamic-Grid-Repräsentation extrahiert, die dabei als Track-Before-Detect Strategie dient und eine generische Detektion beliebiger Objekte primär durch die Erkennung der Bewegung derer Occupancy-Zellen ermöglicht. Um das volle Potenzial dieses Ansatzes zu entfalten, werden zudem neue Konzepte für die Objektzustandsschätzung und -assoziation vorgestellt, die ebenfalls von der Grid-basierten Darstellung profitieren, z. B. wird die Sichtbarkeit der Objekte anhand der zusätzlichen Freiraum-Information des Occupancy Grids bestimmt.

Der Ansatz wurde mit realen Sensordaten und Testfahrzeugen evaluiert; die Ergebnisse demonstrieren eine erfolgreiche Anwendung als Multi-Sensor-Umgebungsschätzung, insbesondere im Kontext des autonomen Fahrens in komplexen urbanen Umgebungen.

1 Introduction

This thesis investigates the multi-sensor environment perception of mobile robots, in particular the detection and tracking of moving objects in the context of automated driving. This introductory chapter first briefly motivates the broad topics of mobile robots, autonomous vehicles, and multi-sensor environment perception in general, followed by a short discussion of the challenges in that context. Finally, an overview of the proposed approach of this thesis is presented, including a summary of the main contributions as well as the structure of the subsequent chapters. Parts of this thesis have also been published separately in [1–5].

1.1 Motivation

Mobile robots have been a vast field of research for several decades. Enormous technological progress has been achieved in the meantime with many robots already being used successfully in various applications. In the household sector, for example, service robots for domestic tasks, such as autonomous vacuum cleaning robots and lawn mower robots, are being used more and more frequently, with a total of 39.5 million units expected to be sold worldwide in the period 2019–2021 and an estimated compound annual growth rate (CAGR) of about 31 % [54]. Similarly, in the logistics industry, warehouses are increasingly automated, even for complex tasks such as object-individual orders with small products or working alongside humans, by mobile piece picking robots [18]. Some supermarkets also already employ first assistance robots, not only for floor cleaning or inventory shelf-scanning tasks, but also for directly guiding and supporting customers [19, 53]. Moreover, several companies have started to use small delivery robots for autonomously bringing parcels and goods such as groceries from warehouses or retail stores to homes and offices within urban areas by small ground-based robots [114] or air drones [9, 10]. Equivalently, in the whole transportation sector, established means of transportation such as trains, trucks, buses, or cars are more and more automated to support and relieve drivers – toward the ultimate vision of *autonomous driving*, i.e., self-driving vehicles without requiring human drivers.

Autonomous vehicles thereby have great potential to fundamentally change our lives in different ways. Most importantly, automated systems generally have the ability to significantly increase the safety of transportation as they permanently remain attentive without getting tired, potentially saving millions of lives. Various advanced driver-assistance systems (ADAS) already support drivers and significantly increase the road safety, but, still, globally about 1.35 million people die every year in road accidents [144], mostly caused by human driving errors that are principally preventable. Another aspect of intelligent autonomous vehicles is an efficient traffic flow without traffic congestion, not only by avoiding accidents, but also by predictive and cooperative speed adaptation rather than the common shock wave phenomenon of stop-and-go reactions [52, 145], possibly further improved by vehicle-to-everything (V2X) communication [140]. This reduces the overall emission

as well as the average travel time, while passengers can also spend their time during the journey on other tasks. Since then no drivers are required anymore, transportation costs for both moving people as well as goods can be further reduced, which also enables higher individual mobility and economically priced on-demand transportation services. Robo-taxi fleets in urban environments may thereby further revolutionize the shape of cities itself, since then the large areas of parking lots would not be necessary anymore [90]. Overall, this wide-ranging vision of autonomous driving is also expressed in economic studies, projecting the global autonomous vehicle market to be valued at $54.23 billion in 2019, increasing to $556.67 billion by 2026 with a CAGR of 39.47 % from 2019 to 2026, according to [68].

With all this great potential, almost all automakers as well as several other tech companies and research institutes are working intensively to make self-driving vehicles a reality. Although there has been tremendous progress in the last decade, intelligent and safe autonomous driving in complex urban environments still remains a big challenge. One of the most essential issues to drive autonomously and safely is a detailed understanding of the surroundings of the corresponding vehicle. To plan interactive maneuvers and avoid collisions in dynamic environments, it is essential to know which parts of the local environment are currently passable, where obstacles have to be avoided, and where other traffic participants are moving. Therefore, sensor measurement data have to be processed in different ways to obtain such information. This *environment perception* includes several tasks, for example, object instances have to be abstracted from the raw measurement data by specific features. A robust object tracking further requires a temporal filtering over time including data association and recursive state estimation. In addition, it is crucial to fuse measurement data of different kinds of sensors, in particular the three types lidar, radar, and camera, to advantageously combine the respective strengths and weaknesses of their individual physical measurement principles.

This thesis presents a novel *multi-sensor environment perception* approach that is suitable for real-time autonomous driving applications. A consistent estimation of moving objects, static obstacles, and passable freespace is proposed, taking into account measurement data of lidar, radar, and camera sensors. The basis of this approach forms a generic low-level occupancy grid representation of the environment, but focus is also given on the subsequent high-level abstraction and tracking of moving objects – overall resulting in the *Grid-Based Object Tracking*. The contribution of this work is discussed in more detail after the following overview of multi-sensor environment perception techniques.

1.2 Challenges of Multi-Sensor Environment Perception

For intelligent mobile robots it is crucial to perceive and understand the current local environment, which overall represents a very broad subject area with many different research approaches and applications. A fundamental and safety-critical task of environment perception, in particular in the context of autonomous driving, is the robust detection and temporal filtering of all kinds of objects and obstacles as well as passable areas of the surroundings. In the following, the sensor-based estimation of such essential parts is discussed for approaches that are solely based on measurement data of sensors mounted on the ego vehicle, without depending on localization, offline road maps, or other external information.

Detecting and tracking objects can involve various aspects and characteristics. In the simplest case, only a single target is tracked with also only one measurement being present at each time instance, i.e., the measurement directly corresponds to the object instance. Since real sensor measurements are generally noisy and may also reflect only a part of the object state, a *state estimation* with a *tracking* over time is required for robustness. For example, only noisy position measurements of an object track may be available, but the temporal filtering enables a smoothed position estimation as well as an implicit estimation of the velocity and acceleration as the corresponding first and second derivatives, respectively. Typically, a recursive strategy is applied with the two alternating steps of prediction and measurement update, which, in the context of probabilistic estimation, corresponds to a Bayes filter. This means that the expected state at the time instance of the measurement has to be calculated based on the previous state combined with a motion model, which is then partly corrected by the actual measurement. Most commonly, different variants of the Kalman filter [57], including the extended Kalman filter (EKF) [32] and the unscented Kalman filter (UKF) [55], or the particle filter [33] are used for almost all applications as robust recursive state estimation methods. Challenges generally occur in selecting proper measurement and state process models as well as an applicable filtering tradeoff between increasing the robustness against noise and minimizing the filtering latency for highly dynamic movements such as emergency braking.

The assumption that an object causes only a single measurement detection is reasonable if that object can be approximated as a point target without a spatial extent, which may be applicable for object tracking in very far distances like maritime or aerospace scenarios. In the context of local environment estimation of mobile robots, however, in particular with recent high-resolution lidar sensors, objects like a car or even a pedestrian may result in a large number of measured detections of a sensor. Hence, the spatial extent as well as the observable shape of an object have to be taken into account, resulting in so-called *extended object tracking* [43]. This also means that measurement points have to be combined to form a hypothesis of an object instance, most commonly using a bounding box shape assumption, which then can be used to update the filtered object track, as exemplarily illustrated in Fig. 1.1a. This object hypothesis extraction represents the actual *object detection* task.

There are various object detection approaches, which are usually optimized for the specific measurement data of the corresponding sensor type used. Established approaches thereby focus on extracting and evaluating specific hand-modeled features. For lidar data, the raw measurement points are thereby generally first clustered to segments, e.g., using (radially bounded) nearest neighbor clustering [63, 79], density-based clustering [11, 40, 129], or a spatial grid structure discretization with a subsequent connected component clustering [46]. Given such clusters, relevant features are analyzed to determine if a cluster represents an object, typically regarding specific object shape assumptions, for example using line, corner, or box fitting [45, 78, 79, 152]. Challenges remain if such features are not very obvious, e.g., due to measurement noise or a lower resolution in larger distances, partial occlusion, or objects with unusual shapes that have not been explicitly modeled. Radar detections are typically clustered similarly using density-based approaches, which can be further enhanced by a polar range/azimuth representation and additionally taking into account measured Doppler velocity and amplitude informa-

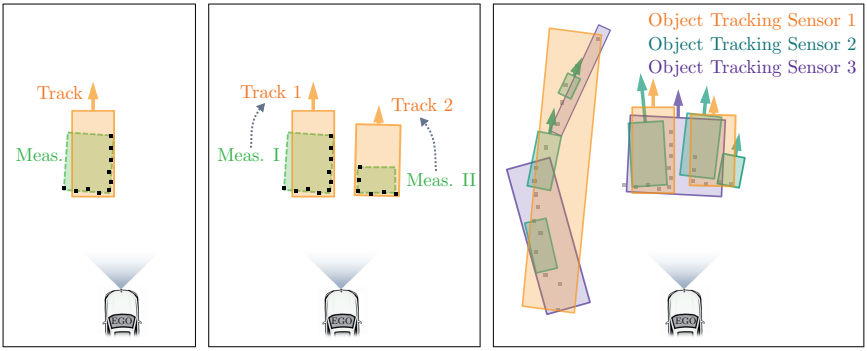


Fig. 1.1: Illustration of different challenges of object tracking. (a) Single object tracking of extended targets: measurement detections (black squares) have to be combined to form an object measurement (green box) for updating the object track (orange box). (b) Multi-object tracking: object measurements have to be associated to the individual tracks. (c) Multi-sensor multi-object tracking: object tracks (or measurements) of different sensors have to be fused, here with additional objects of a static road boundary (left side) and object ambiguities.

tion [62, 75]. In contrast to range-based lidar or radar measurements, camera-based object detection methods require semantic image processing with class-specific features. Popular image-based feature detection approaches are scale-invariant feature transform (SIFT) [72], Haar-like features [132], or histogram of oriented gradients (HOG) [28]. Such classic computer vision methods, however, have been significantly outperformed in the last couple of years by modern deep learning approaches, in particular with different variants of convolutional neural networks (CNN), e.g., [16, 39, 58, 67, 82, 97, 138]. Recently, similar CNN-based approaches have also been successfully applied for object detection with lidar point clouds [23, 153], as well as radar data [139, 151]. Overall, such deep learning approaches enable an end-to-end learning without requiring any hand-modeled features, but in return they depend on the availability of large labeled datasets, with a high variety of scenarios and objects for a robust generalization, and high-performance computation power for sophisticated training and inference. Moreover, neural networks with full end-to-end learning may possibly result in unknown behavior and are hard to analyze in detail, which, when they are not combined with directly interpretable algorithms, is critical for safety-relevant tasks such as object detection in the application of self-driving vehicles.

In addition to multiple measurement reflections that arise from a single object instance, there are typically also multiple objects in the environment that have to be individually tracked and distinguished, commonly known as *multi-object tracking* (MOT). Hence, *data association* between the object hypotheses of the current measurement and the temporally filtered object tracks that have been seen before, predicted to the time instance of the measurement, is required to update each track by the proper measurement. This is exemplarily illustrated in Fig. 1.1b for a simple scenario with two objects and unambiguous measurements. Typically, a validation gate of each predicted track is computed to exclude measure-

ments that differ significantly from the expected measurement likelihood, for example using the Mahalanobis distance [73] of the object state and covariance. Popular data association approaches are different variants of local or global nearest neighbor (NN/GNN) [65] association using fixed one-to-one associations, joint probabilistic data association (JPDA) [41] using association probabilities with a weighted average of all validated measurements, or multiple hypothesis tracking (MHT) [17, 96] using different hypotheses of the data association and track update. Recently, multi-object tracking approaches based on random finite sets (RFS) are increasingly used, e.g., [15, 74, 98–100, 133], modeling not only the object states as random numbers, but also the number of objects and measurements itself. In contrast to common object-individual tracking approaches with separate filters, this enables a more generic estimation of uncertainties about object instances and the corresponding data association. However, due to its high computational complexity, this requires several approximations for real-time applications and also appropriate models for the probabilities of birth, detection, disappearance, and false alarm in terms of clutter that may all be difficult to model. Overall, data association challenges remain when there are association ambiguities between multiple measurements and different tracks, in particular when objects are close to each other. This even gets more complex when some of those measurements do not represent an actual existing object, i.e., clutter, or when multiple measurements are originated by the same track, e.g., due to partial occlusion. Hence, the data association problem is also coupled with the object detection problem as discussed before.

As described so far, multi-object tracking for a single sensor is already a complex task. For a robust and comprehensive environment estimation, covering a full surrounding 360° detection area, however, *sensor data fusion* of multiple heterogeneous sensors is required, with an advantageous utilization of the respective sensor strengths. A popular strategy for such a *multi-sensor multi-object tracking* is a sensor-individual object detection and tracking as described above, followed by a late-stage high-level fusion of the filtered object tracks of all sensors, e.g., [7, 8, 22, 76]. This means that for each individual sensor the raw measurement data are abstracted to object instances, e.g., using a bounding box model. The fusion is then performed afterwards based on that abstracted object representation, which requires a high-level association between the object tracks of the different sensors. Such an approach is suitable when objects can be clearly separated and distinguished from the remaining part of the environment, i.e., when objects can be robustly detected and the object abstraction is feasible without any significant approximation error. For instance, this might be the case for highway scenarios with larger distances between traffic participants. Thereby the early-stage object abstraction reduces the amount of data that have to be transferred and processed in the subsequent steps, especially when it is directly computed within the individual sensor processing units. However, if those object approximations are inaccurate or some object assumptions are invalid, this early-stage abstraction results in information loss of the subsequent processing steps and thus also in a more error-prone fusion due to object ambiguities, as exemplarily illustrated in Fig. 1.1c.

For urban scenarios with densely moving objects, partial occlusion, and complex maneuvers of various traffic participants, such an information loss due to an early object approximation is therefore disadvantageous. This applies in particular to latest autonomous vehicle prototypes with performant computing resources and an increased number of equipped

sensors as well as their higher scanning resolutions itself. Hence, a major challenge of multi-sensor multi-object tracking therefore remains in reducing the information loss of common high-level object fusion architectures by fusing measurement data before object assumptions are made, i.e., an early-stage low-level fusion with a late-stage object abstraction. This is also essential for the object detection itself, as it may be difficult to extract an object by the measurement data of an individual sensor but more feasible to detect that object given the combined data of multiple sensors. The same applies for separating different object types and instances, such as closely moving pedestrians and a small vehicle. A fundamental question thereby is which kinds of objects should be extracted at all – only actually moving objects, all detectable vehicles and traffic participants, or even also all obstacles of the static environment such as guardrails or buildings? For large static obstacles, in particular when they are curved or non-convex, an object box fitting might result in an improper approximation. In particular high-level object fusion approaches are even more error-prone with such an object extraction of the static environment, since different sensors may result in varying numbers and forms of extracted objects of that same static environment, which is also depicted in Fig. 1.1c. An example of that issue with real-world measurement data of lidar and radar sensors is demonstrated in Fig. 1.2. Based solely on the high-level object representation of both sensor types as given in Fig. 1.2e, various object ambiguities occur that prevent achieving an accurate multi-sensor model of the local environment. In fact, it is even hard to understand the scenario at all that way. Details of this scenario, including camera images, are demonstrated together with the processing steps of the proposed approach in the next section in Fig. 1.3.

A popular concept for modeling and estimating the static environment without requiring specific object assumptions is *occupancy grid mapping* [37, 128]. The basic idea here is to spatially discretize the local environment into grid cells and estimate occupancy probabilities of each grid cell. Sensor-specific measurement data can thus be modeled in that uniform occupancy grid representation, enabling to represent obstacles with arbitrary shapes without extracting specific high-level object features, only depending on the resolution of the cell discretization. Furthermore, the concept of occupancy probabilities also enables to model the opposite of obstacles – the freespace, i.e., cells that are free of obstacles, which is important not only for the temporal filtering of clutter measurements but also to determine the observability and clearance within the local environment. Since the discrete grid structure remains fixed in the world and the odometry of the ego-movement is assumed to be known, the association is implicitly performed by a cell-wise combination for both the temporal mapping accumulation and the sensor data fusion. Hence, occupancy grids enable not only a uniform sensor-independent representation and a probabilistic occupancy accumulation, but also a robust sensor data fusion and association, all by the grid cell discretization without requiring object assumptions.

Typically, tracking dynamic objects and mapping the static environment are performed as separate tasks, where measurement data either belong to the object representation or the mapping of the remaining parts. This, however, requires a classification of the measurement data into stationary obstacles and dynamic objects. Thereby it is also possible to perform that classification by evaluating the measurement position in the accumulated grid map, while dynamic parts are characterized by inconsistencies of previously derived freespace

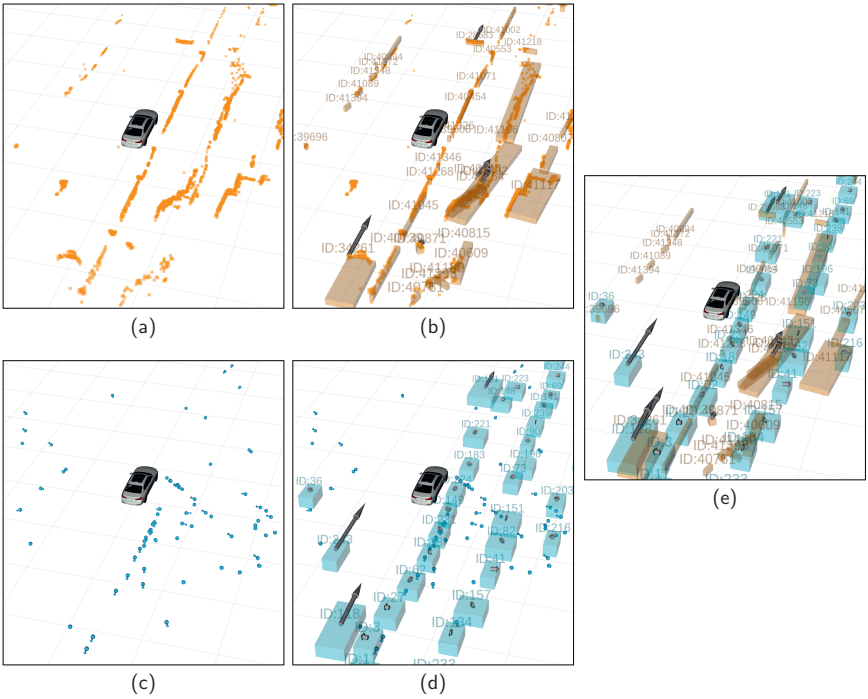


Fig. 1.2: Example of the problem of the late-stage sensor data fusion of abstracted object boxes with real measurement data of lidar and radar sensors in an urban environment. (a/b) Lidar measurements and corresponding object tracking. (c/d) Radar measurements and corresponding object tracking. (e) Visual overlap of the sensor-individually tracked objects of (b) and (d) that have to be fused to achieve a multi-sensor object tracking.

and currently measured occupancy or vice versa [80, 135, 137]. Hence, this enables a moving object detection by clustering measurement data classified as dynamic, e.g., [20, 56, 106, 135]. Such a detection concept directly based on those occupied/free inconsistencies, however, is error-prone to measurement inaccuracies of ranging or odometry sensors.

Therefore, a more robust strategy requires a recursive dynamic state estimation of the grid map, i.e., extending the concept of occupancy grid mapping toward dynamic environments. Latest approaches of such *dynamic occupancy grids* are based on a grid-based particle filtering concept that enables a generic estimation of moving parts by continuous velocity distributions of each grid cell, e.g., [29, 83, 85, 121]. Overall, this promising technique results in an accurate dynamic estimation and thus also in a temporally filtered cell-wise static/dynamic classification. Challenges remain in a robust and efficient combination of the persistent occupancy grid map accumulation and the short-term particle-based dynamic estimation, but also in an advantageous utilization of those dynamic grid mapping concepts as part of a generic object tracking including object detection and association.

1.3 Main Contribution and Outline of This Work

As discussed in the previous section, multi-sensor environment perception is a complex task with various challenges. Thereby the concept of occupancy grids has many advantages: It enables a uniform representation with an implicit data fusion and association by the cell discretization; the additional dynamic grid estimation further enables a temporal filtering of occupancy probabilities for both stationary and moving parts. No object assumptions are required for that low-level estimation – in fact, it can be used instead as a pre-processing step for the moving object tracking itself, which can then further benefit from that grid-based sensor fusion and dynamic estimation. Based on the great potential of such a combined approach, this work investigates a new environment perception strategy that fully utilizes the advantages of dynamic grid mapping with a subsequent object tracking.

Overall Contribution of the Thesis

This thesis proposes the *Grid-Based Object Tracking* – a novel approach for the multi-sensor multi-object tracking of moving targets including a consistent mapping of static obstacles and passable freespace. The fundamental idea is to fuse and pre-process measurement data in advance of the object tracking by the generic estimation of dynamic occupancy grids, overall representing an early-stage sensor fusion and dynamic estimation with a late-stage object abstraction. This architecture is significantly different to popular object tracking approaches with a high-level object fusion, which are commonly based on the reversed processing order in terms of an early-stage sensor-individual object abstraction and a late-stage dynamic estimation and fusion. The resulting object estimation of this work is thus based on the combined and temporally accumulated low-level measurement data of all sensors, which has several decisive advantages.

The principal strength of the proposed strategy is clearly evident when comparing the previously presented high-level object fusion example of Fig. 1.2 with the corresponding processing steps of this work, which are highlighted in Fig. 1.3. The environment is modeled here on the grid cell-level based on the low-level sensor fusion and mapping, where the temporal filtering results in a generic differentiation of static and dynamic occupancy. Surrounding traffic participants are thus directly noticeable as their movement results in areas of dynamic occupancy, which significantly simplifies the subsequent moving object tracking. As a result, all four traffic participants of this scenario are robustly extracted as object instances here, whereas stationary obstacles, such as the construction site fence on the right side, are accurately represented as static occupied cells in the grid map without requiring any further object abstraction – that scenario was not directly obvious based on the object-level representation of Fig. 1.2e.

The key advantage of the proposed approach is that it fully comprises and combines the two tasks of tracking moving objects and mapping the static environment. Objects, obstacles, and passable freespace are all consistently estimated based on the common low-level occupancy grid map representation. No error-prone object abstraction or pre-classification of measurement data into the static world or dynamic objects is required since measurement detections are generally represented as occupancy; the differentiation between static and dynamic occupancy is carried out as part of the temporal filtering over time within

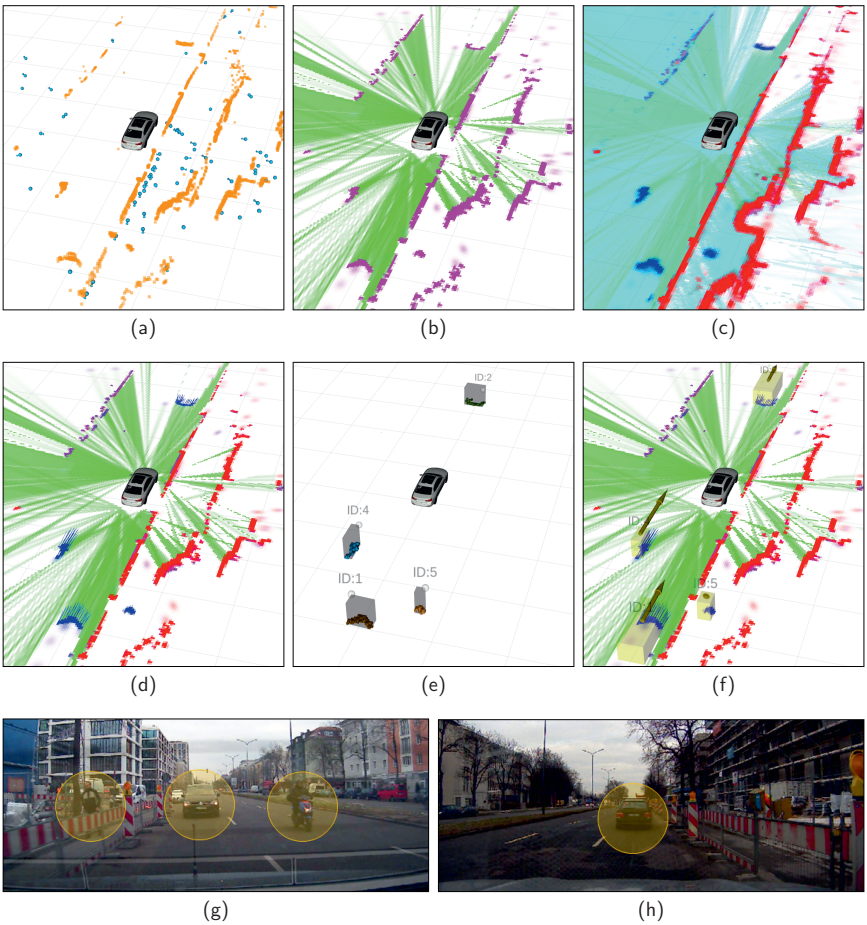


Fig. 1.3: Overview of the main processing steps of the proposed approach. The scenario is the same as in Fig. 1.2, which consists of four traffic participants: a vehicle in the front and another one in the back, an overtaking scooter on the rear left side, and a pedestrian on the sidewalk that is partly occluded by a construction site fence. Color coding of the evidential grid representation: green (freespace), pink (occupancy), blue (dynamic occupancy), red (static occupancy), cyan (passable area). (a) Raw lidar and radar measurement data. (b) Fused 2-D measurement occupancy grid. (c) Temporally accumulated dynamic grid map. (d) Augmented measurement grid (static/dynamic occupancy classification and cell-wise velocity estimates by filtered dynamic grid map). (e) Object extraction and association (only considering dynamic occupied cells). (f) Grid-based object tracking (temporally filtered object state estimation). (g)/(h) Camera images (rear/front) of the scenario (occurring moving objects are highlighted by yellow circles).

the dynamic grid mapping. This conceptual design ensures that no measurement data are neglected due to an unknown object or obstacle type, i.e., all range-based measurement detections result in an occupancy accumulation in this work. Other popular approaches generally lack that essential ability, as they are based on already abstracted object instances or an independent estimation of obstacles and objects. The algorithmic robustness of this work also applies to the subsequent moving object extraction, which is based on the dynamic grid map estimation including the cell-wise velocity estimation that thereby serves as a track-before-detect strategy. That way, arbitrary-shaped moving objects are generally detected by identifying their occupancy motion, overall ensuring a robust object estimation in various environments, even for unknown object types that have never been seen before or sparse sensor data with only a small number of detections, e.g., in far distances or in case of partial occlusions. Popular appearance-based object detection approaches that rely on specific features of the object shape, in contrast, commonly fail in such scenarios. This also includes recent deep learning approaches, which, as opposed to this work, further highly depend on large labeled training datasets. Altogether, the proposed approach results in a generic and consistent environment estimation of arbitrary obstacles and objects, which is especially crucial for the safety of autonomous robots in unstructured environments to ensure avoiding collisions even in complex and unknown scenarios.

The primary grid-based estimation of this work thereby has several other benefits, also for the subsequent object tracking. Due to the uniform occupancy model of all measurement data, the approach does not necessarily depend on sensor-specific characteristics; instead, it is directly adaptable to the number of sensors used, their mounting positions, and the types of sensors themselves. Moreover, the additional freespace information of the grid represents the negative information where currently no objects are present. This is important not only for improving the object estimation itself, but also for determining the overall observability of the scenario, i.e., to distinguish actual collision-free passable areas and unknown occluded areas with potentially undetected objects or obstacles.

In conclusion, this thesis presents a novel environment perception approach that robustly and consistently estimates moving objects, stationary obstacles, and passable freespace, with a full surrounding 360° detection area of multiple lidar, radar, and camera sensors. The approach thereby remains generic and versatile in various environments due to the uniform dynamic grid representation and the movement-based object estimation. Moreover, since it is solely sensor-based, no prior map knowledge is required. Furthermore, it is shown that the algorithm is real-time capable on a conventional modern computer by an efficient implementation with highly parallelized GPU processing. All in all, the outcome of this work is a new real-time multi-sensor environment estimation algorithm that improves the perception capabilities of mobile robots and thus eventually their safety and intelligence, especially for future autonomous driving applications.

Overview of the Individual Chapters of the Thesis

As part of the proposed overall *Grid-Based Object Tracking* strategy, this work addresses all relevant processing steps of it: modeling and fusing the input measurement data, the dynamic grid mapping, the object extraction and association, and the object state estimation. To utilize the full potential of that approach, novel techniques are proposed for each

of those processing steps, primarily based on the characteristics of the grid-based estimation. In the following, a brief summary of the individual chapters of this thesis is given; a detailed discussion of the contribution, including related work, is included in each chapter:

- *Chapter 2* introduces the measurement grid representation for different sensor data of lidar, radar, and camera. Measurement data are abstracted by an evidential occupancy grid model, based on the Dempster–Shafer theory of evidence, to obtain a uniform, sensor-independent environment representation with uncertainties. A generic grid-based fusion strategy of the sensor-individual measurement grids is proposed, including concepts for the spatiotemporal alignment of asynchronous sensors. In contrast to high-level object fusion approaches, no object abstraction or association is required here due to the implicit cell-wise processing of the grid structure. To further prevent information loss of additional sensor-specific measurement data such as radar Doppler velocities and camera classification, as opposed to common grid-based approaches, the measurement grid representation is extended by separate velocity and classification grid layers in this work. The resulting fused measurement grid represents the combined unaccumulated measurement data of all integrated sensors.

- *Chapter 3* addresses the temporal accumulation of the fused measurement grids over time. A new evidential dynamic grid mapping approach with a particle-based dynamic estimation is proposed. The dynamic grid map is modeled using an extended evidential representation and filtering, which enables differentiating static and dynamic occupancy over time, implicitly modeling temporal transitions of hypotheses, and handling conflicts of contradicting hypotheses. A particle-based velocity estimation is used for filtering the dynamic state of the grid map and thereby initialize and predict its dynamic occupancy, whereas static occupancy is directly modeled in the grid map without requiring particles. Compared to other particle-based dynamic grid approaches, this combined estimation with the persistent evidential map accumulation improves both the robustness of the occupancy classification and the efficiency of the particle tracking. Moreover, this approach is further enhanced by evaluating the additional radar velocity and camera classification data of the fused measurement grid for improving the cell velocity estimation as well as the occupancy classification.

- *Chapter 4* presents novel approaches for the moving object detection and data association, both based on the characteristics of the dynamic occupancy grid. New objects are extracted by a density-based clustering of dynamic occupied cells with a subsequent connectivity-based clustering and velocity variance analysis to minimize false positives. In contrast to common appearance-based detection algorithms, arbitrary-shaped moving objects are thus detected by the movement tracking itself, without depending on specific features or training data. Furthermore, for filtered objects that have been extracted in the past, individual grid cells are directly associated before abstracting object hypotheses, i.e., an association on the grid cell-level rather than on the object-level, which is essential for urban scenarios with partial occlusions and dense traffic. That association approach is further enhanced by linking particles of the underlying particle-based dynamic grid estimation with object tracks, as this enables an accurate free-form occupancy likelihood approximation of each object.

- *Chapter 5* proposes new concepts for the object state estimation over time – the final step of a robust object tracking based on already associated measurement data. A common box-model object shape is selected, but, compared to other object tracking approaches, the grid-based measurement representation of this work also contains freespace information that is evaluated in addition to associated occupied grid cells. To fully utilize this extended measurement information, the object state is estimated by two different filtering techniques. For the dynamic motion state, a popular UKF-based estimation is used, with a specific object pose update by robustly selecting the most visible reference point regarding surrounding freespace. The motion estimation is further enhanced by a radar-based Doppler velocity measurement update. For the static object shape, as a novelty, a combination of a histogram filter geometry distribution estimation and an object classification concept is proposed. This enables 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 to use prior class knowledge of the assumed object size when it has not been fully observed yet.

- *Chapter 6* focuses on a detailed qualitative and quantitative evaluation of different aspects of the proposed grid-based object tracking approach. This work has been primarily tested with real sensor data of test vehicles in various complex scenarios in the context of autonomous driving, also using reference measurement systems for ground truth generation, which is further extended by challenging simulation scenarios.

- *Chapter 7* concludes this thesis and presents an outlook of potential improvements and extensions for future work.

2 Measurement Grid Representation and Fusion

This chapter introduces the generic occupancy grid representation of sensor measurement data, which forms the basis of this work. It is shown how different measurement data of lidar, radar, and camera are modeled and fused within the grid-based framework. The occupancy grid is modeled using an evidential representation based on the Dempster–Shafer theory of evidence, resulting in separate occupancy and freespace evidence masses. Additional sensor-specific measurement data, in particular radar Doppler velocities and camera classification information, are as well integrated into the grid-based measurement representation, modeled by separate grid layers, to avoid information loss of such important data. The individual measurement grids are fused robustly without requiring an explicit association by utilizing the spatial grid cell discretization. Thereby additional strategies for the spatiotemporal alignment of asynchronous sensor data and handling the activity and update status of the different sensors are proposed. The resulting fused measurement grid of this chapter represents unaccumulated measurement data, i.e., measurements from a single time instance, which forms the input for the temporal filtering of the dynamic grid mapping and the object tracking that are addressed in the subsequent chapters.

2.1 Introduction

The concept of occupancy grids is a popular way to model sensor-specific measurements in a uniform low-level representation by discretizing the local environment into grid cells and deriving occupancy probabilities for each cell. Various sensor data can thus be fused and temporally accumulated cell-wise without requiring specific object assumptions. In the following, different approaches of occupancy grids are briefly discussed before the contribution of this chapter is summarized. The focus is thereby placed on the representation of the occupancy state and the sensor data fusion strategy. The dynamic grid estimation including the temporal accumulation is addressed separately in Chapter 3.

2.1.1 Related Work

The original idea of *occupancy grid mapping* was proposed by Elfes [37] in the late 1980s. The local environment is discretized by quadratic grid cells in a planar 2-D representation, where the occupancy state of a grid cell is modeled using a Bayesian occupancy probability. It is based on the following key assumptions or approximations:

- *Binary state*: The state of a cell is either fully occupied or fully free, i.e., the area of a cell is not partially occupied and free, requiring a sufficiently fine cell discretization.
- *Cell independence*: Cells are independent of each other, which is essential for simplifying the estimation problem of the map into an efficient cell-individual estimation.

- *Known robot pose*: The pose of the robot is known, e.g., measured by odometry sensors, otherwise that would be a SLAM [127] problem.
- *Static world*: Each cell in the fixed world has a static state that does not change, i.e., no moving objects occur in the local environment.

This occupancy grid representation enables a direct sensor fusion and temporal accumulation with an implicit association, i.e., a cell-wise processing due to the uniform cell discretization. Occupancy probabilities are thereby combined using the Bayes rule, or, equivalently using a logarithmic representation, by the summation of the log-odds ratio of the occupancy probability and its counterpart. Overall, that occupancy grid mapping concept is widely used in various applications, often still using similar approximations and assumptions. A detailed overview of occupancy grid mapping and various extensions is given in [128]; the following brief discussion is limited to approaches that are relevant in the context of this work.

The assumption that grid cells are independent of each other enables an efficient occupancy grid computation by a parallel processing of the individual cells utilizing the high parallelization capability of a graphics processing unit (GPU), as presented in [50, 147]. Thereby the authors model range-based measurement data as occupancy grids that directly use the polar coordinate system of the respective measurement space, i.e., grid cells are discretized by the measurement angle and range. Those polar grids are then transformed to a uniform Cartesian space, as required for the data fusion and temporal accumulation, by a GPU-optimized texture mapping approximation [44]. However, that polar-to-Cartesian cell transformation may result in improper approximations, especially when multiple polar cells overlap with one Cartesian cell or vice versa. This could potentially also cause the disappearance of a small obstacle close to the sensor origin with a high angular resolution due to averaging that occupied polar cell with freespace cells of neighboring beams.

The efficient occupancy grid computation of [50] is utilized for the road boundary detection based on lidar and radar sensor data, which are compared but not fused in that work. Thereby radar measurements with non-zero absolute Doppler velocities are removed, i.e., such measurement information of moving parts is not considered. In [147], the occupancy grid fusion of multiple lidar sensors is demonstrated using the common log-odds representation with the assumption of a static environment. Similarly, a grid fusion of lidar, sonar, and monocular camera data based on the Bayes rule is presented in [115], with a specific occupancy threshold binarization adaptation and a neighborhood cell consideration for filtering the inaccurate sonar measurements.

Pagac et al. [91] introduced an *evidential occupancy grid representation* based on the Dempster–Shafer theory (DST) of evidence [34, 111], which enables a more generic model of uncertainties and conflicts by separate evidence masses of occupancy, freespace, and unknown instead of a single Bayesian occupancy probability. A similar approach is proposed by Moras et al. [80, 81], which further use the evidential hypotheses to determine conflicts and thereby detect moving objects as discussed separately in the dynamic grid estimation of the next chapter. The generic concept of such an evidential grid framework is also used in several other approaches and forms the basis of this work as well.

An evidential grid fusion of lidar and stereo camera is presented by Seeger et al. [109], where small and large obstacles are distinguished as separate hypotheses and the state of

neighboring cells is considered to improve the obstacle detection for narrow vertical field of view sensors. The same authors extend the fusion strategy in [110] using an obstacle association concept between different sensor grid maps by evaluating their individual covariance with a discounting strategy of imprecise sensor grids. However, the obstacle association is primarily intended for highway scenarios, whereas more cluttered urban scenarios might result in wrong associations. They also include radar sensors in that work, but, since only the static environment is considered, equivalently to [50], detections with a non-zero Doppler velocity are ignored, which is further restrained by filtering out all sensor data that can be associated to objects of a separate object tracking module.

In a different way, Nuss et al. [87] explicitly include the information of moving objects in the fusion of lidar and monocular camera data. The measurement data of both sensors are also represented as evidential grid maps, where the lidar derives occupancy evidence with a high position accuracy and the camera derives evidence of the image classification obtained by a scene labeling, with additionally estimating the spatial uncertainty of the camera-to-world projection. Overall, a fused evidential occupancy grid is extracted that contains not only occupancy evidence but also camera classification information. However, the approach remains error-prone to an inaccurate depth estimation of the camera. They also proposed to fuse lidar and radar data in [88], however, focusing on using radar Doppler velocities for the dynamic grid estimation without detailing the measurement grid fusion concept, which also applies to [121], as both discussed in more detail in Chapter 3. Recently, Valente et al. [130] presented a similar approach to fuse lidar and camera data, but using a stereo camera with a more accurate depth estimation, at least in the near range. They also use an evidential occupancy grid representation; however, occupancy evidence is directly derived from the stereo camera data, ignoring freespace, without including any additional camera segmentation information in the fused measurement grid.

Overall, there are various approaches for modeling and fusing sensor data using an occupancy grid representation. The specific sensor models generally depend on the deployed sensor type and are therefore usually not directly comparable. Evidential occupancy grids thereby represent a generic concept for robustly modeling different hypotheses and thus also uncertainties and conflicts. Since several of the grid-based approaches focus on estimating static environments, various measurement data of moving objects are often intended to be filtered out instead of explicitly including such measurement information, which also applies to radar Doppler velocities and camera segmentation information. For fusing measurement data of dynamic environments, however, measurements of moving objects obviously have to be considered. Thereby it is further important to take into account differences of the measurement time instances, since the environment may change in-between, which is not investigated by any of those grid-based fusion approaches. In the following, those different aspects are addressed to achieve a robust measurement grid fusion that forms the input of the subsequent dynamic grid mapping and object tracking.

2.1.2 Contribution and Outline

This chapter presents how measurement data of lidar, radar, and camera are reasonably modeled and robustly fused within the grid-based representation of this work. The main contributions are as follows within the respective sections:

- *Section 2.2* provides a brief introduction of the selected evidential occupancy grid representation including the spatially discretized grid structure and the evidential model that is based on the Dempster–Shafer theory of evidence.

- *Section 2.3* presents the sensor models that are required to derive measurement grids with the uniform evidential occupancy grid representation for the used lidar, radar, and camera sensors of this work. Moreover, as opposed to common grid-based approaches, the measurement grid representation is further extended by separate grid layers for radar Doppler velocities and camera classification information to retain such important additional measurement data in the grid-based framework.

- *Section 2.4* focuses on the grid-based data fusion that combines the individual sensor measurement grids. Thereby evidence masses are generically fused with the respective uncertainties, utilizing the evidential Dempster–Shafer framework, and, due to the spatial grid discretization with the cell-wise processing, no association or object abstraction is required. In contrast to similar grid fusion approaches, the problem of asynchronous sensor data is considered, where new strategies for the spatiotemporal alignment of measurement data and handling the activity status of the different sensors are proposed.

This chapter is completed by demonstrating some qualitative results including a brief summary in Section 2.5. In conclusion, the resulting fused measurement grid represents the combined unaccumulated measurement data of all integrated lidar, radar, and camera sensors with a full surrounding 360° detection area.

2.2 Evidential Occupancy Grid Representation

As motivated before, the basic idea of occupancy grids is to discretize the local environment into grid cells and estimate the occupancy state of each individual cell. This section briefly introduces the used spatial grid structure and the evidential occupancy representation.

2.2.1 Spatial Grid Structure

The environment is spatially discretized by a 2-D Cartesian grid, resulting in the individual grid cells $c \in \mathcal{G}$ of the local grid structure $\mathcal{G}$. As commonly used, all cells are arranged as equally sized squares in this work. The grid is referred to a fixed odometry coordinate system, where the position of the ego vehicle, relative to the starting location that denotes the origin of the coordinate system, is estimated by accumulating the relative ego-movement, measured by motion sensors.

The area of the relevant grid cells changes with the position of the ego vehicle, i.e., only the current local environment is considered, which is illustrated in Fig. 2.1 for two time instances. Thereby only integer translations Δx and Δy regarding the cell size are performed and rotations of the grid structure are avoided. Instead, the pose of the ego vehicle inside the local grid varies over time, as also used in [36, 49, 84, 118]. The ego vehicle can be positioned at the center of the relevant area, but also, for instance, placed

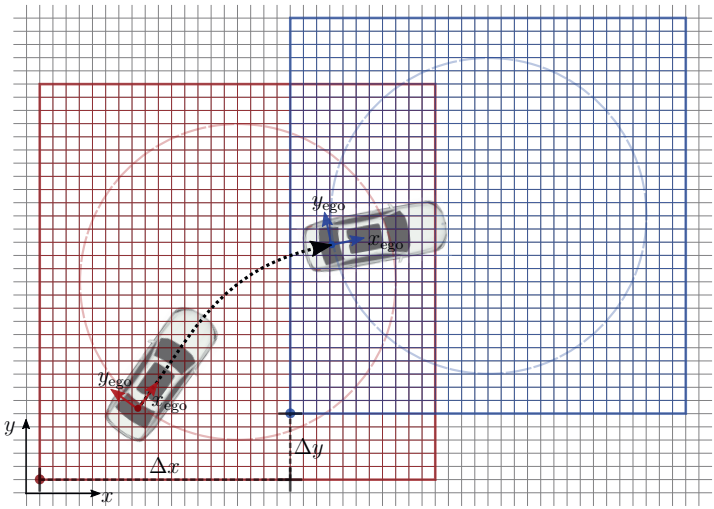


Fig. 2.1: Illustration of the grid structure with the fixed odometry coordinate system. The relevant area of the local grid adjusts with the pose of the ego vehicle, depicted by two time instances shown in red and blue, respectively. The fixed grid structure thereby ensures an exact spatial overlap of the individual cells between different time instances, whereas the pose of the ego vehicle within that local grid changes, here with a radial offset to the center of the local grid.

with a radial offset to that center, which then corresponds to a circular movement of the ego vehicle within the local grid, varying with a changing ego orientation, as also depicted in Fig. 2.1. Hence, this enables a greater perception range to the most critical area in the front without enlarging the local grid resolution. This position offset can also be applied for the local grid of each individual sensor and adapted to the respective sensor field of view, e.g., using an additional tangential offset for sensors that are primarily mounted to the side.

Overall, since occupancy grids of different sensors are also all modeled in that uniform odometry coordinate system, an exact spatial overlap of cells is achieved for both the temporal accumulation and the fusion of measurement data of different sensors. Therefore, the spatial discretization of the grid structure enables an implicit data association and fusion by a cell-wise processing, which is a generic and robust approach since no specific object assumptions are required.

2.2.2 Evidential Occupancy Representation

Based on the spatial grid structure, the occupancy state of each individual cell is estimated. Thereby the Dempster–Shafer theory (DST) of evidence [34, 35, 111] is used, which, in contrast to a single Bayesian occupancy probability, enables modeling separate hypotheses including their combined hypotheses and thus also a more generic representation of uncertainties. In the following, similar to [80, 91, 121], measurement data are modeled in the

occupancy grid by the two hypotheses occupancy (O) and freespace (F), resulting in the frame of discernment

$$\Theta_z = \{O, F\}\,, \tag{2.1}$$

which represents the set of all modeled hypotheses. In Chapter 3, based on [121], the occupancy hypothesis O is further divided into the two hypotheses static occupancy (S) and dynamic occupancy (D), which forms the basis for distinguishing static obstacles and moving objects as part of the temporal filtering. However, this chapter remains the hypotheses superset $O = \{S, D\}$ of occupancy without that static/dynamic differentiation.

In the DST framework, each hypothesis θ of the power set 2^Θ with the frame of discernment Θ, i.e., $\theta \in 2^\Theta$ or equivalently $\theta \subseteq \Theta$, is represented by an evidential basic belief mass $m(\theta)$, also denoted evidence mass in the following, with the basic belief assignment

$$m : 2^\Theta \to [0, 1]\,, \quad m(\emptyset) = 0\,, \quad \sum_{\theta \subseteq \Theta} m(\theta) = 1\,. \tag{2.2}$$

In terms of the defined frame of discernment $\Theta_z = \{O, F\}$, this means that the evidence masses of the individual hypotheses of occupancy, freespace, and the remaining unknown set have to fulfill the equation

$$m(O_z) + m(F_z) + m(\Theta_z) = 1\,. \tag{2.3}$$

The sum of the evidence masses of all subsets of a hypothesis θ is described in the DST framework by the belief

$$\mathrm{bel}(\theta) = \sum_{\tilde{\theta} \subseteq \theta} m(\tilde{\theta}) \;\leq\; p(\theta)\,, \tag{2.4}$$

which represents a lower bound of the Bayesian probability $p(\theta)$, whereas the corresponding upper bound is defined by the plausibility

$$\mathrm{pl}(\theta) = \sum_{\tilde{\theta} \supseteq \theta} m(\tilde{\theta}) \;\geq\; p(\theta)\,. \tag{2.5}$$

This also means that, based on the evidence masses of occupancy $m(O_z)$ and freespace $m(F_z)$ with the frame of discernment Θ_z as defined in (2.1), the Bayesian occupancy probability $p(O_z)$ can be derived, using the pignistic transformation [113], by

$$p(O_z) = \mathrm{bel}(O_z) + \frac{1}{2} m(\Theta_z) = \frac{1}{2} \left(\mathrm{bel}(O_z) + \mathrm{pl}(O_z)\right)\,, \tag{2.6}$$

with $p(O_z) = 0.5$ denoting an unknown state that also represents the prior here. That unknown state, however, cannot be distinguished from contradicting occupancy and freespace measurements in that Bayesian representation.

Overall, the occupancy grid model represents a generic, sensor-independent environment model. The evidential DST framework further enables estimating individual hypotheses as well as their supersets and the conflicts between different hypotheses, which is used for the sensor data fusion and especially later in the temporal filtering with the additional static/dynamic occupancy differentiation.

2.3 Sensor Measurement Grids

The previous section has summarized the concepts of the spatial grid discretization and the evidential occupancy model that result in a generic representation of range-based measurement data. This section discusses how the sensor-individual measurement grids of lidar, radar, and camera data, which form the input data of this work, are derived. Thereby the generic evidential occupancy grid derivation based on position measurements is presented, but also the representation of radar velocity measurements and camera classification information within that grid-based framework, as exemplarily illustrated in Fig. 2.2.

2.3.1 Generic Position-Based Evidential Occupancy Grid Derivation

As discussed before, there are several variants for determining occupancy grids based on sensor data. Those approaches can be divided into the following three aspects. First, a polar grid structure, equivalent to the polar measurement space, can be used with a subsequent transformation to the Cartesian space, e.g., with a texture mapping approximation approach [50, 147], or the Cartesian grid can be filled directly. Second, a forward sensor model can be used that incorporates the dependency between neighboring cells [126], or an inverse sensor model can be used that efficiently evaluates the occupancy of each cell independently. Third, the occupancy state of a cell can be modeled by a Bayesian occupancy probability or by evidence masses using the DST framework.

In the following, occupancy and freespace evidence masses are derived directly within a Cartesian 2-D grid representation with position measurements by an inverse sensor model, similar to [118]. Thereby measurements z_j, with the index set $j \in \mathcal{J}$ defining all currently occurring measurements of the corresponding sensor, are evaluated, but in the following only the measured 2-D position x_z^j of each measurement z_j is considered for the occupancy grid calculation. That position x_z is relative to the local grid structure, i.e., it is referred to the same odometry coordinate system as the position of the center of the cell $c \in \mathcal{G}$ that is defined by x_c. An illustration of a measurement occupancy grid derived from position measurements is depicted in Fig. 2.2a.

Occupancy Evidence

The occupancy evidence is modeled by a 2-D Gaussian distribution

$$\mathcal{N}(x;\mu,\Sigma) = \frac{1}{2\pi\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^{\mathsf{T}}\Sigma^{-1}(x-\mu)\right) \tag{2.7}$$

of the measurement with the measured position as the mean μ and the corresponding spatial uncertainty as the covariance matrix Σ, evaluated at the position x. The covariance Σ can be adapted, e.g., regarding the radial and tangential measurement direction.

In the following, as an approximation, the occupancy evidence mass $m(O_z^c)$ of the grid cell c is evaluated at the cell center x_c rather than calculating the exact 2-D integral of the Gaussian distribution over the area of the cell. Hence, for one measurement point x_z, the occupancy evidence mass $m(O_z^c)$ is proportional to $\mathcal{N}(x_c; x_z, \Sigma_z)$. More generally, for

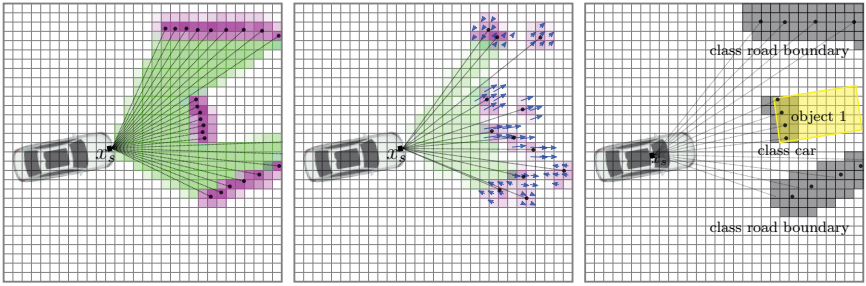


Fig. 2.2: Illustration of the different sensor measurement grids (the sensor origin position is denoted as x_s, measurements are depicted as black dots). (a) Generic evidential occupancy grid derivation from range-based measurements as primarily used for lidar data (pink: occupancy evidence, green: freespace evidence). (b) Adjusted evidential occupancy grid derivation for radar sensors with a higher uncertainty and additionally included radial velocity information (blue arrows). (c) Camera classification grid and camera-detected object (yellow box).

multiple measurement points x_z^j with $j \in \mathcal{J}$ affecting the cell c, the occupancy evidence is derived by a Gaussian mixture of the individual measurements, i.e.,

$$m(O_z^c) = \min\left(m_O^{\max}, \sum_{j\in\mathcal{J}} \alpha_O^{z_j,c}\,\mathcal{N}\left(x_c; x_z^j, \Sigma_z^j\right)\right). \tag{2.8}$$

The variable $\alpha_O^{z_j,c} \in \mathbb{R}_0^+$ additionally enables regulating the impact of the measurement z_j on the cell c, which can depend, for example, on the positive predictive value (PPV) or the signal-to-noise ratio (SNR) of the measurement, but also on the geometric evaluation like the ratio of the beam overlap of that measurement with that cell. This factor also regulates the overall saturation of the occupancy evidence in that cell, while the upper bound of the resulting cell occupancy evidence is limited by the parameter $m_O^{\max} < 1$.

Freespace Evidence

A cell that is currently not occupied or occluded, but visible regarding the current sensor field of view, i.e., not described by the unknown state, is defined as freespace. This freespace is typically not directly measured, but implicitly derived between the sensor origin and the first measured obstacle within that measurement beam. Therefore, a polar relation of the angles is required to derive the freespace evidence of the Cartesian grid. Thereby the relative positions of the measurement and the evaluated cell from the sensor origin x_s are considered in the following, i.e., the two vectors $(x_z^j - x_s)$ and $(x_c - x_s)$. The angle difference between those vectors is calculated by

$$\Delta\phi(x_c, x_z^j, x_s) = \arccos\left(\frac{(x_z^j - x_s)\cdot(x_c - x_s)}{\|x_z^j - x_s\|\cdot\|x_c - x_s\|}\right) \tag{2.9}$$

using the dot product of the two vectors and their corresponding magnitude. Hence, the relevant measurements of a cell are limited to the set

$$\mathcal{J}_c' = \{j \in \mathcal{J} \mid |\Delta\phi(x_c, x_z^j, x_s)| \leq \phi_F^{\max}\} \tag{2.10}$$

with the maximum considered angle difference $\phi_F^{\max}$, which has to be selected at least as the half of the angular sensor resolution, whereas larger values enable further considering neighboring beams. The freespace evidence is finally derived as

$$m(F_z^c) = \begin{cases} \min\left(m_F^{\max}\left(1 - m(O_z^c)\right), \sum\limits_{j \in \mathcal{J}_c'} \alpha_F^{z_j,c}\right), & \text{if } d_F^{\min} \leq \|x_c - x_s\| < \min\limits_{j \in \mathcal{J}_c'} \|x_z^j - x_s\| \\ 0\,, & \text{else}\,, \end{cases} \tag{2.11}$$

between a minimum distance $d_F^{\min} \geq 0$ to the sensor origin and the first measured obstacle in the determined angular interval with the relevant measurements $\mathcal{J}_c'$. Similar to (2.8), the parameter $\alpha_F^{z_j,c}$ further enables regulating the impact of the measurement z_j on the freespace of the cell c, e.g., the ratio of the beam overlap of that measurement with that cell. The maximum freespace evidence of a cell is limited by the parameter $m_F^{\max} < 1$, which is further reduced by the conflict mass with the occupancy evidence, i.e., the occupancy evidence dominates the freespace evidence, such that $m(O_z^c) + m(F_z^c) < 1$ with a certain remaining mass $m(\Theta_z^c) > 0$ of the unknown state to avoid singularities in the measurement grid fusion or the temporal accumulation.

Overall, the measurement occupancy grid is thus generally described by

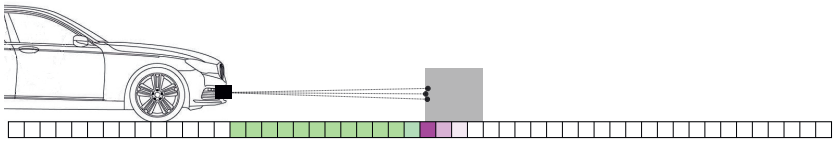
$$\mathcal{Z}_t = \{z_t^c \mid c \in \mathcal{G}\}, \quad z_t^c = \left[m(O_z^c),\, m(F_z^c)\right]^\mathsf{T} \tag{2.12}$$

with the individual cell measurements z_t^c. This occupancy and freespace derivation in the Cartesian grid structure is applicable to arbitrary range-based measurements. In the following, the measurement grids for the used lidar, radar, and camera sensors are briefly discussed, which also contain additional information modeled in separate layers, in particular radar Doppler velocities and camera classification and segmentation information.

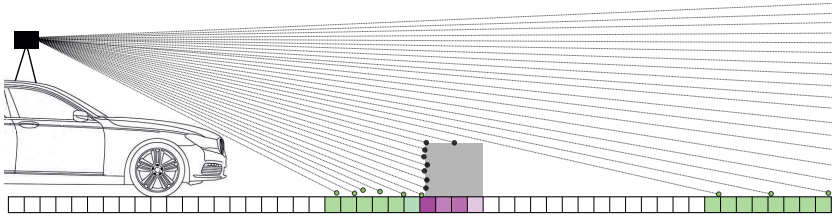
2.3.2 Lidar Measurement Grids

Lidar sensors provide accurate position measurements with a high angular resolution and precisely measured distances. Therefore, those sensors are well suited for deriving occupancy grids with high evidence masses for both occupancy and freespace. In this work, two different lidar configurations are used, which have to be handled slightly different, as briefly discussed in the following. Those two different configurations are illustrated in Fig. 2.3 and shown in an example with real sensor data in Fig. 2.4.

In the first configuration, shown in Fig. 2.3a, low-mounted lidar sensors with only a few vertical layers and a small vertical field of view are used, which are aligned approximately parallel to the roadway. Due to the low-mounted sensor position, the freespace can be derived directly between the sensor origin and the first measured obstacle in each beam segment. Hence, the generic position-based sensor model as defined in (2.7)–(2.11) and depicted in Fig. 2.2a can be used without modifications.



(a) Lidar configuration with a low-mounted sensor with only a few vertical layers.



(b) Lidar configuration with a high-mounted sensor with a high vertical resolution.

Fig. 2.3: Comparison of two lidar configurations with different mounting positions and sensor resolutions. The illustration shows an example with an obstacle (gray box) and the occurring measurement detections (black dots), while segmented ground points of (b) are additionally filled in green. The 1-D sideview perspective of the occupancy grid is shown on the ground of each configuration, respectively.

In the second configuration, shown in Fig. 2.3b, a high-mounted lidar sensor, above the roof of the ego vehicle, with a high vertical resolution and a full 360° horizontal field of view is used. Therefore, the number of measurement points is significantly higher than those obtained in the first configuration, resulting in a more accurate and denser environment representation. However, this also leads to many measurement reflections on the ground surface, and also some reflections of high obstacles that are passable underneath, which both should not be handled as obstacles for the 2-D occupancy grid. Hence, this requires a point cloud segmentation of the 3-D measurement data. For this pre-processing task, an efficient local ground plane estimation as proposed in [46] is used, and invalid obstacles that are too high are determined by a maximum height relative to the estimated ground plane. An example of the resulting point cloud segmentation is shown in Fig. 2.4b.

In that second configuration, due to the high-mounted sensor position, the vertical field of view close to the sensor is limited and partially occluded by the ego vehicle itself. This means that there could be undetected obstacles close to the ego vehicle underneath the vertical field of view. Therefore, the freespace is not directly derived from the sensor origin. Instead, it is only derived above a minimum distance $d_F^{\min}$, as included in the general freespace model in (2.11), which has to be selected large enough such that even a small obstacle is detectable at that distance within that beam segment. The field of view thereby depends on the relative sensor position, the pitch and roll movements of the ego vehicle, and the geometry of the ego vehicle as well as sensor mounting attachments. However, the raw lidar measurement data also contain detections on the ground, which have been segmented by the ground plane estimation as described before. This means that the freespace can

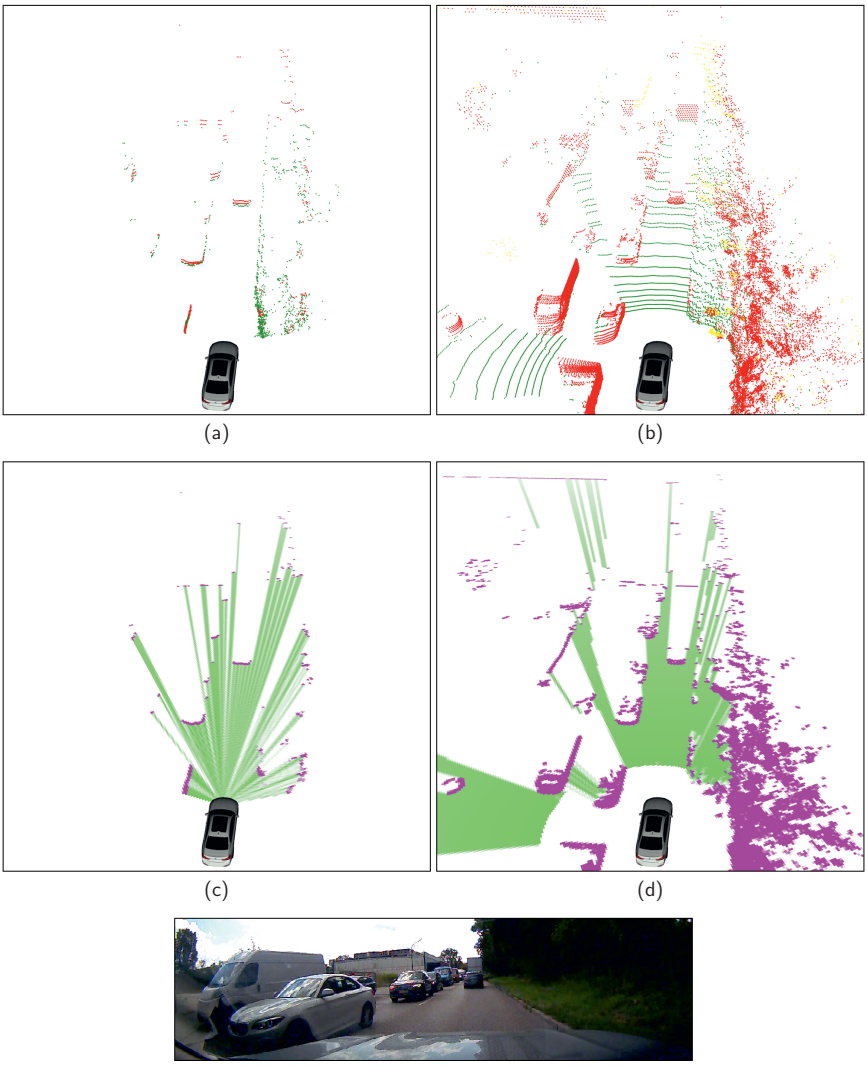


Fig. 2.4: Example of the lidar measurement data and the corresponding measurement grid with the two different sensor configurations. (a) Measurement data of the low-mounted front lidar sensor with only a few vertical layers. (b) Measurement data of the high-mounted lidar sensor with a high vertical resolution. The color coding additionally shows the point segmentation (green: ground, red: valid obstacles, yellow: passable underneath). (c) Measurement grid of the sensor data of (a). (d) Measurement grid of the sensor data of (b). (e) Camera image.

also be directly computed from the measurement without explicitly evaluating the sensor field of view, thereby representing the most conservative choice with actually measured freespace in contrast to an implicit beam derivation. Hence, freespace within one beam segment is derived in distance intervals that contain only ground points, i.e., the lower and upper distance thresholds are defined by the first and last measured consecutive ground point without any in-between obstacles. Those measured ground points further enable deriving freespace behind smaller obstacles, as also illustrated in 2.3b.

Overall, lidar measurement grids, for both sensor configurations, are modeled with high evidence masses for occupancy as well as freespace, since they generally have a high spatial accuracy. They form the main sensors for the occupancy grid mapping and thus also for the detection and the pose and shape estimation of moving objects of this work.

2.3.3 Radar Measurement Grids

Radar sensors are widely used in various perception systems, in particular due to their robustness in different weather conditions, the high detection range, the ability to directly measure velocity components by the Doppler effect, and the advanced level of industrialization with relatively low production costs compared to lidar sensors. Detailed information about the physical measurement principle based on radio waves and different applications of radars can be found, for instance, in [59, 112, 142].

Similar to lidar detections and range-based measurements in general, radar detections, extracted by signal processing of the raw frequency measurement, also represent obstacles measured in the polar space by the corresponding distance and azimuth angle. The radar measurement occupancy grid, as illustrated in Fig. 2.2b, is thus also derived as described before with the position measurements x_z^j, which is directly computed in the Cartesian space of the grid. As with all sensors, thereby the sensor uncertainty has to be considered for the parameterization of the occupancy and freespace derivation. Most commonly, the spatial accuracy of radars, in particular the angular accuracy, is lower in comparison to lidar sensors. In such a case, the occupancy derivation of (2.8) has to be selected more conservatively with a higher spatial uncertainty Σ_z^j and a lower amplitude $\alpha_O^{z_j,c}$, in particular for measurements with a low radar cross section (RCS) [64], which may further depend on the ratio of false positives or additional measurement effects like multi-path reflections. Equivalently, also the freespace derivation of (2.11) has to be adjusted then. In addition, since the angular resolution is not directly fixed for some radar sensors, i.e., radar detections may occur at arbitrary angles without specific beam segments, the freespace limitation regarding the first obstacle of neighboring measurements may be selected more restrictive by considering a higher maximum angle difference $\phi_F^{\max}$ in (2.10). An example of real radar detections with a conservatively-modeled evidential occupancy grid is demonstrated in Fig. 2.5c–2.5d.

Another very important aspect of radar sensors is the unique ability to actually measure velocity components by the Doppler effect. This direct velocity measurement, i.e., not implicitly derived by the position difference, significantly improves the dynamic estimation on both the grid cell-level and the object-level, as discussed in the subsequent chapters. Therefore, it is important to retain that measurement information and include it within the measurement grid representation.

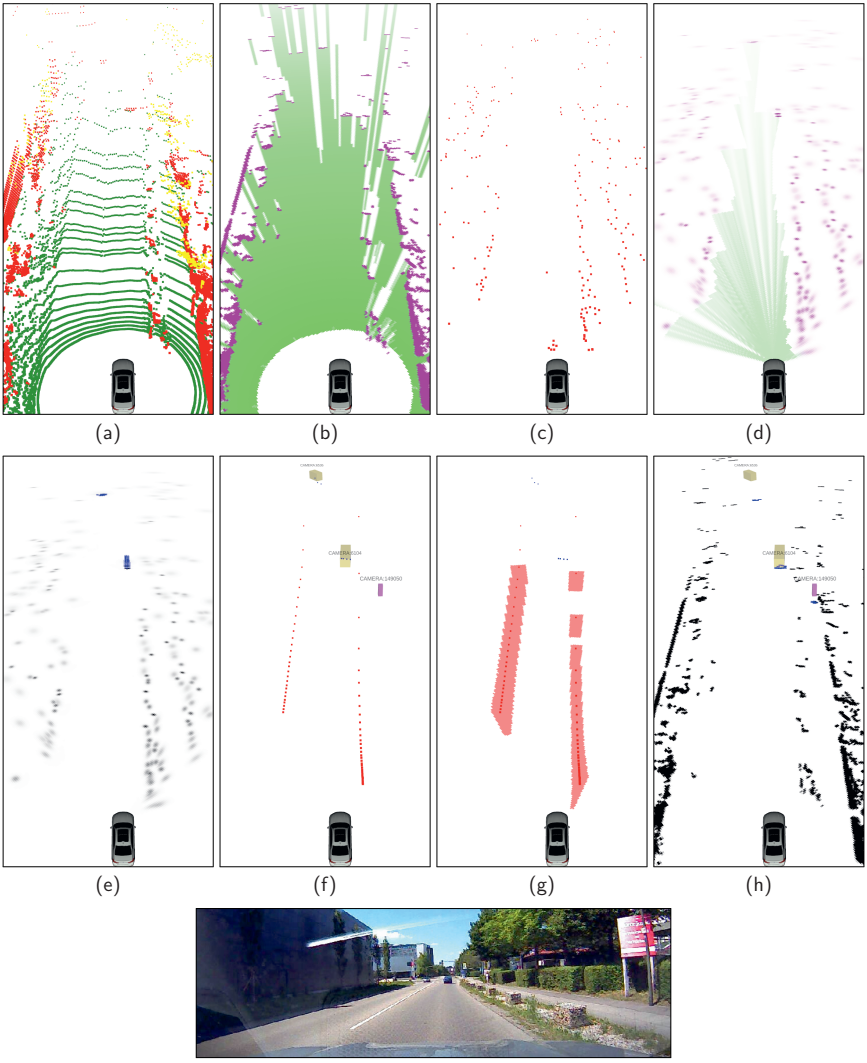


Fig. 2.5: Example of the measurement data of the different sensor types and their corresponding measurement grid representations. (a) Lidar detections (high-mounted sensor). (b) Lidar measurement occupancy grid. (c) Radar detections. (d) Radar measurement occupancy grid. (e) Radar velocity grid (blue lines). (f) Camera detections and objects. (g) Camera classification grid of static detections (red cells). (h) Camera classification grid of objects with additional spatial camera–lidar alignment (blue cells). (i) Image of the scenario.

In the following, a radar velocity measurement is described by the vector

$$z^r = \left[v_z^r,\, \theta_z^r,\, x_s\right] \tag{2.13}$$

with the radial velocity component v_z^r, the corresponding azimuth angle θ_z^r, and the 2-D position x_s of the sensor origin. All components are referred to the fixed odometry coordinate system, i.e., the angle θ_z^r corresponds to the angle between that radial measurement direction and the x-axis of the grid structure. Accordingly, the radial velocity

$$v_z^r = v_{z,\mathrm{rel}}^r + v_{\mathrm{ego},s}^r\,, \tag{2.14}$$

describes the absolute velocity rather than the originally measured relative velocity $v_{z,\mathrm{rel}}^r$, i.e., it is already compensated by the velocity $v_{\mathrm{ego},s}^r$ of the ego vehicle evaluated at the position x_s of the sensor s. The velocity component $v_{\mathrm{ego},s}^r$ of the measuring sensor depends on the longitudinal velocity v_{ego} and the turn rate ω_{ego} of the ego vehicle, which are both referred to the center of the rear axle as the assumed center of rotation. However, the position differences between the sensor s and that reference position of the ego vehicle within the object coordinate system (OCS), i.e., the fixed mounting position within the ego vehicle, have to be taken into account, which are defined by $\Delta x_s^{\mathrm{OCS}}$ and $\Delta y_s^{\mathrm{OCS}}$. The 2-D velocity of the sensor within the OCS of the ego vehicle is then determined by

$$\begin{bmatrix} v_{\mathrm{ego},s}^{x,\mathrm{OCS}} \\ v_{\mathrm{ego},s}^{y,\mathrm{OCS}} \end{bmatrix} = \begin{bmatrix} v_{\mathrm{ego}} - \omega_{\mathrm{ego}}\,\Delta y_s^{\mathrm{OCS}} \\ \omega_{\mathrm{ego}}\,\Delta x_s^{\mathrm{OCS}} \end{bmatrix}\,, \tag{2.15}$$

as described in more detail in [59], which is also similarly discussed in Section 5.4.2 in the context of the radar-based velocity estimation of surrounding target vehicles. The compensating velocity of the sensor is finally calculated as

$$v_{\mathrm{ego},s}^r = \cos(\theta_z^r - \varphi_{\mathrm{ego}})\, v_{\mathrm{ego},s}^{x,\mathrm{OCS}} + \sin(\theta_z^r - \varphi_{\mathrm{ego}})\, v_{\mathrm{ego},s}^{y,\mathrm{OCS}} \tag{2.16}$$

with the angular difference $\theta_z^r - \varphi_{\mathrm{ego}}$ between the azimuth angle of the radial component and the orientation of the ego vehicle.

In order to combine this velocity information with the position measurements of the occupancy grid, each radar velocity measurement z^r is also represented in the discretized grid structure for the individual grid cells $c \in \mathcal{G}$. Equivalent to the occupancy derivation, the velocity information is also spread over several cells to model the spatial uncertainty, but also the assumed similarity of the velocity component in neighboring cells. Thereby at most one radial velocity measurement is represented in a cell, requiring that the best measurement regarding the highest occupancy evidence is selected by

$$j^* = \underset{j \in \mathcal{J}}{\arg\max} \left(\alpha_O^{z_j,c}\, \mathcal{N}\left(x_c; x_z^j, \Sigma_z^j\right) \right), \tag{2.17}$$

using the occupancy derivation of each measurement j as defined in (2.8). The measured velocity information of a cell $c \in \mathcal{G}$ finally results in

$$z_v^c = z_{j^*}^r \quad \forall \left\{ c \in \mathcal{G} \mid m(O_z^c) > m_O^{\min,v} \right\}, \tag{2.18}$$

while only cells above an occupancy evidence $m_O^{\min,v} > 0$ are considered for containing the corresponding velocity measurement.

Overall, radar Doppler sensors enable also to derive measurement velocity grids, which are modeled as two additional layers of each measurement grid to represent the 2-D velocity vectors, i.e., the radial velocity v_z^r and the corresponding measurement direction θ_z^r. Those measured radial velocities are visualized as additional arrows in the illustration of Fig. 2.2b as well as in Fig. 2.5e. Moreover, as used later for the radar-based object state estimation in Section 5.4, the position of the radar sensor origin x_s is also contained in the measurement grid to avoid bias errors of the cell positions due to the grid discretization. In the current processing step, for the measurement grid of one sensor, this sensor position is equal for all cells and thus included independently of the individual cells.

2.3.4 Camera Measurement Grids

Camera sensors are also very popular for various perception applications, as they are well-developed, relatively cheap to produce, and computer vision is essential for a detailed scene understanding that is similar to human vision. The measurement principle of camera sensors, however, is significantly different to lidar or radar sensors, since, in contrast to such generic range-based measurements, relevant information has to be extracted first from the image, by segmentation and classification, before it can be further processed. As computer vision is a very broad and complex field with various approaches, it is not covered in the scope of this thesis. Instead, the image processing parts as well as the projection into the world coordinate system are assumed to be pre-processed in advance and available as input to this work. In the following, two kinds of extracted measurement data of the camera sensors are used. First, primarily for the estimation of surrounding traffic participants, extracted high-level objects with a bounding box shape are available. Second, primarily for the additional estimation of the static environment, detections that represent the first obstacle in terms of non-drivable area in an angular segment using a polar discretization are available. Thereby both objects and detections include classification information, whereby the instance-based segmentation information is only present for objects. Both representations are depicted in Fig. 2.2c and shown with real sensor data in Fig. 2.5f.

Based on such pre-processed camera measurement data, especially using the detection-based representation, occupancy grids can be derived similarly to those of the lidar and radar sensors. However, the depth estimation based on camera images is generally rather inaccurate, which is also the case for the cameras used in this work. This means that the distance error of camera-extracted objects and obstacles can be high, as demonstrated in Fig. 2.5f for the most distant vehicle, for instance. Hence, it would be error-prone to directly derive significant occupancy and freespace evidence masses of that sensor, since the occupancy evidence masses of all sensors are fused cell-wise as discussed in the next section, which requires a high position accuracy. Therefore, occupancy and freespace evidence masses are not directly derived from the camera data in the following. Instead, the camera information is primarily used for the segmentation as well as the classification, which are modeled as separate grid layers by

$$z^{\kappa} = \left[\kappa_{\text{class}}^{\text{cam}},\, \kappa_{\text{id}}^{\text{cam}}\right] , \tag{2.19}$$

with $\kappa_{\text{class}}^{\text{cam}}$ denoting the camera classification and $\kappa_{\text{id}}^{\text{cam}}$ denoting the segmentation in terms of an object identifier or index that links a grid cell to an instance of the camera object list. This additional information is then used to classify occupied cells of other sensors; hence it can be derived with a higher spatial uncertainty here. This classification strategy is directly applied for the camera-based obstacle detections of the static environment, with a modeled high uncertainty in the radial direction as shown in Fig. 2.5g. The camera-detected objects are handled differently, as briefly discussed in the following.

Spatial Camera–Lidar Alignment

To further improve the position accuracy of the camera objects, in particular in larger distances, an existing camera–lidar alignment strategy [119] is applied; other similar approaches can be found, for instance, in [14, 31, 150]. The basic idea here is to correct the distance of camera objects by lidar data if matching lidar detections can be found in that angular camera segment with a maximum allowed relative distance difference. As a result, camera objects are directly linked to lidar detections, where also the camera classification and segmentation of objects is then derived with the same spatial accuracy as the occupancy evidence of the lidar detections, as exemplary demonstrated in Fig. 2.5h. This strategy can also be further extended using an association of the camera objects with radar detections, in particular when no corresponding lidar data were found in the angular camera segment. If no matching range-based sensor data are found for a camera-detected object, this object is assumed to be a false positive, i.e., a wrongly extracted object.

2.4 Measurement Grid Fusion

The previous section has presented how measurement data of different sensors are represented as individual measurement grids using the evidential occupancy and freespace hypotheses and the spatial grid cell discretization. This section focuses on the subsequent sensor data fusion of those different measurement grids to obtain one fused measurement grid, which forms the input to the measurement update of the grid mapping and object tracking of the subsequent chapters.

2.4.1 Basic Cell-Wise Fusion of Evidence Masses

The occupancy grids of all sensors are discretized using the fixed odometry coordinate system, ensuring an exact spatial overlap of all grid cells of different sensors as described in Section 2.2.1. Therefore, the sensor data fusion can be performed for each individual cell of the uniform grid structure by combining the evidence masses of each sensor for that particular grid cell. Hence, no explicit association between the measurements of different sensors is required, since the grid structure implicitly performs that cell-wise association.

General Combination Principle of Evidential Belief Masses

The evidence masses of different sources, e.g., measurements of multiple sensors, can be fused within the Dempster–Shafer evidential framework with various rules of combination,

while the combination is typically denoted by the symbol $\oplus$. The basic idea is formed by the intersection of different hypotheses, resulting in the conjunctive rule of combination

$$m(\theta|\mathcal{Z}_{s_1} \oplus_\mathrm{c} \mathcal{Z}_{s_2}) = \sum_{\alpha\cap\beta=\theta} m(\alpha_{s_1})m(\beta_{s_2})\,, \quad \theta,\alpha,\beta \subseteq \Theta\,, \tag{2.20}$$

based on the evidential data $\mathcal{Z}_{s_1}$ and $\mathcal{Z}_{s_2}$ of the two sources s_1 and s_2, respectively, which is denoted in the following by the conditional term $\mathcal{Z}_{s_1} \oplus_\mathrm{c} \mathcal{Z}_{s_2}$. Thereby θ describes the evaluated hypothesis of the frame of discernment Θ, whereas α and β describe all possible hypotheses of the sources s_1 and s_2, respectively, within that same frame of discernment Θ, such that the intersection of those hypotheses results in the evaluated hypothesis θ of the combined output.

Similar to (2.20), the conflict mass

$$\zeta(\mathcal{Z}_{s_1}, \mathcal{Z}_{s_2}) = \sum_{\alpha\cap\beta=\emptyset} m(\alpha_{s_1})m(\beta_{s_2}) \tag{2.21}$$

describes the evidence of the sum of intersections resulting in the empty set $\emptyset$, which is caused by evidence masses of contradicting hypotheses. This conflict mass can be assigned in different ways toward the evidence masses of the individual hypotheses $\theta \subseteq \Theta$, taking into account that the evidence masses of all subsets of Θ including Θ itself have to sum up to 1 as defined in (2.2). A popular approach is Dempster's rule of combination [111]

$$m(\theta|\mathcal{Z}_{s_1} \oplus_\mathrm{d} \mathcal{Z}_{s_2}) = \frac{m(\theta|\mathcal{Z}_{s_1} \oplus_\mathrm{c} \mathcal{Z}_{s_2})}{1-\zeta(\mathcal{Z}_{s_1}, \mathcal{Z}_{s_2})}\,, \quad \zeta(\mathcal{Z}_{s_1}, \mathcal{Z}_{s_2}) < 1\,, \tag{2.22}$$

which normalizes the conjunctive combination $\oplus_\mathrm{c}$ by an inverse of the conflict mass, i.e., the conflict mass is assigned proportionally to the corresponding evidence masses of the intersections. Another possibility is an assignment of the conflict mass to concrete hypotheses, e.g., the unknown set Θ, the union set $\alpha \cup \beta$, or a hypothesis-specific assignment modeled by expert knowledge [70]. The latter is used in Chapter 3 to solve the prediction and measurement update of the dynamic grid mapping with specific behavior.

For more than two sources, the combination is simply performed iteratively by

$$m(\theta|\mathcal{Z}_{s_1} \oplus \mathcal{Z}_{s_2} \oplus \cdots \oplus \mathcal{Z}_{s_N}) = m\left(\theta|\left((\mathcal{Z}_{s_1} \oplus \mathcal{Z}_{s_2}) \oplus \dots\right) \oplus \mathcal{Z}_{s_N}\right)\,, \tag{2.23}$$

since the used evidential combination approaches fulfill the associative property [123].

Cell-Wise Evidential Measurement Grid Fusion

In the context of the measurement grid fusion of this chapter using the frame of discernment $\Theta_z = \{O, F\}$ with the two individual hypotheses occupied O and freespace F, the conjunctive rule of combination of (2.20) of a cell $c \in \mathcal{G}$ of the grid structure $\mathcal{G}$ results in

$$m(O_z^c|\mathcal{Z}_{s_1} \oplus_\mathrm{c} \mathcal{Z}_{s_2}) = m(O_{z,s_1}^c)m(O_{z,s_2}^c) + m(O_{z,s_1}^c)m(\Theta_{z,s_2}^c) + m(\Theta_{z,s_1}^c)m(O_{z,s_2}^c)\,, \tag{2.24}$$

$$m(F_z^c|\mathcal{Z}_{s_1} \oplus_\mathrm{c} \mathcal{Z}_{s_2}) = m(F_{z,s_1}^c)m(F_{z,s_2}^c) + m(F_{z,s_1}^c)m(\Theta_{z,s_2}^c) + m(\Theta_{z,s_1}^c)m(F_{z,s_2}^c)\,. \tag{2.25}$$

As described before, the conflict mass

$$\zeta(\mathcal{Z}_{s_1}, \mathcal{Z}_{s_2}) = m(O_{z,s_1}^c)m(F_{z,s_2}^c) + m(F_{z,s_1}^c)m(O_{z,s_2}^c)\,, \tag{2.26}$$

which is caused here by contradicting occupied and freespace measurements, can be assigned in different ways. The most conservative choice in the context of collision avoidance would be an assignment to the occupancy mass $m(O_z^c)$, meaning that an obstacle that is measured by at least one sensor would always dominate measured freespace of other sensors. However, such a strategy would also be error-prone to false positives in terms of ghost obstacles and objects, which could cause, for example, critical emergency brake maneuvers. Instead, the evidence masses of the sensor measurement grids, as described in the previous section, are modeled and weighted in a way such that they represent the certainty of that particular sensor measurement. Therefore, the conflict mass is assigned proportionally to both contradicting hypotheses by (2.22), i.e., the measurement grids are fused cell-wise using the common Dempster's rule of combination. The remaining unknown mass of the combination thereby results in

$$m(\Theta_z^c|\mathcal{Z}_{s_1} \oplus \mathcal{Z}_{s_2}) = 1 - m(O_z^c|\mathcal{Z}_{s_1} \oplus_\mathrm{d} \mathcal{Z}_{s_2}) - m(F_z^c|\mathcal{Z}_{s_1} \oplus_\mathrm{d} \mathcal{Z}_{s_2}) > 0\,. \tag{2.27}$$

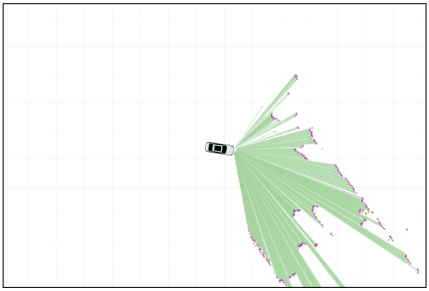
An example of that cell-wise measurement grid fusion with real measurement data of five surrounding lidar sensors is shown in Fig. 2.6, which can be extended with an arbitrary number of additional measurement grids of any sensor type.

2.4.2 Spatiotemporal Alignment of Asynchronous Sensor Data

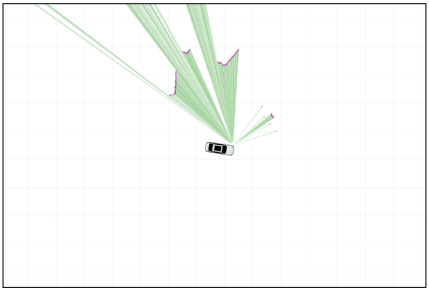
It has been shown how occupancy and freespace evidence masses of sensor-individual measurement grids are fused cell-wise using Dempster's rule of combination, resulting in a fused measurement grid with all measurement data representing the same time instance. However, the measurement time of different sensors is typically not exactly synchronized, e.g., due to varying measurement rates or an intentional offset to minimize mutual interferences of active sensors with the same physical measuring principles and an overlapping field of view. Similarly, also the latency between the measurement time and the time when the data is actually available as input to the sensor fusion algorithm typically varies for different sensor types depending on the sensor-specific signal processing steps and also on the data communication.

For static environments, it is sufficient to evaluate the pose of the ego vehicle and thus the sensor origin at the corresponding measurement time. Since each sensor measurement grid is represented in reference to the fixed odometry coordinate system, this time instance is important to correctly map the raw measurements, which are relative to the sensor origin, to that fixed grid structure. Hence, the measurement grid fusion can be performed for all available measurement grids in static environments without further evaluating their time instances, as they are already represented in that uniform fixed coordinate system.

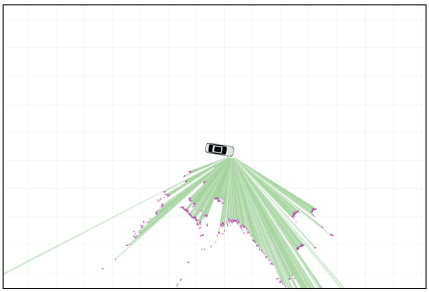
For dynamic environments, however, time differences between individual measurements are problematic, since the environment itself may change between different measurements, e.g., due to surrounding moving objects, which cannot be directly compensated. As illustrated in Fig. 2.7, that problem of asynchronous sensor data is particularly relevant for the low-level cell-wise fusion of this work, since spatial differences of position measurements with varying time instances result in offsets of several grid cells due to the fine grid resolution. Hence, measurements of the same moving object may occur at completely different grid cells, causing multiple occupancy shapes and occupancy/freespace conflicts.



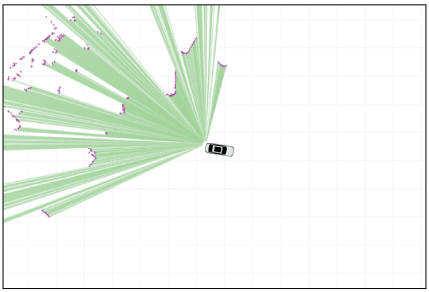
(a) Measurement grid of lidar front.



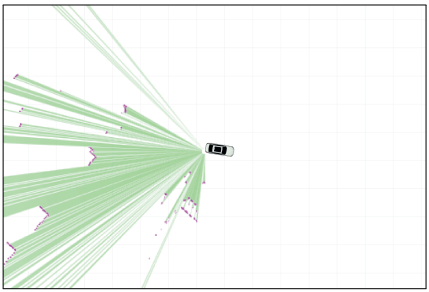
(b) Measurement grid of lidar front-left.



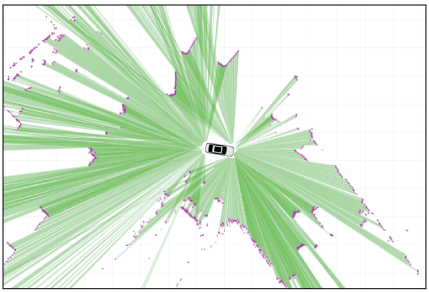
(c) Measurement grid of lidar front-right.



(d) Measurement grid of lidar rear-left.



(e) Measurement grid of lidar rear-right.



(f) Fused measurement grid of the 5 lidar sensors.

(g) Camera rear.

(h) Camera front.

Fig. 2.6: Example of the cell-wise measurement grid fusion of the evidence masses of the sensor-individual measurement grids of the five low-mounted lidar sensors of a real test vehicle.

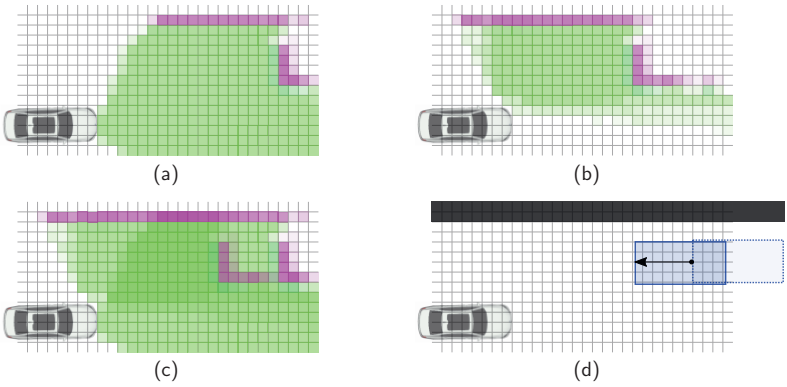


Fig. 2.7: Illustration of the problem of asynchronous sensor data fusion of measurement grids in dynamic environments without considering time differences. (a) Latest measurement grid at t_{s_1} of sensor s_1 (e.g. front-middle). (b) Latest measurement grid at t_{s_2} of sensor s_2 (e.g. front-left), here with $t_{s_2} \gg t_{s_1}$. (c) Fused measurement grid of (a) and (b). (d) Illustration of the actual scenario with a static road boundary (gray) and an oncoming moving object (blue).

A naive sensor data fusion approach is thus to combine the latest measurement data of all sensors, without evaluating their original measurement times, i.e., ignoring the additionally included individual processing times and communication latencies of all sensors. To avoid these mentioned fusion ambiguities or artifacts, however, it is important to consider those time differences and to determine which sensor data have to be included in the current measurement grid fusion. In contrast to high-level fusion approaches that combine sensor-individually tracked objects, e.g., [8], no direct prediction of the individual sensor data to one combined time instance is possible, since no objects are abstracted and no dynamic estimation is available at this low-level processing step. One exception are radar detections that can be predicted by the measured Doppler velocity when the movement direction is known, which, at least for fast-moving objects such as highway scenarios, can often be approximated by the heuristic of a movement direction parallel to the ego vehicle. That concept has been proposed in [5] as preliminary work of this thesis. In this work, however, primarily urban scenarios are considered, where that heuristic is generally not valid. Therefore, the remaining section focuses on a generic measurement time interval alignment strategy, as also published in [5], combined with a newly proposed fusion logic that determines the activity and update status of all individual sensors.

Measurement Time Interval Alignment

The major part of time-caused fusion errors can be minimized by considering the original measurement times rather than the times when the individual measurement grids are available, i.e., without the additional latency. In the following, it is thereby assumed that the measurement times of the individual sensors are referred to one system clock without any sensor clock synchronization errors. This strategy requires a buffering of the

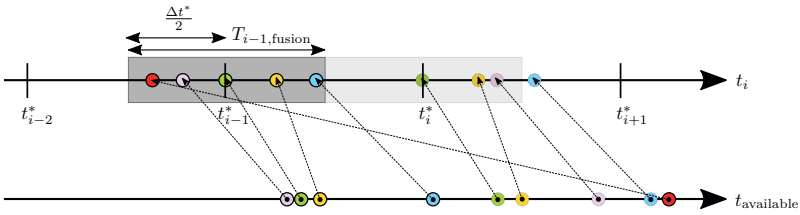


Fig. 2.8: Measurement time interval alignment approach, which evaluates the original measurement times t_i (upper row) rather than the time $t_{\text{available}}$ (lower row) when the measurement data are available for the fusion, delayed by sensor and communication latencies. The fusion interval $T_{i-1,\text{fusion}}$ (dark gray area) is defined by the measurement time t^*_{i-1} of a reference sensor with the corresponding update rate Δt^*. The colored circles indicate the time instances of occurring measurements with individual colors for each sensor.

sensor data, meaning that the fusion has to wait for the sensor data with the highest latency, causing a trade-off between the accuracy of the fusion and the overall latency of the system. As proposed in [5], the basic idea is to use one sensor with measurement times t^*_i as a reference sensor of the fusion, which defines corresponding time intervals

$$T_{i,\text{fusion}} = \left[t^*_i - \frac{t^*_i - t^*_{i-1}}{2},\, t^*_i + \frac{t^*_{i+1} - t^*_i}{2} \right), \tag{2.28}$$

or, if the measurement update rate Δt^* of that reference sensor is constant, by

$$T_{i,\text{fusion}} = \left[t^*_i - \frac{\Delta t^*}{2},\, t^*_i + \frac{\Delta t^*}{2} \right), \quad \text{with } \Delta t^* = t^*_i - t^*_{i-1} \ \forall i\,, \tag{2.29}$$

which then enables to define the fusion interval $T_{i,\text{fusion}}$ without waiting for the next measurement of that reference sensor at t^*_{i+1}.

All measurement grids, where the original measurement times are within that fusion time interval $T_{i,\text{fusion}}$, are included in that fusion and thus approximated by the time t^*_i, while each measurement is included exactly once in the fusion process. By that interval alignment approach, the maximum fusion time error of each included measurement grid is limited to $\Delta t^*/2$, as also illustrated in Fig. 2.8. As noted before, possible fusion time errors thereby only affect moving parts of the environment, since the measurement grid itself, i.e., the mapping of the raw sensor data to the fixed odometry frame, is derived before that fusion step with the original measurement time of that sensor.

As an example, with a reference sensor with a measurement frequency of 25 Hz, the time interval corresponds to $\Delta t^* = 40\,\text{ms}$, limiting the maximum time error to $\Delta t^*/2 = 20\,\text{ms}$. For urban scenarios with typical velocities up to 50 km/h, the maximum time-caused fusion error is thus $0.02\,\text{s} \cdot 13.89\,\text{m/s} = 0.28\,\text{m}$, which, with a grid cell size of 0.15 m, corresponds to an error of 1–2 grid cells. This error is commonly negligible as it is not significantly larger than the spatial uncertainty of most sensors, whereas possible latency differences of about 110 ms would result in errors of up to 10 grid cells for a fusion of the latest measurement grids without any interval alignment, which would significantly distort the fusion result and thus also the subsequent processing steps.

The selection of the reference sensor thereby depends on different aspects. Generally, to minimize that error, the sensor with the highest measurement frequency, i.e., the smallest intervals Δt^*, should be selected as the reference sensor. This further guarantees that at most one measurement of each sensor is included in the fused measurement grid. Additionally, since the area in front of the ego vehicle is typically most time-critical, a sensor with a field of view in that area should be selected. Moreover, the spatial accuracy of a sensor should be taken into account, as the relevance of the time-error reduces when the sensor already has a high measurement uncertainty. Most commonly, a spatially-accurate lidar sensor that is oriented to the front area represents a suitable choice as a reference sensor for the time-interval fusion, which is also the case for the sensors as used in this work.

Overall, this simple measurement time interval alignment already significantly reduces critical time errors caused by high latencies of some sensors. However, this approach requires buffering sensor data and a decision when the fusion should be performed, in particular when measurement data of some sensors are missing within the fusion interval $T_{i,\text{fusion}}$, i.e., a tradeoff between waiting for sensor data and reducing the overall system latency. This sensor data handling is briefly addressed in the following.

Sensor Activity and Update Status

Fusing measurement data with the measurement time interval alignment requires a concept for determining which sensors are active, which sensor data within a fusion interval are already available, and which sensor data are still pending due to a higher latency. The corresponding fusion logic is presented as pseudocode in Algorithm 1.

For each incoming measurement Z with the time t_z of the sensor $s_z \in S$, with S representing the set of all sensors, the function MeasurementGridCallback(s_z, Z, t_z) is called, which first sets the activity status of this sensor s_z to active and separately saves t_z as the latest time instance of it. If the measurement is not too old, i.e., if it is within the next fusion interval or newer ($t_z \geq t_{\text{fusion}} - \Delta t/2$) with $t_{\text{fusion}} \pm \Delta t/2$ representing the next fusion time instance and interval width, that measurement is buffered by adding it to the sensor data queue. After that it is checked if the fusion should be performed, which is the case if no active sensor is pending, or, as a fallback to reduce the overall latency, if the time t_z is significantly newer that the fusion time instance t_{fusion} with the time difference parameter $t_{\text{max}}^{\text{wait}} > \Delta t/2$. If no measurement data of a sensor are available in the current fusion interval, but newer data of that sensor are already present for one of the following fusion intervals, e.g., due to a slower measurement rate, the fusion should also be triggered, since it is assumed that the measurement data of one sensor arrive in time order, i.e., no older data are expected to arrive for that sensor. Hence, active sensors are only interpreted as pending if no buffered data are in their queue, as determined by AnyActiveSensorPending(). If the fusion should be performed, it is evaluated which sensors and which of their measurement data are included in the fusion, where sensors that are not included in the fusion are set to inactive if their latest measurement time is significantly older than the fusion time with the time difference parameter $t_{\text{max}}^{\text{inactive}} > \Delta t/2$.

Overall, this measurement time interval alignment strategy combined with the fusion logic represents a robust approach for minimizing errors of the asynchronous sensor data fusion. Moreover, it enables to automatically detect when sensors are inactive, where the

Algorithm 1 Pseudocode of fusion logic with buffered sensor data and sensor activity status. The fusion logic is updated by sensor data callbacks, given the sensor s_z, measurement data Z, and corresponding measurement time t_z. Further parameters: next fusion time t_{fusion}, fusion interval length Δt, set of all sensors S, max. time $t_{\text{max}}^{\text{wait}} > \Delta t/2$ to wait for sensor data, max. time $t_{\text{max}}^{\text{inactive}} > \Delta t/2$ before sensor is set to inactive.

```
function MeasurementGridCallback(s_z, Z, t_z)
    s_z.active = true                          ▷ Set sensor with incoming meas. to active
    s_z.t_latest = t_z        ▷ To check inactivity, separately save latest meas. time of sensor
    if t_z ≥ t_fusion - Δt/2 then      ▷ Only process meas. within fusion interval or newer
        s_z.queue.push(Z)                              ▷ Add meas. to end of sensor queue
            ▷ Fusion is triggered if time delta is too large or no active sensor is pending
        if (t_z > t_fusion + t_max^wait) OR !AnyActiveSensorPending() then ▷ Apply fusion
            G = {}                                                  ▷ Grids to be fused
            for s ∈ S do          ▷ Determine which sensors contribute to current fusion
                sensorIncludedInFusion = false
                while !s.queue.empty() AND s.queue.front().t < t_fusion + Δt/2 do
                    G.push(s.queue.front())
                        ▷ Alternatively, only use one meas. of each sensor (closest to t_fusion)
                    s.queue.pop()              ▷ Remove evaluated sensor data from queue
                    sensorIncludedInFusion = true
                end while
                if !sensorIncludedInFusion AND s.t_latest < t_fusion - t_max^inactive then
                    s.active = false          ▷ Set sensors without recent meas. to inactive
                end if
            end for
            (...)                          ▷ Fuse grids of G and update subsequent modules
        end if
    end if
end function

function AnyActiveSensorPending( )
    for s ∈ S do                                  ▷ Check the status of each sensor s ∈ S
        if s.active AND s.queue.empty() then  ▷ Pending if active and no data queued
            return true
        end if
    end for
    return false
end function
```

fusion should not further wait for missing sensor data, and if sensor data of the active sensors are still pending in the current fusion interval, with an additionally introduced maximum waiting time to enable a tradeoff between waiting for all pending data and limiting the total system latency.

2.5 Results and Summary

This chapter has presented how measurement data of lidar, radar, and camera sensors are processed and fused within the grid-based framework. Sensor-individual measurement grids are derived for each sensor with specific models to utilize the respective sensor strengths. An evidential occupancy grid representation is applied with separate hypotheses of occupancy and freespace using the Dempster–Shafer theory of evidence. Based on that uniform representation and the same spatial grid cell discretization within one combined odometry coordinate system, the sensor measurement grids are fused cell-wise by an evidential combination strategy without requiring specific features. To deal with inactive sensors and asynchronous sensor data, an additional fusion logic with a measurement time alignment concept has been proposed. Examples of the different sensor types and configurations, their measurement grids, and the measurement grid fusion have been shown with real sensor data in Fig. 2.4–2.6.

Two more scenarios of the resulting fused measurement grid with the different layers are demonstrated in Fig. 2.9 and Fig. 2.10. Thereby the primary evidential occupancy grid layers are shown in (a), which are mainly derived from lidar data due to their high spatial accuracy. Radar detections further include radial Doppler velocity measurements, which are also modeled in the grid representation by separate layers as depicted by blue lines in (b). Those velocity vectors are already compensated by the ego movement, i.e., they represent the absolute velocity in the radial measurement direction. Note that for the static environment that vector is thus close to zero, but some radar detections may also result in an incorrect velocity due to measurement noise, e.g., caused by a wrongly resolved frequency or azimuth angle. The camera classification information of both the static detections, only available in the front area in this work, and surrounding objects is shown in (c). The corresponding segmentation information that links grid cells to object instances is visualized in (d). Thereby the additional camera–lidar alignment enables that the camera-detected objects with their imprecise position estimation, as shown uncorrected in (d), are matched with the spatially accurate lidar data. This means that the lidar detections are partly enriched by the camera information, whereby the camera classification and segmentation of objects are derived at the same precise position as the occurring occupancy evidence of the lidar measurements.

Altogether, this fused measurement grid already represents an accurate model of the current local environment with a full surrounding 360° detection area by multiple lidar, radar, and camera sensors. The generic occupancy grid representation is further extended by the additional radar velocity measurements and camera information, which partly enable a direct detection of moving objects and a classification of the environment. However, to further improve the robustness against noise, occlusion, or missing measurement data, a temporal filtering over time is essential. Moreover, it is important that moving objects and static obstacles are estimated in a generic way by considering all sensor data, which is supported but not fully dependent on the additional radar Doppler or camera data. In other words, moving objects also have to be detected solely from occupancy measurements without any supplemental measurement information of the dynamic state, which has to be derived implicitly by the temporal filtering and accumulation. Those aspects are addressed in the following chapters.

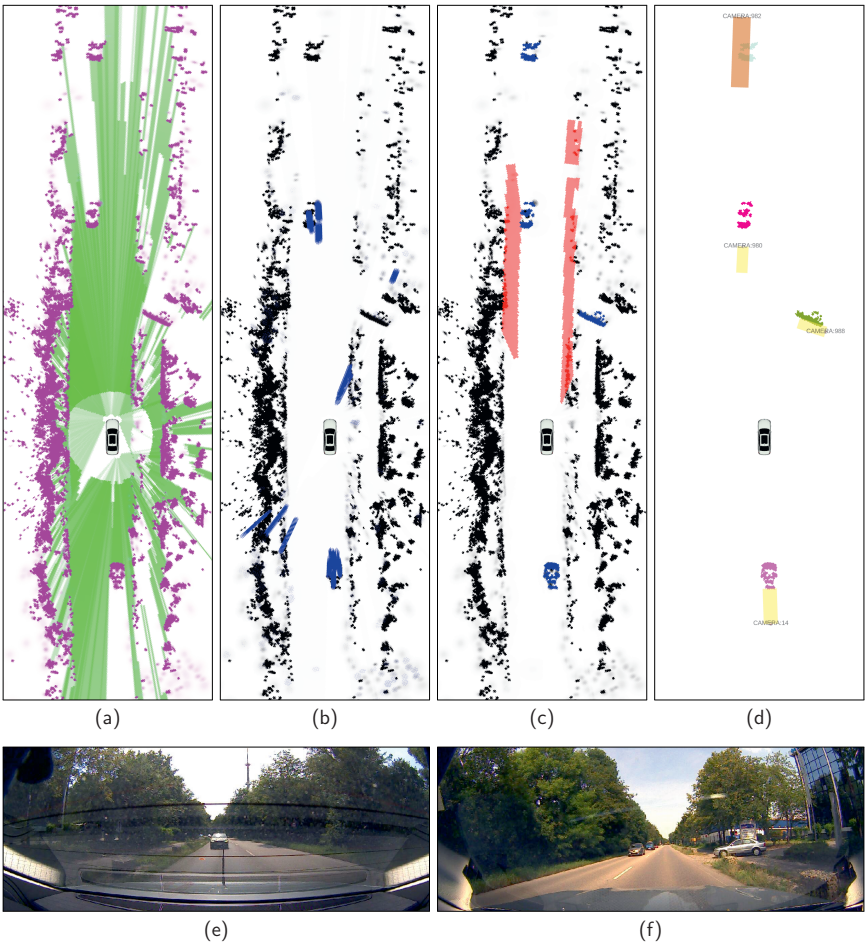


Fig. 2.9: Results of the different grid layers of the fused measurement grid in a real-world scenario. (a) Evidential occupancy grid of all sensors (primarily based on lidar data, but also derived from radar data). (b) Velocity grid (based on measured radial radar Doppler velocities). (c) Camera classification grid (all classes of the static environment are shown in red; all classes of objects are shown in blue). (d) Camera segmentation grid (with spatial camera–lidar alignment) and corresponding camera-detected object instances without distance correction. (e/f) Camera images (rear/front).

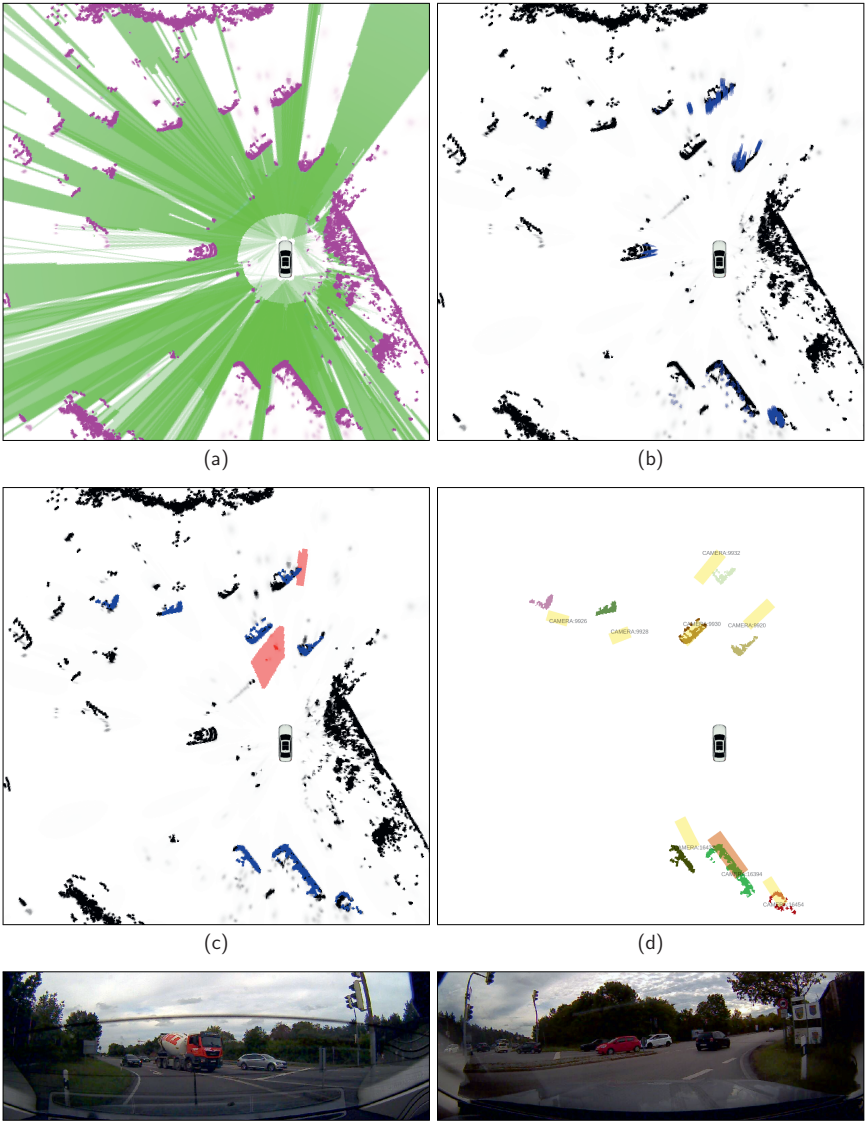


Fig. 2.10: Another example of the fused measurement grid for a junction scenario. (a) Evidential occupancy grid. (b) Radar velocity grid. (c) Camera classification grid. (d) Camera segmentation grid and object instances. (e/f) Camera images (rear/front).

3 Dynamic Grid Mapping and Particle Tracking

The previous chapter has focused on the representation and fusion of sensor measurements as measurement grids. The fused measurement grid thereby contains measurement data that approximately represent a single time instance. This chapter focuses on the subsequent temporal accumulation of the fused measurement grids over time, i.e., the occupancy grid mapping. As the overall application is the estimation of dynamic environments, this requires an additional estimation of the dynamic state of the grid cells and thus a prediction and measurement update of the accumulated occupancy grid map. This is achieved by a newly proposed combination of a dynamic grid mapping and a grid-based particle filter tracking. The dynamic grid map is modeled using an extended Dempster–Shafer evidential representation by dividing the hypothesis of unclassified occupancy of the measurement grid into two separate hypotheses of static and dynamic occupancy. Those hypotheses are differentiated over time by an adapted evidential filtering, which is also used to implicitly model the temporal transitions between the individual hypotheses and their supersets as well as the conflict handling between contradicting hypotheses. A particle-based velocity estimation is used to determine the dynamic state of the grid map and thereby initialize and predict its dynamic occupancy evidence masses. Static occupancy, in contrast, is directly modeled in the grid map without requiring particles, increasing efficiency and improving the static/dynamic occupancy classification due to the persistent map accumulation.

The work presented in this chapter is an extension of [4] that has been published in the context of this thesis, which is not explicitly referenced in the following. Additional contribution results from evaluating radar- and camera-specific measurements for improving the static/dynamic occupancy classification as well as the particle-based velocity estimation.

3.1 Introduction

The generic representation and the discretized cell structure of occupancy grids enable a low-level sensor data fusion, without specific object assumptions, by the implicit cell-wise association of different sensors, which has been discussed in the previous chapter. That strategy is also used for the temporal filtering, resulting in a cell-wise accumulation of the individual evidence masses and thus a mapping of the local environment, also without requiring object assumptions. Equivalent to the measurement grid fusion, the individual cells thereby exactly overlap in the temporal filtering, as depicted in Fig. 2.1, due to the fixed odometry coordinate system and the known relative ego movement within it.

The original concept of occupancy grid mapping [37] requires a static state of the local environment for an accumulation of occupancy probabilities without performing any prediction, since moving objects would cause contradicting occupancy artifacts along their movement without further adjustments. But within the last decade that idea of occupancy grid mapping has been extended toward dynamic environments by additionally estimating

the dynamic state of the grid, typically referred to as *dynamic occupancy grids* or *dynamic grid mapping*. Recent promising approaches use particle filters for a short-term estimation of cell velocities of the occupied environment and thereby distinguish between static obstacles and moving parts, still without forming high-level objects. However, a robust and efficient combination of the persistent occupancy grid map accumulation and that short-term particle-based dynamic estimation still remains a challenge. Those aspects are addressed in this chapter; the contribution and the outline of the proposed approach are summarized after a brief discussion of related work in the following.

3.1.1 Related Work

For occupancy grid mapping in dynamic environments it is essential to estimate the dynamic state of grid cells. A simple strategy is a binary classification of occupied cells into static or dynamic by determining inconsistencies between the occupancy value of the accumulated grid map, i.e., the temporally filtered measurements of the past, and that of the new measurement [135, 137]. For example, the accumulated grid map indicates a high freespace probability of a specific grid cell, whereas the current measurement for that same cell represents a high occupancy probability. Assuming noise-free measurement data, this inconsistency has to be the result of a moving object that currently passes over that cell, i.e., a dynamically occupied cell. Equivalently, using a Dempster–Shafer evidential occupancy grid representation, conflict masses are analyzed in [56, 80], which is further extended in [69] using prior map knowledge with a modified evidential combination and accumulation. Recently, a similar evidential approach was presented in [51] using an adapted transferable belief model. A distance-based kernel function additionally weights the occupancy transfer toward static or dynamic based on the vicinity of a cell. Thereby new static cells are characterized by neighboring static cells of the map, whereas dynamic cells are classified by previously measured occupancy in a distance-based location that depends on the expected average velocity of moving objects and the corresponding measurement time difference. In addition, object hypotheses can be extracted and tracked using distance-based clustering of those detected moving parts, e.g., [20, 24, 56, 104, 134]. However, a static/dynamic classification based solely on analyzing inconsistencies or conflicts is error-prone to measurement inaccuracies of ranging or odometry sensors. Even though results of an object tracking can be used to improve this classification, no robust cell-individual velocity estimation is achieved this way.

In [26, 122] the *Bayesian Occupancy Filter* (BOF) is introduced, an approach that recursively estimates the dynamic state of an occupancy grid through discrete cell velocities modeled as neighborhood cell transition histograms. However, an accurate and wide range cell velocity estimation using these transition histograms of each cell requires enormous computing/memory consumption and the discretization also leads to aliasing problems. This approach has been further extended in various interesting approaches – a detailed review of the BOF including multiple extensions is given in [102].

Most promising approaches use particle filters to estimate the dynamic state of the occupancy grid, partly referred to as *Sequential Monte Carlo BOF* (SMC-BOF). Originally proposed by Danescu et al. [29, 30], each particle represents a point mass of occupancy with continuous position and velocity that moves freely in the entire grid structure. The

number of particles in a cell is adapted proportional to the cell occupancy value. Sampling, weighting, and resampling of the particles results in a robust filtering of both the occupancy probabilities and the multimodal velocity distribution for each grid cell. A cell is classified as either static or dynamic by evaluating the velocity variance of all particles in that cell.

Tanzmeister et al. [118, 120, 121] adapted that grid-based particle filtering strategy of [29] toward a Dempster–Shafer evidential representation with an additional mapping of evidence masses, referred to as evidential *Grid-based Tracking And Mapping* (GTAM) [121]. Occupancy and freespace are modeled as separate evidential hypotheses as in [80], but the occupancy hypothesis is further subdivided into the two hypotheses of static occupancy and dynamic occupancy, which are then estimated continuously in contrast to the discrete classification of [29]. In addition to an initial uniform velocity distribution of new particles, pure static particles with zero-velocity are sampled that improve the convergence of static occupancy. The number of particles and the occupancy evidence in a cell are defined by the measured occupancy belief, while additional particle survival probabilities enable short persistence in occluded areas. Moreover, the output of the particle filtering is temporally accumulated, allowing full persistence of the static environment in a grid map representation without keeping all particles. However, information of the accumulated map is not considered in the particle filtering and dynamic evidence masses are not accumulated, i.e., multiple measurements do not increase the dynamic masses.

Similarly, Nègre et al. [83] introduced the idea of *Hybrid Sampling BOF* (HSBOF) as an adaptation of [122] and [29] toward a decomposition of static and dynamic occupancy. Static occupancy is represented in a typical occupancy grid, whereas dynamic occupancy is modeled by particles, i.e., particles only represents the dynamic environment rather than all occupied areas, which significantly reduces the number of required particles. The same authors improved this idea in [101] by introducing formal states of static occupancy, dynamic occupancy, freespace, and unknown areas in the filtering process, resulting in a more accurate velocity estimation and compact algorithm definition. However, in contrast to the evidential representation of [121], no dependencies of the individual hypotheses and their supersets are modeled by these states. The temporal filtering and the combination of the states of the approach cannot be compared, since they are not described in detail.

Another SMC-BOF approach has been presented by Nuss et al. [88] based on the concepts of [29, 30, 83, 120]. They focused on fusing laser and radar data, and improving the particle velocity estimation using radar Doppler velocity measurements, similarly described in [121]. An extension and reformulation of Nuss et al. [85] models the grid-based dynamic estimation as a random finite set (RFS) problem using a combination of Bernoulli filter and PHD filter. As noted by the authors, this approach is not real-time capable. Therefore, they also proposed an approximation using evidences that is similar to [121], restricting particles to areas of currently visible occupancy. However, static and dynamic occupancy are not individually estimated. Since occupancy is only filtered by the particles without an additional grid map accumulation, occlusion also leads to particle destruction and information loss of static occupancy.

Recently, based on [85], Vatavu et al. [131] proposed the idea of *Multi-Layer Particle Filter-based Tracking* (MLPT), where, in addition to the occupancy grid layer, also a semantic grid layer with camera information is used as input. Particles are grouped as track-

lets, where all particles of a tracklet share common higher-level data, such as a semantic label or an object identifier, and are self-localizing using landmarks with respect to the estimated object shape of each tracklet. The main idea is to enrich particles with extended group information with a more precise particle-to-measurement matching, overall improving the convergence of the dynamic estimation and thereby also requiring less particles. With that extended particle estimation, however, the computational effort increases significantly, and it also introduces dependencies on the accuracy of the object-based tracklet estimation on that particle-level.

3.1.2 Contribution and Outline

The different variants of the dynamic occupancy grids, in particular those that use a particle-based estimation of the dynamic state, already show promising results. However, none of these approaches fully combines the benefits of an evidential representation with multiple hypotheses, a consistent filtering of evidence masses with a persistent grid mapping accumulation of the static environment, and an efficient particle-based filtering. Moreover, additional radar and camera measurement data are not evaluated in detail for improving the static/dynamic occupancy classification and particle velocity weighting.

Those aspects are addressed in the following. As a result, a new concept for the dynamic grid mapping with a particle-based dynamic estimation and an evidential Dempster–Shafer representation is proposed. The main contributions of this approach are as follows:

- The evidential occupancy grid mapping and the particle-based dynamic estimation are combined into one filtered estimation. Rather than performing those two parts as separate processing steps, the particle-based estimation is directly integrated within the evidential map accumulation. The evidential dynamic grid map thereby forms the core component of the temporal filtering, since, in contrast to the short-term particle population, it represents a persistent accumulation. That way, it is ensured that previously aggregated information of the static environment is retained when such areas are outside the current field of view or temporarily occluded, which avoids information loss in such a case. The particle-based estimation is thereby primarily used for the initialization and prediction of dynamic occupancy of the grid map.

- Based on this combined evidential estimation, static occupancy is directly accumulated by the evidential grid map representation without requiring any particles. Therefore, particles solely represent the hypothesis of dynamic occupancy of the grid map. As a result, this significantly reduces the number of required particles and thus directly increases the computational efficiency of the dynamic grid estimation.

- As an extension of the evidential representation of [121], the hypothesis of passable area is introduced as the superset of current freespace and dynamic occupancy. To model the temporal transition, freespace evidence is transferred to the superset of passable area within the prediction. Overall, this enables a consistent prediction of all hypotheses, an implicit dynamic classification for occupancy measurements in cells with previously accumulated passable area, and a differentiation between previously measured freespace that may be temporarily occupied and actual current freespace.

- All evidence masses of the dynamic grid map are consistently filtered within the Dempster–Shafer framework – instead of using individual particle survival probabilities. This also enables the accumulation of dynamic occupancy over time. An explicitly modeled conflict assignment further ensures that the persistent map accumulation dominates the more uncertain particle-based prediction.

- An adapted evidential filtering strategy with a modified assignment of occupancy evidence masses is proposed, which is directly integrated into the evidential measurement update. This enables that measurements of unclassified occupancy can partially increase the subsets of static occupancy or dynamic occupancy of the map, which overall results in a generic and robust static/dynamic differentiation that is solely based on the temporal filtering. Hence, no pre-classification of the measurement data is necessarily required for this dynamic grid estimation – but it is also presented how additional radar velocity and camera classification measurement data, if available, can further improve that evidential filtering and occupancy classification.

- Based on the updated dynamic grid map, the particle population is updated accordingly to retain the consistent dynamic occupancy estimation of both representations and also to enable the temporal filtering of the particle-based velocity estimation. Thereby it is also demonstrated how radar velocity measurements and camera orientation estimates further enhance the particle initialization and weighting.

These different aspects of the proposed approach are structured into the following sections:

- *Section 3.2* introduces the two representations of the evidential dynamic grid map and the low-level particle population. It is also discussed how those two representations are linked to each other regarding the dynamic occupancy evidence mass.

- *Section 3.3* proposes the prediction step of the dynamic grid map. It results from a combination of the particle-based prediction of dynamic occupancy and an adaptation of the non-dynamic hypotheses of the previous map.

- *Section 3.4* addresses the measurement update of the dynamic grid map. It is discussed in detail how static and dynamic occupancy are filtered and differentiated over time – based on unclassified occupancy measurements, but also with additional radar velocity and camera classification information, and with feedback of filtered objects.

- *Section 3.5* focuses on the weighting and resampling of the particle population, including the consistent representation with the dynamic grid map and an enhanced particle filter update with radar and camera measurement data.

- *Section 3.6* presents the augmented measurement grid, an enriched form of the measurement grid with a static/dynamic classification of measured occupancy, robustly differentiated by the filtered dynamic grid map, and particle-based cell velocity estimates. This representation forms the input to the subsequent object tracking.

Finally, qualitative results and a summary of the proposed dynamic grid mapping and particle tracking approach are briefly presented in Section 3.7. Further results of this approach, including a quantitative evaluation, are demonstrated in Chapter 6.

3.2 Dynamic Grid Map and Particle Representation

This section introduces the representation of the dynamic occupancy grid map including the evidential hypotheses and the particle-based estimation of the dynamic state. Based on that dynamic grid map and particle tracking definition, the subsequent sections then discuss in detail the combined temporal filtering of both representations.

3.2.1 Evidential Frame of Discernment for Dynamic Environments

The evidential occupancy grid representation of the measurement grids has been modeled by the frame of discernment $\Theta_z = \{F, O\}$ with the two hypotheses of freespace (F) and occupancy (O) using the Dempster–Shafer evidential framework, as presented in Section 2.2. The dynamic grid mapping of this chapter, however, requires an additional estimation of the dynamic state of the occupied environment and thus a differentiation of static and dynamic occupied areas. Therefore, as proposed by Tanzmeister et al. [118, 120, 121], the occupancy hypothesis (O) is further divided into the two hypotheses static occupancy (S) and dynamic occupancy (D) in the following, i.e.,

$$O = \{S, D\}\,. \tag{3.1}$$

The overall frame of discernment of all possible individual hypotheses thus results in

$$\Theta = \{F, S, D\}\,, \tag{3.2}$$

which is referred to as the FSD-frame in [121]. The corresponding powerset of all combinations of the different hypotheses arises from

$$2^{\Theta} = \big\{\, \{F\}, \{S\}, \{D\}, \{S, D\}, \{F, D\}, \{S, F\}, \{\Theta\}, \emptyset \,\big\} \tag{3.3}$$

with the following possible hypotheses:

$\{F\}$	passable area that is currently free (freespace),
$\{S\}$	not passable area that is statically occupied (static occupancy),
$\{D\}$	passable area that is currently dynamically occupied (dynamic occupancy),
$\{S, D\}$	currently occupied (temporary or permanent occupancy),
$\{F, D\}$	passable area (free or dynamically occupied),
$\{\Theta\}$	unknown state, i.e., either free or occupied.

Equivalent to [121], the superset $\{S, F\}$ is also not used in the following as it is permanently conflicting by definition. However, as a difference, the superset $\{F, D\}$ of passable area is considered in this work. Although that hypothesis is always conflicting for a single measurement, it represents a reasonable hypothesis in the temporal filtering as the corresponding two subsets $\{F\}$ and $\{D\}$ are interchangeable over time. For example, a grid cell that has currently been observed as freespace $\{F\}$ may be occupied at a different time by a moving object, modeled as dynamic occupancy $\{D\}$, which moves across that grid cell, i.e., that area is generally passable and can thus be modeled by $\{F, D\}$. Overall, the superset $\{S, D\}$ primarily models measured occupancy at a single time instance which cannot be directly classified as static or dynamic occupancy, whereas the superset $\{F, D\}$ models the uncertain state of measured freespace predicted to a different time, as it could then be occupied by a moving object.

3.2.2 Dynamic Grid Map Representation

For all hypotheses $\theta \in 2^\Theta$ of the powerset 2^Θ with the frame of discernment Θ, evidence masses $m(\theta)$ are estimated within the Dempster–Shafer framework, as defined before in (2.2). In the following, the argument of the evidence mass, i.e., the corresponding hypothesis set, is abbreviated without brackets or comma for better readability, in particular $m(\{S, D\}) = m(SD)$ and $m(\{F, D\}) = m(FD)$.

The accumulated evidential dynamic occupancy grid map at time t of all modeled hypotheses with $\Theta = \{F, S, D\}$ is thus described by

$$\mathcal{M}_t = \left[m(S_t),\ m(D_t),\ m(SD_t),\ m(F_t),\ m(FD_t)\right]^\mathsf{T}. \tag{3.4}$$

The evidence mass of the unknown state $m(\Theta_t)$ is thereby implicitly included as the remaining evidence mass, since all evidence masses have to sum up to 1, cf. (2.2). As with the measurement grid, those evidence masses are defined for all cells $c \in \mathcal{G}$ of the grid structure $\mathcal{G}$, where the cell index c is omitted for better readability when irrelevant.

Overall, this dynamic grid map $\mathcal{M}_t$ with the Dempster–Shafer evidential model forms the main component of the low-level grid-based environment estimation of this work. It represents the fused and temporally accumulated estimation of the local environment, distinguishing between static occupancy, dynamic occupancy, freespace, and their combined evidential hypotheses. An example of the estimated dynamic grid map $\mathcal{M}_t$ with real sensor data in an urban scenario is shown in Fig. 3.1a. Another representation of that scenario is demonstrated in Fig. 3.1b, highlighting the particle-based velocity orientation estimation of each cell. Before the particle representation is discussed in more detail, the color coding of the grid visualizations is briefly presented in the following.

Red–Green–Blue (RGB) Color Coding of Different Evidential Hypotheses

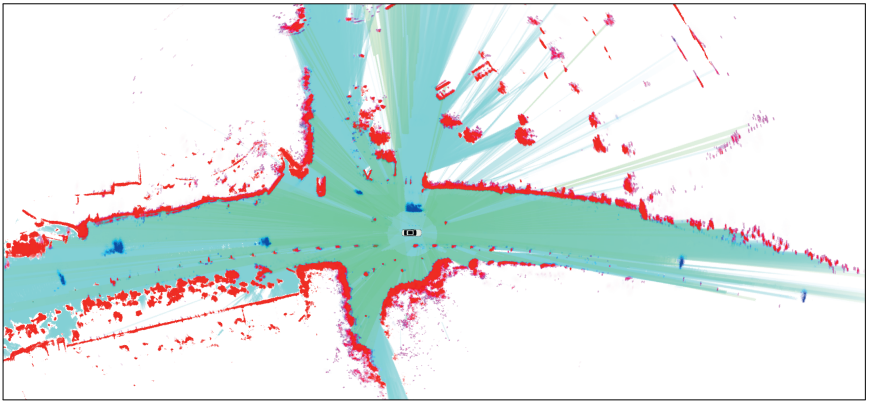
The evidence masses of the different hypotheses of the frame $\Theta = \{F, S, D\}$ are visualized in this work by an additive color mixing

$$\mathrm{RGB} = \left(1 - \sum_{\theta \cap \{S\} = \emptyset} m(\theta),\ 1 - \sum_{\theta \cap \{F\} = \emptyset} m(\theta),\ 1 - \sum_{\theta \cap \{D\} = \emptyset} m(\theta)\right),\quad \theta \subset \Theta\,, \tag{3.5}$$

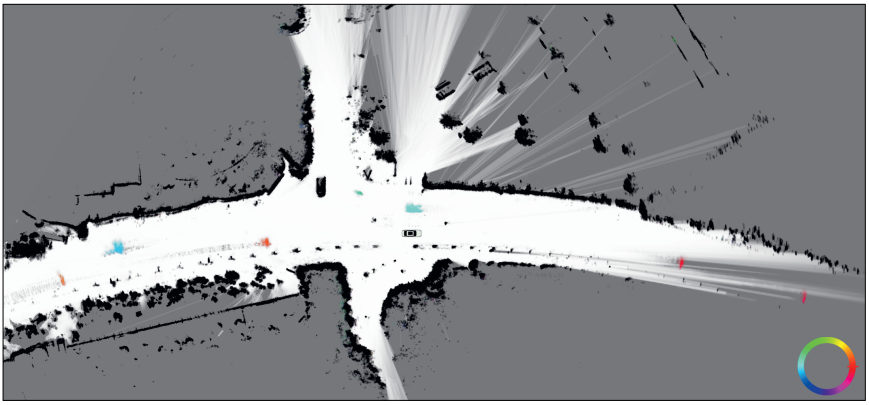
using the red–green–blue (RGB) color model. This means that static occupancy results in red (R), freespace in green (G), dynamic occupancy in blue (B), unclassified occupancy in pink (R + B), passable area in cyan (G + B), and an unknown state in white. The hypotheses supersets thereby correspond to the additive color of the individual subsets, which is also illustrated in Fig. 3.2a. This generic visualization thus also agrees to that of the measurement grids as used in the previous chapter. Note that the yellow color is not used in the evidential occupancy grid visualization, since the conflicting superset $\{S, F\}$ between static occupancy and freespace is not modeled in this work as described above.

Hue–Saturation–Lightness (HSL) Color Coding of Velocity Orientation

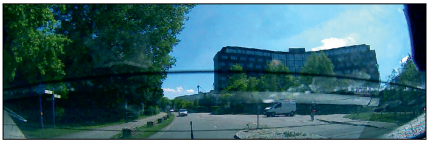
In order to visualize the velocity estimation of a dynamic occupied grid cell, a hue–saturation–lightness (HSL) color coding is selected, similar to [84, 118]. Given a particle-based 2-D cell velocity estimate ν_t^c, the orientation $\phi_c = \arctan(\nu_t^{c,y}, \nu_t^{c,x})$ is encoded as the



(a)



(b)



(c)

(d)

Fig. 3.1: Example of the estimated dynamic grid map in an urban scenario with a differentiation of static occupancy, dynamic occupancy, freespace, and their hypotheses supersets. (a) Representation of the estimated evidence masses of the different hypotheses. (b) Representation of the particle-based velocity orientation estimation (color circle) with an additional grayscale encoding of the occupancy value. (c/d) Camera images (rear/front) of the scenario.

(a) (b) (c)

Fig. 3.2: Illustration of the grid color coding visualizations. (a) RGB mixture of evidential hypotheses: static occupancy $\{S\}$ is shown in red (R), freespace $\{F\}$ in green (G), dynamic occupancy $\{D\}$ in blue (B), unclassified occupancy $\{S, D\} = \{O\}$ in pink (R + B), passable area $\{F, D\}$ in cyan (G + B), and an unknown state $\{\Theta\}$ in white (R + G + B). (b) Hue color circle of the velocity orientation in reference to the red-colored ego orientation. (c) HSL color cone: colors show the orientation of moving parts with an additional grayscale occupancy value encoding (image source: Wikimedia Commons: "HSL color solid dblcone").

H-channel with a color circle as depicted in Fig. 3.2b. The S-channel depends on the velocity magnitude $|\nu_t^c|$, normalized by a function $\alpha_v(|\nu_t^c|) \in [0, 1]$ such that velocities higher than a modeled threshold result in full saturation, and the corresponding dynamic evidence mass $m(D)$ of that cell. As visualized in Fig. 3.2c, the L-channel enables to model the grayscale value, whereby fully saturated colors result with L = 0.5. This channel is further used to model information about the occupancy value similar to an occupancy probability. Overall, the HSL-coding is selected as

$$\mathrm{HSL} = \left(\phi_c,\ \alpha_v(|\nu_t^c|)\, m(D),\ \frac{1}{2}\Big(1 - \big(\mathrm{bel}(O) - \alpha_v(|\nu_t^c|)\, m(D)\big) + \sum_{\theta \in \{F, FD\}} m(\theta)\Big)\right). \tag{3.6}$$

Thereby the L-channel is modeled such that unknown areas ($m(\Theta) = 1$) result in gray (L = 0.5), freespace and passable area ($m(F) + m(FD)$) in white (L = 1), and occupancy of currently non-moving parts in black (L = 0). Note that in this work occupied cells of tracked objects can be modeled to retain as dynamic occupied rather that static occupied even though their velocity converges toward zero, e.g., when they stop at a traffic light, since only dynamic occupied cells are considered in the subsequent object tracking. Hence, depending on the normalized velocity magnitude $\alpha_v(|\nu_t^c|)$, dynamic occupied cells ($m(D) > 0$) can result in an orientation color ($\alpha_v(|\nu_t^c|) \gg 0$, L ≈ 0.5), but also in black as for the static environment when they are currently not moving ($\alpha_v(|\nu_t^c|) \approx 0$). The particle representation, which forms the basis of such a cell-wise velocity estimation of the dynamic grid map, is presented in the following.

3.2.3 Low-Level Particle Representation

The evidential representation of the dynamic grid map enables a differentiation of static and dynamic occupancy as separate hypotheses. To distinguish those two hypotheses, a filtered estimation over time of the dynamic state of the grid map is required, which also

includes a prediction of the dynamic occupancy evidence masses based on the corresponding movement of those occupied cells. As discussed before, most of the recent promising approaches, e.g., [29, 83, 85, 121, 131], use a particle filter for that dynamic estimation of the occupancy grid. The basic idea of all those approaches is that each individual particle of the particle population represents a hypothesis of occupancy at a particular 2-D position within the grid structure and with a specific 2-D velocity for the corresponding movement estimation. Particles that steadily move over time from occupied cells into cells with newly measured occupancy are reinforced as those movement hypotheses are supported by the corresponding measurements, whereas particles that move into grid cells with an unknown or freespace state are reduced and deleted. In other words, the survival of the fittest principle of a particle filter is applied in terms of the current occupancy values of grid cells and their movement over time.

This particle-based dynamic estimation is also used in this work, but that short-term filtering is combined with the persistent evidential filtering of the dynamic grid map to avoid information loss of temporarily occluded parts of the static environment. Hence, as part of the dynamic grid mapping, the particle population is primarily used for the initialization and the prediction of the dynamic evidence mass $m(D_t)$. Therefore, particles are concentrated in areas with moving (or possibly moving) occupancy of the environment, as exemplarily shown in Fig. 3.3.

Each particle $\chi \in \mathcal{X}_t$ of the particle population $\mathcal{X}_t$ at time t is described by a position $x_\chi \in \mathbb{R}^2$, a velocity $\nu_\chi \in \mathbb{R}^2$, and an occupancy evidence amount $o_\chi \in [0, 1)$, i.e.,

$$\chi = \left[x_\chi^x, x_\chi^y, \nu_\chi^x, \nu_\chi^y, o_\chi\right]^\mathsf{T} . \tag{3.7}$$

The population of all particles χ in a grid cell c regarding the individual particle positions x_χ evaluated at time t is described by the population subset

$$\mathcal{X}_t^c = \{\chi \in \mathcal{X}_t \mid x_\chi \in \mathcal{A}(c)\} , \tag{3.8}$$

with the 2-D interval

$$\mathcal{A}(c) = \left[x_c^x - \frac{d_c}{2}, x_c^x + \frac{d_c}{2}\right) \times \left[x_c^y - \frac{d_c}{2}, x_c^y + \frac{d_c}{2}\right) \in \mathbb{R}^2 \tag{3.9}$$

defining the spatial area of the grid cell c given the corresponding center position x_c and a constant cell size of $d_c \times d_c$ that is equal for all grid cells $c \in \mathcal{G}$. The sum of all particle occupancy values o_χ of that population $\mathcal{X}_t^c$ of each grid cell c at time t is selected to be equal to the dynamic evidence mass $m(D_t^c)$ of that cell, which links the dynamic occupancy mass of the dynamic grid map $\mathcal{M}_t$ with the particle population $\mathcal{X}_t$, i.e.,

$$\sum_{\chi \in \mathcal{X}_t^c} o_\chi = m(D_t^c) . \tag{3.10}$$

As depicted in Fig. 3.1b, the particle tracking also enables a robust low-level velocity estimation and thus an estimation of a 2-D grid cell velocity

$$\nu_t^c = \frac{1}{m(D_t^c)} \sum_{\chi \in \mathcal{X}_t^c} o_\chi \nu_\chi \tag{3.11}$$

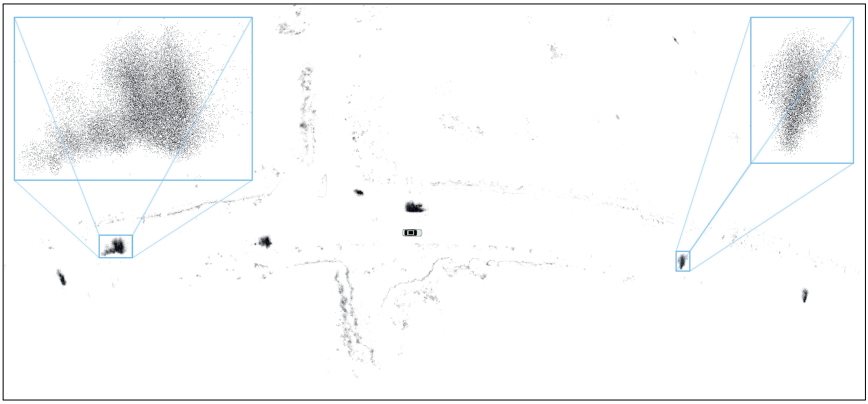


Fig. 3.3: Visualization of the particle population for the scenario of Fig. 3.1. Each particle is represented as a black point, as highlighted by the additionally shown enlarged areas of two moving objects (here with up to 100 particles per grid cell with cell size $d_c = 0.15\,\mathrm{m}$).

by the mean particle velocity of all particles in a cell, weighted by their occupancy values that sum up to the dynamic occupancy evidence mass. The static occupancy mass $m(S_t)$ is not considered in the mean velocity estimation of (3.11) to obtain a bimodal velocity distribution and thus keep the particle-based cell velocity estimate ν_t^c of the dynamic hypothesis and the zero-velocity estimate of the static hypothesis separately.

Overall, the combination of the dynamic grid map with the low-level particle representation enables a robust estimation of dynamic environments. The particle population is thereby linked with the dynamic occupancy evidence mass of the dynamic grid map, as discussed in detail in the following sections.

3.3 Particle-Based Prediction of the Dynamic Grid Map

The temporal filtering of the dynamic grid map requires two steps as part of a recursive state estimation: the prediction of the previous map $\mathcal{M}_{t-1}$ to the time t of the new incoming measurement and the measurement update of the predicted map with the current measurement given by the fused measurement grid as derived in the previous chapter. A schematic overview of the proposed prediction and measurement update of the dynamic grid map, combined with the particle population, is depicted in Fig. 3.4.

The prediction of the dynamic grid map is thereby estimated in two different ways by distinguishing between the dynamic part and the non-dynamic part, which is illustrated in Fig. 3.5. The dynamic part is predicted using the low-level particle tracking, resulting in a predicted particle-based map $\widehat{\mathcal{M}}_t$. The non-dynamic part, in contrast, is directly adapted from the previous map, primarily adjusting the freespace evidence, which results in the predicted non-dynamic map $\mathcal{M}'_t$. Finally, those different predictions are combined

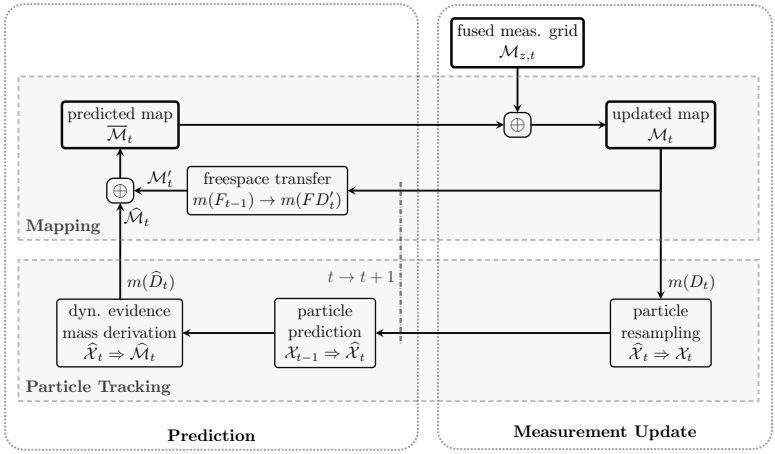


Fig. 3.4: Overview of the proposed dynamic grid mapping and particle tracking approach.

to the predicted dynamic grid map

$$\overline{\mathcal{M}}_t = (1 - \overline{\varepsilon})\, (\widehat{\mathcal{M}}_t \oplus \mathcal{M}'_t)\,. \tag{3.12}$$

Thereby the reduction factor $\overline{\varepsilon}$ additionally models the temporal uncertainty of the prediction, meaning that the evidence masses slowly converge toward the unknown state $m(\Theta)$. This factor, however, is omitted in the following discussion for a better readability. Overall, this prediction thus consists of three steps, which are presented in the following.

3.3.1 Prediction of the Dynamic Evidence Mass

The dynamic occupancy evidence $m(D_t^c)$ of a cell c of the map $\mathcal{M}_t$ is coupled with the occupancy evidence values o_χ of the particles $\chi \in \mathcal{X}_t^c$ in that cell, cf. (3.10). Hence, the basic concept of the prediction of the dynamic mass of the map results from predicting the particle population, with each particle retaining its occupancy value, and then evaluating the new position of the particles and thus summing up the occupancy values of all predicted particles present in a grid cell.

All particles $\chi \in \mathcal{X}_{t-1}$ of the particle population $\mathcal{X}_{t-1}$ at time $t-1$ are predicted to time t, with the time difference Δt in-between, by a simple constant velocity motion model with

$$\hat{x}_{\chi,t} = x_{\chi,t-1} + \Delta t\, v_{\chi,t-1} + w_x\,, \tag{3.13}$$

$$\hat{v}_{\chi,t} = v_{\chi,t-1} + w_v\,, \tag{3.14}$$

where w_x and w_v represent additive zero-mean 2-D Gaussian noise terms for the position and velocity, respectively, with the corresponding distributions $W_x \sim \mathcal{N}(0, \Sigma_{w,x})$ and $W_v \sim \mathcal{N}(0, \Sigma_{w,v})$, and the diagonal covariance matrices $\Sigma_{w,x}$ and $\Sigma_{w,v}$.

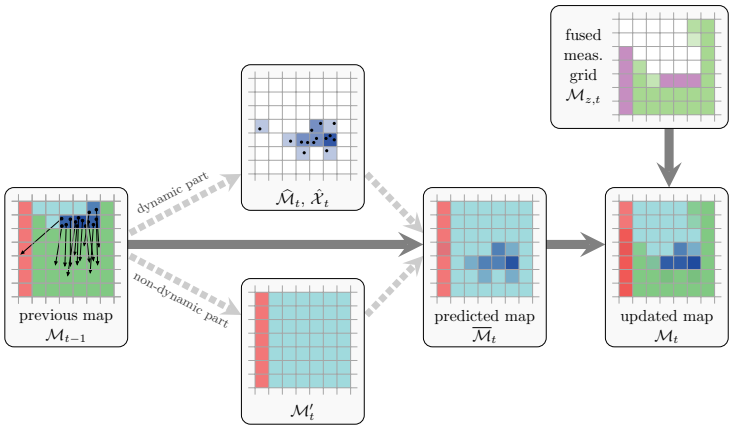


Fig. 3.5: Illustration of the prediction and update steps of the dynamic grid map.

Overall, this results in the predicted particle population $\hat{\mathcal{X}}_t$. Thereby also the set of particles $\hat{\mathcal{X}}_t^c$ that are present in a grid cell c has to be re-evaluated by the predicted particle positions, i.e., equivalent to (3.8),

$$\hat{\mathcal{X}}_t^c = \left\{ \chi \in \hat{\mathcal{X}}_t \mid \hat{x}_{\chi,t} \in \mathcal{A}(c) \right\} . \tag{3.15}$$

Based on those predicted particles $\hat{\mathcal{X}}_t^c$, the predicted dynamic mass finally results in

$$m(\widehat{D}_t^c) = \min \left(1 - \varepsilon_o, \sum_{\chi \in \hat{\mathcal{X}}_t^c} o_\chi \right) , \quad \varepsilon_o \in (0,1) . \tag{3.16}$$

Since particles of various cells could be predicted into the same cell, that sum could possibly exceed the maximum value 1 of an evidence mass. Therefore, that value is limited in (3.16) to $1-\varepsilon_o$ to obtain valid evidence masses and also avoid saturation, i.e., remain an unknown mass $m(\widehat{\Theta}_t) \geq \varepsilon_o > 0$. This forms the predicted particle-based grid map

$$\widehat{\mathcal{M}}_t = \left[0,\ m(\widehat{D}_t),\ 0,\ 0,\ 0 \right]^\mathsf{T} , \tag{3.17}$$

with the same structure of the evidence masses as the dynamic grid map $\mathcal{M}_t$, cf. (3.4), meaning that in this case no other evidences are derived and thus $m(\widehat{\theta}_t) = 0$ for $\theta \subset \Theta \backslash \{D\}$.

3.3.2 Prediction of the Non-Dynamic Evidence Masses

The remaining non-dynamic part of the map is predicted by an adaptation of the evidence masses of the previous map $\mathcal{M}_{t-1}$. Thereby the static and unclassified occupancy evidence masses remain as they were before, the dynamic evidence mass is ignored and set to 0, and the freespace evidence mass is transferred to the superset $\{F, D\}$ of passable area, as the

state $\{F\}$ is time-dependent and may have changed to dynamically occupied at that new time t. Overall, this results in the adapted map

$$\mathcal{M}'_t = \begin{bmatrix} m(S'_t) \\ m(D'_t) \\ m(SD'_t) \\ m(F'_t) \\ m(FD'_t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{1-m(D_{t-1})} & \frac{1}{1-m(D_{t-1})} \end{bmatrix} \mathcal{M}_{t-1} = \begin{bmatrix} m(S_{t-1}) \\ 0 \\ m(SD_{t-1}) \\ 0 \\ \frac{m(F_{t-1})+m(FD_{t-1})}{1-m(D_{t-1})} \end{bmatrix}, \tag{3.18}$$

also structured equivalently to (3.4). The transfer of $m(F)$ toward $m(FD)$ as part of the prediction thus ensures that moving objects with dynamic evidence $m(D)$ that move across a cell with passable area $m(FD)$ are consistently resolved as dynamic occupancy, since $\{D\}$ is a subset of $\{F, D\}$ and the intersection of both hypotheses results in $\{F, D\} \cap \{D\} = \{D\}$. Hence, also the conjunctive rule of combination $\oplus_c$ of those hypotheses, cf. (2.20), is implicitly resolved toward $m(D)$ without any occurring conflicts. However, this also means that the evidence mass $m(FD)$ is reduced when dynamic occupancy $m(D)$ is derived or predicted into that cell, while that reduced evidence mass of $\{F, D\}$ would remain at that lower value even when the dynamic mass is predicted to a different cell in the next time step or that predicted dynamic mass in that cell has not been confirmed by any measurement. To ensure that the accumulated passable area evidence retains the previous value after a dynamic object leaves this cell, the derivation of $m(FD'_t)$ is adjusted in (3.18) and also depends on the previous dynamic evidence mass $m(D_{t-1})$. The denominator thereby represents the reciprocal of the reduction by a dynamic mass and thus represents the normalization to the originally accumulated value.

Note that the previous dynamic mass $m(D_{t-1})$ is not transferred to $m(FD'_t)$ in (3.18), as, in contrast to the measured freespace $m(F_{t-1})$, the dynamic mass is usually not directly measured and is thus primarily derived by a prediction. Hence, that implication would be critical for wrongly predicted dynamic occupancy masses. In fact, it would even form the basis for a self-affirmation of dynamic occupancy, since the combination of subsequently measured unclassified occupancy $\{S, D\}$ with that then falsely derived $\{F, D\}$ in turn results in the intersecting set $\{D\}$ and thus implicitly derives the false hypothesis of dynamic occupancy rather than the actual hypothesis of static occupancy in that case.

3.3.3 Resulting Combined Predicted Dynamic Grid Map

The predicted dynamic grid map $\overline{\mathcal{M}}_t$ finally results from the combination (3.12) of the particle-based prediction of the dynamic part $\widehat{\mathcal{M}}_t$ and the prediction of the non-dynamic part $\mathcal{M}'_t$ that is derived by an adaptation of the previous map $\mathcal{M}_{t-1}$. That evidential combination $\widehat{\mathcal{M}}_t \oplus \mathcal{M}'_t$ of those two sources results in one possible conflict mass

$$\zeta(\widehat{\mathcal{M}}_t, \mathcal{M}'_t) = m(S'_t)\, m(\widehat{D}_t) \equiv \overline{\zeta}\,, \tag{3.19}$$

as defined before in (2.21). As part of the measurement grid fusion, Dempster's rule of combination (2.22) has been used to assign conflicts proportionally to the corresponding hypotheses. In this chapter, in contrast, the assignment of conflicts is explicitly modeled to solve the prediction, and also the subsequent measurement update, with specific behavior.

The static/dynamic conflict mass $\overline{\zeta}$ is assigned to the predicted static mass, meaning that the accumulated static occupancy mass, which results from measurements within that evaluated grid cell, dominates the more uncertain particle-based prediction of the dynamic occupancy mass in case of a conflict. In other words, the evidence of static occupancy remains unchanged in the prediction. Hence, the predicted static mass results in

$$m(\overline{S}_t) = m(S \,|\, \widehat{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}'_t) + \overline{\zeta} = m(S_{t-1})\,, \tag{3.20}$$

with $\widehat{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}'_t$ representing the conjunctive rule of combination (2.20) of those two sources regarding the basic intersection of the hypotheses. All remaining predicted hypotheses are combined without additional conflict assignment, i.e.,

$$m(\overline{\theta}_t) = m(\theta \,|\, \widehat{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}'_t), \quad \theta \subset \Theta \setminus \{S\}\,, \tag{3.21}$$

and, altogether, the predicted dynamic grid map results in

$$\overline{\mathcal{M}}_t = \begin{bmatrix} m(\overline{S}_t) \\ m(\overline{D}_t) \\ m(\overline{SD}_t) \\ m(\overline{F}_t) \\ m(\overline{FD}_t) \end{bmatrix} = \begin{bmatrix} m(S_{t-1}) \\ (1 - m(S_{t-1}))\; m(\widehat{D}_t) \\ (1 - m(\widehat{D}_t))\; m(SD_{t-1}) \\ 0 \\ (1 - m(\widehat{D}_t))\; \frac{m(F_{t-1}) + m(FD_{t-1})}{1 - m(D_{t-1})} \end{bmatrix}. \tag{3.22}$$

Overall, the proposed combined prediction $\overline{\mathcal{M}}_t$ of the dynamic grid map $\mathcal{M}_{t-1}$ utilizes advantageously the Dempster–Shafer evidential framework for implicitly modeling the temporal transitions between current freespace $\{F\}$, passable area $\{F, D\}$, and dynamic occupancy $\{D\}$, as well as the evidential combination with the separate particle-based prediction of the dynamic mass $m(\widehat{D}_t)$, and an explicitly modeled conflict assignment.

3.4 Measurement Update of the Dynamic Grid Map

The dynamic grid map $\overline{\mathcal{M}}_t$ has been predicted to the time t of the incoming fused measurement grid $\mathcal{M}_{z,t}$. Based on that prediction, this section focuses on the second step of the recursive estimation, the measurement update by that fused measurement grid, eventually resulting in the desired updated dynamic grid map

$$\mathcal{M}_t = \overline{\mathcal{M}}_t \oplus \mathcal{M}_{z,t}\,, \tag{3.23}$$

as also schematically shown in Fig. 3.4 and exemplarily illustrated in Fig. 3.5. The fused measurement grid

$$\mathcal{M}_{z,t} = \left[0,\; 0,\; m(SD_{z,t}),\; m(F_{z,t}),\; 0\right]^{\mathsf{T}} \tag{3.24}$$

contains the two measured evidential hypotheses of unclassified occupancy $\{S, D\} = \{O\}$ and freespace $\{F\}$, and represents the grid-based fusion of the sensor-individual measurement grids as discussed in the previous chapter. Since generally only unclassified occupancy is directly measured, an adapted evidential filtering is required to distinguish static occupancy $\{S\}$ and dynamic occupancy $\{D\}$ of the accumulated dynamic grid map $\mathcal{M}_t$, which

is discussed in the following. This static/dynamic occupancy differentiation, however, can also be further improved when additional radar-based Doppler velocity and camera-based classification measurements are available for the evaluated grid cells, which is also addressed in this section after the generic case of distance-only measurements.

3.4.1 Conflict Assignment

The combination (3.23) of the predicted dynamic grid map $\overline{\mathcal{M}}_t$ as derived in (3.22) and the measurement grid $\mathcal{M}_{z,t}$ causes three possible conflicts

$$\zeta(\overline{\mathcal{M}}_t, \mathcal{M}_{z,t}) = \underbrace{m(\overline{S}_t)m(F_{z,t})}_{\zeta_S^F} + \underbrace{m(\overline{D}_t)m(F_{z,t})}_{\zeta_D^F} + \underbrace{m(\overline{SD}_t)m(F_{z,t})}_{\zeta_{SD}^F} \tag{3.25}$$

between the three different occupancy masses of the prediction and the measured freespace. Since the freespace mass $m(F_{t-1})$ of the dynamic grid map is transferred toward $m(\overline{FD}_t)$ as part of the prediction, i.e., $m(\overline{F}_t) = 0$, no further conflicts occur by the measured unclassified occupancy mass $m(SD_{z,t})$, meaning that none of the remaining hypotheses intersections results in the emptyset, i.e., $\{S, D\} \cap \theta \neq \emptyset \;\; \forall \theta \subseteq \Theta \backslash \{F\}$.

Similar to (3.19), the three conflict masses of (3.25) are also individually assigned by

$$\begin{aligned} m(S_t) &\leftarrow \tfrac{1}{2}\zeta_S^F \,, \\ m(F_t) &\leftarrow \tfrac{1}{2}\zeta_S^F + \zeta_D^F + \zeta_{SD}^F \,, \end{aligned} \tag{3.26}$$

based on the following discussion. The static/free conflict ζ_S^F is equally transferred to $\{S\}$ and $\{F\}$, since both hypotheses have been measured at the current time or accumulated previously, thus no further knowledge is available. The dynamic/free conflict ζ_D^F, in contrast, is assigned to $\{F\}$, as $m(\overline{D}_t)$ is only an uncertain particle-based prediction and $m(F_{z,t})$ describes an actual measurement of the current time instance. Equivalently, the occupancy/free conflict ζ_{SD}^F is also transferred to the more recent freespace measurement, since unclassified occupancy $\{S, D\}$ of a previous time, in contrast to $\{S\}$, is also an uncertain state as it could be caused by a dynamic object. Hence, all conflicts of $\zeta(\overline{\mathcal{M}}_t, \mathcal{M}_{z,t})$ are assigned toward static occupancy or freespace of the updated map $\mathcal{M}_t$.

3.4.2 Occupancy Differentiation from Distance-Only Measurements

The occupancy measurement $m(SD_{z,t})$ of the measurement grid $\mathcal{M}_{z,t}$ does not distinguish between static and dynamic obstacles, as this information is generally not directly available for distance-based measurements like lidar detections. In terms of the update of the dynamic grid map, this means that the static and dynamic evidence masses are not increased by the default conjunctive rule of combination $\oplus_\mathrm{c}$, i.e.,

$$\begin{aligned} m(S \,|\, \overline{\mathcal{M}}_t \oplus_\mathrm{c} \mathcal{M}_{z,t}) &\leq m(\overline{S}_t) \,, \\ m(D \,|\, \overline{\mathcal{M}}_t \oplus_\mathrm{c} \mathcal{M}_{z,t}) &\leq m(\overline{D}_t) \,, \text{ if } \; m(\overline{FD}_t) = 0 \,, \end{aligned} \tag{3.27}$$

even if an occupancy measurement $m(SD_{z,t}) > 0$ occurs at that cell. Equivalently, neither of the two hypotheses is derived from $\{S, D\}$ in the prediction step. In other words, without

modifications of the evidential filtering of the measurement update, static occupancy would never appear in the accumulated dynamic grid map and dynamic occupancy would only be initialized if unclassified occupancy is measured in a cell with an accumulated evidence of passable area $m(\overline{FD}_t)$, since $\{D\}$ forms the intersection of the two superset hypotheses. Therefore, the occupancy derivation of the measurement update is adapted in the following to enable the desired convergence toward static and dynamic occupancy over time.

Adaptation of the Unclassified Occupancy Derivation

The unmodified update of the unclassified occupancy mass using the conjunctive rule of combination (2.20) consists of three parts that are given by

$$m(SD \,|\, \overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z,t}) = \underbrace{m(\overline{SD}_t)m(\Theta_{z,t})}_{\lambda_{SD}^{\Theta}} + \underbrace{m(\overline{SD}_t)m(SD_{z,t})}_{\lambda_{SD}^{SD}} + \underbrace{m(\overline{\Theta}_t)m(SD_{z,t})}_{\lambda_{\Theta}^{SD}}. \tag{3.28}$$

These three terms $\lambda_{SD}^{\Theta}, \lambda_{SD}^{SD}, \lambda_{\Theta}^{SD}$ form the basis for the adjustment of the occupancy derivation with the static/dynamic differentiation. The individual interpretation and modified assignment of each of the three summands is discussed in the following.

The first term λ_{SD}^{Θ} depicts the combination of the predicted occupancy mass $m(\overline{SD}_t)$ with the remaining unknown mass of the measurement $m(\Theta_{z,t})$, which is defined as

$$m(\Theta_{z,t}) = 1 - m(SD_{z,t}) - m(F_{z,t}) \tag{3.29}$$

since only $\{S, D\}$ and $\{F\}$ are modeled in the measurement grid $\mathcal{M}_{z,t}$. This means that λ_{SD}^{Θ} equals $m(\overline{SD}_t)$ if no measurement occurs. In such a case, the predicted map $\overline{\mathcal{M}}_t$ should not change, i.e.,

$$m(\Theta_{z,t}) = 1 \Rightarrow \mathcal{M}_t = \overline{\mathcal{M}}_t\,. \tag{3.30}$$

Hence, the summand λ_{SD}^{Θ} remains assigned to $\{S, D\}$ by

$$m(SD_t) \leftarrow \lambda_{SD}^{\Theta}\,. \tag{3.31}$$

The second term λ_{SD}^{SD} represents the mutual support of the occupancy masses of $\overline{\mathcal{M}}_t$ and $\mathcal{M}_{z,t}$. An occupied cell should converge toward static occupancy $\{S\}$ when multiple occupancy measurements for that same cell are accumulated. Thus, the static occupancy mass is increased by λ_{SD}^{SD}, i.e.,

$$m(S_t) \leftarrow \lambda_{SD}^{SD}\,. \tag{3.32}$$

The sum of the first two terms

$$\lambda_{SD}^{\Theta} + \lambda_{SD}^{SD} = m(\overline{SD}_t) \underbrace{\big(\, m(\Theta_{z,t}) + m(SD_{z,t}) \,\big)}_{1-m(F_{z,t})} \leq m(\overline{SD}_t) \tag{3.33}$$

corresponds at most to the predicted occupancy mass $m(\overline{SD}_t)$, i.e., it models the decrease of the predicted occupancy mass by a conflicting freespace measurement $m(F_{z,t})$. Only the third term λ_{Θ}^{SD} models the increase of the occupancy mass.

Dynamic occupancy has to be initialized and increased differently than the convergence of static occupancy derived from accumulated occupancy at the same cell. Instead, dynamic occupancy is primarily characterized by newly occurring occupancy. Hence, an amount $f_D \in [0,1]$ of the newly increased occupancy mass λ_Θ^{SD} is assigned to $\{D\}$, i.e.,

$$\begin{aligned} m(D_t) &\leftarrow f_D\, \lambda_\Theta^{SD}\,, \\ m(SD_t) &\leftarrow \big(1 - f_D\big)\, \lambda_\Theta^{SD}\,. \end{aligned} \tag{3.34}$$

The amount $(1 - f_D)$ remains in $\{S, D\}$, whereby $f_D < 1$ results in slow convergence toward static occupancy with several occupancy measurements of that same cell, cf. (3.32). The variable f_D is modeled as a function of the ratio between the number of predicted particles $|\hat{\mathcal{X}}_t^c|$ in that cell and the defined maximum number of particles $n_{\max}$ in a cell. To achieve fast convergence of dynamic occupancy, f_D is modeled

$$f_D \equiv \sqrt{\frac{|\hat{\mathcal{X}}_t^c|}{n_{\max}}}\,, \quad |\hat{\mathcal{X}}_t^c| \leq n_{\max} \tag{3.35}$$

as the square root of this ratio. Dynamic evidence is initialized by spreading particles in cells with unclassified occupancy as discussed in Section 3.5. These particles initially have no individual dynamic occupancy value, i.e., $o_\chi = 0$, hence no predicted dynamic occupancy mass is derived, cf. (3.16). New dynamic occupancy in (3.34) is thus only derived when predicted particles are confirmed by an occupancy measurement in that cell.

Those different adaptations of the occupancy derivation, in comparison with the original conjunctive combination, are illustratively summarized with multiple examples in Fig. 3.6. All cases are based on an unclassified occupancy measurement $m(SD_{z,t}) = 0.4$. If the state of the predicted map is fully unknown, $m(\overline{\Theta}_t) = 1$, and no predicted particles are within that cell, $f_D = 0$, both results are equal, i.e., $m(SD_t) = m(SD \,|\, \overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z,t}) = m(SD_{z,t})$, as shown in (a). However, if the predicted map already contains accumulated occupancy, $m(\overline{SD}_t) > 0$, static occupancy is newly derived or increased with that measurement, as demonstrated in (b) and (c), respectively, which is not the case with the conjunctive rule. Equivalently, if predicted particles are present in that cell, $f_D > 0$, dynamic occupancy is newly derived or increased with that measurement, depicted in (d) and (e), respectively, which is also not the case with the original combination rule.

Adaptation of the Dynamic Occupancy Derivation

Dynamic occupancy is also implicitly increased by the selected set of hypotheses 2^Θ together with the modeled prediction. Previously measured freespace is transferred to the superset of passable area $m(\overline{FD}_t)$ in the prediction as explained in (3.18). In such a case, newly measured occupancy $m(SD_{z,t})$ results in an increased dynamic mass $m(D_t)$,

$$m(D \,|\, \overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z,t}) = m(\overline{D}_t)\big(1 - m(F_{z,t})\big) \;+\; \underbrace{m(\overline{FD}_t) m(SD_{z,t})}_{\lambda_{FD}^{SD}}\,, \tag{3.36}$$

since the intersection of $\{F, D\}$ and $\{S, D\}$ is $\{D\}$, which is represented by the term λ_{FD}^{SD}. This complies the desired behavior that observed passable area that is currently occupied

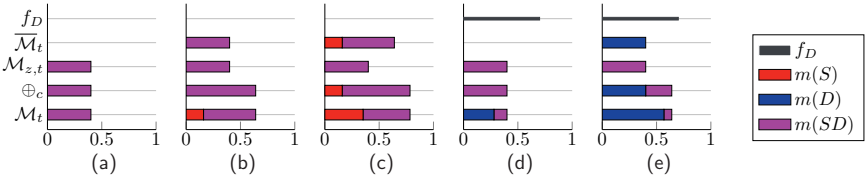


Fig. 3.6: Exemplary illustration of the occupancy derivation. Rows from top to bottom in all examples: factor f_D depending on the number of predicted particles, predicted map $\overline{\mathcal{M}}_t$, measurement grid $\mathcal{M}_{z,t}$, conjunctive combination $\oplus_c$ of $\overline{\mathcal{M}}_t$ and $\mathcal{M}_{z,t}$, and proposed combination of $\overline{\mathcal{M}}_t$ and $\mathcal{M}_{z,t}$ as defined in (3.53). (a)-(e) show different examples of varied predicted values of f_D and $\overline{\mathcal{M}}_t$, whereas $m(SD_{z,t}) = 0.4$ is equal in all examples. Figure initially published in [4], © 2018 IEEE.

has to be of a dynamic obstacle. However, completely resolving this toward $\{D\}$ is prone to measurement and odometry errors as discussed in Section 3.1.1. Measured occupancy is thus partly assigned to $m(SD_t)$ even though freespace has been accumulated before. A parameter $\gamma_D \in [0, 1]$ regulates the modeled uncertainty of the assignment of λ_{FD}^{SD},

$$\begin{aligned} m(D_t) &\leftarrow \big(1-\gamma_D\big)\lambda_{FD}^{SD} + f_D\,\gamma_D\,\lambda_{FD}^{SD}\,, \\ m(SD_t) &\leftarrow \big(1-f_D\big)\gamma_D\,\lambda_{FD}^{SD}\,, \end{aligned} \tag{3.37}$$

while $\gamma_D = 0$ corresponds to a total assignment of λ_{FD}^{SD} to $m(D_t)$ without any modeled uncertainty as in (3.36). The remaining amount $\gamma_D\,\lambda_{FD}^{SD}$ that is not implicitly derived as $\{D\}$ represents an increased unclassified occupancy evidence like the term λ_Θ^{SD}. As in (3.34), a part f_D of that is still assigned to $\{D\}$, which, however, depends on the number of predicted particles and not the accumulated evidence mass of passable area $m(\overline{FD}_t)$.

Resulting Updated Occupancy Evidence Masses

Overall, the measurement update of the occupancy evidence masses for the hypotheses $\{S\}$, $\{D\}$, and $\{S, D\}$, based on the conjunctive combination $\oplus_c$ as defined in (2.20), is adapted by

$$\begin{bmatrix} m(S_t) \\ m(D_t) \\ m(SD_t) \end{bmatrix} = \begin{bmatrix} m(S\,|\,\overline{\mathcal{M}}_t \oplus_c \mathcal{M}_{z,t}) \\ m(D\,|\,\overline{\mathcal{M}}_t \oplus_c \mathcal{M}_{z,t}) \\ m(SD\,|\,\overline{\mathcal{M}}_t \oplus_c \mathcal{M}_{z,t}) \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & f_D & -\big(1-f_D\big)\gamma_D \\ -1 & -f_D & \big(1-f_D\big)\gamma_D \end{bmatrix} \begin{bmatrix} \lambda_{SD}^{SD} \\ \lambda_\Theta^{SD} \\ \lambda_{FD}^{SD} \end{bmatrix}, \tag{3.38}$$

ignoring the additional conflict mass assignment of (3.26) here. That modified assignment of the terms λ_{SD}^{SD} and λ_Θ^{SD}, both defined in (3.28), ensures a robust derivation and thus differentiation of static and dynamic occupancy over time with measurements that only represent the superset of unclassified occupancy. Static occupancy is thereby characterized by repeating occupancy measurements within the same grid cell, whereas dynamic occupancy is initialized in cells with newly measured occupancy by the presence of particle hypotheses, modeled by the factor f_D. In addition, the assignment of the term λ_{FD}^{SD} of (3.36) toward $\{D\}$ is adjusted for robustness by an introduced uncertainty factor γ_D to reduce that implicit derivation of dynamic occupancy, based on the evidential hypotheses intersection, for measured occupancy in cells with an accumulated evidence of passable area.

3.4.3 Additional Radar- and Camera-Based Occupancy Classification

The adapted occupancy filtering of (3.38) has demonstrated the generic case of deriving static and dynamic occupancy as part of the evidential filtering over time from measurements of unclassified occupancy without any additional measurement information. This static/dynamic occupancy differentiation can be further enhanced when radar Doppler velocity measurements or semantic information of a camera sensor are available for the evaluated cells, as that additional information enables a direct occupancy classification of the measurement. Hence, in the following, the unclassified occupancy measurement $m(SD_{z,t})$ is divided into separate measurement occupancy evidence masses of the subsets of static occupancy $m(S_{z^*,t})$, dynamic occupancy $m(D_{z^*,t})$, and the remaining mass of unclassified occupancy $m(SD_{z^*,t})$, i.e.,

$$m(SD_{z,t}) = m(S_{z^*,t}) + m(D_{z^*,t}) + m(SD_{z^*,t})\,. \tag{3.39}$$

Thereby the occupancy belief of the original measurement, as defined in (2.4), remains unmodified by (3.39), i.e., $\mathrm{bel}(SD_{z^*,t}) = \mathrm{bel}(SD_{z,t}) = m(SD_{z,t})$. In other words, this means that the overall evidence of occupancy including its subsets is not increased or decreased, only the subdivision between the superset $\{S, D\}$ and the individual hypotheses $\{S\}$ and $\{D\}$ is adapted in the following.

The occupancy measurement subdivision is modeled by the two factors $\beta_S, \beta_D \in [0, 1]$ with $\beta_S + \beta_D \leq 1$, such that the static mass is derived by $m(S_{z^*,t}) = \beta_S\, m(SD_{z,t})$, the dynamic mass by $m(D_{z^*,t}) = \beta_D\, m(SD_{z,t})$, and the remaining unclassified occupancy mass by $m(SD_{z^*,t}) = (1 - \beta_S - \beta_D)\, m(SD_{z,t})$, overall resulting in the adapted measurement grid

$$\mathcal{M}_{z^*,t} = \begin{bmatrix} m(S_{z^*,t}) \\ m(D_{z^*,t}) \\ m(SD_{z^*,t}) \\ m(F_{z,t}) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \beta_S & 0 & 0 \\ 0 & 0 & \beta_D & 0 & 0 \\ 0 & 0 & (1 - \beta_S - \beta_D) & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \mathcal{M}_{z,t}\,. \tag{3.40}$$

In the following, it is discussed how both factors β_S and β_D are determined from radar- and camera-based measurements, as well as the combination of those two sources.

Radar-Based Occupancy Classification

As presented in Section 2.3.3, the measurement grid of radar sensors additionally contains the measured radial velocities v_z^r as a separate velocity layer. Those direct velocity component measurements are used for the static/dynamic occupancy differentiation using a zero-mean Gaussian model with individual variances $\sigma_{r,S}^2$ and $\sigma_{r,D}^2$, and a maximum value of $\beta_{S,\max} \in [0, 1]$ and $\beta_{D,\max} \in [0, 1]$ for the static and dynamic assignment factors, respectively. Thereby the constraint $\beta_S^{\mathrm{radar}} + \beta_D^{\mathrm{radar}} \leq 1$ is ensured by selecting the variances accordingly by $\sigma_{r,D}^2 \geq \sigma_{r,S}^2$. Altogether, this results in the radar-based coefficients

$$\beta_S^{\mathrm{radar}} = \beta_{S,\max}^{\mathrm{radar}} \exp\left(-\frac{(v_z^r)^2}{2\sigma_{r,S}^2} \right), \tag{3.41}$$

$$\beta_D^{\mathrm{radar}} = \beta_{D,\max}^{\mathrm{radar}} \left(1 - \exp\left(-\frac{(v_z^r)^2}{2\sigma_{r,D}^2} \right) \right), \tag{3.42}$$

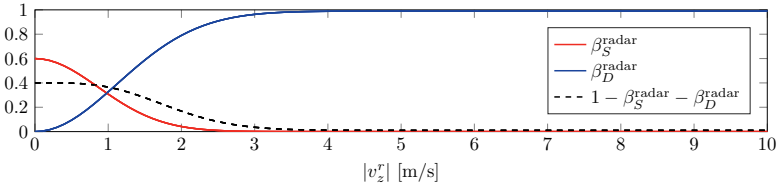


Fig. 3.7: Illustration of the radar-based occupancy classification coefficients β_S^{radar} and β_D^{radar} (here with $\beta_{S,\max}^{\text{radar}} = 0.6$, $\beta_{D,\max}^{\text{radar}} = 0.99$, $\sigma_{r,S}^2 = 0.75\,\text{m}^2/\text{s}^2$, $\sigma_{r,D}^2 = 1.25\,\text{m}^2/\text{s}^2$).

which are exemplarily illustrated in Fig. 3.7. In conclusion, this means that a measured radial velocity, already compensated by the ego motion as discussed in (2.14), with an absolute value $|v_z^r|$ close to zero increases the evidence of static occupancy, whereas the evidence of dynamic occupancy increases with the magnitude of the radial velocity.

The radar-based maximum classification evidence of occupancy toward static $\beta_{S,\max}^{\text{radar}}$ has to be selected lower than the maximum dynamic value β_D^{radar}, since only the radial velocity component is measured, which is also zero for objects moving directly tangential, i.e., orthogonal to the evaluated direction. As discussed later in Section 3.4.4, additional information by a feedback of the subsequent object tracking can be further taken into account to improve the static/dynamic convergence. Thereby the radar-based evaluation of static occupancy by the factor β_S^{radar} can be suppressed for cells within predicted objects to remain dynamic occupancy for tangential object movements as well as stopping objects.

Camera-Based Occupancy Classification

Camera measurements obtain semantic information and thus enable a classification that is also included for the corresponding occupied cells of the measurement grid, cf. Section 2.3.4. This classification is also taken into account to further improve the static/dynamic differentiation of measured occupancy. Therefore, if a cell includes a camera classification z_κ, it is evaluated whether this class is part of the superset $\mathcal{K}_{\text{static}}$ that represents classes of the static environment, or, alternatively, if that class is part of the group $\mathcal{K}_{\text{dynamic}}$ of dynamic objects. This results in the camera-based occupancy classification coefficients

$$\beta_S^{\text{camera}} = \begin{cases} \beta_{S,\max,\kappa}^{\text{camera}} & , \text{ if } z_\kappa \in \mathcal{K}_{\text{static}} \\ 0 & , \text{ else} \end{cases}, \tag{3.43}$$

$$\beta_D^{\text{camera}} = \begin{cases} \beta_{D,\max,\kappa}^{\text{camera}} & , \text{ if } z_\kappa \in \mathcal{K}_{\text{dynamic}} \\ 0 & , \text{ else} \end{cases}. \tag{3.44}$$

The maximum values $\beta_{S,\max,\kappa}^{\text{camera}}$ and $\beta_{D,\max,\kappa}^{\text{camera}}$ can further be modeled individually for each determinable class κ and thereby also depend on the camera classification accuracy of the corresponding class. The first group $\mathcal{K}_{\text{static}}$ of static classes includes, for instance, road boundaries, buildings, or vegetation, whereas the second group $\mathcal{K}_{\text{dynamic}}$ of dynamic classes includes pedestrians, cyclists, animals and so on. Vehicles like cars and trucks are a special case, as they are generally movable, but may also be parked for a longer period

of time, in particular in urban scenarios. The subsequent grid-based object tracking of this work extracts objects in areas with dynamic occupancy. Hence, depending on the decision whether parking vehicles should also be extracted as filtered objects, they may or may not be included in the second group $\mathcal{K}_{\text{dynamic}}$. Thereby classes that cannot be explicitly assigned to $\mathcal{K}_{\text{static}}$ or $\mathcal{K}_{\text{dynamic}}$ are not contained in either of those two groups, as they have to be mutually exclusive as used in (3.43)–(3.44), i.e., $\mathcal{K}_{\text{static}} \cap \mathcal{K}_{\text{dynamic}} = \emptyset$. In other words, not all determined classes of the camera have to be used for the occupancy classification, but only those that allow an unambiguous static/dynamic differentiation.

Combined Sensor-Based Occupancy Classification

The occupancy classification coefficients have been determined for both radar velocity measurements and camera-based semantic information. Those different sources have to be fused to obtain the combined multi-sensor classification into static β_S and dynamic β_D. To perform this combination, the coefficients are interpreted as evidence masses within the frame $\Theta_O = \{S, D\}$ with $m(S) = \beta_S$, $m(D) = \beta_D$, and $m(\Theta_O) = 1 - \beta_S - \beta_D$. The goal is thus the evidential combination

$$\beta = \beta^{\text{radar}} \oplus \beta^{\text{camera}}, \tag{3.45}$$

that is also resolved by the conjunctive combination $\oplus_{\text{c}}$ of the intersecting hypotheses and an explicit assignment of possible conflict masses as performed equivalently in the other evidential combination steps of this chapter. Thereby two possible conflicts occur by the combination (3.45), caused by a contradicting static/dynamic classification of the camera and the radar, i.e., $\beta_S^{\text{camera}}\beta_D^{\text{radar}}$ and vice versa $\beta_D^{\text{camera}}\beta_S^{\text{radar}}$. Both conflicts are assigned toward dynamic, as the dynamic classification of both sensor types is more accurate than the static classification. For camera measurements, the low spatial accuracy is only corrected for movable objects by the camera–lidar alignment, whereas the static environment without specific features remains uncorrected. For radar measurements, detections of moving objects with large Doppler velocities are robustly classified as dynamic, whereas a Doppler velocity of zero could also be caused by a movement tangential to the evaluated radial component. Overall, the combination including that conflict assignment results in

$$\beta_S = \beta_S^{\text{camera}}(1 - \beta_D^{\text{radar}}) + \beta_S^{\text{radar}}(1 - \beta_D^{\text{camera}}) - \beta_S^{\text{camera}}\beta_S^{\text{radar}}, \tag{3.46}$$

$$\beta_D = \beta_D^{\text{camera}} + \beta_D^{\text{radar}} - \beta_D^{\text{camera}}\beta_D^{\text{radar}}, \tag{3.47}$$

which, eventually, enables the adaptation of the measurement grid $\mathcal{M}_{z,t}$ by this additional sensor-based static/dynamic occupancy differentiation.

Resulting Occupancy Adaptation with Additional Measurement Information

The subdivision of measured unclassified occupancy into static occupancy and dynamic occupancy by the additional radar- and camera-based measurement information forms the enhanced measurement grid $\mathcal{M}_{z^*,t}$ as defined in (3.39). This enables a direct measurement update with an implicit evidential combination of the dynamic grid map with that enhanced measurement grid, i.e., $\mathcal{M}_t = \overline{\mathcal{M}}_t \oplus \mathcal{M}_{z^*,t}$, rather than the original measurement grid $\mathcal{M}_{z,t}$ that does not further distinguish measured occupancy.

This generic extension combines the benefits of an optional sensor-based occupancy classification with the adapted evidential filtering for a static/dynamic differentiation of the remaining unclassified occupancy measurements as part of the mapping over time as proposed in (3.38). The introduced adjusted assignment of the terms $\lambda_{SD}^{SD}, \lambda_{\Theta}^{SD}, \lambda_{FD}^{SD}$ can directly be adapted to $\mathcal{M}_{z^*,t}$, it only has to be considered that the remaining mass of unclassified occupancy of the measurement may be reduced due to assignment toward the individual subsets of static or dynamic, i.e., $m(SD_{z^*,t}) = (1-\beta_S-\beta_D)\, m(SD_{z,t}) \leq m(SD_{z,t})$. Hence, the terms $\lambda_{SD}^{SD}, \lambda_{\Theta}^{SD}, \lambda_{FD}^{SD}$ have to be defined accordingly by

$$\begin{bmatrix} \lambda_{SD}^{SD^*} \\ \lambda_{\Theta}^{SD^*} \\ \lambda_{FD}^{SD^*} \end{bmatrix} = \begin{bmatrix} m(\overline{SD}_t) m(SD_{z^*,t}) \\ m(\overline{\Theta}_t) m(SD_{z^*,t}) \\ m(\overline{FD}_t) m(SD_{z^*,t}) \end{bmatrix} = (1-\beta_S-\beta_D) \begin{bmatrix} \lambda_{SD}^{SD} \\ \lambda_{\Theta}^{SD} \\ \lambda_{FD}^{SD} \end{bmatrix} \tag{3.48}$$

to represent the combination of the predicted map $\overline{\mathcal{M}}_t$ with $\mathcal{M}_{z^*,t}$ rather than $\mathcal{M}_{z,t}$, i.e., the conjunctive combination of the evidence masses with $m(SD_{z^*,t})$ rather than $m(SD_{z,t})$.

Overall, equivalent to (3.38), the update of the occupancy evidence masses, ignoring the additional assignment of conflict masses, results in

$$\begin{bmatrix} m(S_t) \\ m(D_t) \\ m(SD_t) \end{bmatrix} = \begin{bmatrix} m(S \,|\, \overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z^*,t}) \\ m(D \,|\, \overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z^*,t}) \\ m(SD \,|\, \overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z^*,t}) \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & f_D & -\big(1-f_D\big)\gamma_D \\ -1 & -f_D & \big(1-f_D\big)\gamma_D \end{bmatrix} \begin{bmatrix} \lambda_{SD}^{SD^*} \\ \lambda_{\Theta}^{SD^*} \\ \lambda_{FD}^{SD^*} \end{bmatrix} . \tag{3.49}$$

Note that (3.49) equals (3.38) if no sensor-based occupancy classification is performed, i.e., if $\beta_S = \beta_D = 0$, meaning that this represents a generic extension of the previous formulation and is thus used in the following, without requiring that such additional radar- or camera-based measurement information actually has to be present.

Additionally Occurring Conflicts by the Sensor-Based Occupancy Classification

The sensor-based occupancy classification and the combination of that enhanced measurement grid $\mathcal{M}_{z^*,t}$ with the predicted dynamic grid map $\overline{\mathcal{M}}_t$ introduce new conflict masses,

$$\zeta(\overline{\mathcal{M}}_t, \mathcal{M}_{z^*,t}) = \zeta(\overline{\mathcal{M}}_t, \mathcal{M}_{z,t}) + \underbrace{m(\overline{S}_t) m(D_{z^*,t})}_{\zeta_S^{D^*}} + \underbrace{m(\overline{D}_t) m(S_{z^*,t})}_{\zeta_D^{S^*}} + \underbrace{m(\overline{FD}_t) m(S_{z^*,t})}_{\zeta_{FD}^{S^*}} , \tag{3.50}$$

in addition to the conflict masses $\zeta(\overline{\mathcal{M}}_t, \mathcal{M}_{z,t})$ caused by the measured freespace evidence as defined in (3.25). All those additionally occurring conflicts are thus caused by the subdivision of measured unclassified occupancy into static and dynamic. However, in contrast to (3.26), there is generally no decisive indication whether those conflicts are caused by a wrong hypothesis of the predicted dynamic grid map or a wrong sensor-based occupancy classification. Therefore, all three additional conflicts are assigned toward the original measurement hypothesis of unclassified occupancy, i.e.,

$$m(SD_t) \leftarrow \zeta_S^{D^*} + \zeta_D^{S^*} + \zeta_{FD}^{S^*} , \tag{3.51}$$

which represents a conservative choice as the static/dynamic occupancy classification has then to be resolved over time in the subsequent measurement update steps and the currently measured occupancy also dominates the previously measured freespace in terms of a more cautious assumption of collision avoidance.

3.4.4 Adapted Occupancy Convergence by Object Tracking Feedback

Another option to improve the dynamic occupancy convergence of the filtered dynamic grid map is to use high-level information of the subsequent object tracking. Without any object information or sensor-specific measurements, the evidential filtering of the dynamic grid map is designed such that one fixed grid cell with repeating unclassified occupancy measurements converges toward static occupancy, as described before in Section 3.4.2. This approach is effective for any type of static obstacle, but may also affect actually moving objects, since a fixed grid cell may also be occupied over a longer period of time by different parts of moving objects, in particular the large side of a slow-driving truck. In other words, the generic grid cell-individual static/dynamic occupancy classification lacks the evaluation of the overall object-level movement estimation in such cases.

If a filtered object track has been extracted before and predicted to the measurement time of the current measurement update of the dynamic grid map, this feedback information can be considered to avoid convergence of measured occupancy toward static occupancy within the areas of the predicted tracks. This convergence behavior is normally regulated by the factor f_D as defined before in (3.35) depending on the ratio of the number of particles present in that cell. By overwriting this ratio and setting $f_D = 1$ for all grid cells within the area $\mathcal{A}(\widehat{\mathcal{T}}_t)$ of the set of predicted object tracks $\widehat{\mathcal{T}}_t$, independent on the number of actually present particles, i.e.,

$$f_D = \begin{cases} 1 & , \text{ if } c \in \mathcal{A}(\widehat{\mathcal{T}}_t) \\ \sqrt{\frac{|\hat{\mathcal{X}}_t^c|}{n_{\max}}} & , \text{ otherwise} \end{cases} , \tag{3.52}$$

the convergence toward static occupancy is suppressed, since all newly measured unclassified occupancy is then derived toward dynamic occupancy of the map. Note that this only affects newly measured unclassified occupancy, whereas previously accumulated static occupancy of the dynamic grid map remains as it is, which is important for robustness when that area $\mathcal{A}(\widehat{\mathcal{T}}_t)$ of the predicted tracks is inaccurate and, for example, overlaps with parts of the static environment that have been accumulated before.

Hence, apart from the factor f_D, no further adaptation of the evidential filtering is required to support the convergence of dynamic occupancy by high-level information of the subsequent grid-based object tracking. This feedback could be limited to object tracks that move above a certain minimum speed such that the dynamic grid map converges toward static occupancy when those tracks move very slow or stop moving. However, in this work this feedback is applied for all tracks in order to actually achieve that behavior, i.e., that occupied cells within previously tracked objects remain as dynamic occupancy even though those objects may stop at a traffic light etc. As discussed in the next chapter, this simplifies and improves the association between occupied cells and the filtered object tracks, since then only dynamic cells have to be considered for the object tracking. That way static occupancy artifacts of temporarily stopping objects are avoided and the particle population within those objects is retained due to the dynamic occupancy, enabling a fast particle-based movement detection in the event of the object continuing to move, and a robust differentiation when an object like a pedestrian stops directly next to a static obstacle.

In addition to supporting dynamic occupancy of the grid map, an evaluation on the object-level can also be used to detect a false dynamic estimation, e.g., caused by newly occurring occupancy at a guardrail that gets visible with the movement of the ego vehicle, which can be handled in different ways. In this work, in such a case, predicted dynamic occupancy $m(\overline{D}_t)$ of the map is reset in terms of transferring that predicted evidence mass toward unclassified occupancy $m(\overline{SD}_t)$ and avoiding the convergence by setting $f_D = 0$, which, however, is only applied when an object hypothesis or filtered track is deleted with a high certainty that it actually represents a false positive, otherwise the cell-wise evidential filtering of the dynamic grid map is not modified by deleted objects.

3.4.5 Overall Resulting Updated Evidence Masses of the Map

This section has discussed in detail the different aspects of the measurement update of the dynamic grid map. To sum it up, the updated dynamic grid map $\mathcal{M}_t$, with the map structure as defined in (3.4), altogether, results in

$$\underbrace{\begin{bmatrix} m(S_t) \\ m(D_t) \\ m(SD_t) \\ m(F_t) \\ m(FD_t) \end{bmatrix}}_{\mathcal{M}_t} = \underbrace{\left(\overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z^*,t}\right)}_{\substack{\text{default conjunctive} \\ \text{combination of} \\ \overline{\mathcal{M}}_t \text{ and } \mathcal{M}_{z^*,t}}} + \underbrace{\begin{bmatrix} 1 & 0 & 0 \\ 0 & f_D & -(1-f_D)\gamma_D \\ -1 & -f_D & (1-f_D)\gamma_D \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \lambda_{SD}^{SD^*} \\ \lambda_{\Theta}^{SD^*} \\ \lambda_{FD}^{SD^*} \end{bmatrix}}_{\text{adapted evidential occupancy filtering}} + \underbrace{\begin{bmatrix} \frac{1}{2}\zeta_S^F \\ 0 \\ \zeta_S^{D^*} + \zeta_D^{S^*} + \zeta_{FD}^{S^*} \\ \frac{1}{2}\zeta_S^F + \zeta_D^F + \zeta_{SD}^F \\ 0 \end{bmatrix}}_{\substack{\text{conflict mass} \\ \text{assignment}}} \tag{3.53}$$

with a division into three terms. The first part depicts the default conjunctive combination $m(\theta|\overline{\mathcal{M}}_t \oplus_{\mathrm{c}} \mathcal{M}_{z^*,t})$, between all hypotheses of the predicted map $\overline{\mathcal{M}}_t$ of (3.12) and the measurement grid with the sensor-based occupancy classification $\mathcal{M}_{z^*,t}$ as introduced in (3.40). The second part represents the adapted occupancy derivation of the evidential filtering as defined in (3.38) and (3.49), i.e., modifications based on the conjunctive combination. The third part models the assignment of occurring conflict masses as discussed in (3.26) and (3.51).

Overall, this measurement update of the dynamic grid map thoroughly utilizes the Dempster–Shafer evidential framework for filtering the different hypotheses of static occupancy, dynamic occupancy, freespace, and their combined hypotheses supersets. Thereby additional radar- and camera-based measurements are evaluated, if available, to classify occupancy of the current measurement into static and dynamic. Furthermore, the evidential filtering based on the conjunctive combination of evidence masses has been adapted to enable the convergence of measured unclassified occupancy toward static and dynamic occupancy over time as part of the mapping. Moreover, evidential conflict masses are individually assigned depending on the contradiction of the corresponding hypotheses.

3.5 Weighting and Resampling of the Particle Population

The previous section has presented the measurement update of the evidence masses of the dynamic grid map $\mathcal{M}_t$. Thereby the dynamic occupancy mass $m(D_t)$ of that filtered

map is linked to the low-level particle population $\mathcal{X}_t$, meaning that the sum of the occupancy values o_χ of all particles $\chi \in \mathcal{X}_t^c$ in a grid cell c corresponds to the filtered dynamic mass $m(D_t^c)$ of the map of that cell, as introduced in Section 3.2.3. This connection enables the particle-based prediction of the dynamic evidence masses of the previous map $\mathcal{M}_{t-1}$ by evaluating the predicted particle population $\hat{\mathcal{X}}_t$, cf. Section 3.3. However, the updated dynamic grid map $\mathcal{M}_t$ thus also requires a subsequent determination of a new particle population $\mathcal{X}_t$ based on the new dynamic evidence mass $m(D_t)$ and that predicted particle population $\hat{\mathcal{X}}_t$. This weighting and resampling of the particle population is essential for achieving the temporal filtering of those particle hypotheses by the survival of the fittest principle, i.e., the robust occupancy movement tracking over time, and also for retaining the consistent representation of the dynamic evidence mass between both the grid map and the occupancy values of the particle population.

3.5.1 Cell-Wise Occupancy-Based Number of Desired Particles

The number of desired particles is adapted within each grid cell based on the occupancy evidence of that cell and a predefined maximum number of particles $n_{\max}$ in a single cell c, i.e., $|\mathcal{X}_t^c| \leq n_{\max}$, as similarly proposed in [29, 121]. In other words, instead of defining a fixed value for the total number of particles of the whole grid structure, the population is adapted individually in each grid cell.

Since particles represent only hypotheses of dynamic occupancy in this work, particles only have to be drawn in areas of the map $\mathcal{M}_t$ with dynamic occupancy $m(D_t)$ or newly occurring unclassified occupancy. The latter is required to enable deriving new dynamic occupancy for measurements of unclassified occupancy based on the number of particles that are present in that cell, represented by the factor f_D, as discussed in (3.34). The increased unclassified occupancy mass of the updated occupancy mass $m(SD_t)$ of the map compared to the predicted evidence $m(\overline{SD}_t)$ thereby arises from

$$m(SD_t) = \underbrace{\left(\lambda_{SD}^{\Theta} + \lambda_{SD}^{SD^*}\right)}_{\leq m(\overline{SD}_t)} + \underbrace{(1-f_D)\left(\lambda_{\Theta}^{SD^*} + \gamma_D \lambda_{FD}^{SD^*}\right) + \zeta_S^{D^*} + \zeta_D^{S^*} + \zeta_{FD}^{S^*}}_{\equiv m(SD_t^+)} \tag{3.54}$$

using (3.53) with $m(SD|\overline{\mathcal{M}}_t \oplus_c \mathcal{M}_{z^*,t}) = \lambda_{SD}^{\Theta} + \lambda_{SD}^{SD^*} + \lambda_{\Theta}^{SD^*}$, cf. (3.28), (3.48). Note that particle hypotheses should only be spawn in areas of such increased unclassified occupancy $m(SD_t^+)$ and thus newly occurring occupancy measurements $m(SD_{z,t})$ rather than all cells with previously accumulated unclassified occupancy $m(SD_t)$, otherwise a large number of particles would permanently be spawn in such cells and finally spread over the whole grid structure. Hence, the number of desired particles, based on the combination of dynamic occupancy and newly occurred unclassified occupancy of the dynamic grid map, has to be selected proportional to the cell density

$$\rho_t^c = m(D_t^c) + m(SD_t^{+,c})\,, \tag{3.55}$$

here explicitly stating the cell index c to avoid ambiguities due to the following usage in combination with the particle population of a specific grid cell c.

To further improve the robustness of the particle initialization, the number of the predicted particle population $|\hat{\mathcal{X}}_t^c|$ is at most decreased by a predefined reduction factor $\kappa_p \in (0, 1)$ in order to prevent an abrupt reduction of the number of predicted particles, i.e., to ensure a certain number of surviving particles even without supporting occupancy measurements. This is particularly relevant for irregularly occurring occupancy measurements, e.g., from sensors with a low measurement rate that are not included in every fused measurement grid.

Overall, the new number of particles in a cell c is defined by

$$|\mathcal{X}_t^c| = \left\lfloor \max\left(\rho_t^c \, n_{\max},\, \kappa_p |\hat{\mathcal{X}}_t^c| \right) \right\rfloor , \tag{3.56}$$

and the new occupancy values of all particles in that cell are uniformly set to

$$o_\chi = \frac{m(D_t^c)}{|\mathcal{X}_t^c|} , \quad \chi \in \mathcal{X}_t^c ,\, |\mathcal{X}_t^c| > 0 , \tag{3.57}$$

such that their sum corresponds to the updated dynamic mass $m(D_t^c)$ of the dynamic grid map $\mathcal{M}_t$ as defined in (3.10) and thus both representations retain consistent. In the following, it is discussed how the predicted particle population $\hat{\mathcal{X}}_t^c$ within a cell can be weighted based on radar velocity measurements and camera orientation estimates.

3.5.2 Radar- and Camera-Based Particle Velocity Weighting

The occupancy evidence regulates the number of particles of a grid cell, but that measurement does not enable an individual weighting of the particles within the cell due to the spatially discretized grid structure. In other words, the position hypotheses of all particles of the same grid cell are equally likely. The velocity hypotheses, however, can be weighted when additional radar- or camera-based measurements are available.

Radar-Based Radial Velocity Component Weighting

Radar detections include a direct measurement of the radial velocity component v_z^r along the measurement direction θ_z^r, as described in Section 2.3.3 and also used in the occupancy classification in Section 3.4.3. This enables a weighting of the radial component of the velocity of a particle χ, as similarly performed in [88, 121]. The projection of the 2-D Cartesian particle velocity to the radial direction θ_z^r is described by

$$v_\chi^r = \cos(\theta_z^r)\, v_\chi^x + \sin(\theta_z^r)\, v_\chi^y . \tag{3.58}$$

The corresponding particle velocity weight results in

$$w_\chi^{\text{radar}} \propto \mathcal{N}(v_\chi^r; v_z^r, \sigma_{v,z}^2) , \tag{3.59}$$

which represents a Gaussian distribution with the measurement mean v_z^r and the variance $\sigma_{v,z}^2$, evaluated at the radial velocity component v_χ^r of the particle χ.

Camera-Based Orientation Weighting

The camera measurements include detected objects with the corresponding object orientation $\varphi_z^{\text{camera}}$. This orientation information can also be used to weight the movement orientation of the particles, which is determined by

$$\varphi_\chi = \arctan\left(\frac{v_\chi^y}{v_\chi^x}\right) \tag{3.60}$$

with the 2-D Cartesian particle velocity (v_χ^x, v_χ^y). Thereby it has to be taken into account that only the orientation of the object is determined here, but not whether the actual movement itself is forward or backward. Hence, the evaluated difference between the particle movement direction and the camera orientation is given by

$$\Delta\varphi_\chi = \min\left((|\varphi_\chi - \varphi_z^{\text{camera}}| \bmod \pi),\ \pi - (|\varphi_\chi - \varphi_z^{\text{camera}}| \bmod \pi)\right), \tag{3.61}$$

resulting in the corresponding camera-based particle weight

$$w_\chi^{\text{camera}} \propto \mathcal{N}(\Delta\varphi_\chi; 0, \sigma_\varphi^2) \tag{3.62}$$

that is modeled equivalently to (3.59), here with a zero-mean Gaussian distribution since directly the orientation difference $\Delta\varphi_\chi$ is evaluated.

Combined Weighting

The two weights are combined by a simple multiplication, i.e.,

$$w_\chi = w_\chi^{\text{radar}}\, w_\chi^{\text{camera}}, \tag{3.63}$$

which enables a full 2-D velocity weighting of the corresponding particles when both the radial velocity w_χ^{radar} and the orientation weight w_χ^{camera} are present. To increase the overall robustness against wrong measurements, the minimum value of both weights w_χ^{radar} and w_χ^{camera} can additionally be limited to a value greater than zero such that particles still have a chance to survive in that case.

3.5.3 Initialization of New Particles

If no particles are present in an evaluated grid cell, or when additional random particles should be drawn that are independent of the predicted population, new particles have to be initialized, similar to the aspects of the particle weighting. For the particle position x_χ no preferred position within the discretized grid cell can be evaluated by the measurements, as also no position weight is determined. Hence, the position of all new particles is distributed uniformly within the spatial extent of that grid cell c, i.e., within the area $\mathcal{A}(c)$ as defined in (3.9). The initial particle velocity sampling, however, can be enhanced toward the real velocity by evaluating radar- and camera-based measurements, similar to the weights w_χ^{radar} and w_χ^{camera} as well as the combination of both. This results in different cases of the particle velocity sampling, which are illustrated Fig. 3.8 and discussed in the following.

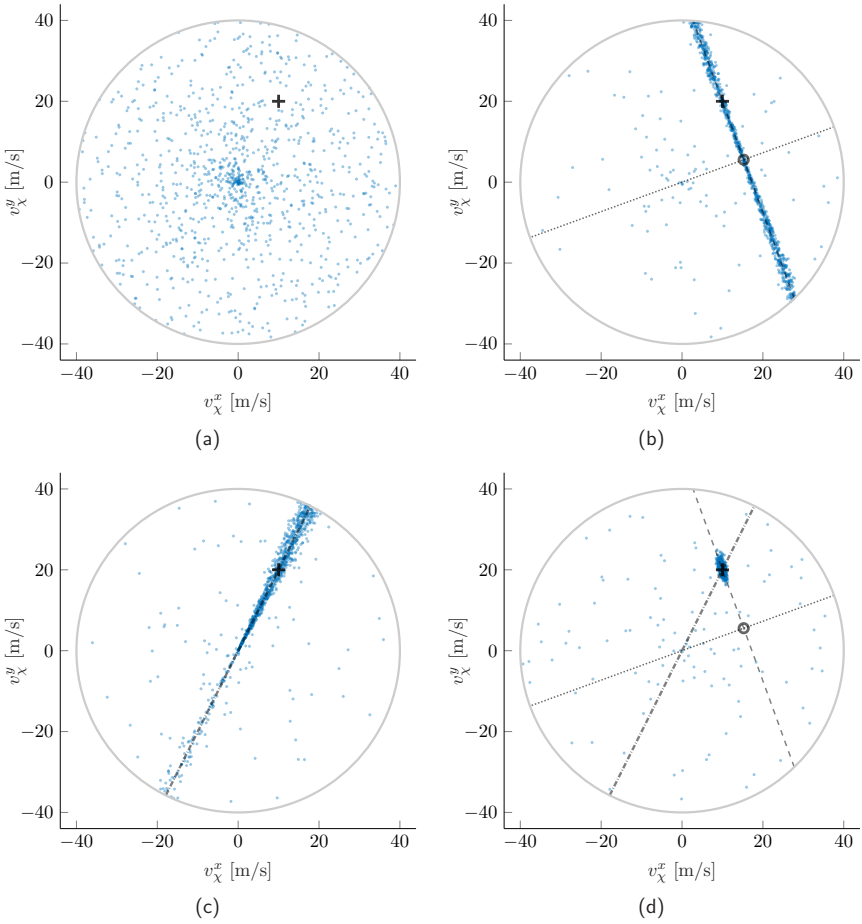


Fig. 3.8: Illustration of the initial particle velocity sampling with differently available measurements. Each figure shows a possible realization with 1000 sampled particle hypotheses (blue points) with $v_{\max} = 40\,\mathrm{m/s}$ and the real velocity $v = [10, 20]^\mathsf{T}\ \mathrm{m/s}$ (plus symbol). The measurements are selected noise-free here for simplification. Independent of the measurement, each particle is drawn uniformly with probability $p_{v,u} = 0.1$ here. (a) Sampling without measurement information. (b) Sampling with radar-based radial velocity measurement $v_z^r = 16.24\,\mathrm{m/s}$ (small circle), with radial direction $\theta_z^r = 20\,\mathrm{deg}$ (radial: dotted line; tangential: dashed line), $\sigma_{v,z}^2 = 0.5\,\mathrm{m^2/s^2}$, $\sigma_{\theta,z}^2 = 0.15$. (c) Sampling with camera-based orientation measurement $\varphi_z^{\mathrm{camera}} = 63.43\,\mathrm{deg}$ (dash-dotted line), $p_{\mathrm{forward}} = 0.9$, $\sigma_{\varphi,z}^2 = 0.2$. (d) Sampling with combination of radar- and camera-based measurements.

Velocity Sampling Without Measurement Information

When no measurement information is available for the particle velocity sampling, a uniform distribution is selected. Thereby the maximum velocity magnitude of the initial particle velocity v_χ is limited by a maximum modeled value $v_{\max}$, i.e., $\|v_\chi\| \leq v_{\max}$. To ensure that the sampled 2-D Cartesian velocity $v_\chi = [v_\chi^x, v_\chi^y]^\mathsf{T}$ is within the circle with the radius $v_{\max}$, values are sampled in the polar space with the random variables

$$\Phi_u \sim \mathcal{U}(-\pi, +\pi)\,, \tag{3.64}$$

$$V_u \sim \mathcal{U}(0, v_{\max})\,, \tag{3.65}$$

and the corresponding concrete realizations of the orientation ϕ_u and the velocity magnitude v_u, respectively. Hence, the initial particle velocity

$$v_\chi = v_u \begin{bmatrix} \cos(\phi_u) \\ \sin(\phi_u) \end{bmatrix} \tag{3.66}$$

is determined by the transformation from the polar space to the Cartesian space.

The particle velocity sampling can be enhanced when radar or camera measurements are present, which is presented in the following. However, to remain robust against completely wrong measurements, similar to possible lower bounds of the particle velocity weights, each newly drawn particle thereby still remains a probability $p_{v,u}$ of being sampled uniformly as defined in (3.66) and thus ignore any corresponding velocity measurement information. To simplify the following discussion, however, this probability $p_{v,u}$ is not further stated below.

Velocity Sampling with Radial Velocity Measurement

The initial sampling of the velocity can be significantly improved when a radar-based radial velocity measurement is available, since then only the tangential component remains unknown. Therefore, the Cartesian particle velocity

$$v_\chi = \underbrace{\begin{bmatrix} \cos(\theta_R) & -\sin(\theta_R) \\ \sin(\theta_R) & \cos(\theta_R) \end{bmatrix}}_{R_{\theta_R}} \begin{bmatrix} v_R \\ v_T \end{bmatrix} \tag{3.67}$$

is determined in this case by the two components of the radial velocity v_R and the tangential velocity v_T using the rotation matrix R_{θ_R} with the evaluated radial direction θ_R. That radial direction is given by

$$\theta_R = \theta_z^r + q_{\theta,z}\,, \tag{3.68}$$

with the measurement angle θ_z^r and additive zero-mean Gaussian noise $q_{\theta,z}$ with variance $\sigma_{\theta,z}^2$. As described before, the maximum particle velocity is limited by the value $v_{\max}$,

$$\|v_\chi\|^2 = v_R^2 + v_T^2 \leq v_{\max}^2\,, \tag{3.69}$$

which also applies to the combination of the radial and tangential component. Therefore, as also proposed in [5], the possible range of the tangential component is restricted by

$$|v_T| \leq v_{T,\max}(v_R) = v_{\max}\sqrt{1 - \frac{v_R^2}{v_{\max}^2}}\,, \tag{3.70}$$

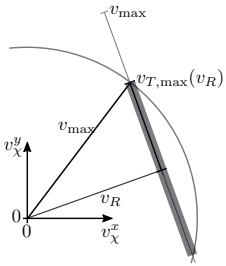


Fig. 3.9: Illustration of the maximum tangential velocity $v_{T,\max}(v_R)$ (gray interval) depending on the radial velocity v_R and a fixed limitation $\|v_\chi\| \leq v_{\max}$ (circle) of the overall Cartesian particle velocity v_χ. Figure similarly published in [5].

with $v_{\max} \geq |v_R|$, which is also illustrated in Fig. 3.9. Despite that limitation, no further information about the tangential component is available. Overall, the radial and tangential velocity components of (3.67) underlie the distributions

$$V_R \sim \mathcal{N}(v_z^r, \sigma_{v,z}^2)\,, \tag{3.71}$$

$$V_T \sim \mathcal{U}(-v_{T,\max}(v_R),\, +v_{T,\max}(v_R))\,, \tag{3.72}$$

i.e., the radial velocity v_R is given by the measurement v_z^r with additive Gaussian noise by the variance $\sigma_{v,z}^2$, equivalent to (3.59), and the tangential velocity v_T is modeled by a uniform distribution using the restriction of the maximum possible value $v_{T,\max}(v_R)$.

Velocity Sampling with Orientation Measurement

Similar to the velocity sampling of radar measurements, the camera orientation measurement $\varphi_z^{\text{camera}}$ also directly measures one component of the target 2-D velocity distribution. Hence, equivalent to (3.66), the particle velocity distribution is sampled by

$$v_\chi = s_v\; v_u \begin{bmatrix} \cos(\varphi_{\chi,z}) \\ \sin(\varphi_{\chi,z}) \end{bmatrix} \tag{3.73}$$

Thereby the orientation $\varphi_{\chi,z}$ is given by the measurement with additive Gaussian noise q_z^φ that is modeled with variance $\sigma_{\varphi,z}^2$, i.e.,

$$\varphi_{\chi,z} = \varphi_z^{\text{camera}} + q_z^\varphi\,. \tag{3.74}$$

However, as discussed before, the camera-based extraction only describes the object orientation, but not whether that object actually moves forward or backward. Hence, in addition to the uniform distribution of the velocity magnitude $v_u \in [0, v_{\max}]$, an additional discrete binary random variable $s_v \in \{+1, -1\}$ is introduced to model the sign of that velocity magnitude, with the probability $P(s_v = +1) = p_{\text{forward}}$ of the forward movement and the remaining probability $P(s_v = -1) = p_{\text{backward}} = 1 - p_{\text{forward}}$ of the backward movement. If no model assumption of the binary movement direction is available, both directions are modeled with an equal probability, i.e., $p_{\text{forward}} = p_{\text{backward}} = 0.5$, which then corresponds to the Rademacher distribution [47].

Velocity Sampling with Combination of Measurements

The most accurate velocity initialization that is close to the actual object velocity is achieved by combining the measurements of the radar Doppler velocity and the camera object orientation estimation. As stated in [5], the full 2-D velocity can be determined from the measured radial velocity when the movement orientation is known, since then the unobserved tangential velocity component is calculated by

$$v_T = v_R \tan(\theta_R - \varphi_{\chi,z}) \,, \quad (\theta_R - \varphi_{\chi,z}) \bmod \pi \neq 0 \,, \tag{3.75}$$

using the measured orientation $\varphi_{\chi,z}$ of the camera with the additional uncertainty as modeled in (3.74) and equivalently the radial direction θ_R of (3.68) as well as the distribution of the radial velocity component of (3.71). The particle velocity v_χ is thus eventually determined by (3.67). Note that the movement direction $\varphi_{\chi,z}$ in (3.75) only has to define the movement axis, but not whether the movement along that axis is forward or backward, i.e., the camera-based forward movement probability p_{forward} is not evaluated in this case, as that direction is implicitly resolved by the sign of the measured radial velocity v_R. However, the equation (3.75) is only defined if the radial direction and the object orientation are not orthogonal to each other, i.e., $(\theta_R - \varphi_{\chi,z}) \bmod \pi \neq 0$ or above a small value to avoid singularities, as then the radial velocity component is zero. In such a case, the particle velocity has to be determined by the camera orientation as in (3.73). Additionally, errors between the measured radial velocity and the camera orientation can be detected if the magnitude of the determined tangential component v_T of (3.75) and thus the combined velocity is unreasonably high, i.e., if $|v_T|$ is above the maximum value $v_{T,\max}(v_R)$ as defined in (3.70). In such a case, if the cause of the error is unknown, the uniform particle velocity sampling of (3.66) is used as the most conservative choice without evaluating those measurements.

3.5.4 Resampling of the Particle Population

It has been shown how velocity or orientation measurements of radar and camera sensors are evaluated to determine particle weights w_χ of the predicted population $\hat{\mathcal{X}}_t^c$ as well as the velocity distributions for the sampling of new particle hypotheses. Moreover, the new number of particles $|\mathcal{X}_t^c|$ within the cell c has been determined in (3.56). On this basis, the new particle population $\mathcal{X}_t^c$ has to be drawn, i.e., the resampling of the particle population.

For the selection of the new particle population, the low variance sampling as proposed in [128] is used. Hence, instead of drawing particles independently of each other, they are drawn proportionally based on their weights and only a single random number is chosen to draw the remaining fractional part with the probability given by the particle weight. For example, for 4 existing particles with equal weights and a total of 9 desired particles without new initialization, each existing particle is selected at least twice and one is selected a third time, rather than drawing 9 times any of the 4 existing particles.

Relative Adjustment of the Particle Population

To minimize the creation and copying of particles, only the difference of the particles

$$\Delta n_{\mathcal{X}} = |\mathcal{X}_t^c| - |\hat{\mathcal{X}}_t^c| \tag{3.76}$$

that have to be added or deleted based on the predicted population can be applied as an approximation to increase the computational performance. Hence, particles of the predicted population have to be copied or new random particles have to be initialized if $\Delta n_{\mathcal{X}} > 0$, whereas the particle population has to be reduced, i.e., individual particles have to be deleted, if $\Delta n_{\mathcal{X}} < 0$. This also means that for $\Delta n_{\mathcal{X}} = 0$ no modification of the particle population $\hat{\mathcal{X}}_t^c$ is applied, even though the particle weights w_χ might be different.

New random particles, i.e., independent of the predicted population, are sampled as presented in Section 3.5.3 when the particle population has to be increased ($\Delta n_{\mathcal{X}} > 0$). This is obviously necessary when no predicted particles are present in a cell ($|\hat{\mathcal{X}}_t^c| = 0$) and thus no particles can be copied. In addition, as proposed in [118], even with existing particles in that cell, some new random particles are sampled in order to reduce the particle deprivation problem [128], i.e., to remain the variety of different particle hypotheses and thus the possibility that the present particles may not represent the real object movement. The number of random particles thereby depends on the weighting of the predicted particle population, such that more random particles are sampled when those predicted particles do not match the velocity measurements and vice versa [118]. This corresponds to a fixed weight $w_{\chi'}^{\text{random}}$ for the initialization of a new random particle, which has to be selected with respect to the particle weights w_χ as determined in (3.63). In other words, this value has to be modeled as $w_{\chi'}^{\text{random}} > w_\chi$ if the probability of sampling a random particle should be higher that a particle with the weight w_χ and vice versa.

Deletion of Particles with Particle Weights

The low variance sampling [128] as used in the copying of particles can also be applied for the reduction of the particle population, i.e., this way only those particles are selected that have to be deleted from the existing population. The following approach for the deletion of particles is based on preliminary work of this thesis as initially presented in [116].

The measurement-based weights w_χ as determined in (3.63), which are assumed to be normalized by $w_\chi \in [0, 1]$ in the following, have to be reversed for the deletion, meaning that the weight for a particle being selected for deletion has to be selected as

$$\tilde{w}_{\chi,\text{del}} = \max(1 - w_\chi, w_{\text{del}}^{\min}) \,, \tag{3.77}$$

such that particles that do not correspond with the measurement result in a high probability of being deleted, while each particle receives at least a minimum weight for deletion $w_{\text{del}}^{\min} > 0$. However, if more than one particle has to be deleted ($\Delta n_{\mathcal{X}} < -1$), particles with high deletion weights $\tilde{w}_{\chi,\text{del}}$ may be selected several times for the deletion, which, in contrast to the copying of particles, is senseless, as each particle can be at most deleted once. A particle is at most chosen once for the deletion, if the corresponding weight is smaller than or equal to the limit

$$w_{\text{del}}^{\lim} = \frac{\sum \tilde{w}_{\chi,\text{del}}}{|\Delta n_{\mathcal{X}}|} \,, \quad \chi \in \hat{\mathcal{X}}_t^c \,, \tag{3.78}$$

as this value corresponds to the step width of the low variance selection. To ensure this limitation of the deletion, the difference of the weights $\tilde{w}_{\chi,\text{del}}$ above that limitation $w_{\text{del}}^{\lim}$

Algorithm 2 Modification of the particle weights for the deletion.
Require: Number of particles to delete n_{del} and initial weights $w_{\chi,\text{del}} = \tilde{w}_{\chi,\text{del}}$.
Ensure: Modify weights $w_{\chi,\text{del}}$ such that each particle is at most selected once for deletion.

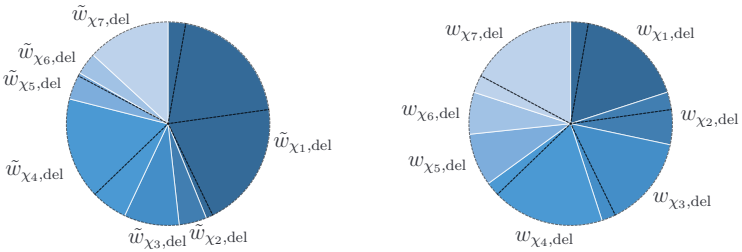
$w_{\text{del}}^{\text{lim}} \leftarrow \dfrac{\sum w_{\chi,\text{del}}}{n_{\text{del}}}$ ▷ Overall limit of all individual weights
while $\max(w_{\chi,\text{del}}) > w_{\text{del}}^{\text{lim}}$ **do**
  $\Delta S_a \leftarrow 0$ ▷ Sum of the difference of the weights *above* $w_{\text{del}}^{\text{lim}}$
  $S_b \leftarrow 0$ ▷ Sum of the weights *below* $w_{\text{del}}^{\text{lim}}$
  for all $w_{\chi,\text{del}}$ **do** ▷ Determine the sums $\Delta S_a, S_b$
    if $w_{\chi,\text{del}} > w_{\text{del}}^{\text{lim}}$ **then**
      $\Delta S_a \leftarrow \Delta S_a + w_{\chi,\text{del}} - w_{\text{del}}^{\text{lim}}$
      $w_{\chi,\text{del}} \leftarrow w_{\text{del}}^{\text{lim}}$ ▷ Limit weights above $w_{\text{del}}^{\text{lim}}$
    else if $w_{\chi,\text{del}} < w_{\text{del}}^{\text{lim}}$ **then**
      $S_b \leftarrow S_b + w_i$
    end if
  end for
  for all $w_{\chi,\text{del}}$ **do**
    if $w_{\chi,\text{del}} < w_{\text{del}}^{\text{lim}}$ **then**
      $w_{\chi,\text{del}} \leftarrow w_{\chi,\text{del}} \left(1 + \dfrac{\Delta S_a}{S_b}\right)$ ▷ Increase weights below w_{lim}
    end if
  end for
end while

has to be distributed to the weights below the limitation to remain the overall sum of the weights and the number of particles that are selected for deletion. Therefore, the particle weights for the deletion are iteratively adjusted as presented in Algorithm 2.

This modification is illustrated in an example in Fig. 3.10 with $|\hat{\mathcal{X}}_t^c| = 7$, $\Delta n_{\mathcal{X}} = -5$, and the following deletion weights:

originals weights $\tilde{w}_{\chi,\text{del}}$:	[1.0,	0.1,	0.200,	0.5,	0.1,	0.080,	0.300]
first iteration:	[0.456,	0.175,	0.351,	0.456,	0.175,	0.140,	0.526]
modified weights $w_{\chi,\text{del}}$:	[0.456,	0.19,	0.38,	0.456,	0.19,	0.152,	0.456]

Thereby the limitation results in $w_{\text{del}}^{\text{lim}} = \frac{2.28}{7} = 0.456$; values above that limitation are highlighted in gray. For the original weights with a possible random selection shown by the black dotted lines in Fig. 3.10a, the particle χ_1 with the deletion weight $\tilde{w}_{\chi_1,\text{del}}$ would be selected three times, but also with each other low variance sampling configuration, which visually corresponds to a rotation of the dotted circular structure, it would be at least selected twice. After the second iteration of the proposed weight adaptation, the final weights $w_{\chi,\text{del}}$ are determined, ensuring that each particle is at most selected once for the deletion, i.e., $w_{\chi,\text{del}} \leq w_{\text{del}}^{\text{lim}}$. In this example, the particles χ_1, χ_4, and χ_7 are definitely deleted, and two more particles are selected from the remaining particles, in the configuration of Fig. 3.10b the particles χ_2 and χ_3 are chosen.



(a) Original weights for the deletion. (b) Modified weights for the deletion.

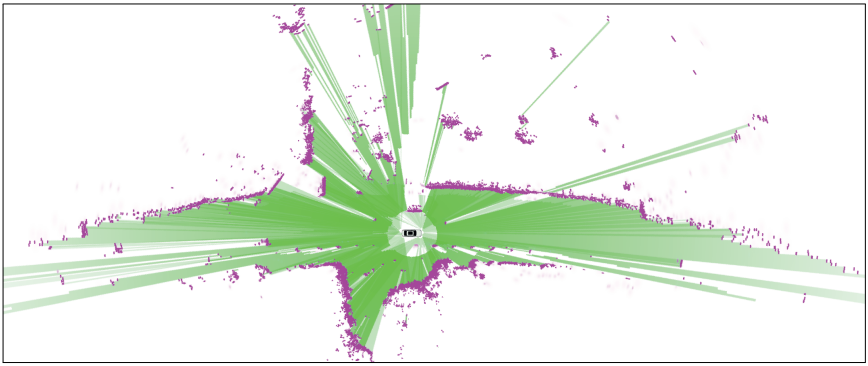
Fig. 3.10: Modification of the particle weights for the deletion using the low-variance sampling concept. The weights are adapted such that each particle is at most selected once for the deletion. The example shows 7 existing particles, from which 5 particles have to be deleted.

Overall, this section has presented how the particle population $\mathcal{X}_t$ is updated such that the combination with the dynamic evidence mass $m(D_t)$ of the updated map $\mathcal{M}_t$ remains consistent and the convergence of the particle tracking is achieved in terms of the resampling of the most promising particle hypotheses given the measurement data. Thereby the velocity hypotheses of the predicted particle population $\hat{\mathcal{X}}_t$ are weighted by radar and camera measurements, which are also used for the initial velocity distribution of newly sampled particles. The number of particles is adjusted cell-wise, primarily proportional to the dynamic occupancy evidence mass, and only the desired difference of the cell-individual particle population is efficiently updated by adding or deleting particles.

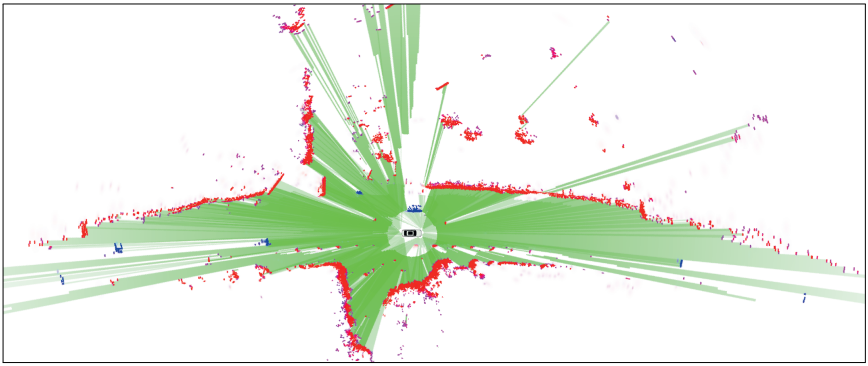
3.6 Augmented Measurement Grid

The dynamic grid map $\mathcal{M}_t$ as introduced in this chapter represents an accumulated estimation of the local environment including a differentiation of static and dynamic occupancy. This classification is also important for the subsequent object tracking, as only occupied cells classified as dynamic have to be considered for the moving object tracking, whereas the static environment does not require a further abstraction. To avoid multi-filtering of the object tracking, however, it is important to update objects only with occupied cells that are currently measured rather than the filtered occupancy accumulation of the dynamic grid map. Therefore, in the following, a new representation is introduced – the augmented measurement grid – which is briefly presented in this section. A comparison of the fused measurement grid without further occupancy differentiation, as derived in the previous chapter, and the adjusted augmented representation is demonstrated in Fig. 3.11, which is based on the dynamic grid map of Fig. 3.1.

The main idea of this augmented measurement grid representation is to use the filtered evidence masses of the dynamic grid map $\mathcal{M}_t$ for distinguishing static and dynamic occupancy of the measured occupancy belief of the fused measurement grid $\mathcal{M}_{z,t}$. This strategy is similar to the radar- and camera-based occupancy classification as proposed in Section 3.4.3. However, the subsequent object extraction should not fully depend or rely on



(a)



(b)

Fig. 3.11: Example of the fused measurement grid as originally measured and with the additional static/dynamic occupancy classification based on the estimation of the dynamic grid map of Fig. 3.1a. (a) Fused measurement grid. (b) Augmented (fused) measurement grid.

the additional radar or camera measurement data of that current fused measurement grid. Therefore, it is essential that this static/dynamic differentiation is generically estimated and robustly filtered, which is directly estimated by the dynamic grid map. Overall, this results in a pseudo measurement $\tilde{z}$, where the occupancy classification is selected by

$$m(S_{\tilde{z},t}) = \min\left(m(SD_{z,t})(1-m(D_t)),\ m(S_t)\right), \tag{3.79}$$

$$m(D_{\tilde{z},t}) = \min\left(m(SD_{z,t})(1-m(S_t)),\ m(D_t)\right), \tag{3.80}$$

$$m(SD_{\tilde{z},t}) = m(SD_{z,t}) - m(S_{\tilde{z},t}) - m(D_{\tilde{z},t}). \tag{3.81}$$

Thereby the evidence mass of the static or dynamic occupancy hypothesis achieves at most the filtered value of the dynamic map, but it is also limited by the measurement belief itself and the conflict with the contradicting hypothesis. Another strategy is to multiply the

corresponding evidence mass of the filtered map with the measurement, which, however, results in a more conservative classification of the pseudo measurement when the value of the map is not fully accumulated yet.

Since the measurement belief remains unchanged and equals the original occupancy measurement, i.e., $\mathrm{bel}(O_{\tilde{z},t}) = \mathrm{bel}(O_{z,t}) = m(SD_{z,t})$, this classified measurement grid still represents an uncorrelated and currently measured occupancy belief, only the subdivision between $\{S\}$, $\{D\}$, and $\{S, D\}$ is adapted by the filtered map. Therefore, this augmented measurement grid can be used for a measurement update to the subsequent object tracking without resulting in a multi-filtering of the occupancy-based position filtering.

The particle tracking as part of the dynamic grid mapping does not only support distinguishing static and dynamic occupancy, it further enables a robust low-level velocity estimation of dynamic parts without requiring specific object assumptions. As introduced in (3.11), the 2-D cell velocity is estimated by the mean particle velocity of all particles in a cell, weighted by the individual particle occupancy values. The corresponding velocity grid representation has been demonstrated in Fig. 3.1b. This cell-wise velocity estimation forms also an important input to the object extraction, since it further enables a more sensible clustering of objects that move closely to each other and velocity estimates of the object instances by calculating the mean velocity of all associated grid cells. Therefore, also included in the augmented measurement grid representation. Note that the particle-based velocity estimation is filtered and thus correlated. Hence, as discussed in more detail in the object state estimation in Chapter 5, this velocity estimate is primarily used for the initialization of the object velocity but not for the subsequent measurement updates, for which the directly measured and thus uncorrelated radar Doppler velocities, which are also still included in the measurement grid, are then primarily used.

3.7 Results and Summary

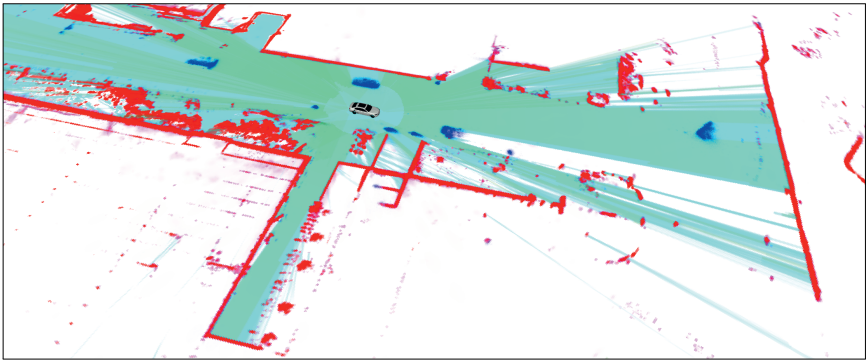
This chapter has discussed how the local environment is robustly modeled and estimated by a temporal filtering of different measurement data. A novel approach has been proposed which combines the benefits of a persistent occupancy grid mapping accumulation with a short-term grid-based particle tracking for the dynamic state estimation in changing environments with moving objects. The presented evidential dynamic grid map forms the core component that processes all incoming fused measurement grids by an adapted evidential filtering within the Dempster–Shafer framework. Thereby the evidential occupancy hypothesis as used in the fused measurement grid of the previous chapter is subdivided into separate hypotheses of static occupancy and dynamic occupancy, which is essential for differentiating moving and stationary parts of the occupied environment. The evidential filtering of the dynamic grid map is further adapted to model the temporal transition between previously measured freespace toward the superset of passable area, specific assignment of occurring conflicts between the different hypotheses, and an adjusted occupancy differentiation to achieve the convergence from unclassified occupancy into the subsets of static or dynamic occupancy. Dynamic occupancy masses of the grid map are directly coupled with the particle tracking, i.e., the temporally filtered particle population corresponds to position and velocity hypotheses of potentially moving parts of the environment. Hence,

this enables a robust velocity distribution estimation of individual grid cells and thus also a robust prediction of their dynamic occupancy masses. Static occupancy, in contrast, is directly accumulated in the persistent map representation without requiring particles, which avoids information loss during temporary occlusion of the static environment and further reduces the computational effort of the particle estimation.

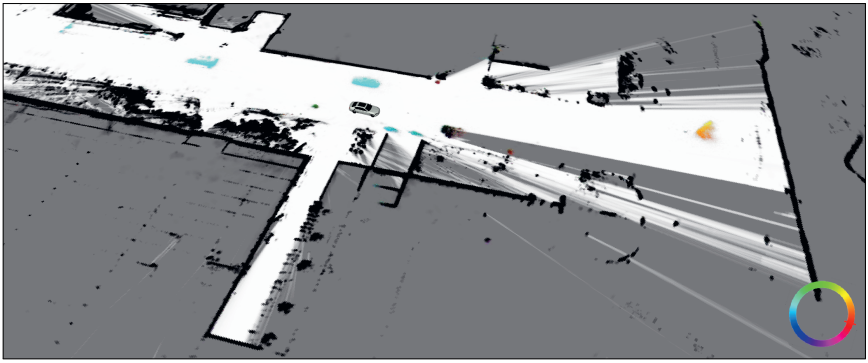
Overall, this approach results in an efficient and consistent temporal filtering of the different evidential hypotheses of static occupancy, dynamic occupancy, freespace, and their combined hypotheses supersets, but also in a robust low-level velocity estimation, all without requiring specific object assumptions. An example of the dynamic grid map including the velocity estimation has been demonstrated with real sensor data in Fig. 3.1. This generic estimation strategy primarily requires measured evidence masses of unclassified occupancy and freespace as input, i.e., the fused measurement occupancy grid as shown in Fig. 3.11a. Additional information of the measurement data such as a static/dynamic classification is not necessarily required, as this is implicitly estimated by the filtering over time including the particle-based movement estimation. Hence, even if that measurement classification is not directly available, the measurement data can then be indirectly classified into static and dynamic based on the filtered estimation of the dynamic grid map, resulting in the augmented measurement grid as demonstrated in Fig. 3.11b. That representation forms the main input of the subsequent detection and tracking of moving objects, where only areas of dynamic occupancy have to be evaluated. But it has also been discussed in detail how additional radar Doppler velocity measurements, camera classification and object orientation information, and feedback of the high-level object tracking further contribute to the dynamic grid mapping and particle tracking as optional inputs.

Two more challenging urban scenarios are demonstrated in Fig. 3.12–3.13. The first scenario includes several oncoming vehicles and cyclists, a vehicle turning left, a stopped vehicle in front of the ego vehicle, and multiple surrounding pedestrians, which are all robustly estimated as dynamic occupied with their respective velocities. Thereby the stopped vehicle in the front has been observed moving before and an object track has been extracted, therefore, even though the velocity correctly converges toward zero, it remains dynamic occupied rather than converging toward static by the feedback of that predicted object to the dynamic grid map as presented in (3.52). The particle-based velocity estimation is further demonstrated in Fig. 3.12c for the left-turning vehicle, highlighting the robust cell-wise velocity estimation. As shown in that example, the velocity vectors are accurately estimated for the individual grid cells, solely based on the particle movement estimation, which, in the case of such a turning object, differ due to its turn rate. The second scenario in Fig. 3.13 also demonstrates the successful dynamic grid map estimation for a complex urban scenario with several crossing cars, a bus, and multiple pedestrians. There are also two boom barriers in the front area, which are robustly estimated as static obstacles, and pedestrians directly next to the static environment that are correctly distinguished by the dynamic occupancy classification and the respective velocity estimation.

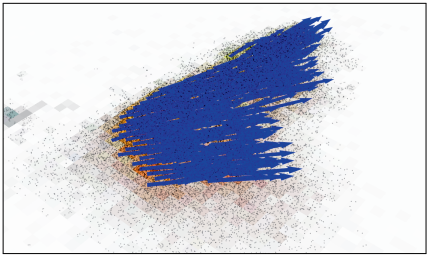
Based on this robust low-level dynamic grid mapping, moving objects have to be abstracted in the following from those individual occupied grid cells, i.e., a spatial representation of object instances, as that object abstraction is essential for a scene understanding and maneuver planning for autonomous driving with surrounding traffic participants.



(a)



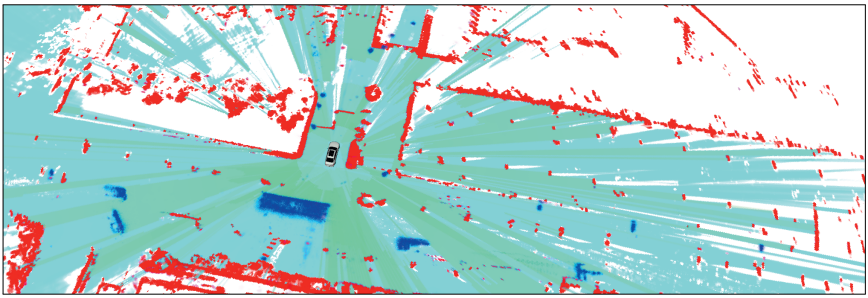
(b)



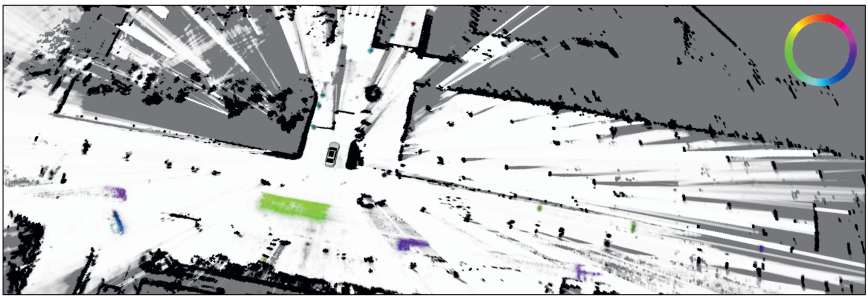
(c)

(d)

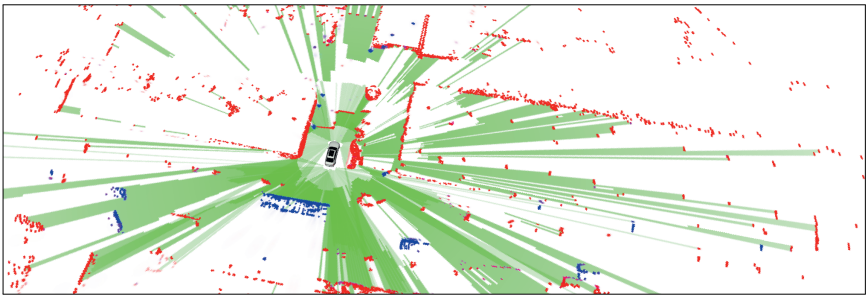
Fig. 3.12: Results of the dynamic grid mapping and particle tracking with real sensor data for a challenging urban scenario. (a) Evidential dynamic grid map. (b) Particle-based velocity orientation visualization. (c) Detailed representation of the particle population and resulting cell-wise velocity estimation for the left-turning vehicle. (d) Camera images (rear/front).



(a)



(b)



(c)

(d)

(e)

Fig. 3.13: Further real data results of the dynamic grid mapping. (a) Evidential dynamic grid map. (b) Particle-based velocity orientation visualization. (c) Augmented measurement grid representation. (d/e) Camera images (rear/front) of the scenario.

4 Object Extraction and Association

The dynamic grid mapping of the previous chapter enables a robust occupancy grid mapping with a temporal accumulation of the fused measurement grids in dynamic environments including a differentiation of static and dynamic occupancy. The applied particle-based estimation thereby also enables a robust low-level tracking of the dynamic state of the occupied environment including cell-wise velocity estimates. This chapter investigates the subsequent extraction of moving objects as well as the association with predicted object tracks that have been extracted at a previous time. In other words, the cell-individual estimation of the dynamic occupancy grid is abstracted in the following toward the set or cluster of dynamic grid cells that belong together as they are originated by the same moving extended object instance. New objects are extracted by a density-based clustering of dynamic occupied cells with similar cell velocities, which is further extended by a subsequent connectivity-based clustering of neighboring occupied cells and a velocity variance analysis to minimize false positives of the object abstraction. Moreover, a new association approach is proposed that directly evaluates the individual grid cells before forming object cell clusters. Thereby the particles of the underlying low-level particle tracking, as introduced in the previous chapter, are linked with the high-level object tracks, i.e., an object label is attached to each particle. Occupied cells are thus associated to objects by evaluating the particle label distribution within each grid cell. In addition, a subsequent clustering is performed, in which multiple sub-clusters of an object are extracted and finally checked for plausibility to further increase the robustness of the association. The object state estimation, including the abstraction of the bounding box shape, is discussed separately in Chapter 5 based on the set of associated occupied cells of this chapter.

The work presented in this chapter is based on [2] and [3] that have been published in the context of this thesis, which are not explicitly referenced in the following.

4.1 Introduction

Object detection is an essential task with the aim of abstracting measurement data toward a more generic high-level representation of object instances, which is crucial for any kind of scene understanding or interpretation. For mobile robots, it is important to know where other traffic participants are in the current local environment, for example surrounding vehicles, cyclists, or pedestrians. This object extraction forms the basis for planning interactive maneuvers and avoiding collisions with other traffic participants, i.e., for enabling safe and intelligent autonomous driving applications. Those objects thereby have to be extracted from different sensor data, such as camera images, range-based detections of lidar or radar sensors, or some other pre-processed or fused representation of measurement data. Due to the wide range of sensors with their respective characteristics as well as different types of considered objects, this results in a broad field of research with various

approaches and applications. The same applies to the temporal filtering of objects for a robust object estimation over time, requiring an association between current measurement data and the predicted objects of the previous time instance, which also depends on the measurement data and the corresponding object detection or abstraction.

In this work, measurement data of lidar, radar, and camera are modeled within the generic occupancy grid representation, followed by the grid-based sensor fusion as well as the temporal accumulation of the evidential dynamic grid mapping, as presented in the previous chapters. The derived augmented measurement grid, as introduced in Section 3.6, additionally combines the unaccumulated sensor data of the fused measurement grid with the filtered static/dynamic occupancy differentiation of the dynamic grid map and the particle-based cell velocity estimation. Based on that enhanced occupancy grid representation, the goals of this chapter are the extraction of newly occurring moving objects and the association of the current measurement with the predicted object tracks that have been extracted before. The contribution and outline of this chapter are presented after the following brief discussion of related work in that context.

4.1.1 Related Work

Based on the occupancy grid representation of measurement data, there are various strategies for distinguishing static and dynamic occupied parts of the environment, as discussed in detail in the previous chapter. Thereby a simple concept for classifying occupied cells as dynamic is to determine free/occupied inconsistencies between the accumulated map and new measurements, e.g., [80, 135–137]. This concept is used in multiple approaches to classify parts of the measurement data as dynamic, which are then clustered to extract object hypotheses. In [135, 136], a distance-based clustering is applied to abstract objects, which are temporally filtered by a multiple hypothesis tracking (MHT) with a consideration of multiple matching association hypotheses. Object hypotheses are extracted similarly in [20, 106] and associated to the filtered tracks using a global nearest neighbor (GNN) approach. In addition, the object tracking and the occupancy grid mapping are linked in [20] to improve each other by extending the object states with a list of associated and accumulated grid cells. In [56], cell clusters are extracted from dynamic grid cells using a density-based clustering and a joint probability data association (JPDA) is used.

Extracting object hypotheses by clustering dynamic grid cells is essential for newly occurring objects since no previous information is available. For the association with existing tracks that have been extracted before, in contrast, information of predicted tracks is available that can be taken into account in advance of a clustering step. This is crucial for solving ambiguities of a clustering in dense environments, where, for example, one object instance can result in several clusters due to partial occlusions, but also the other way, i.e., one combined cluster can also be originated by different objects that move too close to each other such that they cannot be clearly separated in the clustering. In [77], information of existing object tracks is considered by forming a region of interest of each predicted track, which restricts possible starting points of a clustering with an association of the clustered cells to the corresponding track. Ambiguities of multiple possible objects of a cluster are solved using an additional k-means re-clustering based on the number of predicted tracks within the area of that cluster. However, their main goal is a more efficient

clustering and association, without addressing a detailed prediction of expected areas of occupancy or object shape information in those regions of interest.

As discussed in the previous chapter, recent approaches of dynamic occupancy grids, including this work, use a grid-based particle filter for robustly estimating the dynamic state of the occupied environment. Each particle thereby represents a hypothesis of occupancy at a specific position with a particular velocity. The low-level particle tracking thereby does not only improve the generic filtering of dynamic occupancy, it also results in accurate cell velocity distribution estimates. Hence, those cell velocities can be further evaluated for the clustering and association of objects moving closely to each other with different velocities. Based on the particle-based dynamic grid estimation, filtered objects can also be directly extracted by clustering dynamic cells of the filtered dynamic grid map, i.e., without a subsequent filtering on the object-level, since the particle population already represents a filtering of occupancy and velocity distributions. As stated in [29], those particles could be extended by a unique identifier (ID) to reconstruct trajectories of the extracted objects and thus perform the association over time. Similarly, in [101], an object ID of those particles is used to extract object clusters including their mean velocity and gravity center. This ID is propagated by the particle resampling, i.e., duplicated particles inherit the ID, whereas randomly drawn new particles are initialized with a new unique ID. However, the resulting object clusters tend to have similar problems as described before, i.e., a large object can result in several small clusters of estimated objects, whereas objects moving closely to each other converge toward one single cluster.

Other recent approaches [48, 95, 117, 143] use deep learning concepts for the object detection based on dynamic occupancy grids. But, in contrast to the generic cell clustering extraction of arbitrary-shaped moving objects, those machine learning-based approaches require large labeled training datasets for a robust generalization with various kinds of objects. Furthermore, those approaches do not address combining the object detection with the data association by evaluating information of predicted tracks before extracting object hypotheses to further improve both the association as well as the object detection as discussed before. Another recent strategy is to perform not only the detection of objects but also the tracking over time with deep learning approaches using recurrent neural networks (RNN), which is presented in [38] based on a dynamic grid representation. The demonstrated results, however, show that the generalization in complex urban environments is still very difficult, which is therefore simplified in [38] by a stationary ego vehicle that does not move itself.

4.1.2 Contribution and Outline

There are already multiple approaches that use similar concepts of dynamic occupancy grids as part of a high-level object tracking. The main idea thereby is that the additional static/dynamic occupancy differentiation directly simplifies the moving object detection of the environment, since only occupied areas classified as dynamic have to be evaluated. However, none of these approaches further combines in detail the advantages of the particle-based dynamic grid estimation, which results in accurate cell-wise estimates of the velocity and occupancy likelihood, with the object detection and data association tasks. Those aspects are addressed in this chapter, with the main contributions as follows:

- A new clustering-based extraction of moving objects is proposed that generically detects arbitrary-shaped moving objects by the movement itself, without requiring sensor-specific features or labeled training data. In contrast to other similar approaches, a robust combination of a density-based and a connectivity-based clustering is presented, where also the particle-based cell velocity estimates as well as the resulting velocity variance are evaluated. This improves both the ability to distinguish closely moving objects with different velocities and the reduction of false positives.

- A novel cell-level data association concept is presented, where individual grid cells of the measurement are directly associated with predicted objects even before they have to be abstracted to the object-level. This reduces information loss and association ambiguities, especially in scenarios with partial occlusions or densely moving traffic. Two strategies are presented: A simplified approach using the predicted state of object tracks and a detailed approach, the particle labeling association, where particles of the underlying low-level particle tracking are linked to object tracks, which enables an accurate occupancy likelihood approximation and thus a more robust association. Moreover, a subsequent clustering and verification concept further improves the association by forming and evaluating sub-clusters of the individual cells of an object.

In contrast to common approaches that first extract object hypotheses of the measurement and then perform the association afterwards, the processing order of those two steps is also handled differently in this work due to the proposed cell-level association strategy where measurement data are associated to predicted objects in advance of the object abstraction. This also applies to the structure of this chapter:

- *Section 4.2* first provides a brief overview of the object detection and association concepts based on the characteristics of the dynamic grid representation of the previous chapter. It is discussed how the dynamic occupancy classification simplifies the moving object detection and which strategies arise for the data association with object tracks that have been extracted before.

- *Section 4.3* presents the cell-level data association concept where individual grid cells are directly evaluated before forming object cell clusters. First, the simplified approach based on the predicted object state with a box-shape is demonstrated, which is then further enhanced by the proposed particle labeling association. Based on this cell-individual association, in a subsequent clustering step, multiple subclusters of an object are evaluated and finally checked for plausibility.

- *Section 4.4* addresses the extraction of newly occurring object tracks, which is performed after the association step and thus only applied to the remaining set of unassociated grid cells. Thereby initial hypotheses are extracted by a density-based clustering of dynamic cells, which is then further enhanced by a subsequent connectivity-based clustering with a velocity variance analysis of adjacent occupied cells.

Finally, the chapter is concluded in Section 4.5 by different qualitative results in challenging urban scenarios and a summary.

4.2 Overview of the Extraction and Association Strategies

This overview section briefly discusses how objects can be generally detected and associated based on the characteristics of the dynamic occupancy grid representation of this work. The proposed approaches for the association and extraction of objects are then presented in detail in the subsequent sections.

4.2.1 Object Detection Based on Dynamic Occupancy Classification

Object detection is generally a challenging task, in particular for a sensor-individual object detection when the evaluated features of that sensor are not very significant, e.g., due to a low number of measurement detections in larger distances, unstructured object shapes, partial occlusion, or clutter noise. The fused measurement grid of this work, as presented in Chapter 2, already simplifies the object detection task since all measurement data of a single time instance are fused in advance without requiring object assumptions. Objects can thus be detected based on the measurement information of all used sensors, which, in addition to occupancy and freespace evidence masses, can optionally also contain further radar Doppler and camera classification information. Observable objects with corresponding measurement detections can thereby only occur in occupied areas. Hence, based on the measured occupancy evidence $\mathrm{bel}(O^c_{z,t}) = m(O^c_{z,t})$ of the fused measurement grid $\mathcal{M}_{z,t}$, the set of relevant grid cells c of the grid structure $\mathcal{G}$ can be limited to

$$\mathcal{G}_{O,t} = \{c \in \mathcal{G} \mid \mathrm{bel}(O^c_{z,t}) \geq \Gamma^O_{\min}\} \subseteq \mathcal{G}\,, \tag{4.1}$$

with a required minimum occupancy belief $\Gamma^O_{\min} > 0$.

One possibility is thus to extract high-level objects for all, arbitrary-shaped, occupied areas of the environment given by the set of cells $\mathcal{G}_{O,t}$. However, as motivated before, it is not very effective to abstract object instances of the whole static environment, e.g., using a bounding box approximation. Instead, static obstacles can be modeled and estimated more accurately and efficiently by the occupancy grid map accumulation using the same grid representation as the input without further object abstraction. The dynamic grid mapping as proposed in the previous chapter enables that filtered accumulation including the differentiation of static and dynamic occupancy. Therefore, the object detection and tracking should be limited to dynamic parts of the environment, i.e., moving objects. Overall, this significantly simplifies the object detection and tracking, since only the subset

$$\mathcal{G}_{D,t} = \{c \in \mathcal{G} \mid m(D^c_{\tilde{z},t}) \geq \Gamma^D_{\min}\} \subseteq \mathcal{G}_{O,t} \tag{4.2}$$

of currently occupied cells with a dynamic occupancy mass greater or equal than a defined threshold $\Gamma^D_{\min} \geq \Gamma^O_{\min} > 0$ have to be considered for this purpose. Note that $m(D^c_{\tilde{z},t})$ represents the dynamic occupancy mass of the augmented measurement grid as introduced in (3.81) and not directly the dynamic occupancy mass $m(D^c_t)$ of the dynamic grid map $\mathcal{M}_t$, since only cells with currently measured occupancy are considered to avoid multi-filtering on the object-level.

Altogether, this multi-sensor object detection based on the dynamic grid estimation has several advantages. The dynamic grid estimation significantly simplifies the object

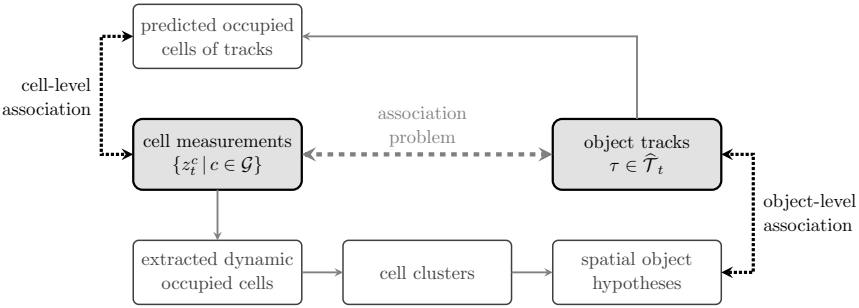


Fig. 4.1: Illustration of the association problem and a comparison of the different abstraction levels including the required processing steps.

detection task and serves as a track-before-detect strategy that enables detecting arbitrary-shaped moving objects solely based on the generic particle-based movement tracking on the grid cell-level. Objects can thus be extracted by clustering dynamic grid cells of $\mathcal{G}_{D,t}$ without requiring specific feature extraction or labeled training data. Furthermore, as radar Doppler measurements and camera information are also evaluated within the dynamic grid mapping, those measurements also implicitly improve the object detection. Note that by the additional feedback of the object tracking toward the dynamic grid mapping as introduced in (3.52), occupied cells within that predicted area remain as dynamic occupied, which enables to continue tracking objects that temporarily stop moving. Moreover, static obstacles and dynamic objects are consistently estimated by that combined approach.

4.2.2 Measurement Abstraction Levels of the Association Problem

The association between the current measurement data and the temporally filtered objects, called tracks, which have been already extracted before, can be generally performed on different levels of abstraction. Most commonly, object hypotheses are abstracted from the measurement data without considering the predicted tracks and the association step is then performed afterwards on the object-level. In terms of the dynamic grid estimation of this work, spatial objects hypotheses can be extracted by clustering dynamic occupied cells of the measurement with a subsequent abstraction of the object shape, resulting in the bottom path of Fig. 4.1 with an abstraction of the cell measurements toward the object-level representation. For new occurring objects that clustering strategy is expedient as no prior knowledge of predicted tracks is available in such a case.

However, clustering without additional information can be error-prone in dense urban environments, as demonstrated in Fig. 4.2. For example, multiple objects that temporarily move closely to each other may result in one large cluster, whereas partial occlusions of one object may result in several small clusters, which is the case in Fig. 4.2b due to a limited sensor field of view. This potentially leads to association errors on the object-level, which are preventable when information of predicted tracks is available and considered in advance of the clustering to help solving these ambiguities, as shown in Fig. 4.2e. Therefore, based

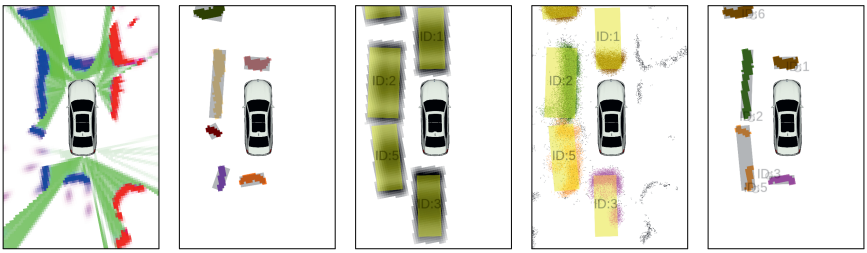


Fig. 4.2: Example of the clustering and association with multiple surrounding vehicles. The vehicle on the rear left results in two separate areas of measured occupancy here due to a limited sensor field of view of that test vehicle. (a) Classified measurement grid. (b) Clustering without prior knowledge. (c) Predicted tracks and occupancy likelihood approximation from predicted track state. (d) Predicted tracks and labeled particle population. (e) Resulting cell-wise association by considering predicted tracks. Figure initially published in [2], © 2018 IEEE.

on the cell measurements z_t^c with the individual grid cells $c \in \mathcal{G}_{D,t}$ of the set of dynamic occupied cells as defined in (4.2), another strategy is to directly associate those cells to the predicted tracks without a preceding clustering. The predicted object tracks thereby have to be evaluated in the grid cell representation, i.e., the association is then performed on the grid cell-level, as depicted by the upper path of Fig. 4.1.

Overall, both the extraction of new objects and the association with existing ones, based on the cell-level association, result in a set of dynamic occupied grid cells

$$\mathcal{C}_{\tau,t} = \{c \in \mathcal{G}_{D,t} \mid f_a(c) = \tau\} \tag{4.3}$$

that are associated to the corresponding track $\tau \in \mathcal{T}_t$, with

$$\mathcal{T}_t = \widehat{\mathcal{T}}_t \cup \mathcal{T}_t^{\text{new}} \tag{4.4}$$

describing the set of all currently tracked objects, both the predicted tracks $\widehat{\mathcal{T}}_t$ that have been extracted before and the newly extracted ones $\mathcal{T}_t^{\text{new}}$. As this work handles extended objects rather than point objects, multiple cells can be associated to a track. However, each cell is at most associated with one track in this work, i.e.,

$$\mathcal{C}_{\tau_1,t} \cap \mathcal{C}_{\tau_2,t} = \emptyset \quad \forall \tau_1, \tau_2 \in \mathcal{T}_t,\ \tau_1 \neq \tau_2\,. \tag{4.5}$$

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

$$f_a : \mathcal{G}_{D,t} \to \mathcal{T}_t \cup \{\zeta_\emptyset\} \ : \ c \mapsto \tau\,, \tag{4.6}$$

while cells that are not associated to a track are mapped to an auxiliary variable $\zeta_\emptyset$.

That cell-to-track assignment is divided into two steps in the following: The association of cells with predicted tracks that have been extracted before is performed first, presented in Section 4.3, whereas the clustering-based extraction of newly occurring objects is applied afterwards in Section 4.4, where only the remaining dynamic cells that have not already been associated are considered. The extraction of the spatial object shape and other high-level attributes are discussed separately as part of the object state estimation in Chapter 5.

4.3 Cell Association for Existing Object Tracks

As motivated before, object tracks that have been extracted before have to be associated with dynamic occupied grid cells of the current measurement. The association should thereby be performed on the grid cell-level rather than on the object-level, i.e., without requiring a high-level object hypotheses extraction to avoid information loss of such a measurement abstraction before additional information of the predicted track is taken into account. In the following, two different strategies for such a cell-wise association are presented, followed by a subsequent clustering concept of those associated cells where multiple potential sub-clusters of a track are formed and eventually checked for plausibility.

4.3.1 Association Based on Predicted High-Level Object Track

A simple strategy for an association of individual grid cells with predicted tracks is to evaluate the similarity between a dynamic occupied grid cell $c \in \mathcal{G}_{D,t}$ and the object state of a predicted track $\tau \in \widehat{\mathcal{T}}_t$ projected to that cell-level representation. The main idea thereby is to compute the expected area of measured occupancy of a predicted track, i.e., its occupancy likelihood, by an approximation based on the predicted object pose and box model shape. This generic concept has been published in more detail in [3], which is briefly summarized in the following. The low-level particle representation as available in this work, however, enables an even more accurate occupancy likelihood approximation and association strategy, which is presented in detail hereafter.

The expected area of measured occupancy of a predicted track is defined by a position-based association probability $\alpha_x(c, \tau)$, which represents the probability that a cell c with center position x_c is occupied by a predicted track τ, similarly described as a track cover probability in [86]. The probability α_x can be approximated with different degrees of detail. In the simplest case, a binary validation gate can be selected, i.e., whether a cell is inside or outside the predicted spatial object shape, e.g., using a bounding box model. A more advanced probabilistic model can additionally evaluate, for instance, the distance between a cell and the predicted object area, uncertainties of the predicted track pose and extent, and the current visibility of the bounding box edges regarding the surrounding freespace evidence masses as proposed in [3]. That position association probability is determined for each possible track, as the source of an occupied cell may be ambiguous due to multiple objects close to each other or clutter measurements. An example of the resulting position association probabilities as determined in [3] is demonstrated in Fig. 4.2c.

Since the dynamic grid estimation also includes particle-based cell velocity estimates, an additional velocity-based association probability $\alpha_v(c, \tau)$ models the likelihood of the

cell velocity estimate ν_t^c based on the mean of the estimated track velocity v_τ and the corresponding covariance matrix $\Sigma_{\tau,v}$ of the track state s_τ, modeled by an unnormalized Gaussian distribution evaluated at ν_t^c, i.e.,

$$\alpha_v(c,\tau) \propto \mathcal{N}(\nu_t^c; v_\tau, \Sigma_{\tau,v})\,,\ \alpha_v \in [0,1]\,. \tag{4.7}$$

Those two association probabilities of the position and velocity are finally combined to a joint association probability

$$\alpha(c,\tau) \propto \alpha_x(c,\tau)\,\left(\lambda_v \alpha_v(c,\tau) + 1 - \lambda_v\right), \tag{4.8}$$

here with an additional weighting coefficient $\lambda_v \in [0,1]$ that regulates the impact of the velocity association probability α_v to the total association probability α. Since the cell velocity estimate or predicted track velocity may be wrong and differ from each other, a coefficient $\lambda_v > 0$ ensures that cells still remain a position-based association probability, for example for cells located in the middle of a predicted track.

As stated in (4.5), a cell is at most associated to one track, meaning that only the maximum association probability $\alpha(c,\tau_c^*)$ of a cell c to the best fitting track

$$\tau_c^* = \underset{\tau \in \widehat{T}_t}{\arg\max}\ \alpha(c,\tau) \tag{4.9}$$

is considered, i.e., a nearest neighbor association for each individual cell regarding $\alpha(c,\tau)$. Overall, a cell c is finally associated to the best fitting track τ_c^*,

$$f_a(c) = \begin{cases} \tau_c^*, & \text{if } \alpha(c,\tau_c^*) \geq \Gamma^c_{\min,\alpha} \\ \zeta_\emptyset, & \text{else} \end{cases} \quad \forall\, c \in \mathcal{G}_{D,t}\,, \tag{4.10}$$

if the association probability is greater or equal than a minimum threshold $\Gamma^c_{\min,\alpha} \in (0,1)$. This threshold can be set to a fixed value, or, alternatively, also be selected inversely proportional to the current measurement occupancy belief $\mathrm{bel}(O^c_{z,t})$. The latter then further combines the uncertainty of the association with the measurement uncertainty if a cell is actually occupied, so that a lower association probability requires a higher certainty that a cell is occupied in order to be extracted and vice versa. Thereby only dynamic occupied cells $c \in \mathcal{G}_{D,t}$ that are inside the gating area regarding the position association probability $\alpha_x(c,\tau)$ of at least one predicted track have to be considered here.

4.3.2 Particle Labeling Association

The cell-wise association based on the predicted high-level track state already enables to associate individual grid cells to a track without requiring a clustering or object hypothesis abstraction. However, a major drawback of that approach is that the occupancy likelihood approximation of cells extracted from the box representation is not accurate, even if the current visibility of the box edges is considered. This is critical in environments with dense traffic, since overlapping areas of multiple objects close to each other cause conflicts of multiple similar cell association probabilities that cannot be robustly solved that way.

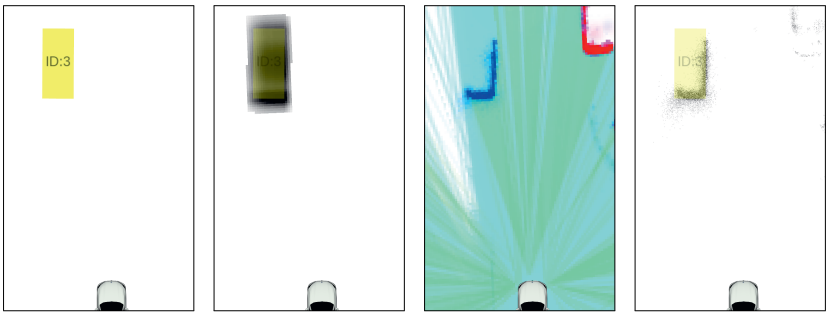


Fig. 4.3: Illustration of the different concepts of the occupancy likelihood approximation of a preceding object. (a) Predicted object track. (b) Approximated position association probability based on the predicted object state. (c) Estimation of the dynamic grid map. (d) Representation of the corresponding low-level particle population, which could be partly linked to that track.

Therefore, in order to further improve the association strategy, a more accurate occupancy likelihood representation of the predicted tracks is required. Independent of object instances, the prediction of the dynamic grid map with the low-level particle population, as discussed in the previous chapter, already presents the actual occupancy likelihood estimation of the dynamic environment. Hence, the main motivation of the following approach is to link parts of the particle population with each object track. In other words, the basic idea is that particles can be linked to an object by attaching an object label to each particle, where the predicted population of all particles linked to an object represents the prediction of its occupancy likelihood. That idea is also illustrated by an example in Fig. 4.3, where the occupancy likelihood of a preceding vehicle is characterized by the currently visible parts of that object, here similar to an L-shape, which is accurately represented by the particles, in contrast to the rough approximation that is derived from the predicted object state with the box-model. This enhanced association approach is denoted as the *particle labeling association*, which has also been published in [2] as part of this thesis.

Particle Label Extension as Object Identifier

The main idea of this association strategy is to extend the track attributes by linking a set of low-level particles $\chi \in \mathcal{X}_t$, as defined in (3.7), to each high-level object track $\tau \in \mathcal{T}_t$. For this purpose, an object label index

$$l_\chi \in \mathcal{T}_t \cup \{\zeta_\emptyset\} \tag{4.11}$$

is attached to each particle χ that defines the possible connection to a corresponding track. The set of particles linked to a track τ is thus defined by

$$\mathcal{X}_t^\tau = \left\{ \chi \in \mathcal{X}_t \mid l_\chi = \tau \right\}, \tag{4.12}$$

where each particle is at most linked to one track at time t, i.e.,

$$\mathcal{X}_t^{\tau_1} \cap \mathcal{X}_t^{\tau_2} = \emptyset \quad \forall \tau_1, \tau_2 \in \mathcal{T}_t,\ \tau_1 \neq \tau_2\,. \tag{4.13}$$

An example of the labeled particle population is shown in Fig. 4.2d with a color-coding of the particle labels, also demonstrating its free-form occupancy likelihood approximation.

This label is inherited in the particle resampling step to remain the linking to the corresponding track, while completely new drawn particles are initialized with $l_\chi = \zeta_\emptyset$, i.e., without any linked track. However, the particle label is not evaluated in the particle filter itself, i.e., no object relation is used for the prediction or weighting of the particles to retain a robust low-level tracking without modeling specific extended object assumptions for the point mass particles. Hence, the filtering behavior of the particle tracking remains unchanged. In the following, the population $\mathcal{X}_t$ represents updated particles after resampling, meaning that a particle velocity weighting and update using radar Doppler measurements and camera orientation estimates as presented in Section 3.5.2 are performed in advance. Measured velocities are thus implicitly considered by the updated particle population.

Cell Association Using Particle Label Evaluation

Based on the set $\mathcal{X}_t^\tau$ of particles linked to a track $\tau \in \mathcal{T}_t$, dynamic occupied grid cells $c \in \mathcal{G}_{D,t}$ are associated to that track τ by evaluating the particle label distribution in a cell c. The origin of an occupancy measurement in terms of the corresponding object instance is thereby indicated by the ratio

$$r_c^\tau = \frac{|\mathcal{X}_{t,c}^\tau|}{|\mathcal{X}_{t,c}|} \in [0,1] \tag{4.14}$$

of the number of particles $|\mathcal{X}_{t,c}^\tau|$ linked to a track τ compared to the total number of particles $|\mathcal{X}_{t,c}|$ in that cell c. This ratio r_c^τ can also be interpreted as an association probability, equivalent to the notation $\alpha(c,\tau)$ as introduced in (4.8). Hence, in the following, $\alpha(c,\tau) = r_c^\tau$ is used, with the result that also the remaining steps of the association can be applied as defined before with the association probability $\alpha(c,\tau)$. This means that only the best fitting track τ_c^* of that cell c is considered, as determined in (4.9), which corresponds to the track with the highest particle labeling ratio r_c^τ in that case. The set of associated cells $\mathcal{C}_{\tau,t}$ then finally results from the function $f_a(c)$ as defined in (4.10), i.e., a cell c is only associated to its best fitting track τ_c^* if the corresponding particle labeling ratio fulfills a minimum threshold.

Particle Label Update / Reselection of Linked Particles

After the measurement update of the tracks, the linking between the particles and tracks has to be updated. On the one hand, particles that are too far away from the bounding box of the corresponding track should not retain linked to that track. On the other hand, particles without a corresponding track, i.e., $l_\chi = \zeta_\emptyset$, but that are inside the bounding box of a track and have a similar velocity, may be linked to that track. This concept basically corresponds to a gating, i.e., defining which area of a track is generally valid and which not.

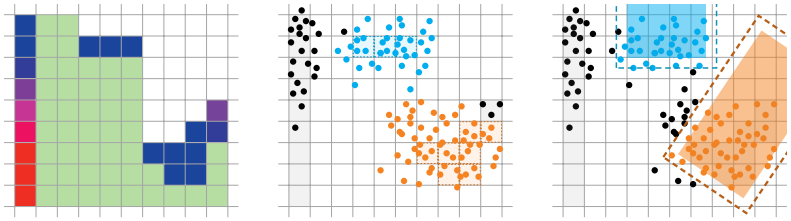


Fig. 4.4: Illustration of the particle labeling association concept. The scenario shows a guardrail on the left side, partly wrongly classified as dynamic, and two existing tracks with one turning right. (a) Classified measurement grid. (b) Labeled particle population and associated cells ($\zeta_\emptyset$: black, τ_1: cyan, τ_2: orange). (c) Update of particle labels (filled rectangles: bounding boxes of updated tracks; dashed lines: valid gating areas of tracks). Figure initially published in [2], © 2018 IEEE.

The first case, i.e., unlinking all particles that do not match with the linked track anymore, is described by

$$l_\chi \leftarrow \zeta_\emptyset \quad \forall\, \chi \in \mathcal{A}_1^\tau\,,\ \tau \in \mathcal{T}_t\,, \tag{4.15}$$

where the set of all particles that have to be unlinked from the track τ is determined by

$$\mathcal{A}_1^\tau = \left\{\chi \in \mathcal{X}_t^\tau \mid x_\chi \notin b(s_\tau, \Delta_b)\right\}. \tag{4.16}$$

Thereby the position x_χ of a particle χ is compared with the bounding box $b(s_\tau, \Delta_b)$ of the track τ which depends on the track pose and length/width that are part of the object state s_τ, where the valid area is enlarged by a distance parameter Δ_b. This unlinking strategy implicitly includes resetting labels of deleted tracks.

The second case, i.e., linking new particles to a track, including newly extracted tracks $\mathcal{T}_t^{\text{new}}$, is described by

$$l_\chi \leftarrow \tau \quad \forall\, \chi \in \mathcal{A}_2^\tau\,,\ \tau \in \mathcal{T}_t\,, \tag{4.17}$$

with the set of corresponding particles

$$\mathcal{A}_2^\tau = \left\{\chi \in \mathcal{X}_t^{\zeta_\emptyset} \,|\, x_\chi \in b(s_\tau) \wedge v_\chi \in v_g(s_\tau) \wedge \left(\nexists \tilde{\tau} \in \mathcal{T}_t \backslash \{\tau\} : x_\chi \in b(s_{\tilde{\tau}})\right)\right\}. \tag{4.18}$$

A particle is only newly linked to a track if it is directly inside the not enlarged bounding box $b(s_\tau)$ of exactly one track. In addition, the particle velocity v_χ is compared to a defined valid velocity gating interval $v_g(s_\tau)$ of the track. In overlapping areas of multiple tracks, no new particles are linked to any track. In such a case, particles that remained their label, without being unlinked by (4.15), are used to solve these ambiguities.

The different steps of the particle labeling concept are illustratively summarized in Fig. 4.4. Dynamic occupied cells of the classified measurement grid, shown in Fig. 4.4a, are associated to the two existing tracks by evaluating the particle label distributions, cf. Fig. 4.4b. After the measurement update of the tracks, the particle labels are updated, while particles outside the gating area are unlinked from that track and particles without a label that are inside the bounding box of a track are added to that track, see Fig. 4.4c.

Altogether, the particle labeling approach enables a robust association with an accurate free-form occupancy likelihood representation. Thereby the linkage between the filtered particle population and the tracked objects is retained over time, which further implicitly helps to solve association ambiguities in overlapping areas of object tracks, since particles keep their label from the past in such a case. This ensures, for instance, that a cell remains associated to a track if it was clearly associated to that track before, even if another object then approaches that area and moves closely to it.

4.3.3 Additional Clustering with Verification

In the following, the proposed cell-wise particle labeling association is further extended by a subsequent clustering step, which has also been published as part of [2]. This concept can also be equivalently applied to the simplified association based on the predicted object state. The main idea of this additional step is to extract multiple sub-clusters of a track that are then checked for plausibility, while only the best sub-clusters are finally selected for the association and the object measurement update. The motivation is thereby twofold:

- Dynamic cells that have not been associated to any track by the particle population but are directly adjacent to associated cells may be added to that cluster. This may occur for cells that are outside the gating area of the track by an inaccurate bounding box shape $b(s_\tau)$; for example, a preceding vehicle that turns quickly where a wrongly estimated box orientation does not fully cover the long side of the detected L-shape.

- The association of individual cells close to a track, i.e., inside its gating area, may be wrong. Possible scenarios, for instance, are road boundaries next to a track, with some parts wrongly classified as dynamic, or parts of newly appearing objects that have not been extracted but are also directly next to an existing track.

Hence, the aim of the additional clustering and verification strategy is to reduce both false negatives and false positives of the association. This concept is divided into three steps, which are discussed in the following.

Adaptive Dynamic Evidence Threshold Selection

In the first step, an adaptive threshold $\gamma(\tau)$ of the dynamic evidence mass $m(D_{\tilde{z},t})$ is determined based on the cells of $c \in \mathcal{C}_{\tau,t}$ that are associated to a track τ by the particle labels. Cells below that threshold are considered as insignificant or too uncertain regarding the dynamic evidence mass of that cluster. Therefore, those cells are discarded to improve the robustness of the association, which reduces the set $\mathcal{C}_{\tau,t}$ to the subset

$$\mathcal{C}'_{\tau,t} = \left\{ c \in \mathcal{C}_{\tau,t} \mid m(D^c_{\tilde{z},t}) \geq \gamma(\tau) \right\}. \tag{4.19}$$

The adaptive threshold is thereby selected as

$$\gamma(\tau) = \max\left(\Gamma^D_{\min},\, \Gamma^{D,a}_{\min} \frac{1}{|\mathcal{C}_{\tau,t}|} \sum_{c \in \mathcal{C}_{\tau,t}} m(D^c_{\tilde{z},t}) \right), \tag{4.20}$$

which depends on the mean dynamic evidence mass of the set $\mathcal{C}_{\tau,t}$ and a coefficient $\Gamma^{D,a}_{\min} \in [0,1]$, where the minimum value of $\gamma(\tau)$ is limited by the fixed dynamic evidence mass threshold $\Gamma^{D}_{\min}$ as defined before in (4.2).

In contrast to selecting a higher threshold of $\Gamma^{D}_{\min}$ itself, where cells below that value would not be considered in the initial cell-wise particle labeling association, this adaptive threshold selection compares the dynamic evidence mass of a cell with the mean of all cells associated to that track. Hence, this enables that cells with a lower dynamic evidence remain associated if the other cells of that track have as well a similarly low dynamic evidence mass, e.g., for objects in larger distances, but also that those more uncertain cells are ignored if the mean dynamic evidence mass of the cells of that track is significantly higher. Overall, this adaptive concept reduces false positives of the association and enables a more sensitive clustering of neighboring dynamic cells as described in the following.

Clustering of Associated Cells and Neighbors

Based on the set $\mathcal{C}'_{\tau,t}$ of cells that fulfill the minimum dynamic evidence threshold $\gamma(\tau)$, sub-clusters $\mathcal{C}^k_{\tau,t}$ of each track are formed, where each sub-cluster represents a set of directly connected cells. Thereby each track τ has a total of $|\mathcal{K}_\tau|$ sub-clusters, with $k \in \mathcal{K}_\tau$ denoting the respective index of the sub-cluster. The basic idea of the clustering is that either the full sub-cluster $\mathcal{C}^k_{\tau,t}$ is actually part of the track τ, i.e., all cells $c \in \mathcal{C}^k_{\tau,t}$ of the sub-cluster should be associated to the track, or that the full sub-cluster represents a false hypothesis of the association, i.e., none of the cells of $\mathcal{C}^k_{\tau,t}$ should remain associated to that track. In conclusion, a track can consist of several sub-clusters that are correctly associated, and potentially some rejected sub-clusters that are not necessarily originated from that track.

In the following, in addition to the cells of the set $\mathcal{C}'_{\tau,t}$, also adjacent unassociated dynamic occupied cells of the set

$$\mathcal{C}^{\zeta_\emptyset}_{\tau,t} = \left\{ c \in \mathcal{G}_{D,t} \mid f_a(c) = \zeta_\emptyset \ \wedge \ m(D^c_{\tilde{z},t}) \geq \gamma(\tau) \ \wedge \ (r^\tau_c + r^{\zeta_\emptyset}_c) \geq (1 - \Gamma^{r,\text{other}}_{\max}) \right\} \tag{4.21}$$

can be added to the sub-clusters of the track τ. Thereby only cells with a dynamic mass of at least the adaptive threshold $\gamma(\tau)$ are considered, which further have no significant indication of a different track origin, i.e., unassociated cells with an ambiguous particle label distribution are excluded in $\mathcal{C}^{\zeta_\emptyset}_{\tau,t}$. Therefore, the particle label ratio, as introduced in (4.14), for the case of no object label or the label of the track itself ($l_\chi \in \{\zeta_\emptyset, \tau\}$), i.e.,

$$r^\tau_c + r^{\zeta_\emptyset}_c = \frac{|\mathcal{X}^\tau_{t,c}| + |\mathcal{X}^{\zeta_\emptyset}_{t,c}|}{|\mathcal{X}_{t,c}|} = 1 - \sum_{\tilde{\tau} \in \mathcal{T}_t \setminus \{\tau\}} r^{\tilde{\tau}}_c \tag{4.22}$$

has to be above a selected threshold $(1 - \Gamma^{r,\text{other}}_{\max})$, with $\Gamma^{r,\text{other}}_{\max} \ll 1$ representing the maximum allowed ratio of particles linked to a different track. That consideration of the additional neighboring cells improves the association for scenarios where some adjacent cells are outside the gating area of the track, as motivated before.

The different sub-clusters $\mathcal{C}^k_{\tau,t}$ are then formed by a single-linkage hierarchical clustering with a distance threshold $\Gamma^d_{\max}$ between the cell centers x_c, i.e., two cells c_1 and c_2 are associated to the same cluster if

$$\|x_{c_1} - x_{c_2}\| \leq \Gamma^d_{\max}, \ c_1 \in \mathcal{C}'_{\tau,t}, \ c_2 \in \mathcal{C}'_{\tau,t} \cup \mathcal{C}^{\zeta_\emptyset}_{\tau,t}. \tag{4.23}$$

This constraint can be further restricted if camera segmentation information is available, i.e., if cells are linked to camera object instances, denoted by κ_{id}, as introduced in (2.19). Therefore, in addition to (4.23), cells can only be added to a cluster if also the camera object instance of the cell does not contradict with that of the cluster. For instance, the two cells c_1 and c_2 are only combined if

$$(\kappa_{\mathrm{id}}^{c_1} = \kappa_{\mathrm{id}}^{c_2}) \ \vee \ (\kappa_{\mathrm{id}}^{c_1} = \zeta_{\emptyset}) \ \vee \ (\kappa_{\mathrm{id}}^{c_2} = \zeta_{\emptyset})\,, \tag{4.24}$$

which is fulfilled when at least one of the two cells has no camera information in terms of no associated object instance, denoted by $\zeta_{\emptyset}$, or if both cells are linked to the same camera object. Overall, the camera segmentation information is thus used to achieve a finer division of the sub-clusters, but not to combine sub-clusters with a larger position distance in-between. That way it is ensured that the association does not fully rely on the camera segmentation including the transformation from the image space to the occupancy grid model.

In conclusion, each sub-cluster $\mathcal{C}_{\tau,t}^{k}$ contains at least one cell of the set $\mathcal{C}'_{\tau,t}$, whereas the union of all sub-clusters of a track includes all cells of the set $\mathcal{C}'_{\tau,t}$, and a cell is at most associated to one cluster, i.e.,

$$\mathcal{C}_{\tau,t}^{k} \subseteq \mathcal{C}'_{\tau,t} \cup \mathcal{C}_{\tau,t}^{\zeta_{\emptyset}}, \quad \mathcal{C}_{\tau,t}^{k} \cap \mathcal{C}'_{\tau,t} \neq \emptyset, \quad \mathcal{C}'_{\tau,t} \subseteq \bigcup \mathcal{C}_{\tau,t}^{k} \quad \forall k \in \mathcal{K}_{\tau}\,. \tag{4.25}$$

Cluster Score Verification

So far, all sub-clusters of a track only represent candidates for the association, but they have to be checked for plausibility before they are finally associated. Therefore, each sub-cluster is compared with the predicted object state $\hat{s}_{\tau}$ of the corresponding track to form a score. Finally, only the sub-clusters with the highest scores remain associated to that track.

In the following, a simple approach is presented that evaluates the conformity of the position and the geometry of the sub-cluster and the predicted track. First, a position score $\lambda_{\tau,k}^{x}$ is determined, which results from the ratio of the number of cells that are inside the predicted bounding box, i.e.,

$$\lambda_{\tau,k}^{x} = \frac{\left|\left\{c \in \mathcal{C}_{\tau,t}^{k} \mid x_c \in b(\hat{s}_{\tau})\right\}\right|}{\left|\mathcal{C}_{\tau,t}^{k}\right|}\,. \tag{4.26}$$

Second, the length $b_l(\mathcal{C}_{\tau,t}^{k})$ and width $b_w(\mathcal{C}_{\tau,t}^{k})$ of each cluster are calculated, i.e., an oriented minimum bounding box is extracted using the predicted track orientation. This extracted bounding box geometry is compared

$$\Delta b_{\tau,k}^{l} = \frac{|b_l(\mathcal{C}_{\tau,t}^{k}) - b_l(\hat{s}_{\tau})|}{b_l(\hat{s}_{\tau})}, \quad \Delta b_{\tau,k}^{w} = \frac{|b_w(\mathcal{C}_{\tau,t}^{k}) - b_w(\hat{s}_{\tau})|}{b_w(\hat{s}_{\tau})} \tag{4.27}$$

with the length $b_l(\hat{s}_{\tau})$ and width $b_w(\hat{s}_{\tau})$ of the bounding box $b(\hat{s}_{\tau})$ of the predicted track. This relative geometry conformity is mapped to a geometry score factor

$$\lambda_{\tau,k}^{b} = \exp\left(-\eta_b \min(\Delta b_{\tau,k}^{l}, \Delta b_{\tau,k}^{w})\right) \in [0,1]\,, \tag{4.28}$$

with a scaling parameter $\eta_b \in \mathbb{R}^+$. Only the more fitting side of the box is considered in $\lambda^b_{\tau,k}$, i.e., it is sufficient if only the length or the width of the track is represented by a cluster. Both scores of the sub-cluster are finally combined by

$$\lambda_{\tau,k} = \lambda^x_{\tau,k}\, \lambda^b_{\tau,k} \in [0,1]\,. \tag{4.29}$$

This concept can be further extended by analyzing additional aspects like the particle-based cell velocities, adjacent static occupied evidence $m(S_{\tilde{z},t})$, the included freespace evidence $m(F_{z,t})$, or the camera segmentation information κ_{id} as also evaluated in (4.24). Moreover, it can be further evaluated if the combination of sub-clusters results in a higher score, in which case they should be considered jointly.

To compare those cluster score relatively to each other, all cluster scores of a track are normalized with respect to the best cluster score of that track, i.e.,

$$\overline{\lambda}_{\tau,k} = \frac{\lambda_{\tau,k}}{\max\limits_{k' \in \mathcal{K}_\tau} \lambda_{\tau,k'}}\,. \tag{4.30}$$

Altogether, sub-clusters are finally associated to the corresponding track

$$f'_a(c) = \begin{cases} \tau\,, \text{if } c \in \mathcal{C}^k_{\tau,t} \wedge \overline{\lambda}_{\tau,k} \geq \Gamma^{\lambda}_{\min} & \forall c \in \mathcal{G}_{D,t}\,, \\ \zeta_\emptyset\,, \text{else} \end{cases} \tag{4.31}$$

if the normalized score of that sub-cluster is above a threshold $\Gamma^{\lambda}_{\min} \in [0,1]$. Due to the normalization, the best fitting sub-cluster of a track is always associated to that track, whereas additional sub-clusters are only added if they have a similar score.

All in all, the proposed cell-level association approach enables an association of individual grid cells by evaluating the predicted tracks that have been extracted before, which is essential for scenarios with objects moving closely to each other. Hence, occupied cells area associated before they are abstracted to the object-level representation, which avoids information loss. Based on this cell-individual association, the subsequent clustering and verification concept further improves the robustness of the association by analyzing those associated cells and thereby detecting and rejecting wrongly associated cells of neighboring dynamic areas that actually do not belong to that track.

4.4 Extraction of Newly Occurring Object Tracks

The previous section has demonstrated the association between dynamic grid cells and predicted object tracks that have been extracted before. Thereby information of those existing tracks is used to directly perform the association on the grid cell-level before clustering those cells and without requiring an object-level abstraction. But in addition to that association, newly occurring moving objects have to be detected and extracted, which is presented in this section.

Therefore, the unassociated dynamic cells of the remaining grid set

$$\mathcal{G}^{\zeta_\emptyset}_{D,t} = \mathcal{G}_{D,t} \setminus \bigcup_{\tau \in \widehat{\mathcal{T}}_t} \bigcup_{k \in \mathcal{K}_\tau} \mathcal{C}^k_{\tau,t} \tag{4.32}$$

have to be analyzed for potential new object tracks. Here all potential sub-clusters $\mathcal{C}_{\tau,t}^{k}$ of the existing tracks $\tau \in \widehat{\mathcal{T}}_t$ are excluded from $\mathcal{G}_{D,t}$, i.e., new object tracks are only extracted when they are clearly separated from existing tracks.

The proposed object extraction approach is divided into two steps, as also published in [3] in the context of this thesis. First, a density-based clustering of dynamic occupied cells is performed to extract the initial hypotheses of newly occurring moving objects. Second, those clusters are extended by a subsequent region growing of neighboring occupied cells, which are then finally evaluated by a velocity variance analysis to decide whether a cluster forms a valid new track or not. Those two steps are discussed in the following.

4.4.1 Density-Based Clustering of Dynamic Occupied Cells

Potential object instances of the unassociated dynamic occupied cells of the set $\mathcal{G}_{D,t}^{\zeta_\emptyset}$ are formed by a density-based clustering. In this work, the DBSCAN algorithm [40] is used, with the ε-neighborhood

$$N_\varepsilon^c = \left\{ c' \in \mathcal{G}_{D,t}^{\zeta_\emptyset} \;\middle|\; \left(\|x_c - x_{c'}\| \leq \Gamma_{\varepsilon,x}^{c,c'} \right) \wedge \left(\|\nu_t^c - \nu_t^{c'}\| \leq \Gamma_{\varepsilon,v}^{c,c'} \right) \wedge \left(\sum_{\tilde{c} \in \mathcal{D}(c,c')} m(F_t^{\tilde{c}}) \leq \Gamma_{\varepsilon,F}^{c,c'} \right) \right\} \tag{4.33}$$

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: a maximum position distance $\Gamma_{\varepsilon,x}^{c,c'}$ of the 2-D cell centers $x_c \in \mathbb{R}^2$, a maximum cell velocity difference $\Gamma_{\varepsilon,v}^{c,c'}$, and a maximum value $\Gamma_{\varepsilon,F}^{c,c'}$ of the sum of the in-between freespace evidence mass $m(F_{z,t})$. Thereby $\mathcal{D}(c,c')$ defines the set of in-between cells of c and c'. A simple and conservative choice for it is the rectangular area between two cells, e.g., using a summed area table [27]; alternatively, cells of the connected line can be extracted by Bresenham's line algorithm [21]. 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. Thereby also the position distance threshold $\Gamma_{\varepsilon,x}^{c,c'}$ can be enlarged if measurements occur on different parts of one object.

Moreover, if additional semantic segmentation information of the camera sensors is available, similar to (4.24), the ε-neighborhood N_ε^c of the cell c can be further improved. The basic idea thereby is that if a cell c' has the same camera object instance as the evaluated cell c ($\kappa_{\mathrm{id}}^c = \kappa_{\mathrm{id}}^{c'}$), it is more likely that that cell c' is a neighbor of the cell c. Equivalently, if those two cells belong to different camera object instances ($\kappa_{\mathrm{id}}^c \neq \kappa_{\mathrm{id}}^{c'}$), it should be more unlikely that they belong to the same ε-neighborhood. However, those two cases should only be evaluated if both cells actually contain a linked camera object instance, i.e., to improve the overall robustness, it is assumed that the set of all cells linked to a camera object generally represents only a subset of all occupied cells of the actual object. This means that the ε-neighborhood N_ε^c remains unchanged and correspond to the case of no camera information if $\kappa_{\mathrm{id}}^c = \zeta_\emptyset$ or $\kappa_{\mathrm{id}}^{c'} = \zeta_\emptyset$, which can be summarized by $\zeta_\emptyset \in \{\kappa_{\mathrm{id}}^c, \kappa_{\mathrm{id}}^{c'}\}$. Overall, also with camera information the generic ε-neighborhood definition of (4.33) is used, only the corresponding thresholds are adapted depending on the camera segmentation

matching, i.e., the maximum position difference threshold is modeled by

$$\Gamma^{c,c'}_{\varepsilon,x}(\kappa^{c}_{\mathrm{id}} \neq \kappa^{c'}_{\mathrm{id}}) \quad \leq \quad \Gamma^{c,c'}_{\varepsilon,x}(\zeta_{\emptyset} \in \{\kappa^{c}_{\mathrm{id}}, \kappa^{c'}_{\mathrm{id}}\}) \quad \leq \quad \Gamma^{c,c'}_{\varepsilon,x}(\kappa^{c}_{\mathrm{id}} = \kappa^{c'}_{\mathrm{id}})\,, \tag{4.34}$$

which also applies to the respective thresholds of the velocity $\Gamma^{c,c'}_{\varepsilon,v}$ and the freespace $\Gamma^{c,c'}_{\varepsilon,F}$.

Altogether, the density-based clustering results in multiple cell clusters $\mathcal{C}_i \subseteq \mathcal{G}^{\zeta_{\emptyset}}_{D,t}$, where each index i describes a potential new object track instance. Those cell clusters are further extended and evaluated in the following.

4.4.2 Additional Region Growing with Velocity Variance Analysis

Based on the density-based clusters $\mathcal{C}_i$ of dynamic occupied cells with similar velocities, the local neighborhood of each cluster is analyzed in terms of the adjacent occupied cells. The main motivation is to determine false positives of the object hypotheses by evaluating the velocity variance of that extended cluster, since the cell velocities and thus the dynamic masses of the evidential dynamic occupancy grid may be partly incorrect, as demonstrated in Fig. 4.5. Moreover, this strategy is also helpful for correctly detected objects that have just started to move and which have not been fully converged toward dynamic occupancy, e.g., the long side of a vehicle, where then all connected occupied cells of that track can thus be extracted.

Region Growing of Adjacent Occupied Cells

In the following, each density-based cluster $\mathcal{C}_i$ is expanded to a cluster $\mathcal{C}^{+}_i \supseteq \mathcal{C}_i$ by iteratively adding all neighboring unassociated and unclustered occupied cells in a seeded region growing [6] with $\mathcal{C}_i$ as the starting seed points. Hence, not only dynamic occupied cells of $\mathcal{G}^{\zeta_{\emptyset}}_{D,t}$ can be added in this step, but also all occupied cells of $\mathcal{G}_{O,t}$ as introduced in (4.1), i.e., cells where the current measurement occupancy belief $\mathrm{bel}(O^{c}_{z,t})$ is greater or equal than a threshold $\Gamma^{O}_{\mathrm{min}}$. For the connectivity criterion of the region growing, the directly adjacent cells in terms of the 8-connected neighborhood is used. To overcome possible small gaps in the current measurement data, e.g., due to temporal occlusion or missing measurement data, that connectivity analysis can be further enhanced by also considering the accumulated dynamic grid map $\mathcal{M}_t$, meaning that cells of the current measurement data can also be connected if there is a connection of occupied cells in the filtered map.

An example of the region growing is demonstrated in Fig. 4.5b–4.5c. In that example, the maximum number of iterations is limited in terms of an efficient implementation, since the main motivation is to reduce false positives. Velocity differences are ignored in this connectivity-based clustering step; instead, the velocity variance of the new cluster is analyzed as discussed in the following.

Velocity Variance Analysis

If any cell has been added to the initial cluster ($\mathcal{C}^{+}_i \neq \mathcal{C}_i$), the velocity variance $\sigma^2_{v_i}$ of the extended cluster $\mathcal{C}^{+}_i$ is evaluated. First, the weighted mean velocity $\overline{\nu}_i$ of the cluster is calculated, but only the cells of the initial dynamic cluster $\mathcal{C}_i$ are considered and the particle-based cell velocity ν^{c}_{t} is only weighted by the respective dynamic evidence

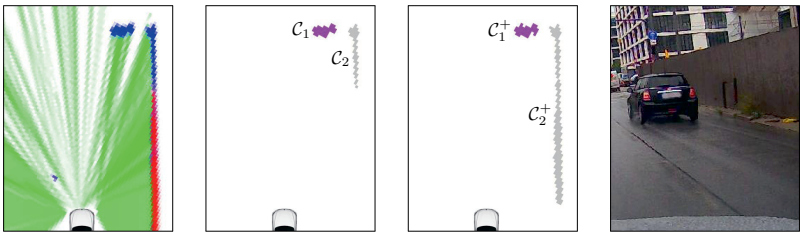


Fig. 4.5: Object extraction for a scenario with an incorrect dynamic estimation at a wall. Both clusters $\mathcal{C}_1$ and $\mathcal{C}_2$ represent potential objects moving north. But only $\mathcal{C}_1$ describes a real vehicle, $\mathcal{C}_2$ is a false positive. (a) Augmented measurement grid. (b) Density-based clustering of dynamic occupied cells. (c) Additional region growing step of neighboring occupied cells (the maximum number of iterations was limited to 40 here). (d) Camera image.

mass $m(D^c_{\tilde{z},t})$, since the evaluated variance is related to the initial dynamic cluster of a moving object. Hence, the weighted mean velocity of each cluster $\mathcal{C}_i$ is determined by

$$\overline{\nu}_i = \left(\sum_{c \in \mathcal{C}_i} m(D^c_{\tilde{z},t}) \right)^{-1} \sum_{c \in \mathcal{C}_i} m(D^c_{\tilde{z},t})\, \nu^c_t \,. \tag{4.35}$$

Based on that mean velocity of the initial cluster $\mathcal{C}_i$, the velocity variance is evaluated for all cells of the extended cluster $\mathcal{C}_i^+$. Since the particle-based cell velocity ν^c_t only represents the velocity hypothesis of dynamic occupancy $m(D^c_{\tilde{z},t})$, as introduced in (3.11), the static occupancy evidence mass $m(S^c_{\tilde{z},t})$ with the hypothesis of a zero-velocity has to be considered separately in the variance calculation, which results in

$$\sigma^2_{v_i} = \left(\sum_{c \in \mathcal{C}_i^+} \left(m(D^c_{\tilde{z},t}) + m(S^c_{\tilde{z},t}) \right) \right)^{-1} \sum_{c \in \mathcal{C}_i^+} \left(m(D^c_{\tilde{z},t}) \|\nu^c_t - \overline{\nu}_i\|^2 + m(S^c_{\tilde{z},t}) \|0 - \overline{\nu}_i\|^2 \right) . \tag{4.36}$$

Hence, rather than evaluating the mean of a cell as a combination of the static hypothesis and the dynamic hypothesis, both are compared separately with the cluster mean $\overline{\nu}_i$ for the variance evaluation here. For instance, with a particle-based cell velocity of the dynamic evidence that is twice the cluster mean ($\nu^c_t = 2\overline{\nu}_i$), but with $m(D^c_{\tilde{z},t}) = m(S^c_{\tilde{z},t}) = 0.5$, the combined velocity hypothesis of this cell would correspond to the cluster mean ($0.5\nu^c_t + 0.5 \cdot 0$), but (4.36) ensures that the variance of that cell is still high, since neither the static hypothesis with velocity zero nor the dynamic hypothesis with velocity ν^c_t really corresponds to the mean cluster velocity $\overline{\nu}_i$ in that case.

Overall, valid clusters that form new tracks have to be below a maximum allowed velocity variance, i.e., $\sigma^2_{v_i} \leq \sigma^2_{v,\max}$. In conclusion, 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, since the association is handled differently as described before.

4.5 Results and Summary

This chapter has presented a novel concept for the object extraction and association based on the characteristics of the dynamic occupancy grid representation of the previous chapter. The fundamental idea of the proposed approach is summarized in an example in Fig. 4.6. The fused measurement grid of the current measurement data, shown in Fig. 4.6a, is classified by the filtered dynamic grid map, resulting in the augmented measurement grid representation with a differentiation of static and dynamic occupancy, cf. Fig. 4.6b. This directly simplifies the moving object detection, since only occupied cells classified as dynamic have to be considered, which represents a generic track-before-detect strategy without requiring specific object assumptions. The approach is further enhanced by evaluating the corresponding low-level particle population of the dynamic grid estimation, shown in Fig. 4.6c, as well as the predicted object tracks that have been extracted before, depicted in Fig. 4.6d. Hence, knowledge of the tracked objects is considered before abstracting object hypotheses and thus losing information of the current measurement data, i.e., dynamic grid cells are directly associated on the grid cell-level rather than on the object-level. The resulting association of the occupied cells is depicted in Fig. 4.6e–4.6f, demonstrating that also measurement data of largely occluded objects that are barely visible can thus be robustly associated, which is here the case for the left-most vehicle (ID 8) that is mostly occluded by another vehicle (ID 4). In order to achieve an accurate free-form occupancy likelihood representation of the predicted object tracks, particles of the low-level particle population are linked to the object tracks as shown in Fig. 4.6g, which, overall, results in a robust and generic association strategy.

As motivated before in Fig. 4.2, it is also essential to consider the predicted object tracks for solving association ambiguities when an object is split into separate areas of occupancy measurements, e.g., due to a restricted sensor field of view, which cannot be directly handled by a clustering. Two approaches have been proposed for the association on the grid cell-level: a simplified strategy using the predicted object bounding box model and, as also depicted in Fig. 4.6g, a detailed occupancy likelihood approximation using the low-level particle population. Both concepts enable a cell-individual association that helps to solve association ambiguities, which has also been demonstrated in Fig. 4.2c–4.2e. In scenarios with multiple tracks moving closely to each other, however, this box model occupancy likelihood approximation is error-prone to inaccuracies of the predicted track state. Fig. 4.7 shows such a scenario with two vehicles closely moving in parallel. Their correct estimation is illustrated in Fig. 4.7b. To demonstrate the error-proneness in case of an incorrect state estimation, an orientation offset of 15° is added to the predicted track on the right side (ID 1). Hence, the front left corner of that track overlaps with the rear right corner of the other vehicle (ID 2), as shown in Fig. 4.7c. Ambiguous cells in the overlapping area can thus be associated to the wrong track, leading to wrong measurement updates, which then further reinforce the incorrect state estimation and thus the association of the next time instance, cf. Fig. 4.7g. In contrast, the particle population linked to a track represents a detailed occupancy likelihood approximation of that track, see Fig. 4.7d. This approach is robust against overlapping areas of the box models, cf. Fig. 4.7e and Fig. 4.7h, since the particles remain their label from the previous time instance in such a case, i.e., the label is only updated when no ambiguities occur.

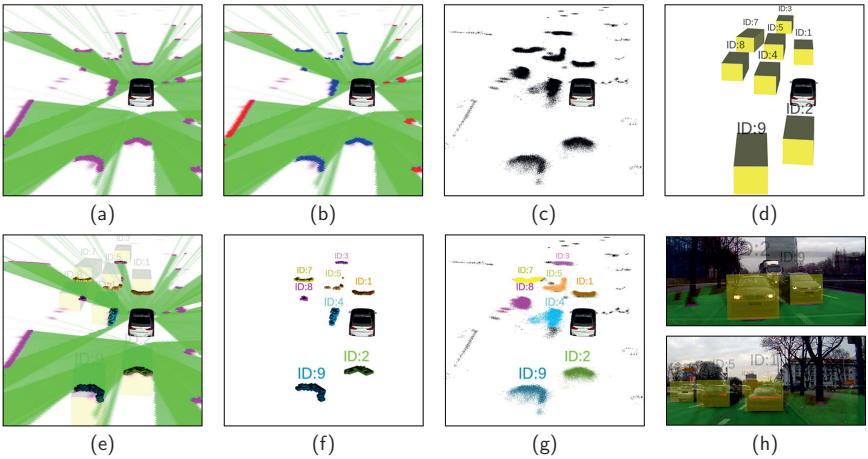


Fig. 4.6: Proposed grid-based object detection and association in an urban environment with dense traffic and partial occlusion. (a) Fused measurement grid. (b) Augmented measurement grid. (c) Low-level particle population. (d) Predicted object tracks. (e/f) Resulting associated occupied cells $\mathcal{C}_\tau$ of each track (with/without grid overlay). (g) Proposed labeled particle population. (h) Camera images. Figure initially published in [2], © 2018 IEEE.

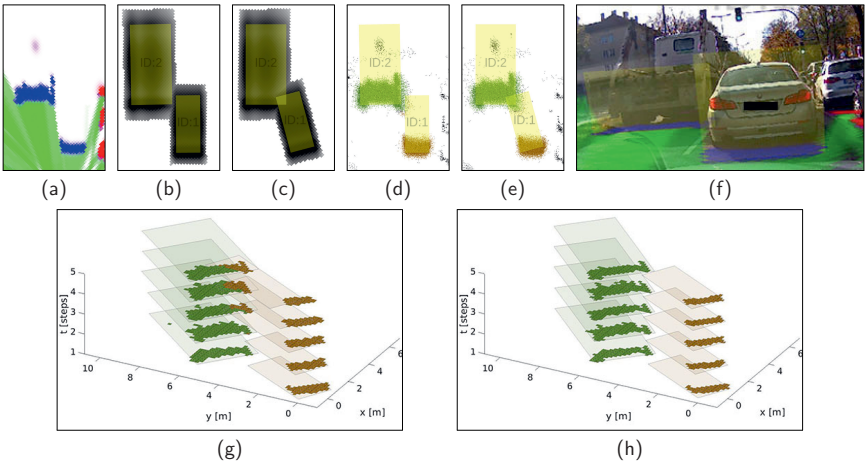


Fig. 4.7: Scenario with two closely moving vehicles with a modified orientation estimation. Orientation of track 1 is modified in (c), (e), (g), (h) by an offset of $15°$. (a) Augmented measurement grid. (b)+(c) Occupancy association probabilities. (d)+(e) Labeled particle population. (f) Camera image. (g)+(h) Associated cells over time (z-axis) and predicted track box model with $15°$ orientation offset of track 1, (g) corresponds to (c); (h) to (e). Figure initially published in [2], © 2018 IEEE.

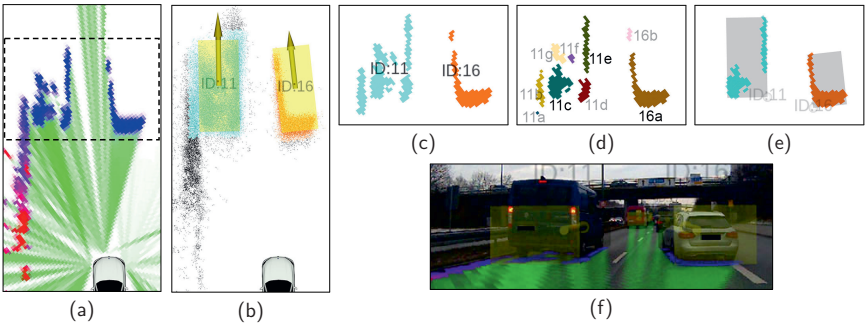


Fig. 4.8: Scenario with a vehicle moving closely to a road boundary with a curb that is wrongly classified as dynamic. (a) Augmented measurement grid (dashed area shows section of (c)-(e)). (b) Labeled particle population and predicted tracks. (c) Particle labeling association. (d) Additional clustering process (color of a sub-cluster is not related to the corresponding track here). (e) Finally associated clusters and resulting measurement minimum bounding boxes. (f) Camera image. Figure initially published in [2], © 2018 IEEE.

In order to further enhance the particle labeling association approach, a subsequent sub-clustering and verification concept has been proposed. This is demonstrated in Fig. 4.8, where the cell-individual particle labeling association results in some wrongly associated cells. In that scenario, cells of a nearby curb, which are incorrectly classified as dynamic, but are not extracted as a separate track, are associated to the closely moving track. The additional clustering process, however, distinguishes those cells of the curb as a separate sub-cluster, as demonstrated in Fig. 4.8d. Hence, by evaluating the individual scores of each cluster, that outlying cluster results in a low cluster score and is thus excluded.

Note that the selection of valid clusters is chosen conservatively here, i.e., the score of a cluster has to be similar to the best cluster of that track. As presented in the next chapter, an object track is updated by an extracted minimum bounding box of the associated cells. Hence, false negatives, i.e., occupied cells of that track which have been omitted in the association, only lead to a smaller, more conservative minimum bounding box that still correctly updates the track. False positives, in contrast, result in an overestimation of the bounding box and thus also of that filtered track, which is more critical. As discussed before, this cluster score verification concept can also be extended by evaluating multiple hypotheses and different combinations of the individual sub-clusters of a track, which, however, also increases the computational effort.

Altogether, this chapter has presented a robust strategy for the grid-based object extraction and association, resulting in a set of occupied grid cells of the current measurement grid that are assigned to an object track. In the following, it is discussed how the corresponding object state estimation is updated based on those associated occupied grid cells, i.e., the temporal filtering on the object-level, which forms the final step of the object tracking. Further results of the overall grid-based object tracking approach are demonstrated afterwards in Chapter 6.

5 Object State Estimation

The previous chapters have presented how measurement data are processed on the grid cell-level using the discretized grid representation with the uniform evidential occupancy and freespace hypotheses. This includes the sensor data fusion and the dynamic grid estimation by the individual grid cells, but also the object extraction and association that both result in a set of associated occupied grid cells for each object track instance. This chapter focuses on the subsequent high-level object state estimation, i.e., the temporal filtering on the object-level. Thereby the associated occupied grid cells are abstracted to a bounding box model, but also the freespace evidence as well as radar velocity measurements and camera classification information are evaluated to further improve the object state estimation. A UKF-based estimation of the dynamic motion state is presented, where the object pose is robustly updated by selecting the most visible reference point in terms of the surrounding freespace and a box orientation optimization in terms of minimizing the included freespace. The motion state estimation is further enhanced by utilizing radar Doppler velocity measurements with a robust update directly in the Doppler measurement space. Moreover, a new combined approach for the object shape estimation and classification is proposed using a histogram filter geometry distribution estimation and a Naive Bayes classifier. Thereby the histogram-based representation enables to model non-Gaussian distributions of the length and width and hence to distinguish lower and upper bounds of the measurement box, also determined by the surrounding freespace of the box edges. The estimated object box size is finally extracted by evaluating both the accumulated measurements of the histogram filter distribution and the likelihood of the most likely object class, which further enables to consider prior class knowledge of the assumed object size when it has not been fully observed yet.

The work presented in this chapter has been published in [1] in the context of this thesis, which is not explicitly referenced in the following.

5.1 Introduction

Mobile robots, especially autonomous vehicles, require a robust object estimation of surrounding traffic participants to predict their individual trajectories, plan interactive maneuvers, and avoid collisions. Therefore, object instances have to be detected and abstracted from the measurement data of different sensors. However, it is further essential to track those objects over time with a temporal filtering of the object state estimation. Only then a real robustness against measurement noise is achieved, previously obtained measurement information is aggregated, and parts of the object state that are not directly observable can be implicitly derived based on the motion tracking. That object state estimation, including the measurement abstraction and the recursive filtering, is addressed this chapter.

Object tracking approaches, and thus also the corresponding object state estimation, generally directly depend on the input data and the different processing steps as well as their order. As motivated in the introduction of this thesis, high-level object fusion approaches that are based on a sensor-individual object tracking with an early-stage object abstraction, e.g., [8, 22, 76], are critical as they cause error-prone information loss regarding the subsequent processing step. In contrast to that, low-level early-stage sensor fusion approaches with a late-stage object abstraction and tracking, as performed in this work, result in a consistent multi-sensor object tracking, where the object state estimation forms the last processing step based on the combined measurement data of all sensors. In the following, new strategies for the object state estimation in the context of the grid-based measurement representation of this work are proposed. Details of the contribution and outline of this chapter are presented after the following brief discussion of related work.

5.1.1 Related Work

Object tracking is a broad field of research with various approaches and applications; a general overview is given, for instance, in [148]. The following discussion of related work focuses on similar generic grid-based approaches with some additional relevant work on radar Doppler-based motion estimation.

As discussed in the previous chapter, a simple strategy for the grid-based object extraction is to classify measurement data as dynamic by determining free/occupied inconsistencies with the temporally accumulated grid map, which are then clustered to form object instances. This is applied in [135], where objects are tracked by a common Kalman filter using a simple bounding box shape model, without providing further details on the filtering. Their approach is extended in [136] by a multiple hypothesis tracking (MHT) [17] using an interacting multiple model (IMM) [146] approach with a fixed set of sixteen different motion models along eight possible movement directions. However, such discrete motion models limit the overall estimation accuracy and also require a Kalman filter instance for each motion model. In [20], oriented bounding boxes of lidar data are extracted with an L-shape orientation fitting, based on [78], which are then also filtered by a Kalman filter. A constant 2-D acceleration model is used by decoupling the x- and y-direction, without directly filtering the object orientation, whereas the object length and width are estimated by a low-pass filter. A similar Kalman filtering approach is presented in [106], where the estimated object state is further extended by the orientation and steering angle. Moreover, the object state is adapted to the best observable reference point at the box corners, edge centers, or the center of the box, based on the extracted features of the lidar data, which enables an accurate position estimation. But their approach also requires a transformation of the full object state and a recalculation of the respective error covariance matrix. A different filtering strategy is presented in [56], where clusters of dynamic cells are tracked as a free-form object model by a particle filter using a constant turn rate and velocity motion model. The object state representation thereby directly includes a set of grid cells with an additional hit counter that combines the hypotheses of the individual particles. However, since all those approaches directly rely on the dynamic occupancy classification of occupied/free inconsistencies, they remain error-prone to measurement inaccuracies of ranging or odometry sensors, which thus also affects the object state estimation.

A more robust filtering strategy requires a recursive dynamic state estimation of the occupancy grid. As presented in Chapter 3, and similarly used in other recent approaches [29, 83, 85, 121, 131], a grid-based particle filter is used for efficiently estimating continuous velocity distributions of each grid cell, which is further combined with an adapted evidential filtering of the occupancy masses. As presented in [3], also published in the context of this thesis, the mean of the cell velocity estimates of all associated cells can then also be used for a direct measurement update of the object velocity and orientation with a subsequent object state filtering. However, this results in correlated input data of an object-level Kalman filter and thus in a multi-filtering of the velocity and orientation, which can cause higher filtering latencies in nonlinear scenarios. A similar object velocity measurement update by the particle-based cell velocity estimates is presented in [149], which is further extended by a de-autocorrelation scheme to whiten the correlated velocity input of the Kalman filter object tracking. However, they use a simplified linear approximation that is prone to deviations of that model, in particular in relevant scenarios with critical nonlinear movements.

Another strategy based on such dynamic occupancy grids is to avoid a subsequent filtering on the object-level and directly extract tracked objects from the filtered particle population by grouping the low-level particles to object instances. This is applied in [29] and similarly in [101], as also briefly discussed in Section 4.1.1. However, then the object estimation primarily results from the short-term particle estimation without a persistent high-level shape estimation. This is error-prone, also for the association, in scenarios with different viewing angles of a moving object, temporal occlusions, or densely moving traffic. Moreover, by averaging the cell-wise low-level particle estimation, no detailed dynamic state estimation on the object-level can be robustly derived, e.g., the velocity or acceleration of the long side of a truck or the turn rate of a turning vehicle, as this requires an evaluation for the full spatial object context rather than the local cell-individual evaluation.

A different approach is proposed in [94] with virtual rays of a laserscanner that are similar to a polar grid representation, which are used to compute the measurement likelihood with a pre-defined cost-function of a box model object shape, also considering 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 no static/dynamic classification is used.

The object shape can also be estimated using an object local grid map [12, 108], 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, in particular for highly dynamic movements, further requires Doppler radars that enable directly measuring velocity components. A multi-radar object tracking is proposed in [60, 61] 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 [103], the velocity profile is used for a single-radar tracking, where linear and nonlinear motions are distinguished and in the latter case the orientation is separately estimated based on the extracted contour of all radar detections. In [125], a Gaussian process is used for the radar-based shape estimation, combined with an extended Kalman filter (EKF) for the motion state 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 those radar-only approaches have difficulties in estimating the object pose and size, which in turn affects the motion estimation due to the inaccurately determined center of rotation.

5.1.2 Contribution and Outline

There are already several promising object tracking approaches that partly use benefits of the dynamic grid estimation, the freespace information contained in the grid, or radar Doppler velocity measurements. However, none of these approaches fully combines the different benefits, which is addressed in the following to achieve a generic and robust multi-sensor object tracking. Based on already associated occupied grid cells including the extraction of new objects, presented in the previous chapter, this chapter focuses on the subsequent object state estimation, with the main contributions as follows:

- In addition to the associated occupied grid cells, the proposed object state estimation also evaluates the freespace evidence of the measurement grid representation. In contrast to sensor-specific approaches that process only the raw sensor measurement detections, this work thus further determines the current visibility of object parts in a generic way based on the fused information of all sensors with different fields of view.

- A new dynamic state estimation strategy with an unscented Kalman filter (UKF) is presented that is based on the characteristics of the dynamic grid representation. Thereby the particle-based cell velocity estimates of the dynamic grid enable a direct initialization of the object velocity and movement direction. The object bounding box orientation is further locally optimized by minimizing the included freespace for different box orientations. Moreover, also the object position is robustly updated by selecting the most visible reference point based on the surrounding freespace.

- The dynamic state estimation is further extended by a generic radar velocity-based UKF motion estimation strategy. In contrast to other similar approaches that perform the measurement update in the object state space by abstracting the radar Doppler measurements to one estimated object velocity along a specific orientation, the proposed approach directly performs the update in the radar Doppler velocity measurement space by projecting the UKF sigma points to the expected radial velocities. This concept thus avoids the error-prone radar measurement abstraction and thereby further enables that the full dynamic state is implicitly updated by the UKF.

- A novel concept for the object shape estimation and classification is proposed. A histogram filter geometry distribution estimation is presented, that, in contrast to the UKF estimation, enables to model non-Gaussian distributions of the length and width and thus also to distinguish lower and upper bounds of the measurement box, also evaluated by the freespace-based visibility. Objects are then classified by a Naive Bayes classifier based on the geometry distribution and maximum velocity, but also considering the camera classification information. The estimated object box size is finally extracted by evaluating both the accumulated measurements of the histogram filter distribution and the likelihood of the most likely object class, i.e., prior class knowledge of the assumed object size is used when it has not been fully observed yet.

Those different aspects of the proposed approach are divided into the following sections:

- *Section 5.2* introduces the object state representation that is divided into the dynamic motion state and the static object shape due to the two different filters used.
- *Section 5.3* presents the generic UKF-based dynamic state estimation, including the object box abstraction of the associated occupied cells, the freespace-based box edge visibility evaluation with the reference point selection, the particle-based object velocity and orientation initialization, and the freespace-based orientation optimization.
- *Section 5.4* extends the UKF dynamic state estimation by the additional utilization of the radar Doppler velocity measurements.
- *Section 5.5* proposes the combined object shape estimation and classification strategy, based on the generic histogram filter geometry distribution estimation for the length and width of the bounding box and the corresponding Naive Bayes classifier.

Finally, some qualitative results and a summary of this chapter are presented in Section 5.6, which is then continued by a detailed evaluation of the overall grid-based object tracking approach in Chapter 6.

5.2 Object State Representation

Before the object state estimation is discussed in detail, the object state representation of each object track $\tau \in \mathcal{T}_t$ of all currently tracked objects $\mathcal{T}_t$ at time t is briefly introduced in this section. Since the object state is estimated by two different filtering techniques in this work, the object state representation is divided into two separate parts: the dynamic motion state $s_{\tau,t}$ and the static object shape g_τ.

The dynamic state of a track τ at time t is defined by the vector

$$s_{\tau,t} = \left[\, x_t^x, \, x_t^y, \, v_t, \, a_t, \, \varphi_t, \, \omega_t \, \right], \tag{5.1}$$

which is composed of the 2-D position $x = [x^x, \, x^y]$, the longitudinal velocity v with the corresponding acceleration $a = \dot{v}$, and the orientation φ with the turn rate $\omega = \dot{\varphi}$.

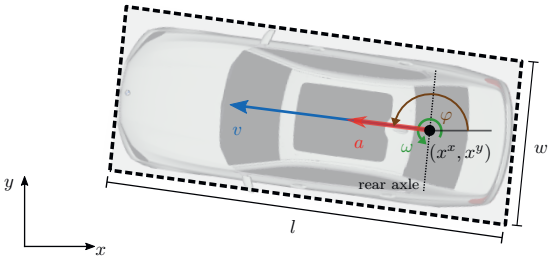


Fig. 5.1: Illustration of the object state representation. Each object is modeled by a bounding box with length l and width w; the dynamic state is composed of the position x, the longitudinal velocity v and acceleration a, and the orientation φ with the corresponding turn rate ω. Figure initially published in [1], © 2019 IEEE.

The corresponding object shape is modeled by an oriented bounding box with length l and width w, which is summarized by the geometric state

$$g_\tau = \left[\, l, \; w \, \right]. \tag{5.2}$$

The length axis of the bounding box of the track is oriented along the track orientation φ_t and the width axis is defined orthogonal to it. This state is assumed to be static, i.e., it does not change over time, thus the time index of g_τ is omitted.

Equivalent to the grid representation, the accumulated odometry is used as the uniform reference frame of the position of all tracks. Thereby the position x of an object is fixed at its estimated center of rotation, which is approximated between the center point and the middle rear point of the selected box representation for all object types, i.e., along the object length axis at $0.25\,l$. 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 summarized in an illustration in Fig. 5.1.

5.3 Dynamic State Estimation

This section focuses on the recursive estimation of the dynamic state $s_{\tau,t}$ of the object track τ, as defined in (5.1), which is temporally filtered by an unscented Kalman filter (UKF) [55]. First, the prediction step of the UKF state with the corresponding covariance is briefly described, followed by different aspects of the measurement update step, which focus on the required abstraction of the measurement data of the grid structure toward the object state representation.

5.3.1 Prediction

The recursive UKF filtering requires a prediction of the previous value at time $t-1$ to the time t of the current measurement, with the in-between time difference Δt. This in turn requires a motion model of the mean state $s_{\tau,t}$ and a process noise model of the corresponding covariance $\Sigma_{\tau,t}$, which are both briefly described in the following.

Motion Model

The motion model describes the state transition of the previous object state $s_{\tau,t-1}$ to the predicted object state $\hat{s}_{\tau,t}$. In the following, all predicted variables are characterized by a hat symbol. The motion model of this work is based on a non-linear constant turn rate and acceleration (CTRA) motion model [105], a popular vehicle model that enables clothoid trajectories. To increase the overall robustness, the CTRA motion model is slightly modified in this work by ensuring slow convergence of the turn rate ω and acceleration a toward zero, which is modeled by two reduction factors $\varepsilon_\omega, \varepsilon_a \in (0,1)$, i.e.,

$$\hat{\omega}_t = (1-\varepsilon_\omega)\,\omega_{t-1}\,, \tag{5.3}$$

$$\hat{a}_t = \underset{a_1,a_2}{\arg\min}\left(|a_1|,\,|a_2|\right),\quad a_1 = (1-\varepsilon_a)\,a_{t-1}\,,\; a_2 = \frac{-v_{t-1}}{t_{\text{horizon}}}\,. \tag{5.4}$$

Moreover, the acceleration is additionally limited by the term $-v_{t-1}/t_{\text{horizon}}$ here 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 significant negative velocity. The parameter $t_{\text{horizon}} \geq \Delta t$ thereby defines the time horizon and hence the smoothness of the limitation.

The remaining state is updated by the default CTRA motion model [105] with

$$\hat{\varphi}_t = \varphi_{t-1} + \omega\Delta t\,, \tag{5.5}$$

$$\hat{v}_t = v_{t-1} + a\Delta t\,, \tag{5.6}$$

$$\hat{x}_t^x = x_{t-1}^x + \frac{1}{\omega^2}\left[\omega\hat{v}_t\sin(\hat{\varphi}_t) + a\cos(\hat{\varphi}_t) - \omega v_{t-1}\sin(\varphi_{t-1}) - a\cos(\varphi_{t-1})\right], \tag{5.7}$$

$$\hat{x}_t^y = x_{t-1}^y + \frac{1}{\omega^2}\left[-\omega\hat{v}_t\cos(\hat{\varphi}_t) + a\sin(\hat{\varphi}_t) + \omega v_{t-1}\cos(\varphi_{t-1}) - a\sin(\varphi_{t-1})\right], \tag{5.8}$$

using $\omega = \hat{\omega}_t$ and $a = \hat{a}_t$. To avoid singularities, the position update in (5.7)–(5.8) is replaced by a constant acceleration motion model for $\omega \approx 0$.

Process Noise Model

The applied process noise model is based on a Wiener-sequence acceleration model [13, 71],

$$Q_w = \sigma_{w,\dddot{\zeta}}^2 \begin{bmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^3}{2} & \Delta t^2 & \Delta t \\ \frac{\Delta t^2}{2} & \Delta t & 1 \end{bmatrix}, \tag{5.9}$$

where, for a state $[\zeta, \dot{\zeta}, \ddot{\zeta}]$, a change of the acceleration $\ddot{\zeta}$ is assumed to be an independent white noise process with variance $\sigma_{w,\dddot{\zeta}}^2$. Applying (5.9) independently to the tangential acceleration a and the radial acceleration $\dot{\omega}$ of the state as defined in (5.1), the UKF process noise matrix results in

$$Q = \begin{bmatrix} Q_1 & 0 \\ 0 & Q_2 \end{bmatrix}, \tag{5.10}$$

with the two symmetric submatrices

$$Q_1 = \sigma_{w,a}^2 \begin{bmatrix} \frac{\Delta t^4}{4}\cos^2\varphi & \frac{\Delta t^4}{4}\sin\varphi\cos\varphi & \frac{\Delta t^3}{2}\cos\varphi & \frac{\Delta t^2}{2}\cos\varphi \\ \frac{\Delta t^4}{4}\sin\varphi\cos\varphi & \frac{\Delta t^4}{4}\sin^2\varphi & \frac{\Delta t^3}{2}\sin\varphi & \frac{\Delta t^2}{2}\sin\varphi \\ \frac{\Delta t^3}{2}\cos\varphi & \frac{\Delta t^3}{2}\sin\varphi & \Delta t^2 & \Delta t \\ \frac{\Delta t^2}{2}\cos\varphi & \frac{\Delta t^2}{2}\sin\varphi & \Delta t & 1 \end{bmatrix}, \tag{5.11}$$

$$Q_2 = \sigma_{w,\dot{\omega}}^2 \begin{bmatrix} \frac{\Delta t^4}{4} & \frac{\Delta t^3}{2} \\ \frac{\Delta t^3}{2} & \Delta t^2 \end{bmatrix}. \tag{5.12}$$

Overall, these motion and process noise models are used for the prediction of the dynamic state and the corresponding covariance, which represents the first step of the recursive state estimation of the UKF. For the second step, the measurement update, measurement data of the grid structure have to be abstracted to the object state representation, which is addressed in the following.

5.3.2 Transformation of Associated Cells to the Box Representation

The set $\mathcal{C}_{\tau,t}$ of occupied cells of the current measurement grid that have been associated to the track τ, as defined in (4.3) and discussed in detail in the previous chapter, is transformed to an oriented minimum bounding box in the following. Hence, occupied grid cells are abstracted to the same object box representation as the filtered tracks in this processing step. Given the orientation φ, the length and width of the measurement box

$$l_z = \max_{c \in \mathcal{C}_{\tau,t}} \left(\tilde{x}_c^{\varphi,x}\right) - \min_{c \in \mathcal{C}_{\tau,t}} \left(\tilde{x}_c^{\varphi,x}\right) + \delta_d^{\varphi} \tag{5.13}$$

$$w_z = \max_{c \in \mathcal{C}_{\tau,t}} \left(\tilde{x}_c^{\varphi,y}\right) - \min_{c \in \mathcal{C}_{\tau,t}} \left(\tilde{x}_c^{\varphi,y}\right) + \delta_d^{\varphi} \tag{5.14}$$

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

$$\left[\tilde{x}_c^{\varphi,x},\, \tilde{x}_c^{\varphi,y}\right]^{\mathsf{T}} = R_{\varphi} \, \left[x_c^x,\, x_c^y\right]^{\mathsf{T}} \tag{5.15}$$

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

$$\delta_d^{\varphi} = d_c \left(|\sin(\varphi)| + |\cos(\varphi)| \right) \in [d_c, \sqrt{2}\, d_c] \tag{5.16}$$

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 5.3.5. A geometric illustration of the transformation to the box model is shown in Fig. 5.2.

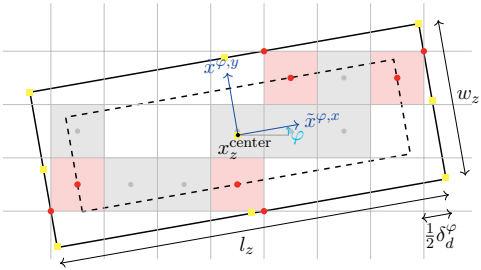


Fig. 5.2: Transformation of associated grid cells to a bounding box model. Associated cells are filled gray and crucial cells are highlighted red. The inner dotted rectangle fits the cell centers, the solid rectangle also considers the cell extent by adding δ_d^{φ}. All possible reference points are visualized by yellow squares. Figure initially published in [1], © 2019 IEEE.

5.3.3 Position Measurements with Reference Point Selection

The associated occupied grid cells have been abstracted to an oriented measurement box as described above, which enables a position update. The position of the center of the measurement box in the odometry frame is calculated as

$$x_z^{\text{center}} = R_{\varphi}^{\mathsf{T}} \begin{bmatrix} \min\limits_{c \in \mathcal{C}_{\tau,t}} (\tilde{x}_c^{\varphi,x}) + \frac{1}{2}(l_z - \delta_d^{\varphi}) \\ \min\limits_{c \in \mathcal{C}_{\tau,t}} (\tilde{x}_c^{\varphi,y}) + \frac{1}{2}(w_z - \delta_d^{\varphi}) \end{bmatrix} . \tag{5.17}$$

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, cf. (5.1). 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 (5.17). 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, also with noise-free measurement data.

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, as also used in [7, 106]. The position in terms of a reference point is thus defined as

$$x_z^{\text{ref}} = x_z^{\text{center}} + R_{\varphi}^{\mathsf{T}} \begin{bmatrix} \delta_{z,l}^{\text{ref}} \, l_z \\ \delta_{z,w}^{\text{ref}} \, w_z \end{bmatrix} , \; \delta_{z,l}^{\text{ref}}, \delta_{z,w}^{\text{ref}} \in \left\{0, +\tfrac{1}{2}, -\tfrac{1}{2}\right\}, \tag{5.18}$$

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, the surrounding freespace evidence of the measurement grid is considered to determine the best reference point. Hence, this concept is suitable for a multi-sensor setup

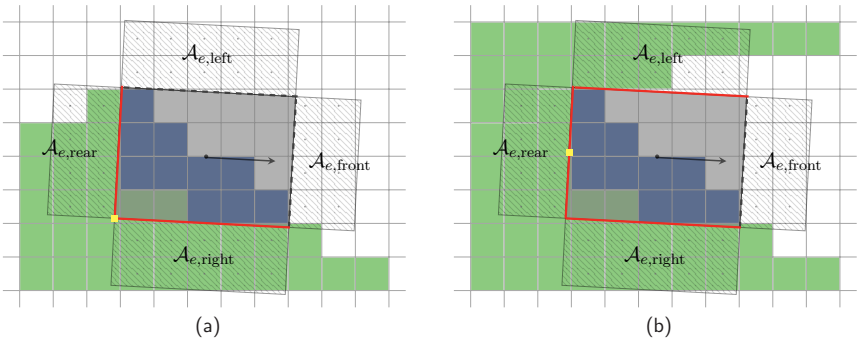


Fig. 5.3: Analysis of track edge visibility considering the freespace evidence in the surrounding striped rectangle areas $\mathcal{A}_{e,\tau}$. Visible edges are highlighted red, a yellow square denotes the selected reference point. (a) Scenario where the rear and right edge are visible. (b) Scenario with additional freespace at the left edge such that the extracted width represents also an upper bound. Figure initially published in [1], © 2019 IEEE.

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 regarding 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

$$r_F(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \sum_{c \in \mathcal{A}} m(F_{z,t}^c), \quad \mathcal{A} \subseteq \mathcal{G} \tag{5.19}$$

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

$$\vartheta_z^e \approx r_F(\mathcal{A}_e), \tag{5.20}$$

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. 5.3.

The position along the length axis is finally selected as

$$\delta_{z,l}^{\text{ref}} = \begin{cases} +\frac{1}{2}, & \text{if } \vartheta_z^{\text{front}} \geq \vartheta_{\min} \ \wedge \ \vartheta_z^{\text{rear}} < \vartheta_{\min} \\ -\frac{1}{2}, & \text{if } \vartheta_z^{\text{front}} < \vartheta_{\min} \ \wedge \ \vartheta_z^{\text{rear}} \geq \vartheta_{\min} \\ 0, & \text{else} \end{cases}, \tag{5.21}$$

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. If no object edge is fully visible regarding surrounding freespace, e.g., due to partial occlusion, the reference point can optionally be selected by evaluating the closest distance to the ego vehicle. This edge visibility ϑ_z^e is also considered later for the geometry estimation, which is discussed in Section 5.5. Overall, this reference point selection with the freespace evaluation enables a robust position update of the track.

5.3.4 Velocity and Orientation Estimation by the 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 (3.11). Hence, equivalent to (4.35), the 2-D weighted mean velocity vector

$$\overline{\nu} = \left(\sum_{c \in \mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \right)^{-1} \sum_{c \in \mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \nu_t^c \tag{5.22}$$

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.,

$$v_{\tilde{z}} = \| \, \overline{\nu} \, \| \, , \tag{5.23}$$

$$\varphi_{\tilde{z}} = \arctan(\overline{\nu}) \, . \tag{5.24}$$

The corresponding weighted variances are calculated as

$$\sigma_{v,\tilde{z}}^2 = \eta_\sigma \sum_{c \in \mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \left(\| \nu_t^c \| - v_{\tilde{z}} \right)^2 , \tag{5.25}$$

$$\sigma_{\varphi,\tilde{z}}^2 = \eta_\sigma \sum_{c \in \mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \left((\arctan(\nu_t^c) - \varphi_{\tilde{z}}) \bmod 2\pi \right)^2 , \tag{5.26}$$

with an unbiased weight normalization [42]

$$\eta_\sigma = \frac{\sum\limits_{c \in \mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c)}{\Big(\sum\limits_{c \in \mathcal{C}_{\tau,t}} m(D_{\tilde{z},t}^c) \Big)^2 - \sum\limits_{c \in \mathcal{C}_{\tau,t}} \big(m(D_{\tilde{z},t}^c) \big)^2} \, . \tag{5.27}$$

That 2-D velocity $\overline{\nu}$ 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 velocity and orientation 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 previous work of this thesis [3]. However, these estimates are already filtered by the particle tracking, causing 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 to any deviations of the motion model, which is critical for scenarios with a fast-changing turn rate or acceleration. Therefore, in this work, the velocity and orientation estimates of the particles 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

$$\mathcal{I}_{\varphi}^{\mathcal{X}} = \left\{ \varphi \in [-\pi, \pi] \;\middle|\; |\varphi - \varphi_{\tilde{z}}| \bmod \pi \leq \gamma_{\sigma}\, \sigma_{\varphi,\tilde{z}} + \sigma_{\varphi,\min}^{\mathcal{X}} \right\} \tag{5.28}$$

of the assumed object orientation based on the movement of the point-mass particles. The parameter γ_{σ} 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.

5.3.5 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 (5.19) 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, cf. (5.13)–(5.14). 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

$$\varphi_z^{*} = \underset{\varphi \in \mathcal{I}_{\varphi}^{\mathcal{X}}}{\arg\min}\; \kappa(\varphi), \quad \kappa(\varphi) = r_F(\mathcal{A}_i^{\varphi})\,, \tag{5.29}$$

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 (5.28).

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 ($\varphi_0 = \varphi_{\bar{z},\tau}$). 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.,

$$\varphi_z^* \approx \left(\sum_{\varphi'} \frac{1}{\kappa(\varphi')} \right)^{-1} \sum_{\varphi'} \frac{1}{\kappa(\varphi')} \cdot \varphi', \tag{5.30}$$

$$\varphi' \in \{\varphi_{i^*},\, \varphi_{i^*} + \delta_{\varphi},\, \varphi_{i^*} - \delta_{\varphi}\}\,. \tag{5.31}$$

The corresponding variance

$$\sigma_{\varphi,z}^2 = (\nabla_r - \nabla_l)^{-2} \tag{5.32}$$

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

$$\nabla_r = \frac{\kappa(\varphi_i^{\max,r}) - \kappa(\varphi_{i^*})}{\varphi_i^{\max,r} - \varphi_{i^*}}, \quad \varphi_i^{\max,r} = \max_{i>i^*} \varphi_i\,. \tag{5.33}$$

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, hence, resulting in an approach that is also robust against deviations from the box model shape.

Overall, 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.

5.4 Additional Radar-Based Dynamic Estimation

The dynamic state of each track is robustly estimated based on the occupancy and freespace evidence of the measurement grid as described in the previous section. In the following, that concept is further extended by evaluating radar Doppler measurements that may occur at a track, which significantly improves the dynamics estimation for nonlinear movements.

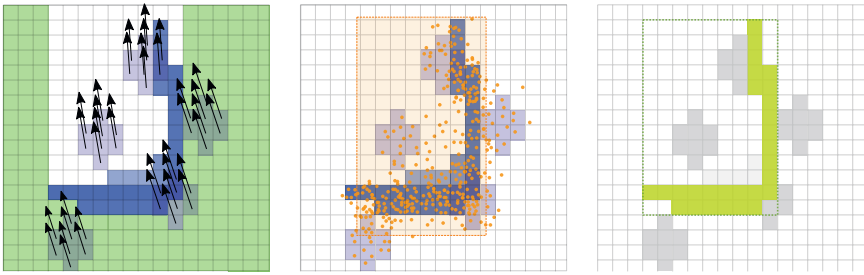


Fig. 5.4: Extended particle labeling association with radar measurements. (a) Classified measurement grid with radar measurements illustrated by arrows. (b) Dynamic cells with labeled particles and predicted track. (c) Extracted minimum bounding box formed by cells with a high dynamic mass. The remaining cells with radar velocities are considered for the dynamic state estimation, but not utilized for the extraction of the bounding box. Figure initially published in [1], © 2019 IEEE.

5.4.1 Association of Radar Doppler Measurements

As defined in (2.18), the radar velocity measurements $z^c_{v,t}$ are represented in the measurement grid structure as separate layers, containing the radial velocity v^r_z, the azimuth angle θ^r_z, and the position x_s of the corresponding sensor origin at the measurement time t. The radar measurements thereby typically have a higher spatial uncertainty than those of 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, as discussed before. The presented cell association strategy, however, discards cells with a low dynamic occupancy evidence mass, as defined in (4.19), as only cells with a higher occupancy evidence should be taken into account to robustly form the measurement minimum bounding box of the object.

Hence, all occupied cells of the extended set of associated cells $\mathcal{C}^+_{\tau,t} \supseteq \mathcal{C}_{\tau,t}$, still extracted by evaluating 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.,

$$\mathcal{C}^r_\tau = \{\, c \in \mathcal{C}^+_{\tau,t} \mid \exists\, z^c_{v,t} \,\}\,. \tag{5.34}$$

This extended association is illustrated in Fig. 5.4, which forms the basis for updating the tracks by the associated radar Doppler measurements.

5.4.2 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, cf. Fig. 5.1. The 2-D Cartesian velocity v_i in the odometry reference frame at an arbitrary point x_i of the

object is calculated as

$$\begin{bmatrix} v_i^x \\ v_i^y \end{bmatrix} = v_\tau \begin{bmatrix} \cos(\varphi_\tau) \\ \sin(\varphi_\tau) \end{bmatrix} + \omega_\tau \begin{bmatrix} -(x_i^y - x_\tau^y) \\ x_i^x - x_\tau^x \end{bmatrix}, \tag{5.35}$$

as described in [61] using the instant center of rotation (ICR) and the assumption of the Ackermann steering geometry with a drift-free driving state.

The radial velocity component regarding the angle θ_i^r results in

$$\begin{aligned} v_i^r =& \begin{bmatrix} \cos(\theta_i^r),\, \sin(\theta_i^r) \end{bmatrix} \begin{bmatrix} v_i^x \\ v_i^y \end{bmatrix} \\ =& v_\tau \left(\cos(\theta_i^r)\cos(\varphi_\tau) + \sin(\theta_i^r)\sin(\varphi_\tau)\right) + \omega_\tau \begin{bmatrix} \cos(\theta_i^r),\, \sin(\theta_i^r) \end{bmatrix} \begin{bmatrix} -(x_i^y - x_\tau^y) \\ x_i^x - x_\tau^x \end{bmatrix}, \end{aligned} \tag{5.36}$$

which depends on the evaluated position x_i if a turn rate occurs ($\omega_\tau \neq 0$), i.e., a turning object results in different velocity vectors for different positions of that object. Using the cell centers x_c of $z_{v,t}^c$ as the position, however, is critical, as it introduces a bias error due to the grid discretization and the additional spatial uncertainty over multiple grid cells as shown in Fig. 5.4. Instead, the position of the original radar detection can also be described

$$\begin{bmatrix} x_i^x \\ x_i^y \end{bmatrix} = d_i \begin{bmatrix} \cos(\theta_i^r) \\ \sin(\theta_i^r) \end{bmatrix} + \begin{bmatrix} x_s^x \\ x_s^y \end{bmatrix} \tag{5.37}$$

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_i^r of the measurement. Since (5.37) also fulfills the relation

$$\begin{bmatrix} \cos(\theta_i^r),\, \sin(\theta_i^r) \end{bmatrix} \begin{bmatrix} -x_i^y \\ x_i^x \end{bmatrix} = \begin{bmatrix} \cos(\theta_i^r),\, \sin(\theta_i^r) \end{bmatrix} \begin{bmatrix} -x_s^y \\ x_s^x \end{bmatrix}, \tag{5.38}$$

the radial velocity of (5.36) can also be evaluated without any additional bias error, which finally results in

$$v_i^r = v_\tau \cos(\theta_i^r - \varphi_\tau) + \omega_\tau \big(\sin(\theta_i^r)(x_s^x - x_\tau^x) - \cos(\theta_i^r)(x_s^y - x_\tau^y) \big) \tag{5.39}$$

using the trigonometric addition formula [141]

$$\cos(\theta_i^r - \varphi_\tau) = \cos(\theta_i^r)\cos(\varphi_\tau) + \sin(\theta_i^r)\sin(\varphi_\tau). \tag{5.40}$$

The different geometric relations of the velocity parts are summarized in Fig. 5.5.

5.4.3 Radar-Based Motion Estimation

Based on the relation between the radial velocity component and the object state as derived in (5.39), the object state estimation can be updated by the radar measurements $v_{z,i}^r$ with the corresponding measurement azimuth angle $\theta_{z,i}^r$ and the sensor position x_s. A common

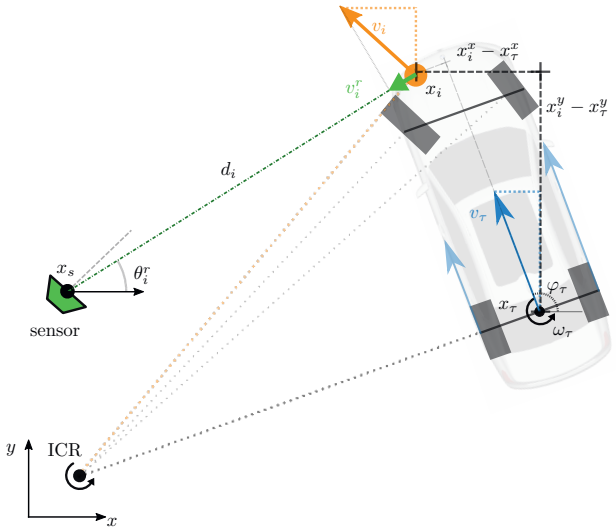


Fig. 5.5: Geometric relations of the velocity components of an object with a turn rate using the instant center of rotation (ICR) as the stationary rotation point of the object. The radial velocity v_i^r depends on the angle θ_i^r of the radial component, the dynamic object state s_τ, and the position difference between the track x_τ and the evaluated position x_i or the sensor position x_s. Figure initially published in [1], © 2019 IEEE.

approach is to summarize the relation of all radar measurements to an equation system, resulting in a least-squares (LSQ) optimization problem in the form of

$$\min_{v_z^x, v_z^y, \omega_z} \left\| \begin{bmatrix} \cos(\theta_{z,1}^r) & \sin(\theta_{z,1}^r) & \Omega_{z,1} \\ \vdots & \vdots & \vdots \\ \cos(\theta_{z,i}^r) & \sin(\theta_{z,i}^r) & \Omega_{z,i} \\ \vdots & \vdots & \vdots \\ \cos(\theta_{z,N}^r) & \sin(\theta_{z,N}^r) & \Omega_{z,N} \end{bmatrix} \begin{bmatrix} v_z^x \\ v_z^y \\ \omega_z \end{bmatrix} - \begin{bmatrix} v_{z,1}^r \\ \vdots \\ v_{z,i}^r \\ \vdots \\ v_{z,N}^r \end{bmatrix} \right\|_2^2 \tag{5.41}$$

by transforming (5.39) to the equivalent relation

$$v_{z,i}^r = \begin{bmatrix} \cos(\theta_{z,i}^r) & \sin(\theta_{z,i}^r) & \underbrace{\left(\sin(\theta_{z,i}^r)(x_s^x - x_\tau^x) - \cos(\theta_{z,i}^r)(x_s^y - x_\tau^y)\right)}_{\Omega_{z,i}} \end{bmatrix} \begin{bmatrix} v_z^x \\ v_z^y \\ \omega_z \end{bmatrix} \tag{5.42}$$

using the Cartesian representation of the object state velocity rather than the polar representation, i.e., $v_z^x = v_z \cos(\varphi_z), v_z^y = v_z \sin(\varphi_z)$. That representation can then be transformed to the object state representation by $v_z = \sqrt{(v_z^x)^2 + (v_z^y)^2}$ and $\varphi_z = \arctan(v_z^y, v_z^x)$, with a respective uncertainty error propagation. Hence, this generally enables to derive measurements of the scalar object velocity v_z, the object orientation φ_z, and the turn

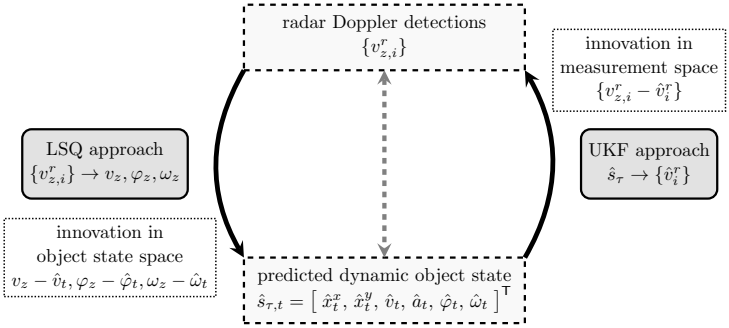


Fig. 5.6: Overview of possible radar-based motion estimation concepts. The left path represents an innovation in the object state space by an object state abstracting of the radar Doppler measurements, whereas the right path represents the proposed concept where the measurement update is performed directly in the measurement space by transforming the predicted object state to predicted radial velocity components.

rate ω_z, which can then be directly used to update the dynamic object state s_τ of the object track τ. However, in order to derive the full three degrees of freedom of (5.41), the rank of the matrix also has to be greater or equal than three, which, as denoted by Kellner et al. [59–61], is fulfilled if at least two radar sensors provide a total of at least three detections. Otherwise, the number of degrees of freedom has to be reduced, e.g., by avoiding to extract an orientation measurement and using the predicted track orientation instead, or, in the case of only one detection, by further assuming that no turn rate occurs, resulting in different cases and error-prone assumptions.

But the UKF estimation also enables a different strategy, 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 $h(\xi)$. Hence, rather than abstracting the measurements from the measurement space to the object state space, it is also possible to transform the predicted object state $\hat{s}_\tau$ to the radar Doppler velocity measurement space and then determine directly the innovation between the measured radial velocity component $v^r_{z,i}$ and the expected value $\hat{v}^r_i$ for the measurement update. This concept thus avoids the error-prone radar measurement abstraction of the LSQ optimization and thereby further enables that the full dynamic state is implicitly updated by the UKF. A comparison of those two different strategies is illustrated in Fig. 5.6.

In order to determine the predicted radial velocity component $\hat{v}^r_i$ for each measured velocity $v^r_{z,i}$ with azimuth angle $\theta^r_{z,i}$ and the corresponding sensor position $x_{s,i}$, each predicted sigma point $\hat{\xi}_{\tau,t}$ of the track τ is projected

$$h^r_i(\hat{\xi}_{\tau,t}) = \hat{v}_{\tau,t,\xi} \cos(\theta^r_{z,i} - \hat{\varphi}_{\tau,t,\xi}) + \hat{\omega}_{\tau,t,\xi} \big(\sin(\theta^r_{z,i})(x^x_{s,i} - \hat{x}^x_{\tau,t,\xi}) - \cos(\theta^r_{z,i})(x^y_{s,i} - \hat{x}^y_{\tau,t,\xi}) \big) \tag{5.43}$$

to the measurement space of the radial velocity using the relation of (5.39).

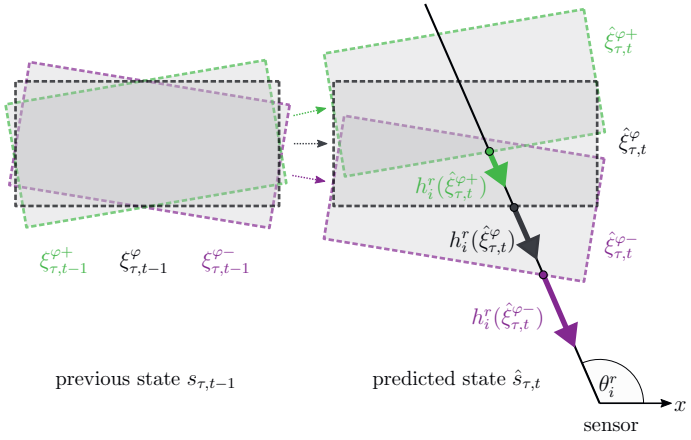


Fig. 5.7: Illustration of different state orientations φ of the UKF sigma points ξ, their prediction $\hat{\xi}$, and the different expected radial velocity components $h^r_i(\hat{\xi}_{\tau,t})$ given the observation angle θ^r_i of the radar sensor. Figure initially published in [1], © 2019 IEEE.

The predicted radial velocity then results from the weighted sum of all sigma points, i.e.,

$$\hat{v}^r_i = \sum_{\xi \in \Xi} w_\xi h^r_i(\hat{\xi}_{\tau,t}) \,, \tag{5.44}$$

which is used to determine the innovation $v^r_{z,i} - \hat{v}^r_i$ as well as the state-measurement cross-covariance matrix and the Kalman gain, finally resulting in the updated state and covariance, as described in more detail in [55]. For the dynamic state $s_{\tau,t}$ as defined in (5.1), there is a total of $|\Xi| = 2 \cdot |s_{\tau,t}| + 1 = 13$ sigma points.

The measured azimuth angle $\theta^r_{z,i}$ is assumed to be noise-free in (5.43), i.e., errors of this parameter are not considered by the predicted UKF measurement in this work. However, implausible measurements, e.g., caused by micro-Doppler [25] or a wrong association, are discarded by evaluating the innovation with the predicted measurement covariance using a 3σ-range as the valid gating area.

Overall, this UKF-based measurement update by the radar Doppler velocities enables a robust and generic dynamic state estimation. Not only the velocity is 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 $\hat{v}^r_i$. For example, a different track orientation results also in a different angle difference to the azimuth angle of the measurement, which leads to a different radial velocity as illustrated in Fig. 5.7. An explicit determination of the velocity, orientation, and turn rate by an LSQ fitting of the equation set (5.39) for all measured radial velocities, in contrast, requires at least three detections of two sensors at the same measurement time and with varying azimuth angles.

5.5 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 prior class knowledge of the assumed length and width when either of them has not been fully observed yet.

5.5.1 Histogram Filter Geometry Distribution Estimation

The box shape is updated iteratively by the geometry measurement of the current time instance t

$$z^g_{\tau,t} = \left[\, z^l_{\tau,t},\, z^w_{\tau,t},\, z^{\vartheta,\text{front}}_{\tau,t},\, z^{\vartheta,\text{rear}}_{\tau,t},\, z^{\vartheta,\text{left}}_{\tau,t},\, z^{\vartheta,\text{right}}_{\tau,t} \,\right]^\mathsf{T} . \tag{5.45}$$

However, the real length and width are usually not fully observable, which means that the measured length $z^l_{\tau,t}$ and width $z^w_{\tau,t}$ generally describe only a minimum bounding box, i.e., a lower bound of the geometric shape. The visibility $z^\vartheta_{\tau,t} \in [0,1]$ of each edge of the current measurement indicates whether $z^l_{\tau,t}$ and $z^w_{\tau,t}$ also represent upper bounds and thus the real size of that track. For example, an upper bound of the length l_τ is measured if both edges at the front and the rear are currently visible.

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.,

$$p(l_\tau, w_\tau \,|\, z^g_{\tau,1:t}) = p(l_\tau \,|\, z^g_{\tau,1:t})\; p(w_\tau \,|\, z^g_{\tau,1:t}) . \tag{5.46}$$

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

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

$$p(l_\tau \,|\, z^g_{\tau,t}) \propto \exp\left(-\frac{(l_\tau - z^l_{\tau,t})^2}{2\varsigma^\vartheta_{l_\tau}\sigma_l^2}\right), \quad \varsigma^\vartheta_{l_\tau} = \begin{cases} \left(z^{\vartheta,\text{front}}_{\tau,t}\; z^{\vartheta,\text{rear}}_{\tau,t}\right)^{-1}, & \text{if } l_\tau \geq z^l_{\tau,t} \\ 1\,, & \text{else} \end{cases}, \tag{5.47}$$

i.e., a normal distribution in which the variance below and above the measured mean $z^l_{\tau,t}$ differs by a scaling variable $\varsigma^\vartheta_{l_\tau}$. This scaling variable is used to model the impact of the edge visibility. If both edges are fully visible, i.e., $z^{\vartheta,\text{front}}_{\tau,t} = z^{\vartheta,\text{rear}}_{\tau,t} = 1$, then the probability $p(l_\tau \,|\, z^g_{\tau,t})$ corresponds to an unmodified normal distribution with $z^l_{\tau,t}$ representing an estimate of the real length. If at least one edge is not visible, then $z^l_{\tau,t}$ is only a lower bound of l_τ, in which (5.47) is modeled similar to a sigmoid function. This sensor model is illustrated in Fig. 5.8 for varying edge visibilities.

The probability distribution of the length $p(l_\tau \,|\, z^g_{\tau,1:t})$ is estimated by a 1-D histogram filter, requiring a decomposition of the continuous state space $l_\tau \in \mathbb{R}^+$ into discrete values $\{l_i\}_{i=1}^{I_l} = [l_1, l_2, \ldots, l_{I_l}]$. The probability density function of the random variable L_τ

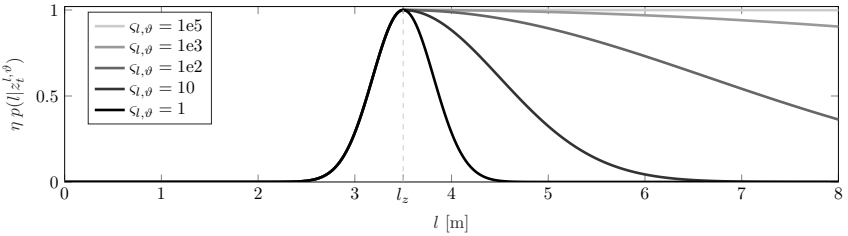


Fig. 5.8: Inverse sensor model with varying boundary visibilities. The example shows the length distribution with a measured length of $l_z = 3.5\,\mathrm{m}$ (dashed line). Figure initially published in [1], © 2019 IEEE.

is thus approximated by the discrete probability mass function $\{p^i_{l_\tau,t}\}$ in that interval, i.e.,

$$p(l_\tau \,|\, z^g_{\tau,1:t}) \approx \frac{p^i_{l_\tau,t}}{\delta_i}\,, \quad l_\tau \in \left(l_i - \tfrac{\delta_i}{2}, l_i + \tfrac{\delta_i}{2}\right]\,, \tag{5.48}$$

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 [128] to increase the histogram approximation. However, a constant interval size is used 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,

$$p^i_{l_\tau,t} = \frac{p(l_i \,|\, z^g_{\tau,t})\; p^i_{l_\tau,t-1}}{\sum_{j=1}^{I_l} p(l_j \,|\, z^g_{\tau,t})\; p^j_{l_\tau,t-1}}\,, \tag{5.49}$$

by applying Bayes rule and using the inverse sensor model as defined in (5.47). 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 (5.49).

5.5.2 Classification Based on Geometry and Velocity Information

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

$$\mathcal{K} = \{\text{car}, \text{truck}, \text{pedestrian}, \text{cyclist}, \text{motorcycle}, \text{other}\}\,. \tag{5.50}$$

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

$$p(k \,|\, l, w, v) = \frac{p(k)\, p(l, w, v \,|\, k)}{p(l, w, v)}\,. \tag{5.51}$$

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

$$p(l, w, v \,|\, k) = p(l \,|\, k)\, p(w \,|\, k)\, p(v \,|\, k)\,, \tag{5.52}$$

which simplifies (5.51) to a Naive Bayes classifier with

$$p(k \,|\, l, w, v) \propto p(k)\, p(l \,|\, k)\, p(w \,|\, k)\, p(v \,|\, k)\,. \tag{5.53}$$

The modeled likelihood functions of the length $p(l \,|\, k)$, the width $p(w \,|\, k)$, and the velocity $p(v \,|\, k)$ are shown in Fig. 5.9 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 classes car and truck 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

$$p(l \,|\, k) \propto S(l, l^k_{\min}, \alpha^k_{l,\min}) \left(1 - S(l, l^k_{\max}, \alpha^k_{l,\max})\right)\,, k \in \{\text{car}, \text{truck}\} \tag{5.54}$$

using a combination of two sigmoid (logistic) functions with

$$S(x, x_0, \alpha) = \frac{1}{1 + e^{-\alpha(x - x_0)}}\,. \tag{5.55}$$

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

$$p(v \,|\, k) \propto 1 - S(v, v^k_{\max}, \alpha^k_v)\,. \tag{5.56}$$

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, hence, the likelihood in (5.56) is not normalized over the individual classes here. For example, a velocity of $v = 1\,\text{km/h}$ 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 (5.56) and adjusting the individual class priors $p(k)$ in (5.53) accordingly.

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

$$k^* = \underset{k \in \mathcal{K}}{\arg\max}\; p(k \,|\, l, w, v) = \underset{k \in \mathcal{K}}{\arg\max}\; p(k)\, p(l \,|\, k)\, p(w \,|\, k)\, p(v \,|\, k)\,. \tag{5.57}$$

However, rather than classifying the individual measurements and recursively filtering that classification, the classification is directly performed on the filtered track τ. This means, 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 as described above. Therefore, instead of evaluating a single length l of the likelihood $p(l \,|\, k)$, the expectation

$$E_L[\, p(L_\tau \,|\, k)\,] = \sum_{i=1}^{I_l} p(l_i \,|\, k)\, p^i_{l_\tau, t} \tag{5.58}$$

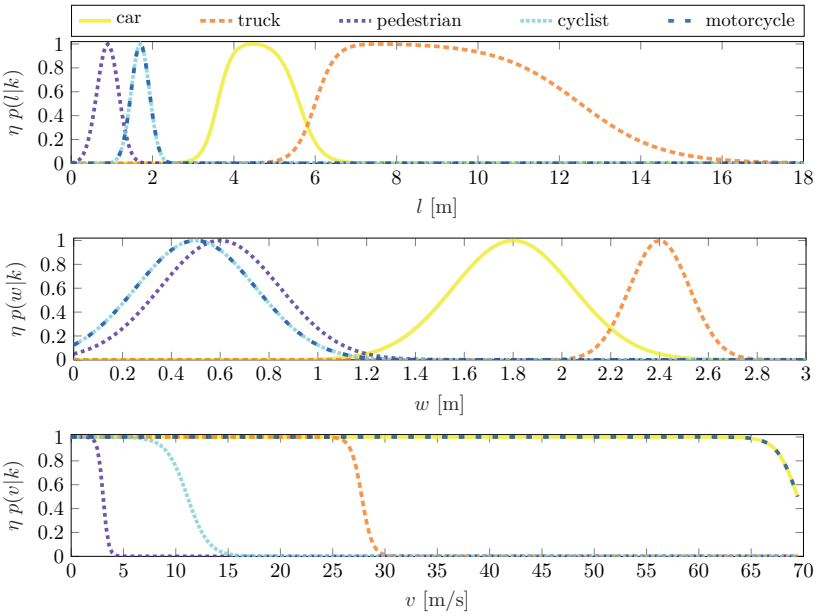


Fig. 5.9: Modeled likelihood $p(\cdot|k)$ of the features length l, width w, and velocity v given the individual classes $k \in \mathcal{K}$. The distributions are illustrated unnormalized with η indicating different normalization factors. Figure initially published in [1], © 2019 IEEE.

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 $E_W[\,p(W_\tau \,|\, k)\,]$ 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.,

$$v_\tau^{\max} = \max_{t'=1,\dots,t} (v_{\tau,t'})\,. \tag{5.59}$$

Overall, based on (5.57), the best fitting class k^*_τ of the track τ regarding the filtered geometry state and the maximum of the filtered velocity results in

$$\begin{aligned} k^*_\tau &= \arg\max_{k\in\mathcal{K}} E_{L,W}[\,p(k \,|\, L_\tau, W_\tau, v_\tau^{\max})\,] \\ &= \arg\max_{k\in\mathcal{K}} p(k)\, p(v_\tau^{\max} \,|\, k)\, E_{L,W}[\,p(L_\tau \,|\, k)\, p(W_\tau \,|\, k)\,] \\ &= \arg\max_{k\in\mathcal{K}} \left(p(k)\, p(v_\tau^{\max} \,|\, k) \sum_{i=1}^{I_l} p(l_i \,|\, k)\, p^i_{l_\tau,t} \sum_{i=1}^{I_w} p(w_i \,|\, k)\, p^i_{w_\tau,t} \right). \end{aligned} \tag{5.60}$$

This simple classification concept can further be combined with other extracted features, especially direct camera classification information, which is described in the following.

5.5.3 Combined Object Classification with Camera Information

One of the major strengths of camera sensors is their ability to perceive the semantic of scenes and especially the classification of objects. The measurement grid representation of this work also contains such camera classification information $\kappa^{\text{cam}}_{\text{class}}$, as introduced in (2.19), which has also been used before for improving the dynamic grid mapping as well as the object extraction and association. As derived in (5.60), the generic classification strategy enables a robust object classification even if no camera information is available, based on the features length, width, and velocity. But if the associated occupied cells $\mathcal{C}_{\tau,t}$ of a track include such additional camera information, the object classification should primarily use that camera information as input.

Generally, this corresponds to extending the Naive Bayes classifier of (5.53) by another feature, the camera classification $\kappa^{\text{cam}}_{\text{class}}$, i.e.,

$$p(k \,|\, l, w, v, \kappa^{\text{cam}}_{\text{class}}) \propto p(k \,|\, l, w, v)\, p(\kappa^{\text{cam}}_{\text{class}} \,|\, k) \tag{5.61}$$

which is also assumed to be conditionally independent of the other features. Equivalently, the best fitting class k^*_τ of a track τ regarding the maximum a posteriori classification of (5.60) extends in that case to

$$k^*_\tau = \underset{k \in \mathcal{K}}{\arg\max}\; p(k)\, p(v^{\max}_\tau \,|\, k)\, E_{L,W}[\, p(L_\tau \,|\, k)\, p(W_\tau \,|\, k)\,]\; p(\kappa^{\text{cam}}_{\text{class}} \,|\, k)\,, \tag{5.62}$$

i.e., by the likelihood $p(\kappa^{\text{cam}}_{\text{class}} \,|\, k)$ of the camera classification given the object class k. This likelihood can be modeled based on the characteristics and the influence of the camera to the classification. In the simplest case it corresponds to a binary assignment where only the camera class is supported and all other classes ($k \neq \kappa^{\text{cam}}_{\text{class}}$) are excluded, for example, $p(\kappa^{\text{cam}}_{\text{class}} = \text{pedestrian} \,|\, k = \text{car}) = 0$ when a car would never result in a camera classification of a pedestrian. However, the likelihood can also be selected to allow modeling the correlation or misclassification between similar classes, for example, $p(\kappa^{\text{cam}}_{\text{class}} = \text{motorcycle} \,|\, k = \text{cyclist}) > 0$ still allows the class cyclists, even if the camera classification has the class motorcycle.

Overall, this extended strategy represents a robust object classification approach that combines the advantages of the camera classification with the generic estimation based on the histogram filter geometry distribution as well as the maximum observed velocity of that track. In the following, this classification is further used to extract estimates of the real length and width of the bounding box shape of the track.

5.5.4 Extraction of Estimated Length and Width of Box Model

The histogram-based geometry distribution 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 used 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 (5.62) as the extended classification strategy with camera information if available. 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. 5.9 rather than extracting that observed minimum bounding box.

The finally extracted length l_τ^*, and equivalently the width w_τ^*, is determined by the length of the histogram bin with the highest combined probability

$$p_{\text{comb}}^{i,l} \propto p(l_i \mid k_\tau^*)\, p_{l_\tau,t}^{i}\,. \tag{5.63}$$

regarding the length likelihood $p(l_i \mid k_\tau^*)$ of the best class k_τ^* and the length histogram distribution $p_{l_\tau,t}^{i}$, i.e.,

$$l_\tau^* = l_{i_{l_\tau}^*}\,, \quad i_{l_\tau}^* = \underset{i=1,\dots,I_l}{\arg\max} \left\{ p_{\text{comb}}^{i,l} : p_{\text{comb}}^{i,l} > \left((1+\epsilon) \max_{j=1,\dots,i} p_{\text{comb}}^{j,l}\right)\right\}. \tag{5.64}$$

In the case of multiple lengths with similarly high probabilities, the smallest length is chosen by (5.64) 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 (5.63) rather than all classes weighted by their individual class probability as determined in (5.60), 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 $p_{\text{comb}}^{i,l}$, i.e.,

$$\bar{l}_{\tau,k^*} = \left(\sum_{i=1}^{I_l} p_{\text{comb}}^{i,l}\right)^{-1} \sum_{i=1}^{I_l} p_{\text{comb}}^{i,l}\, l_i\,. \tag{5.65}$$

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 using (5.64).

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, 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, cf. (4.15)–(4.18), which is selected more conservatively that way.

Overall, this object shape extraction strategy combines the histogram filter geometry distribution estimation of the length and width with the corresponding likelihood of the modeled classes, which, altogether, results in a reasonable object size estimation.

5.6 Results and Summary

This chapter has presented new approaches for a robust object state estimation, primarily based on the characteristics of dynamic occupancy grids. Thereby the associated occupied grid cells, as derived in the previous chapter, are abstracted to an oriented minimum bounding box model for each object instance. In addition, the freespace layer of the grid is

evaluated, where the current visibility of each edge of the measurement box is determined by the surrounding freespace. The position of the bounding box is thereby referred to the most visible reference point regarding the edge visibility, enabling a robust position estimation for scenarios with occlusions and changing sensor fields of view. Similarly, a generic object orientation optimization strategy has been proposed, where the freespace within the box is evaluated and minimized. Moreover, the particle-based cell velocity estimates of the dynamic grid estimation allow a direct velocity and orientation initialization on the object-level by forming the mean velocity vector of all associated cells. This particle movement direction, combined with the corresponding variance, further represents a confidence interval in which the freespace-based object orientation optimization is evaluated.

The object state estimation has been divided into two separate probabilistic filtering strategies. The dynamic state is estimated by a UKF using a common Gaussian state representation, where the UKF sigma points, in contrast to a standard Kalman filter, allow nonlinear transformations of the state transition and observation models. Together with the selected CTRA motion model, this enables tracking nonlinear movements including the combined estimation of the acceleration and turn rate. For a precise dynamic estimation, radar Doppler velocity measurements have been further integrated into the UKF approach. Thereby the projection of the UKF sigma points enables a direct update in the measurement space of the radial velocity components, i.e., no error-prone abstraction of the Doppler measurements toward the object state space is necessary that way.

The object shape is estimated differently based on a histogram filter, as it is not limited to Gaussian distributions. Hence, arbitrary distributions of the length and width can be modeled, allowing one to distinguish lower and upper bounds of the object size, which are differentiated by the freespace-based box edges visibility. Moreover, an object classification approach has been presented based on a Naive Bayes classifier. It primarily utilizes the camera classification if such information is available and associated to the track; but even without camera information, the object is generically classified based on the histogram distributions of the length and width as well as the maximum observed track velocity. The finally extracted object size combines the histogram geometry distributions with the classification and the corresponding class geometry likelihood, where the assumed object size is supported by prior class knowledge if it has not been fully observed yet.

Altogether, the presented approach thereby also combines the individual sensor strengths of lidar, radar, and camera for the object-level estimation. Lidar sensors with their high spatial accuracy primarily contribute to the object pose and shape estimation, since occupancy and freespace of the grid are mainly derived from lidar in this work; radar sensors primarily contribute to the motion estimation by their Doppler velocity measurements; and camera sensors primarily contribute to the object classification. Nevertheless, the proposed approach is still modeled generically, such that the full object state and classification can still be estimated solely based on occupancy measurements, without necessarily depending or fully relying on additional sensor-specific measurement data.

This chapter is concluded by some qualitative results in different urban scenarios that are demonstrated in Fig. 5.10. The upper scenario in Fig. 5.10a 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

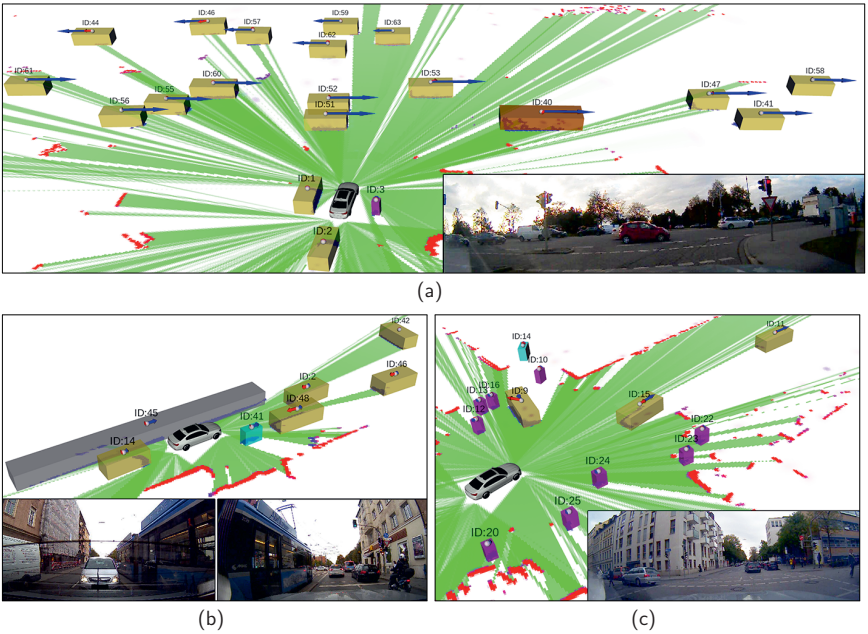


Fig. 5.10: Results of the object-level estimation in complex urban scenarios. Object classification color coding: car (yellow), truck (orange), pedestrian (purple), cyclist (cyan), motorcycle (blue), unknown (gray). Blue arrow: scalar object velocity; red arrow: object acceleration and yaw rate (transformed to lateral acceleration $a_{\text{lat}} = v \cdot \omega$). (a) Intersection with a lot of cross traffic. (b) An unknown object (tram) and surrounding vehicles. (c) Several surrounding pedestrians and a turning car. Figure initially published in [1], © 2019 IEEE.

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, 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. 5.10b shows a situation with a tram, which is classified as unknown as the large length and small width does not fit the modeled classes. 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. 5.10c includes several slow-moving pedestrians crossing the road as well as a turning vehicle, among others, which are all correctly extracted, tracked, and distinguished. Further qualitative and quantitative results of the overall grid-based object tracking approach are demonstrated in the next chapter.

6 Evaluation

The previous chapters have proposed novel methods for the dynamic grid estimation as well as the subsequent object extraction and tracking. Thereby first qualitative results have been demonstrated in combination with the respective summary of each chapter. In the following, a more detailed evaluation of the proposed grid-based object tracking approach is presented for various scenarios in the context of urban autonomous driving. First, an overview is given in Section 6.1, which addresses details of the sensor setup, the grid configuration and implementation, and a brief summary of the different processing steps of this work. Afterwards, the dynamic occupancy grid estimation is evaluated in Section 6.2. Results of the object-level estimation and tracking are then subsequently presented in more detail in Section 6.3. Finally, this chapter is summarized in Section 6.4 together with a brief discussion of the overall application of the proposed approach.

Parts of the evaluation have been presented in [1–4] in the context of this thesis, which are not explicitly referenced in the following.

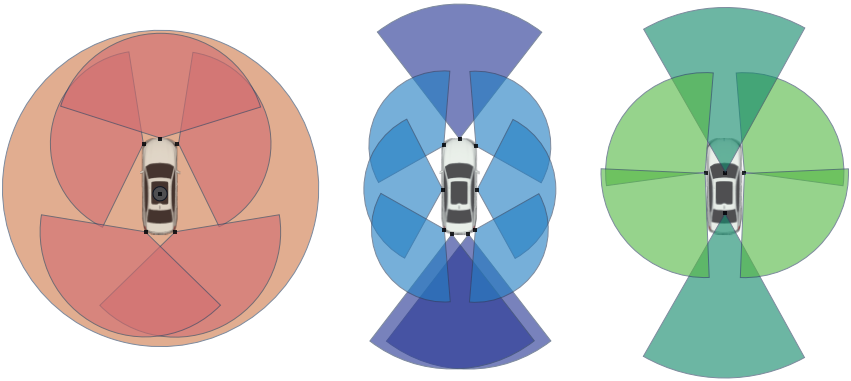
6.1 Overview

This section presents an overview of the sensor setup of the test vehicles, followed by a brief summary of the main processing steps of the proposed approach for one exemplary scenario and a discussion of the primary grid configuration and implementation.

6.1.1 Sensor Setup

The presented methods of this work have been tested with real sensor data in various traffic scenarios of multiple test vehicles during the last years. Thereby the most comprehensive sensor setup of a test vehicle is composed as follows:

- Lidar sensors:
  - 5 low-mounted lidar sensors with a low vertical field of view (FOV).
  - 1 high-mounted lidar sensor with a high vertical FOV and 360° horizontal FOV.
- Radar sensors:
  - 3 long-range radar sensors (1 to the front, 2 to the rear).
  - 6 short-range radar sensors (4 corners, 2 to the left/right side).
- Camera sensors:
  - 2 high-resolution cameras to the front and rear.
  - 4 wide-angle cameras at the side.



(a) Lidar. (b) Radar. (c) Camera.

(d) Test vehicle.

Fig. 6.1: Illustration of the sensor setup with an indication of the respective sensor field of views (not to scale). (a) Lidar sensors (red: low-mounted sensors, orange: high-mounted 360° sensor at the roof). (b) Radar sensors (light blue: short-range radar sensors, purple: long-range radar sensors). (c) Camera sensors. (d) Image of a BMW 7 Series test vehicle with the corresponding sensor setup (photos courtesy: BMW Group, part of an official video: https://twitter.com/i/status/1219211789822504961).

An illustration of that sensor setup including images of the corresponding test vehicle is depicted in Fig. 6.1. Altogether, this enables full 360° coverage with a multi-sensor fusion and partial redundancy of lidar, radar, and camera sensors. The maximum detection range of those sensors is about 150–200 m, with measurement rates between 16–25 Hz. Details of the two different low- and high-mounted lidar configurations have been presented before in Fig. 2.3. For the odometry estimation of the ego movement, the test vehicle further consists of high-precision inertial measurement units (IMU).

6.1.2 Main Processing Steps of this Work

Before this work is evaluated in more detail, the main processing steps of the proposed grid-based object tracking are briefly recapitulated in the following, which are further illustrated by an example in Fig. 6.2. As presented above in the sensor setup, different

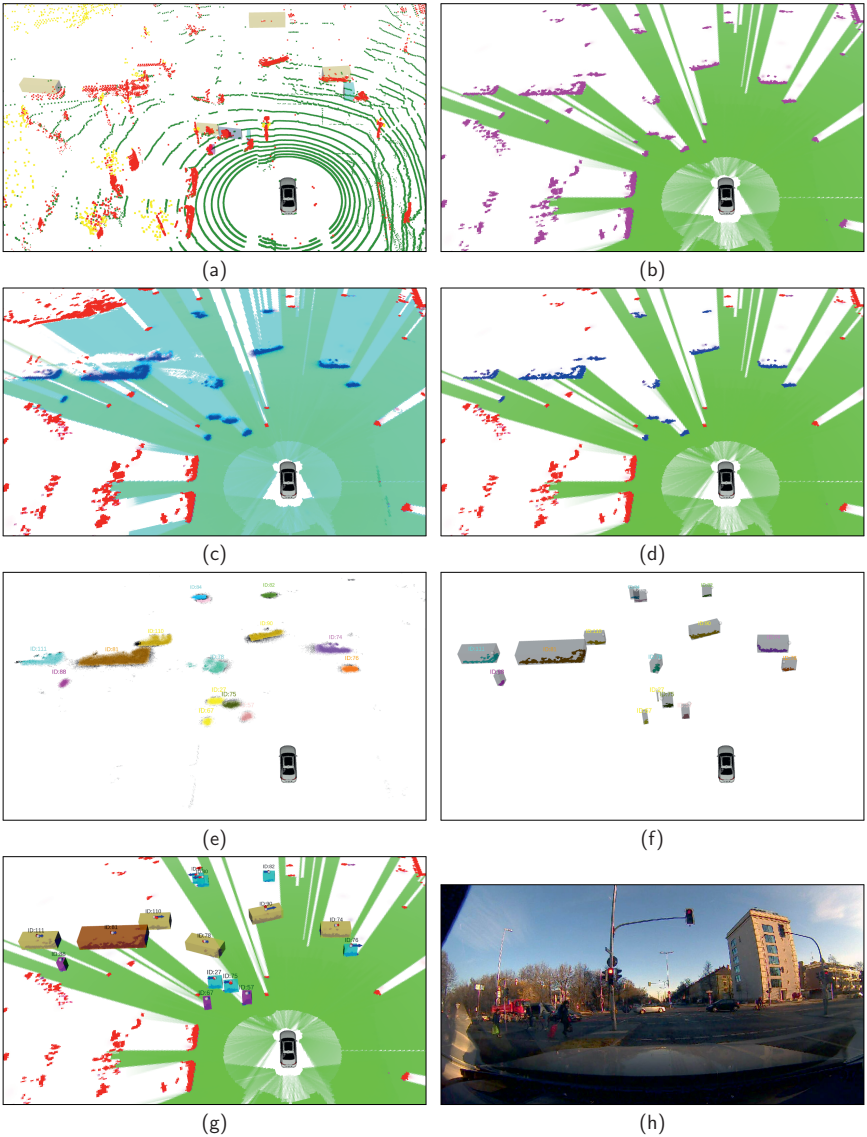


Fig. 6.2: Overview of the main processing steps of the proposed approach. (a) Raw measurement data of lidar, radar, camera. (b) Fused measurement grid. (c) Dynamic grid map. (d) Augmented measurement grid. (e) Labeled particle population. (f) Associated grid cells and measurement boxes. (g) Overall resulting object tracking. (h) Camera image of the scene.

measurement data of multiple lidar, radar, and camera sensors form the input of this work. These measurement data are illustrated in Fig. 6.2a, containing the range-based lidar and radar detections as well as the camera-based object instances. The measurement detections of each individual sensor are represented as a measurement occupancy grid using the evidential Dempster–Shafer representation with the two hypotheses of occupancy and freespace. All those sensor-individual measurement grids are fused cell-wise, with a temporal alignment of asynchronous sensor data, overall resulting in the fused measurement grid as depicted in Fig. 6.2b. Thereby the measurement grid representation additionally contains radar velocity measurements and camera classification information, which are not highlighted here in that example.

The fused measurement grid forms the input of the dynamic grid mapping, which performs the temporal filtering and accumulation of evidence masses that enables the differentiation of static and dynamic occupancy over time, as demonstrated in Fig. 6.2c. The low-level particle tracking thereby further enables a robust cell-wise velocity estimation. The augmented measurement grid representation of Fig. 6.2d combines the static/dynamic occupancy differentiation of Fig. 6.2c with the current measurement grid of Fig. 6.2b, resulting in unfiltered occupancy measurements that are only extended by the additional static/dynamic classification, avoiding multi-filtering for the subsequent object tracking.

Based on this generic low-level dynamic grid estimation, high-level object instances are extracted and tracked, where for moving objects only dynamic occupied cells have to be considered, as the static environment remains in the grid map representation. The association of measurement data to tracked object instances is thereby achieved using the particle labeling association, shown in Fig. 6.2e, where the low-level particle population is partially linked to object instances. Newly occurring objects are formed by a density-based clustering of dynamic occupied cells with similar cell velocities. The resulting association of occupied cells as well as the corresponding measurement minimum bounding boxes are illustrated in Fig. 6.2f. The finally resulting object tracking, including the object state estimation as well as the object classification, is summarized in Fig. 6.2g.

6.1.3 Primary Grid Configuration and Algorithm Implementation

In the following, the main configuration of the occupancy grid representation is briefly discussed, as it forms the core of the proposed environment perception. The fundamental concept of the grid structure and discretization with the fixed odometry coordinate system has been presented before in Fig. 2.1. This grid configuration can be adjusted in different ways, primarily by the grid resolution in terms of the total number of evaluated grid cells, the grid cell size in terms of the discretization size, and the offset of the ego vehicle within the grid structure. The selection of the grid configuration in turn generally depends on different aspects, for instance, the desired application and arising scenarios, the accuracy and range of the sensor setup, and the available computation power and runtime.

Most results of this work have been achieved with the configuration as shown in Fig. 6.3. In that configuration, the cell size is set to $0.15\,\mathrm{m} \times 0.15\,\mathrm{m}$ with a total of 1536×1536 cells, which is selected here as a multiple of $128 = 2^7$, i.e., $1536 = 12 \cdot 128$, for an efficiency-optimized memory allocation of the implementation. The selected cell size ensures a high spatial accuracy of obstacles as well as the pose and shape of object tracks, required for

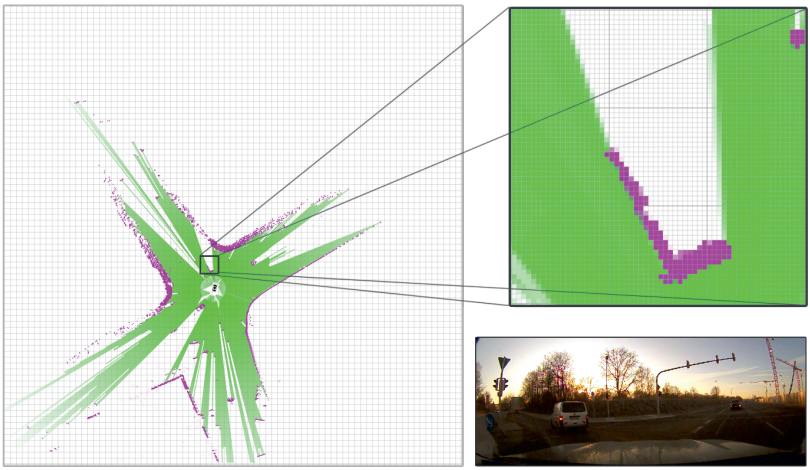


Fig. 6.3: Illustration of the grid configuration of the fused measurement grid. The local environment is discretized by a total of 1536×1536 grid cells with a cell size of $0.15\,\mathrm{m} \times 0.15\,\mathrm{m}$ and a radial position offset of the ego vehicle of about $30\,\mathrm{m}$ backwards from the grid center. The left part shows the full grid area (here shown with a raster size of 20×20 cells); the right part shows a section of the grid in detail where the individual grid cells are visible.

a precise path planning in narrow streets of urban environments with closely stopped or parked vehicles. But this also ensures a detailed movement estimation with a fine-granular tracking of occupancy position changes, while the cell size is still sufficiently large enough to remain robust against measurement noise and ensure the occupancy accumulation within one cell over time. Combined with the large amount of grid cells, this results in a total grid area of the local environment of $230.4\,\mathrm{m} \times 230.4\,\mathrm{m}$. The ego vehicle is thereby positioned with a radial offset of about $30\,\mathrm{m}$ backwards from the grid center. Altogether, this ensures a detection range with at least $100\,\mathrm{m}$ to the rear and $130\,\mathrm{m}$ to the front of the ego vehicle, where, depending on the orientation of the ego vehicle within the odometry coordinate system, the maximum possible range to the front in that configuration is about $\sqrt{2} \cdot 130\,\mathrm{m} \approx 184\,\mathrm{m}$ in case of a diagonal alignment.

For the measurement grid fusion, the roof-mounted lidar sensor is selected as the main trigger sensor due to the high spatial accuracy and the full 360° coverage. The measurement rate of this sensor is about $50\,\mathrm{ms}$, which thus also represents the update rate of the grid fusion and all subsequent processing steps of the dynamic grid mapping as well as the object tracking. Since the primary application of this work is an online real-time environment perception as part of autonomous driving test vehicles, the overall runtime is also crucial and has to be considered for the algorithm configuration, as opposed to offline post-processing strategies. Hence, to ensure that all relevant measurement data are permanently processed without skipping measurement update cycles, the overall runtime should be below the measurement rate of $50\,\mathrm{ms}$.

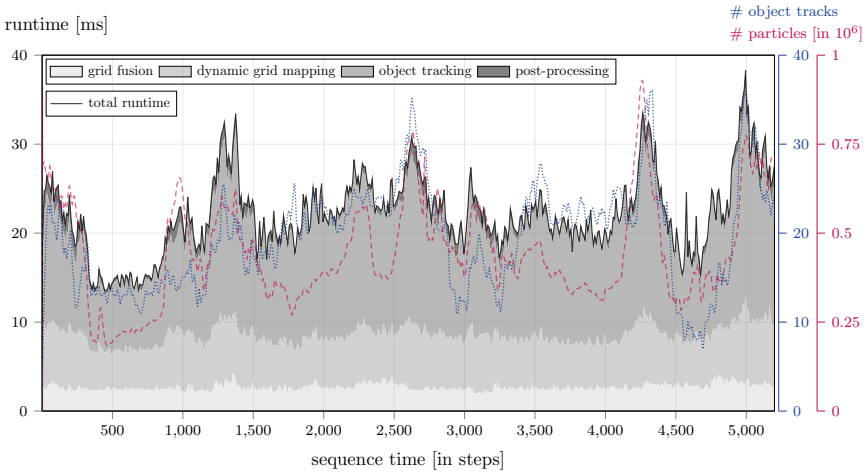


Fig. 6.4: Evaluation of the algorithm runtime for a longer sequence of a test drive in a highly dynamic urban environment with a varying number of occurring traffic participants, indicated by the number of extracted object tracks (blue dotted line) and low-level particles (red dashed line). Results are averaged over 10 time steps here. (Hardware specs: CPU: *Intel Xeon Gold 6134* [cores/threads: 8/16, 3.2 GHz]; GPU: *Nvidia GeForce GTX 1080 Ti*).

To achieve a fast algorithm computation, the implementation of this work is based on *C++*, where most parts are further optimized using *Nvidia CUDA* [89] for an efficient processing on the GPU that is highly parallelized, especially for the individual grid cells and low-level particles. The runtime is evaluated in Fig. 6.4 for a longer sequence of a test drive in a changing dynamic urban environment. This evaluation shows the algorithm runtime based on already derived measurement grids of the individual sensors. Hence, it is composed of the four following processing steps: the grid fusion of all incoming measurement grids, the dynamic grid mapping including the particle tracking, the object tracking including the object extraction and association, and a minor post-processing part that can be processed after the current output where primarily the relation between the object tracks and the particle labeling is updated. In that scenario, the full sensor setup with the grid configuration as discussed above is used. For the particle tracking, a maximum of $n_{\max} = 100$ particles per grid cell is selected, which is used for all results of this work.

The runtime evaluation is processed on a modern common development computer, details of the hardware are given in the caption of Fig. 6.4. Overall, the resulting runtime for that configuration is commonly between 15–30 ms, which also depends on the dynamics of the scenario, as indicated by the additionally shown number of currently tracked objects and low-level particles. But even for such highly dynamic scenarios with a large number of occurring traffic participants and thereby increased number of objects tracks and particles, the runtime is still significantly below the measurement update rate of 50 ms here. Thereby the hardware resources of the test vehicle prototype are even more powerful, but

also require processing several other autonomous driving tasks that have not been considered here. Altogether, those results demonstrate that the algorithm of this work, with the extensive grid configuration and respective sensor setup as discussed above, is efficiently implemented and generally real-time capable.

6.2 Dynamic Occupancy Grid Estimation

In this section, the dynamic grid mapping as proposed in Chapter 3 is evaluated in more detail. First, the accumulation over time is briefly presented by an example. Second, the proposed approach is compared to the approach of Tanzmeister et al. [120, 121], focusing on a scenario with partial occlusion of the road boundary. Third, the occupancy classification of this work is analyzed for a challenging highway scenario with a guardrail and a long truck, where different results with and without additional information of radar velocity measurement, camera classification information, and feedback of the subsequent object tracking are presented. Finally, the camera-based dynamic occupancy convergence is further evaluated for two more urban scenarios that contain stationary objects.

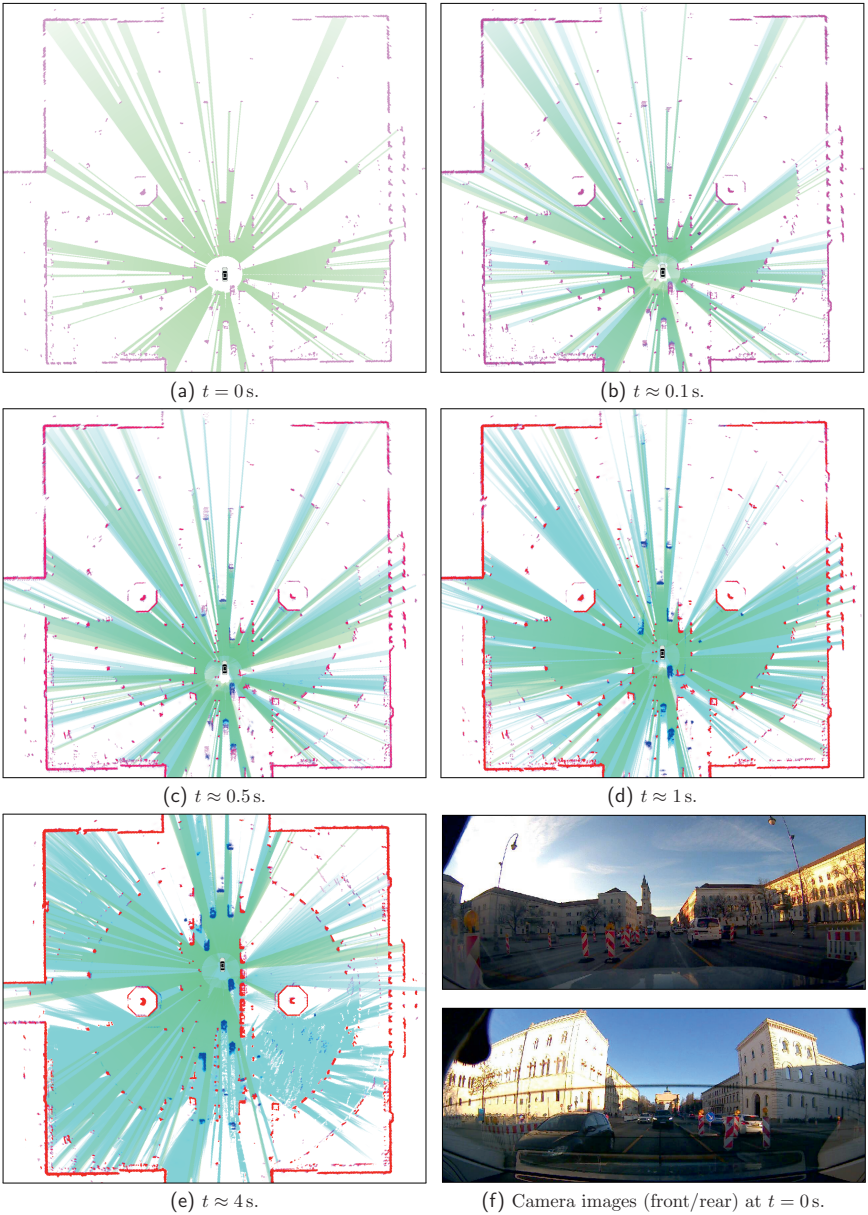
6.2.1 Accumulation over Time

The temporal evolution of the dynamic grid map is illustrated for an exemplary scenario in Fig. 6.5. In the initial update step, shown in Fig. 6.5a, the dynamic grid map corresponds to the fused measurement grid without any accumulation. Thereby no additional radar velocity or camera classification information is included here, hence, also no direct occupancy classification of that measurement grid is derived.

Note that in this work, to remain robust against measurement errors, the impact of the measured evidence masses of the fused measurement grid in the measurement update of the accumulated map is limited to the interval $[0, \eta_z]$ with $\eta_z \leq 1$ regulating the maximum amplitude of a measurement. However, the grid visualization of the measurement grid, and correspondingly the augmented measurement grid, is referred to the maximum amplitude without considering the actual impact of the dynamic grid map update. In this work, the maximum evidence mass of the measurement is limited to $\eta_z = 0.4$.

The dynamic grid map after the second measurement update is shown at Fig. 6.5b, already indicating the accumulation as well as the slight convergence of the static environment. Static and dynamic occupancy are clearly differentiated in this scenario at $t \approx 0.5\,\mathrm{s}$, as demonstrated in Fig. 6.5c. This is improved in Fig. 6.5d at $t \approx 1\,\mathrm{s}$, where areas of the moving vehicles fully converge toward dynamic occupancy, which is further enhanced here by extracted moving objects of the subsequent object tracking and the corresponding feedback to the dynamic grid mapping, as presented in Section 3.4.4. The accumulation over time with different fields of view, due to the moving ego vehicle, is further depicted in Fig. 6.5e, e.g., visible at the parked vehicles or the two fountains with the octagonal bases.

Overall, the proposed approach generally enables an accurate map accumulation of the local environment with a robust temporal filtering including a differentiation of static and dynamic occupancy as well as passable freespace. In the following, the dynamic grid mapping is evaluated in more detail in different scenarios.



(a) $t = 0\,\mathrm{s}$.

(b) $t \approx 0.1\,\mathrm{s}$.

(c) $t \approx 0.5\,\mathrm{s}$.

(d) $t \approx 1\,\mathrm{s}$.

(e) $t \approx 4\,\mathrm{s}$.

(f) Camera images (front/rear) at $t = 0\,\mathrm{s}$.

Fig. 6.5: Dynamic map accumulation with static/dynamic occupancy differentiation over time.

6.2.2 Comparison with Original Approach

The proposed dynamic grid mapping approach including the particle tracking concept is primarily based on the work of Tanzmeister et al. [120, 121], denoted as evidential *Grid-based Tracking And Mapping* (GTAM), as briefly presented in Section 3.1.1. In the following, both approaches are compared in a scenario with multiple fast-moving vehicles that temporarily occlude parts of the static environment – such scenarios with temporary occlusion depicted the main motivation for further enhancing the GTAM approach regarding the information loss in that case. This comparison has also been presented as part of [4], where, as noted before, the main concept of the proposed dynamic grid mapping has been originally published as part of this thesis.

The following scenario is evaluated without radar Doppler velocity measurements, camera information, or feedback of the subsequent object tracking, i.e., only occupancy measurements of lidar sensors are used as input data to highlight more clearly the different occupancy filtering and classification of both approaches solely based on occupancy measurements. Moreover, the sensor setup includes only the low-mounted lidar sensors without the high-mounted roof lidar sensor, thereby further highlighting the effect of partial occlusion for objects moving closely to the ego vehicle.

A qualitative comparison of the resulting evidential grids at a fixed time is demonstrated in Fig. 6.6. The fused measurement grid, representing the input to both approaches, is visualized in Fig. 6.6a. On both the left and the right side of the ego vehicle there is one car each in that scenario, both moving at a similar speed as the ego vehicle, which temporarily occlude parts of the road boundaries behind them. The proposed evidential dynamic grid map of this work is shown in Fig. 6.6b. Static and dynamic occupancy are robustly filtered and correctly distinguished. The occluded area of the road boundary persistently remains as static occupancy, since it is accumulated and temporally filtered in the grid map representation. Accordingly, the corresponding augmented measurement grid, as demonstrated in Fig. 6.6c, represents correct occupancy classification results.

In contrast, the GTAM particle map in Fig. 6.6d, representing the particle-based occupancy classification, fails in such a scenario and points out a significant systematic drawback of that approach. Although particles at the road boundaries show a slow but correct convergence behavior from wrong dynamic occupancy to static occupancy, the edges of the temporarily occluded road boundaries are classified incorrect, as highlighted by the dotted areas. The correctly filtered static information is lost when static particles are destructed due to occlusion, which then results in a wrong dynamic classification again when the area is not occluded anymore. Although the output of the particle map including static occupancy is accumulated in the DST map representation in Fig. 6.6e, which remains when particles are destructed, it is not used as a feedback to the particle filtering.

This behavior has also been observed in urban scenarios where the sensor field of view can change quickly due to the ego motion and obstacles passing closely to the sensors. For example, similar results occur at house walls with densely parked vehicles in front, where gaps between those near obstacles cause a fast-changing sensor field of view and thus fast changing occlusions of the areas behind.

Another decisive advantage of the proposed approach is the significantly reduced number of used particles, as the static environment is directly modeled in the map without requiring

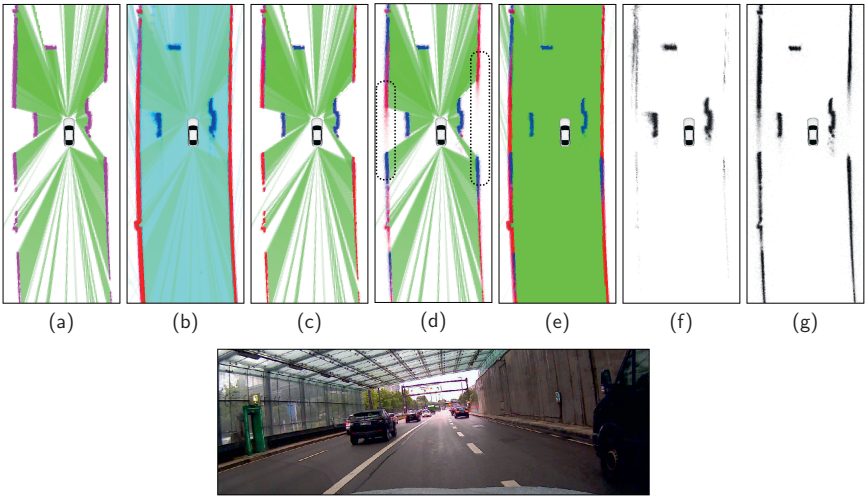


Fig. 6.6: Comparison of the dynamic grid estimation of the proposed approach and [121] in a scenario with guardrails temporary occluded by vehicles. The ego vehicle and three vehicles on its left, right, and front left are moving north. Input data are exclusively from lidar measurements. (a) Fused measurement grid. (b) Dynamic map. (c) Augmented measurement grid. (d) Particle map of [121], not limited to measurement grid evidences here to visualize particle destruction. The dotted areas highlight the critical destruction of static particles due to occlusion. (e) DST map of [121]. (f) Particle population of proposed approach. (g) Particle population of [121]. (h) Camera image. Figure initially published in [4], © 2018 IEEE.

static particles in contrast to [121], as demonstrated in Fig. 6.6f–6.6g. In that exemplary scene, using only the low-mounted lidar sensors with a grid configuration of 680×680 cells, a cell size of $0.2\,\mathrm{m} \times 0.2\,\mathrm{m}$, and a maximum of $n_{\max} = 100$ particles per cell, as evaluated in [4], the population of the entire grid consists of 48.787 particles, whereas the GTAM approach uses 158.719 particles, i.e., more than three times as many particles in that case.

A further advantage of the proposed dynamic map in comparison to the DST map representation is the introduced hypothesis of $\{F, D\}$. Hence, passable area $\{F, D\}$ that may be temporarily occupied and actual current freespace $\{F\}$ are modeled separately, which can be used to analyze the current occlusion of a traffic scenario.

The evidential filtering of the scenario of Fig. 6.6 over time for one fixed cell of the right road boundary is further analyzed in Fig. 6.7. This illustrates in detail the different consequences of the temporary occlusion, here at $t \in [70, 91]$, of a static cell caused by the right vehicle as described above. Note that in this work, as discussed before, the maximum evidence mass of the fused measurement grid, and correspondingly the augmented measurement grid as well as the particle map of [121], is within the interval $[0, \eta_z]$ and thus

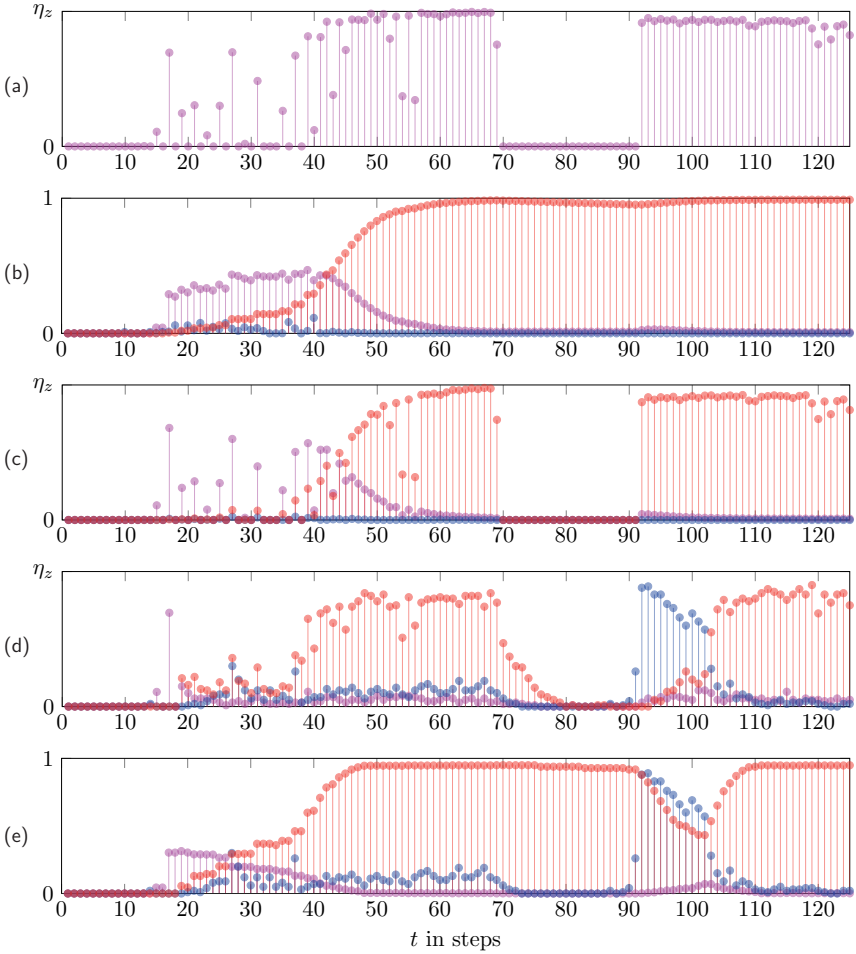


Fig. 6.7: Temporal filtering of occupancy evidence masses at a static cell on the right guardrail of the scenario of Fig. 6.6. Occupancy masses are visualized as unstacked stems using the same color coding as for the grid visualization, i.e., pink $\widehat{=}\ m(SD)$, red $\widehat{=}\ m(S)$, and blue $\widehat{=}\ m(D)$. The measurement grid impact is $\eta_z = 0.4$. (a) Fused measurement grid. (b) Proposed dynamic map. (c) Augmented measurement grid. (d) Particle map of [121]. (e) DST map of [121], blue shows not limited dynamic evidence of particle map without normalization here. Figure initially published in [4], © 2018 IEEE.

limited by the factor $\eta_z \leq 1$, whereas the grid visualization is referred to the maximum amplitude without considering the actual impact.

The comparison of both approaches for the scenario of Fig. 6.6 is finally evaluated quantitatively in terms of the static/dynamic occupancy classification and the number of required particles. A binary classification

$$\Lambda = \begin{cases} \{S\}, & \text{if } m(S_t) \geq \psi_\Lambda \\ \{D\}, & \text{else} \end{cases} \tag{6.1}$$

is performed for all relevant occupied cells of the fused measurement grid that have at least an occupancy mass of $m(SD_{\Lambda,\min})$, i.e., $m(SD_{z,t}) \geq m(SD_{\Lambda,\min})$. This classification, with a variation of the threshold $\psi_\Lambda \in [0, 1]$, is evaluated with labeled ground truth data, resulting in a receiver operating characteristic (ROC) curve. The results of the proposed approach in comparison to [121] for a small sequence of 250 frames of the scenario of Fig. 6.6 are demonstrated in Fig. 6.8a. This quantitative evaluation confirms the qualitatively shown results. Especially for scenarios with temporary occlusion of road boundaries, the false dynamic classification is significantly reduced by the proposed architecture in comparison to [121].

The number of required particles $|\mathcal{X}_t|$ of the evaluated sequence is shown in Fig. 6.8b. At initialization, both approaches create new particles proportional to the sum of measured occupancy, indicated by the pink dotted line. The proposed approach uses the filtered occupancy evidence of the map, thus the impact of one measurement is limited to the interval $[0, \eta_z]$, here $\eta_z = 0.4$ is used, as described above. The approach of [121], in contrast, directly initializes particles proportional to the not limited measured occupancy evidence. After the initialization phase, the proposed approach significantly reduces the number of particles, as the static environment converges toward $\{S\}$ which is modeled without particles. Since the ego vehicle is moving forward in that scenario, new cells of unclassified occupancy are detected every step. Around $t \approx 100$, the detected occupied area enlarges significantly, and new dynamic objects occur in the sensor field of view, indicated by the dashed blue line that corresponds to $n_{\max}$ times the number of ground truth dynamic grid cells of the fused measurement grid. Hence, the number of particles increases for both approaches. The number of particles of [121] remains proportional to the measured occupancy evidence with a large amount of required static particles. Overall, in comparison to [121], the proposed approach thereby requires significantly less particles to accurately represent the local environment, in that scenario approximately only a third, which thus also significantly reduces the corresponding memory and computation resources.

Altogether, this comparison has demonstrated that the proposed approach enables a more robust and efficient dynamic grid mapping than the original work of [121]. Due to the persistent map accumulation rather than using short-term static particles, information loss is avoided for scenarios with temporary occlusion of the static environment, which significantly increases the static/dynamic occupancy classification, while at the same time the number of required particles is significantly reduced. In this comparison, no additional information of radar Doppler measurements, camera information, or feedback of the object tracking has been used to focus on the fundamental filtering differences of both approaches. In the following, it is discussed how such additional information further improve the occupancy classification of the proposed dynamic grid mapping approach.

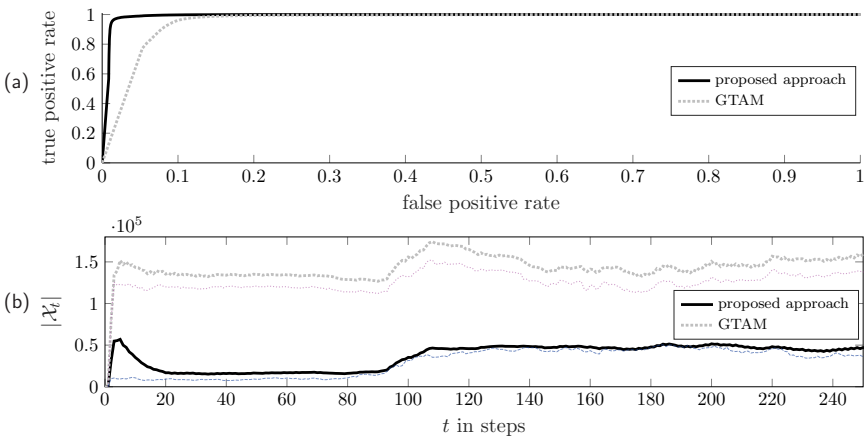


Fig. 6.8: Quantitative comparison of the proposed approach and GTAM [121] for the scenario of Fig. 6.6. (a) Receiver operating characteristic with a variation of the classification threshold ψ_Λ. Positive denotes the binary classification of dynamic here. (b) Number of total particles $|\mathcal{X}_t|$ with $n_{\mathsf{max}} = 100$ particles per cell. Pink dotted line indicates $n_{\mathsf{max}} \sum m(SD_{z,t})$ without η_z, blue dashed line indicates number of ground truth dynamic cells of the fused measurement grid times n_{max}. Figure initially published in [4], © 2018 IEEE.

6.2.3 Occupancy Classification with Additional Information

As discussed so far, the dynamic grid mapping generally converges toward static or dynamic occupancy over time, even for occupancy measurements without additional measurement information, solely based on the adapted evidential occupancy filtering as presented in Section 3.4.2. Thereby the temporal filtering is designed such that occupied grid cells converge toward static occupancy when multiple occupancy measurements occur at the same position over time, whereas they converge toward dynamic occupancy when they are newly occurring and correspond with the predicted low-level particle population, i.e., when the movement is properly tracked by the particle filtering. However, there exist scenarios in which the movement and thus the static/dynamic occupancy differentiation cannot be properly solved for the individual grid cells without any further information, which is discussed in the following before the extension with additional information is presented.

Critical Scenarios of Occupancy Classification without Additional Information

A complex highway scenario that is particularly challenging for the occupancy classification is demonstrated in Fig. 6.9. Generally, as shown in Fig. 6.9a, static and dynamic parts of the environment are properly accumulated and distinguished here, even though the detection area is quite large and includes several fast-moving vehicles that are also partly occluded here. However, when focusing on the dashed image section, as shown enlarged in Fig. 6.9b, this scenario also highlights the two prominent problem cases of any dynamic

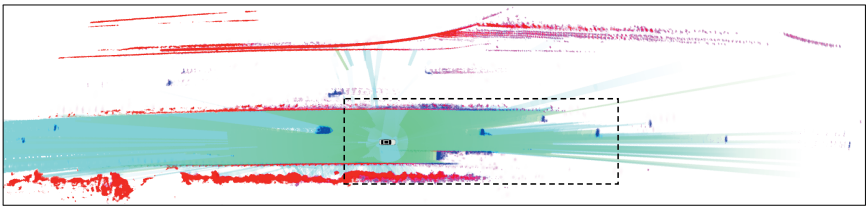
grid mapping approach – newly observed parts of straight guardrails of the static environment and an observed long side of moving trucks. From the cell-individual perspective, those two cases cannot be really distinguished, as discussed in the following.

On the one hand, both cases exhibit a (pseudo) movement of newly occurring occupancy in the front edge area that corresponds to the ego-speed and the respective movement of the sensors field of view, thus converging toward dynamic occupancy. In that scenario, the newly observed parts of the static environment are further limited by occlusion of another moving vehicle, thus additionally influenced by the relative movement of that vehicle.

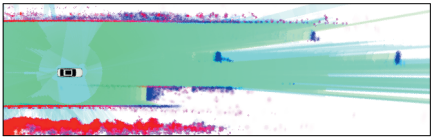
On the other hand, an initially newly occurring occupied cell of both cases then remains occupied at that fixed position, with respect to the fixed odometry coordinate system, for multiple subsequent measurements and thus converges toward static occupancy. That can be caused either by the fact that the cell actually results from the fixed static environment, but also by different parts of the moving truck, especially the observed long left side of the truck in that scenario, where the movement of the truck across that fixed cell remains temporarily occupied for several time instances. As an example, with an object length of 20 m and a velocity of 20 m/s = 72 km/h, a fixed cell can be occupied for 1 s; with a measurement update rate of 20 Hz this corresponds to an occupancy accumulation of 20 cycles.

A reasonable trade-off between the wrong dynamic estimation of the guardrail and the wrong convergence of the moving truck toward static occupancy is shown in Fig. 6.9b; depicting the relevant section of the enlarged image of Fig. 6.9a. In that case the evidential occupancy filtering as proposed in (3.38) is used without modifications. For the parameter γ_D that models the uncertainty of the implicit derivation of dynamic occupancy $m(D_t)$ based on the accumulated passable area evidence $m(\overline{FD}_t)$ of previously measured freespace and currently measured occupancy $m(SD_{z,t})$, as introduced in (3.37), a value of $\gamma_D = 0.7$ is selected. This also represents the default value of this work based on the characteristics of the equipped sensors and the respective measurement models, overall resulting in a robust behavior in various scenarios, also in the case of improperly derived freespace across static obstacles, as it is slightly the case in that scenario for far away parts of the small guardrail.

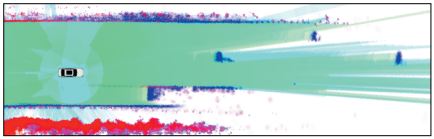
The convergence of the proposed dynamic grid mapping can also be adapted to achieve a faster convergence toward static or dynamic occupancy, as demonstrated in Fig. 6.9b–6.9g. In Fig. 6.9c, the case without the introduced uncertainty parameter γ_D is depicted, i.e., $\gamma_D = 0$, generally resulting in a faster dynamic convergence if evidence of passable area has been accumulated before for those cells. However, this does not directly improve the result of the large truck here, where, based on the sensors field of view of the ego vehicle, no freespace was observable in the past in front of the truck, which just completed a lane change to the right in that sequence. A slight improvement is noticeable for the preceding car, one lane left to the ego vehicle, which moves across previously accumulated passable area and thereby results in full convergence toward dynamic occupancy here. However, that parametrization also results in a more error-prone dynamic occupancy estimation of the static guardrail due to inaccuracies of derived freespace. The opposite extreme case of this parameter with $\gamma_D = 1$ is shown in Fig. 6.9d, avoiding any implicit derivation of dynamic occupancy for previously accumulated passable area, resulting in a better convergence of static occupancy in the case of wrongly derived freespace, but also a reduced convergence toward dynamic occupancy for correctly accumulated freespace. The case of Fig. 6.9d is



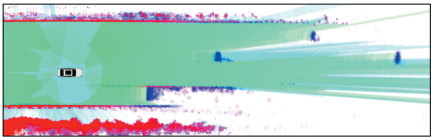
(a) Dynamic grid map (without additional information) with default parametrization.



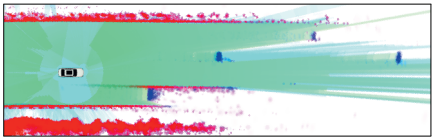
(b) $\gamma_D = 0.7$ (image section of (a)).



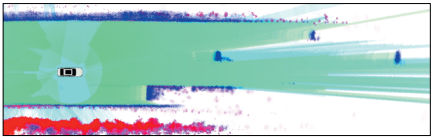
(c) $\gamma_D = 0$.



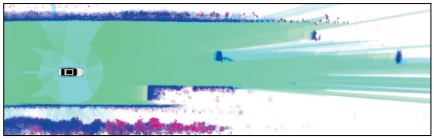
(d) $\gamma_D = 1$.



(e) $\gamma_D = 1$, f_D selected linear.



(f) Prediction conflict assigned to SD.



(g) Prediction conflict assigned to D.

(h) Camera image of the scenario.

Fig. 6.9: Evidential dynamic grid mapping and respective occupancy classification in a highway scenario with a straight guardrail and a long truck. This example shows the issues of the cell-individual static/dynamic classification for those two cases without using any additional information. The overall dynamic grid mapping result with the default parametrization is shown in (a), the critical section (dashed rectangle) is highlighted in (b); while results with different modifications of the occupancy filtering are shown in (c)–(g).

further adjusted in Fig. 6.9e by selecting the factor f_D for the convergence toward dynamic based on the number of particles, as introduced in (3.35), linearly instead of the square root, resulting in a very conservative estimation of dynamic occupancy and a fast convergence toward static occupancy, for both the guardrail and the long side of the moving truck.

As discussed before, the proposed occupancy filtering is generally designed such that an occupied cell converges toward static with multiple occupancy measurements occurring at that fixed cell over time. This is primarily ensured by the dominance of static occupancy over the particle-based dynamic occupancy in the prediction step, where static occupancy remains unchanged by assigning the corresponding prediction conflict $\overline{\zeta}$ toward $m(\overline{S}_t)$ as discussed in (3.20). Hence, the convergence of static occupancy is significantly reduced, with a respective enhanced convergence toward dynamic occupancy, when that prediction conflict is assigned to $m(\overline{SD}_t)$, shown in Fig. 6.9f, or even directly to the dynamic mass $m(\overline{D}_t)$ as depicted in Fig. 6.9g. This represents the extreme case toward dynamic occupancy where no static artifacts occur at the truck, however, it clearly limits the robustness of the accumulation of the static environment or even completely suppresses it that way.

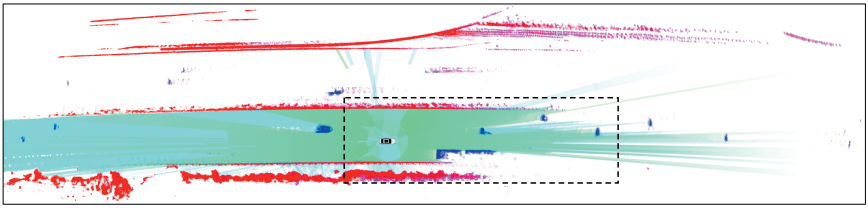
Altogether, the static/dynamic occupancy filtering can be adjusted in different ways; however, for the cell-wise estimation solely based on occupancy measurements, it is not possible to fully distinguish the two cases of guardrails and the long side of moving trucks properly without additional information. From the cell-individual perspective, both exhibit the same characteristics of newly occurring occupancy at the edge of the sensor field of view and multiple subsequent occupancy measurements at a fixed cell.

Improved Occupancy Classification with Additional Information

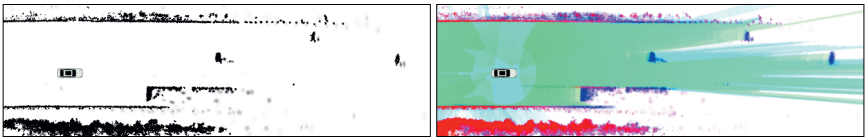
In the following, the same scenario as in Fig. 6.9 is evaluated, but using additional information for the dynamic grid mapping, based on radar Doppler velocity measurements, camera classification information, and using the feedback of the subsequent object tracking. This significantly enhances the dynamic grid mapping for the critical parts of the guardrail as well as the long truck, as presented in Fig. 6.10. The resulting dynamic grid map based on the combination of all additional sources is depicted in Fig. 6.10a, details are presented in Fig. 6.10b–6.10f, showing the enlarged relevant image section as before. The left column thereby further illustrates the occupancy classification coefficients based on the respective additional information of the latest measurement update, i.e., not accumulated over time. For comparison, the result of without any additional information is shown again in Fig. 6.10b, which is equivalent to Fig. 6.9b.

The dynamic grid map with additional radar Doppler velocity measurements, with the respective occupancy classification coefficients for static β_S^{radar} and dynamic β_D^{radar} as introduced in (3.41)–(3.42), is highlighted in Fig. 6.10c. Based on those direct velocity measurements, the wrong pseudo movement of the guardrails as well as the wrong static convergence of the moving truck can be directly solved, eliminating almost all wrong occupancy classification artifacts. Only some minor errors remain, especially slight dynamic occupancy for the guardrails on both the left and the right side, since the respective radar measurement data are not fully dense and also contain some measurement errors.

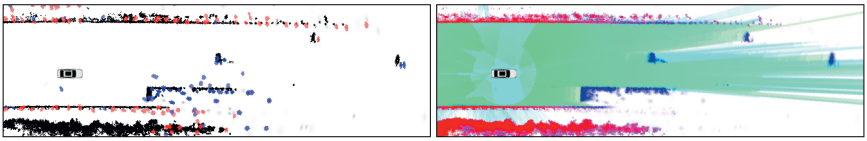
Results based on additional camera classification information are shown in Fig. 6.10d. Note that for the camera classification, the static environment is based on the detection



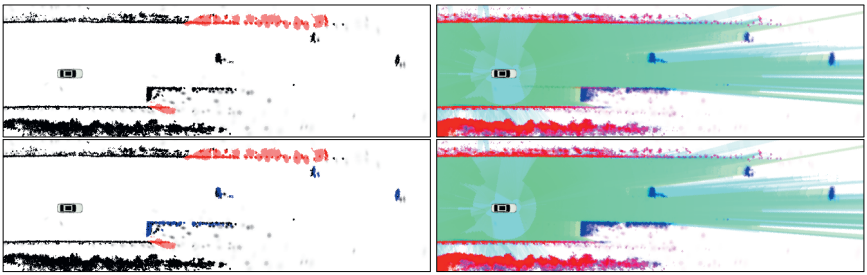
(a) Dynamic grid map with additional information of radar, camera, and object tracking.



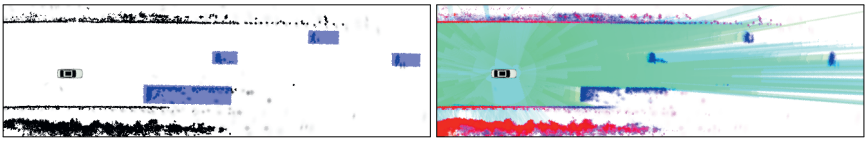
(b) Results without additional information (as in Fig. 6.9b).



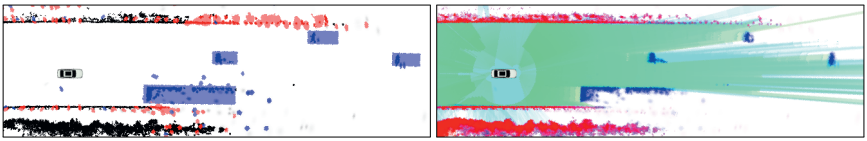
(c) Results with radar velocity information.



(d) Results with camera classification (top: only static obstacles, bottom: camera obstacles+objects).



(e) Results with feedback of subsequent object tracking.



(f) Results with combination of radar, camera, and object tracking information (image section of (a)).

Fig. 6.10: Enhanced dynamic grid mapping with additional information (same scenario as in Fig. 6.9).

representation, whereas dynamic parts are based on the object instances with the additional spatial alignment with lidar data, as presented in Section 2.3.4. Two different results of the camera data are shown in Fig. 6.10d – the upper row results from using only the static classification with the coefficient β_S^{camera}, the lower row includes also the dynamic classification with β_D^{camera} of camera-detected vehicles. As discussed with the definition of β_D^{camera} in (3.44), that coefficient can be modeled differently depending on the class of the camera object, where vehicles are a special case that can be modeled to converge toward dynamic, but do not necessarily have to be, as in the case of parked vehicles in urban environments, which can also be modeled as part of the static environment. This is further discussed in the next scenario. If no camera object movement information is directly available for such vehicles, this can also be modeled depending on the occurring scenarios, possibly automatically adapted by the current speed of the ego vehicle, such that for highway-like scenarios as in Fig. 6.10, where no parked vehicles are expected, all camera-detected objects should converge toward dynamic occupancy and therefore be extracted in the subsequent object tracking. Overall, the camera classification significantly improves the dynamic grid mapping in that scenario. The static environment is improved by the camera obstacle detections that are only available in the field of view of the front camera, but thereby highly depends on the respective camera distance estimation. Dynamic parts are improved by the camera object instances, if the respective object class is selected to be modeled as dynamic, which further depends on the spatial camera–lidar alignment.

Another possibility for improving the dynamic occupancy convergence is to utilize information of the subsequent object tracking as feedback to the dynamic grid mapping, as presented in Section 3.4.4. This is demonstrated in Fig. 6.10e, where all occurring moving objects are already correctly tracked as demonstrated by the respective bounding box shapes used for the feedback that are shown in the left image. That way, the movement is thereby further evaluated on the object-level in addition to the estimation of the individual grid cells, which enables to suppress all static artifacts of moving objects, especially in the case of the long truck. Moreover, this strategy further allows one to keep track of objects in urban environments that temporarily stop at a traffic light. For a more conservative approach, the feedback where dynamic occupancy is retained can thereby be limited to those object tracks that are quite certain and have been seen by multiple sensors and thereby be confirmed by radar Doppler measurements or corresponding camera object instances. The static environment remains unchanged that way, therefore, the guardrail of Fig. 6.10e still exhibits wrong dynamic occupancy as before in Fig. 6.10b.

Finally, the combination of those different inputs, as summarized in (3.53), is highlighted in Fig. 6.10f, which depicts the most comprehensive results. Altogether, even though the proposed dynamic grid mapping generally achieves promising results without additional information, some scenarios such as newly occurring parts of straight guardrails and the long side of trucks cannot be fully solved from the cell-individual perspective and therefore require further information. Each of the presented additional inputs thereby significantly improves the resulting dynamic grid map for such scenarios, but, by that combination, the approach is further robust against partly missing or wrong sensor data and thereby utilizes the different advantages of the radar velocity measurements, the camera classification, and the feedback of the subsequent object tracking.

Camera-Based Dynamic Occupancy Convergence and Object Extraction

In the following, the influence of camera object instances as additional classification information for the dynamic convergence of the dynamic grid mapping as well as the subsequent object extraction is addressed. As discussed for the scenario of Fig. 6.10d, the classification information of camera-detected objects can optionally be used to enforce an occupancy classification convergence toward dynamic, especially in highway scenarios. For urban environments, however, camera-detected vehicles can generally also be permanently parked, in which case an extraction of object tracks is not absolutely necessary, since the representation of static occupied cells generally enables a more accurate position and shape representation of obstacles by the persistent grid map accumulation, which further is more efficient since no particles or object track instances are required.

Such an example with a large number of parked vehicles is presented in Fig. 6.11. Thereby Fig. 6.11a represents the case where all camera-detected objects are included in the group of dynamic classes $\mathcal{K}_{\text{dynamic}}$ that are modeled by an occupancy classification coefficient of $\beta_D^{\text{camera}} = 1$ here, as defined in (3.44). As a result, all occupied cell that contain camera object information converge toward dynamic occupancy in the dynamic grid map, even if those objects are not moving. The corresponding object extraction is demonstrated in Fig. 6.11b, with a total of 25 vehicles (cars and trucks/vans) that are all parked here, in addition to 19 moving vulnerable road users (VRUs), i.e., pedestrians and cyclists. The respective object boxes class color coding has been introduced in Fig. 5.10. Hence, object tracks are correctly extracted for all stationary vehicles here. However, the object track estimation of non-moving objects fully relies on the camera information, not only for the object occurrence itself, but also for the initialization of the orientation estimation, as no robust particle-based movement direction estimation is possible in that case.

A different configuration of that scenario is shown in Fig. 6.11c, where camera-detected vehicles are excluded from the group $\mathcal{K}_{\text{dynamic}}$ and thus have no effect on the occupancy classification. Therefore, all parked vehicles properly converge toward static occupancy in the grid map, in which case the grid cell representation can be directly used for an accurate collision avoidance. Nevertheless, camera-detected pedestrians and cyclists are retained in the group $\mathcal{K}_{\text{dynamic}}$ and are thus always extracted as object tracks, since, even if they are currently not moving, that category of traffic participants is generally highly dynamic and vulnerable. Hence, that additional information about the presence of surrounding VRUs is directly safety-relevant for autonomous driving applications, where the speed and safety distance have to be selected even more conservatively than to stationary obstacles.

Such a scenario is highlighted in Fig. 6.12, where a pedestrian with a bike stands close to the road, waiting at pedestrian lights. Without any camera information, solely based on range-based measurements, that pedestrian is modeled as part of the static environment without any additional object information, as shown in Fig. 6.12a–6.12b. Nonetheless, the non-moving pedestrian is still robustly detected as an obstacle that way, as all measurements are considered and consistently accumulated in the dynamic grid map. But by utilizing additional camera information, the dynamic grid mapping is adapted such that the occupied cells of the pedestrian converge toward dynamic occupancy and an object track with the respective classification is properly extracted, as presented in Fig. 6.12c–6.12d.

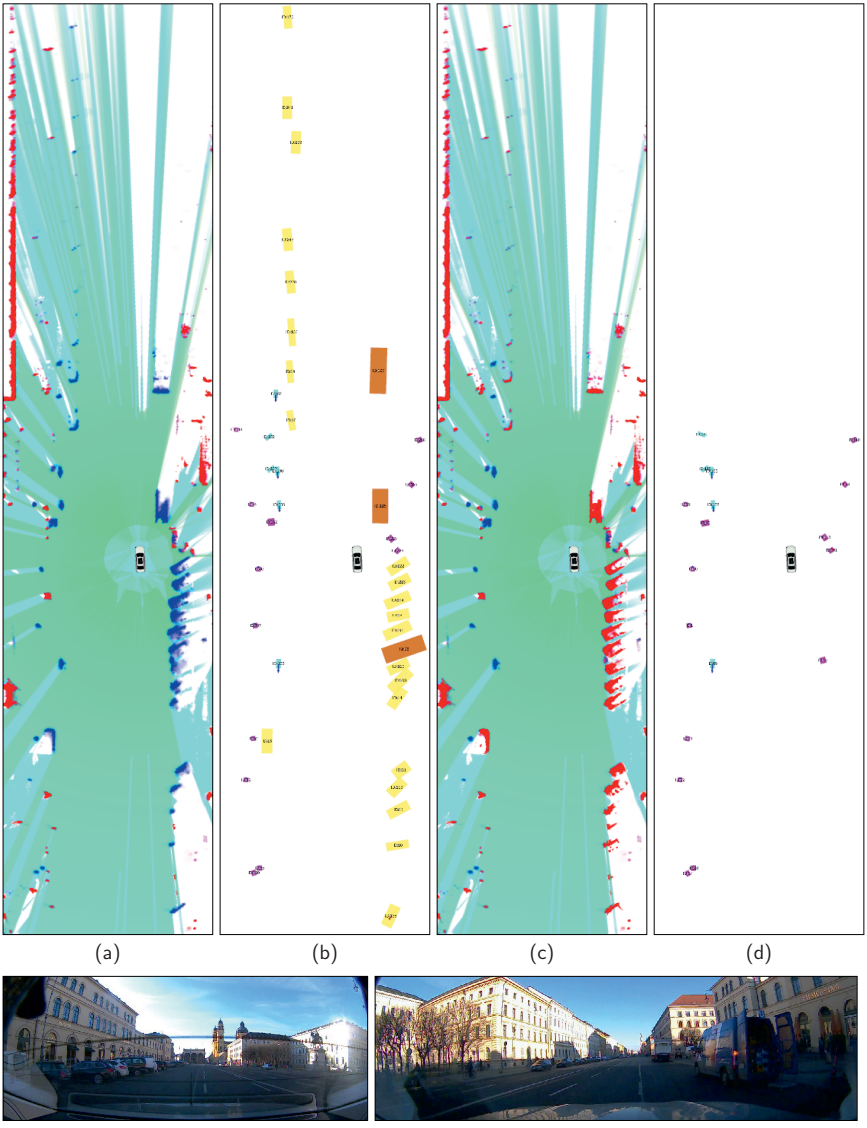


Fig. 6.11: Influence of the camera classification for the dynamic occupancy convergence and subsequent object extraction. (a/b) Dynamic convergence for all camera-detected object types. (c/d) Dynamic convergence only for camera-detected VRUs. (e/f) Camera images.

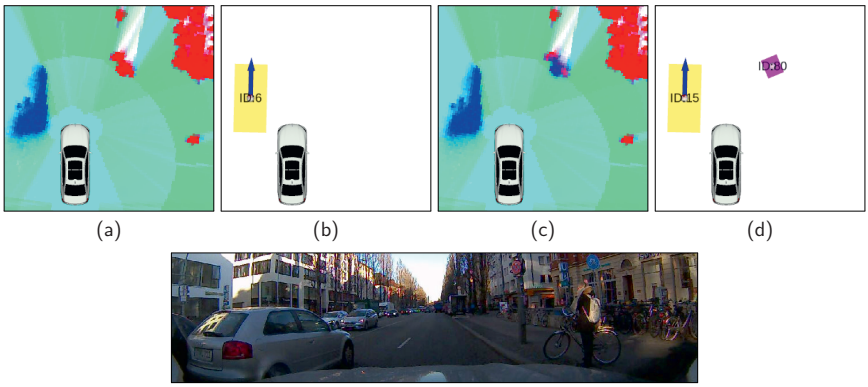


Fig. 6.12: Influence of the camera classification for the estimation of stationary VRUs. On the right side there is a pedestrian with a bike standing directly next to the road. (a/b) Results without any camera information. (c/d) Results with camera information. (e) Camera image.

Altogether, this section has evaluated different aspects of the dynamic grid mapping. This work has been compared to the approach of [120, 121], demonstrating that the proposed concept is more robust and more efficient for estimating the static environment due to the persistent grid map accumulation rather than using short-term static particles, especially for scenarios with temporarily occluded areas of the static environment. Moreover, it has been demonstrated how the static/dynamic occupancy classification is further enhanced by additional measurement data of radar and camera sensors as well as using feedback of the subsequent object tracking, which is essential for solving challenging scenarios with newly observed parts of straight guardrails and moving trucks with an observed long side. Thereby it has also been analyzed how the occupancy filtering can be modified by the camera classification such that even non-moving object tracks are able to be extracted due to an enforced dynamic convergence. This adapted filtering has been restricted to pedestrians and cyclists, whereas parked vehicles remain as static occupied in the grid map representation without a subsequent object extraction, as the static occupancy accumulation ensures a high spatial accuracy without relying on the camera object estimation, which is also more efficient since no particles or object tracks are required.

6.3 Object Detection and Tracking

Based on the presented dynamic grid estimation, this section evaluates the subsequent object-level estimation in more detail. First, the object extraction and association are demonstrated for different challenging scenarios. Afterwards, the UKF-based dynamic state estimation is quantitatively evaluated for challenging full braking and turning maneuvers. Finally, the combined object shape estimation and classification concept is analyzed for a simulation scenario with different configurations.

6.3.1 Object Extraction and Association

In the following, the robustness of the moving object extraction and cell-wise association as well as the overall object tracking is demonstrated qualitatively for different challenging urban and highway scenarios.

Extraction of Arbitrary-Shaped Moving Objects

The generic extraction of arbitrary-shaped moving objects is highlighted in a challenging scenario in Fig. 6.13, which contains a truck with a specific low-mounted pipe construction on the rear. To demonstrate the robustness of the object extraction and association even with sparse sensor measurement data, only the low-mounted lidar sensors with the low vertical field of view are used here, i.e., no information of the radar, the camera, or the roof-mounted lidar sensor is evaluated in that scenario, as also presented in [1]. Hence, the lidar detection pattern at the sensing height does not fit a rectangular shape and results in an unstructured shape here, as depicted in Fig. 6.13a. However, since the proposed dynamic grid estimation with the low-level particle tracking generally estimates the movement of the individual occupied cells without depending on any object assumptions, those moving parts robustly converge toward dynamic occupancy with an accurate velocity distribution estimation of the respective cells, as shown in Fig. 6.13b. As the proposed moving object extraction concept results from a density-based cell clustering of dynamic occupied cells with similar velocities, which in turn is based on the dynamic grid estimation, even such rare objects are robustly extracted, which also applied to the cell-level association when that object track has been extracted before, as demonstrated in Fig. 6.13c–6.13d. Hence, this approach is also suitable for lower resolution sensor data, whereas, in contrast, an L-shape feature detection or a box fitting on the raw measurement data, without considering the particle-based movement direction estimation or the predicted object track, are prone to such deviations of the box shape model.

Extraction of Far-Distant Moving Objects

Another challenging scenario of the moving object extraction is demonstrated in Fig. 6.14, where the ego vehicle is approaching an oncoming vehicle that is at far distance. Thereby the oncoming vehicle enters the detection area of the grid structure in the upper right corner, whereas the ego vehicle is approximately aligned diagonal within the grid due to the ego orientation in the odometry coordinate system. Hence, this example enables the maximum possible range to the front, which is up to 184 m with the grid configuration as introduced in Fig. 6.3. In that scenario, the oncoming vehicle results only in a single lidar detection, with an additional camera object detection of that vehicle, whereas no radar information is available. The resulting dynamic grid mapping over time is visualized in Fig. 6.14b, illustrating the fixed area within the odometry frame that is independent of the ego movement within that sequence. The results demonstrate the correct convergence toward dynamic occupancy over time, while still including artifacts of unclassified occupancy of the previous time instances until the full convergence toward dynamic occupancy is achieved. The respective augmented measurement grid with the additional representation of the particle-based cell-wise velocity estimates is highlighted in Fig. 6.14c, indicating

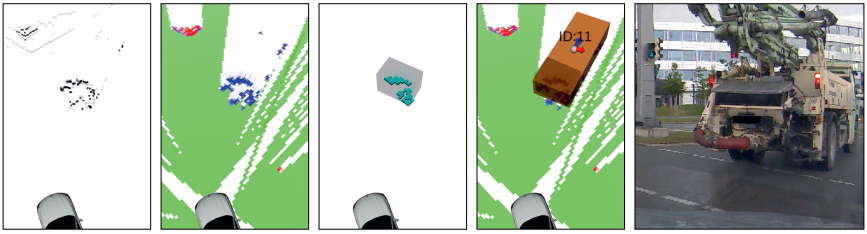
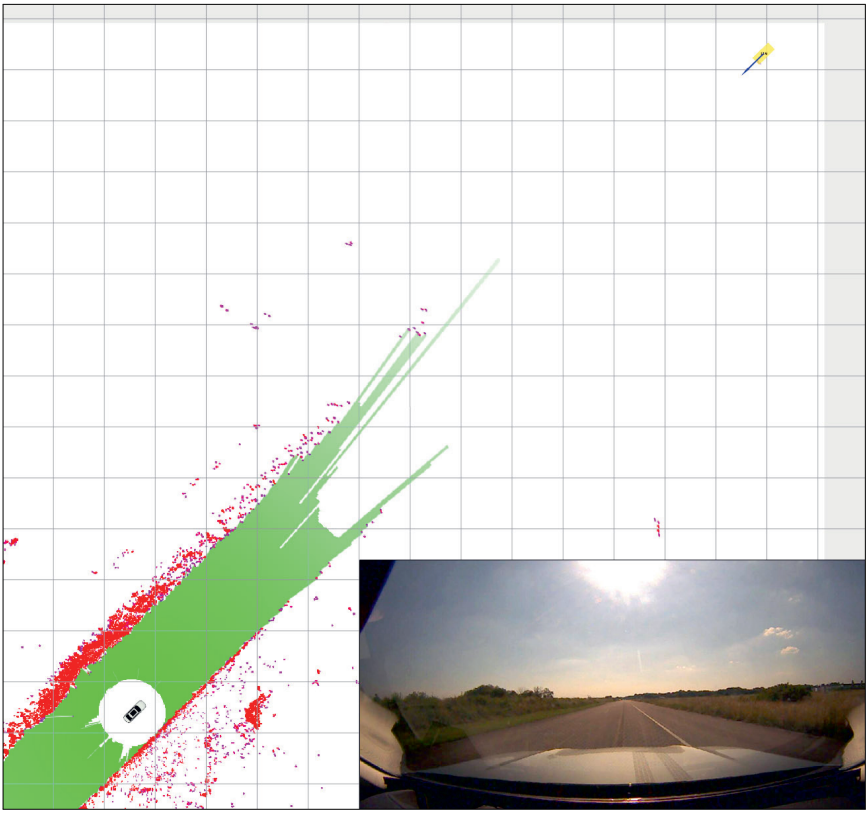


Fig. 6.13: Moving object with an unstructured shape; the detection pattern of the low-mounted lidar sensors at the sensing height does not fit a rectangular shape. (a) Lidar detections of front sensor (black: obstacles, gray: ground). (b) Augmented measurement grid (blue lines indicate cell velocities). (c) Associated cells and measurement bounding box. (d) Resulting object tracking. (e) Camera image. Figure initially published in [1], © 2019 IEEE.

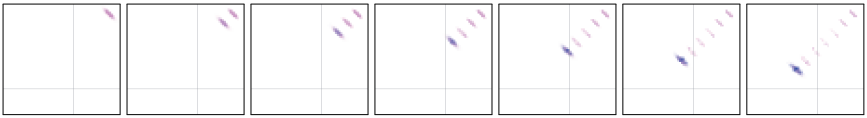
that a first convergence of the velocity estimation is already achieved after three time instances. Since the augmented measurement grid is only derived for the corresponding measurement data of the current time instance, the grid mapping artifacts of the initial frames have no effect here. Even though the object track is extracted conservatively here after the full convergence of seven time instances, the distance of the object detection is still more than 160 m in that example, as shown in Fig. 6.14a. Overall, this scenario has demonstrated that the presented grid-based object tracking is generally also capable of detecting objects in far distance, also with very sparse sensor data.

Association for Densely Moving Traffic Participants and Occurring Occlusion

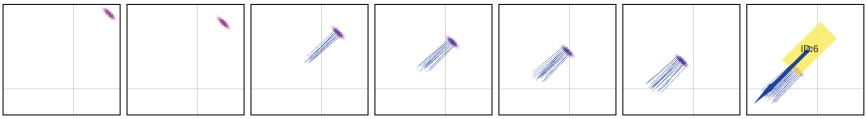
A challenging urban scenario with multiple surrounding vehicles and an approaching scooter is shown in Fig. 6.15. Multiple time instances are visualized that illustrate the movement of the scooter, which is very close to the other vehicles and the ego vehicle and thus also causes occlusions. As with the example before in Fig. 6.13, also no measurement data of the camera or the roof-mounted lidar sensor are available in this scenario, as initially presented in [2], to focus on the robustness of the approach without those enhanced measurement data. The scooter is robustly extracted as a new object track, even though the initial position is in-between two extracted vehicles. As demonstrated by the different processing steps and the successive time steps, the proposed particle labeling association with the subsequent clustering and verification of the sub-clusters of a track, as presented in Section 4.3, correctly associates the measured occupied cells to the corresponding tracks and thus successfully distinguishes those different objects. This is even the case in the final sequence where the scooter stops directly next to a vehicle and the dynamic occupancy of both objects is directly adjacent, shown in the bottom image of Fig. 6.15a, which could not be separated by a common clustering extraction without prior knowledge of the predicted tracks here. Overall, this scenario has demonstrated the robustness of the object extraction and association for scenarios with densely moving or stopped vehicles as well as fully or partial occlusions of the surrounding traffic participants.



(a) Scenario with the resulting object extraction at t_7.



(b) Dynamic grid mapping for the different time instances t_1–t_7.



(c) Augmented measurement grid with the particle-based cell velocity estimation for t_1–t_7.

Fig. 6.14: Object extraction for an oncoming vehicle at far distance on a test ground. The visualized raster corresponds to a size of $10\,\mathrm{m} \times 10\,\mathrm{m}$ here.

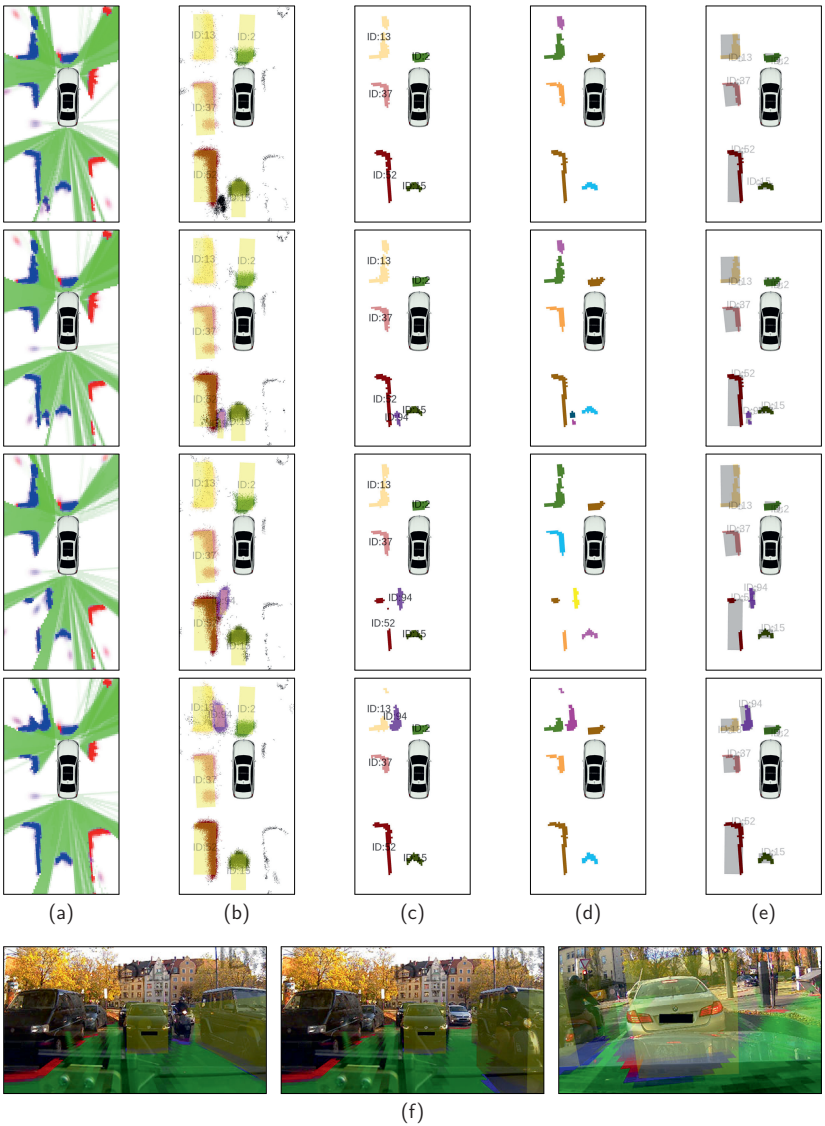


Fig. 6.15: Scenario with multiple surrounding stopped vehicles and an approaching scooter. Rows show evaluation over time. (a) Augmented measurement grid. (b) Predicted tracks and labeled particle population. (c) Particle labeling association. (d) Additional clustering step (cluster colors are not related to the corresponding tracks here). (e) Finally associated clusters and measurement bounding boxes. (f) Camera images. Figure initially published in [2], © 2018 IEEE.

Another urban scenario with temporary occlusions of crossing traffic is demonstrated in Fig. 6.16. In this sequence, a blue bus (ID 189) on the right side of the intersection waits for oncoming traffic to pass by in order to turn left, which is also the case for multiple cars in the opposite direction. Thereby the bus as well as other stopped vehicles are significantly occluded several times from the perspective of the ego vehicle due to the crossing traffic in front of them. The proposed object extraction and association approach, however, robustly keeps track of those partly occluded vehicles and thereby also retains their individual ID. This scenario further demonstrates the robustness of the overall object tracking in urban environments in general, where the object pose and shape are also accurately estimated for partly occluded objects due to the evaluated visibility in terms of the freespace of the measurement grid, which in turn is essential for the consistent association here.

Object Tracking for Highway Scenarios with Fast-Moving Vehicles

A different type of challenging scenarios for the object estimation is presented in Fig. 6.17, demonstrating the successful application of the grid-based object tracking also for highway environments. As opposed to urban environments with the complexity of densely moving traffic and occlusions, highway scenarios are primarily challenging due to the generally fast speed of both the ego vehicle and surrounding vehicles as well as the resulting required larger detection range. As highlighted in Fig. 6.17a, all of those fast-moving vehicles around the ego vehicle are detected and robustly tracked in that scenario, including a far-distant car in the front area that is barely visible due to the occlusion of the more close overtaking cars as well as the preceding truck in front of the ego vehicle.

A variation of such highway environments is presented in Fig. 6.17b, depicting the beginning of a construction site scenario in which the lane course is redirected by delineator posts and thus deviates from the normal course here. Since the object tracking approach of this work has no dependencies on road map information that might be incorrect here, the objects are accurately represented at their actual position without relying on any lane assignment or prior map knowledge. Another major benefit of the proposed strategy is that not only moving objects are estimated here but also the static environment in terms of the temporally filtered dynamic grid map, which is shown here on the ground rather than the augmented measurement grid to highlight the accumulated grid map representation. Further details of the dynamic grid mapping in such highway scenarios have been discussed before in Fig. 6.9–6.10. Hence, the overall consistent and combined environment estimation ensures that in this scenario, in addition to the other traffic participants, also such construction site delineator posts are robustly extracted here as accumulated static occupancy. This is crucial for obstacle avoidance and an adapted path planning of autonomous vehicles in such scenarios, but possibly also for updating road map information.

6.3.2 Dynamic State Estimation for Highly Dynamic Maneuvers

In this section, the dynamic state estimation of the object tracking is evaluated quantitatively using real sensor data of a test vehicle with a reference measurement system. Two challenging scenarios are presented, the first scenario addresses high negative accelerations of a full braking maneuver, whereas the second scenario focuses on turning maneuvers with

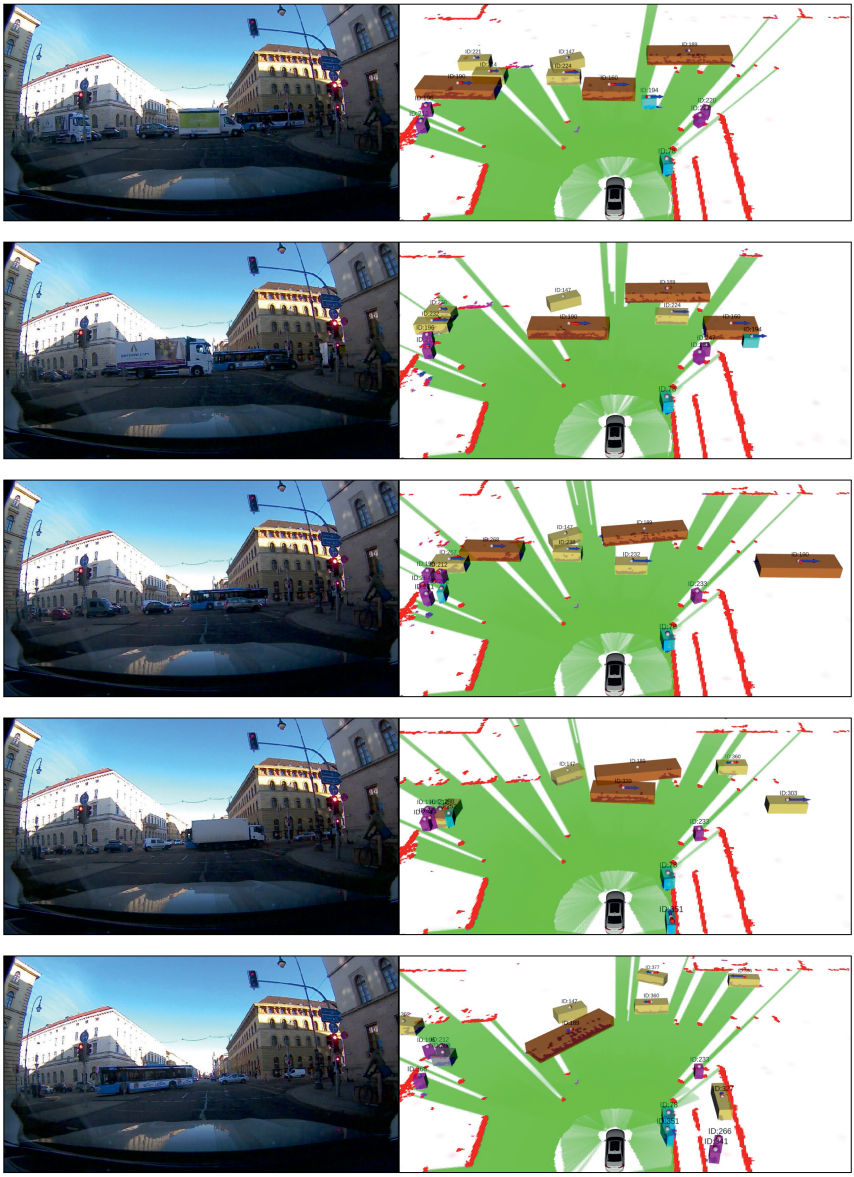
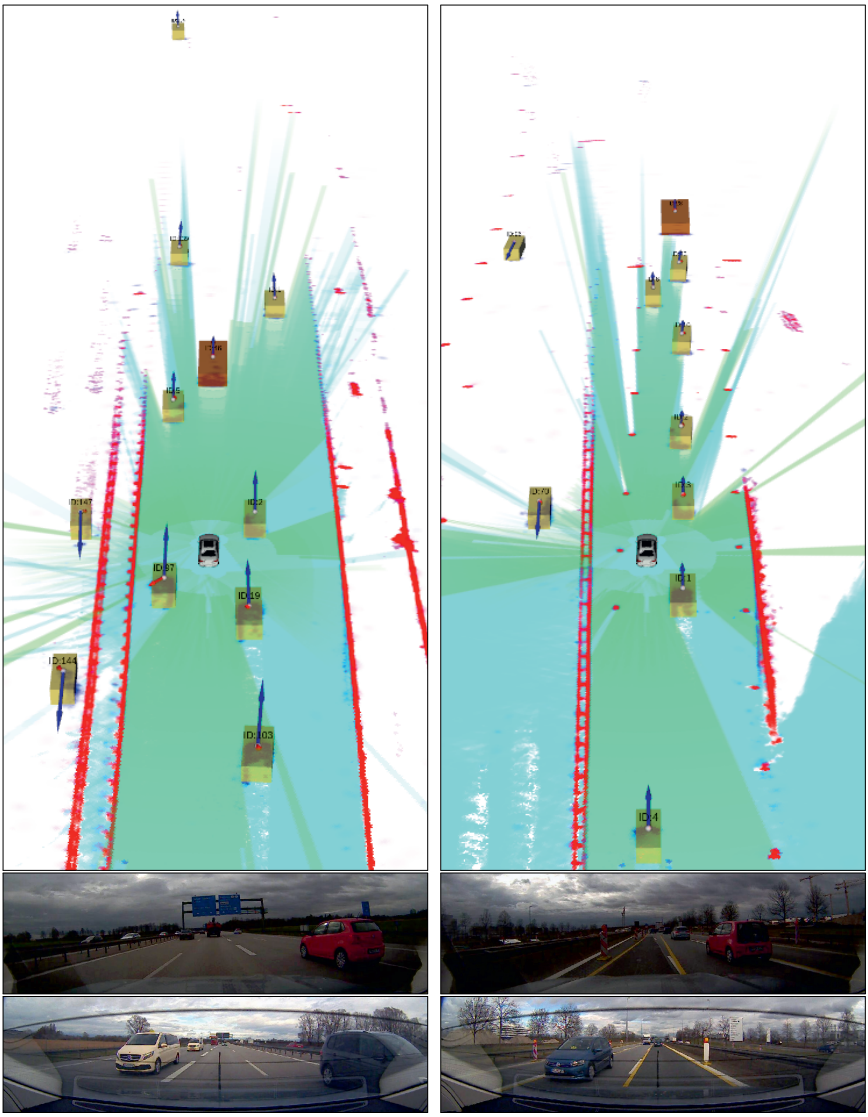


Fig. 6.16: Object tracking at an urban intersection with crossing traffic and occurring temporary occlusions. The rows from top to bottom show the evaluation over time of that sequence.



(a) (b)

Fig. 6.17: Results of the proposed approach in highway scenarios. Here the dynamic grid map is shown on the ground instead of the augmented measurement grid to further highlight the accumulation of the static environment. (a) Highway scenario with fast-moving vehicles. (b) Highway construction site with delineator posts and a modified road course.

multiple "figure-eight" trajectories. The following evaluation has been originally published in [1] in the context of this thesis, together with the proposed object state estimation.

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. 6.18. 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 thereby 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 (5.4), 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 $t_{\mathrm{horizon}} = 0.25\,\mathrm{s}$, shown by the red solid line, converges fast toward zero by that additional constraint.

The multi-filtering approach, where the mean velocity $\overline{\nu}$ of all associated cells as defined in (5.22), based on the cell-wise particle velocity estimates ν_t^c, 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 shown case. That approach is robust against wrong object assumptions and position updates, e.g., due to an incorrect association, but in 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 accurately 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

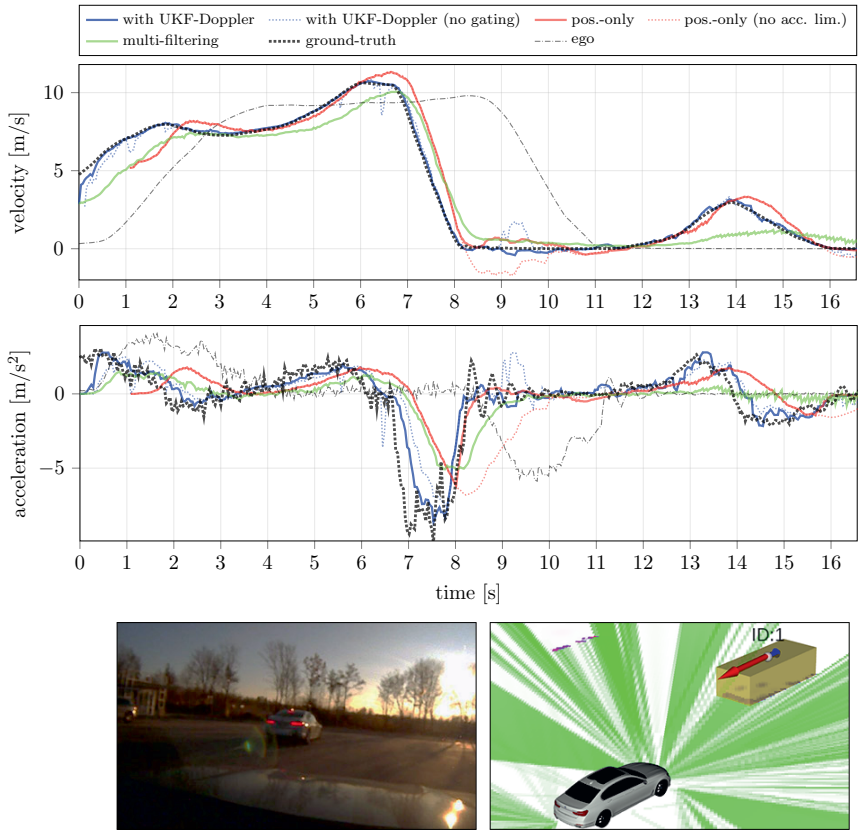


Fig. 6.18: Full braking scenario with real sensor data. The shown time of all data corresponds to the original measurement time to focus on the filtering behavior, i.e., sensor and system latencies as well as the processing time are excluded in the comparison with the ground truth data. Figure initially published in [1], © 2019 IEEE.

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 6.1. 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 for the underlying grid-based estimation either. For the other two approaches, in contrast, the dynamic grid estimation is enhanced by radar measurements, resulting

Tab. 6.1: Root mean square error (RMSE) analysis of dynamic state estimation for highly dynamic movements with real sensor data. Results initially published in [1].

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 acceleration limitation)	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 estimation	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.

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.

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. 6.19. This represents an extreme case with high and fast changing turn rates with up to 65 deg/s occur in that scenario, much higher than that of common turning vehicles.

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 tracked for the complete sequence, the estimation of the orientation and in particular the turn rate is rather inaccurate for that highly nonlinear movement, as the low-level particle tracking only estimates 2-D velocities on the cell-level without any considered high-level object instance or turn rate.

The orientation estimation based on the currently measured freespace, as proposed in Section 5.3.5, 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 here.

Best results are achieved by combining that freespace-based orientation estimation with the UKF-Doppler state estimation, which is shown in blue. In that combined approach, also the radar Doppler radial velocity components at different parts of the object are taken into account, which implicitly enable measuring the object orientation and turn rate.

The different orientation estimation approaches of the turning scenario are also summarized in Table 6.1 by the RMSE analysis, equivalent to the full braking scenario as discussed before. Altogether, those results for both scenarios have demonstrated that the presented dynamic state estimation strategy results in an accurate object state estimation, even for such highly dynamic maneuvers.

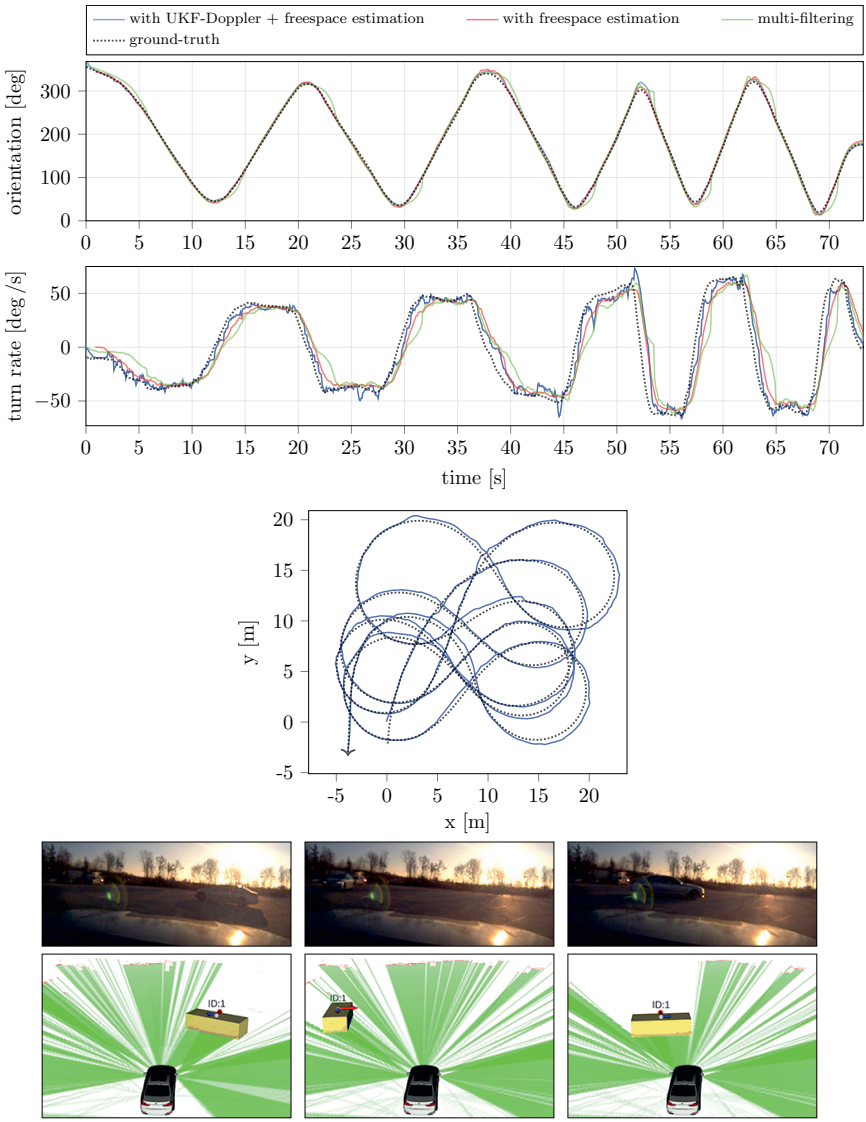


Fig. 6.19: Turning scenario with multiple "figure-eight" trajectories and real sensor data. As with Fig. 6.18, the shown time of all data corresponds to the original measurement time by excluding sensor and system latencies as well as the processing time in the comparison with the ground truth data. Figure initially published in [1], © 2019 IEEE.

6.3.3 Object Shape Estimation and Classification

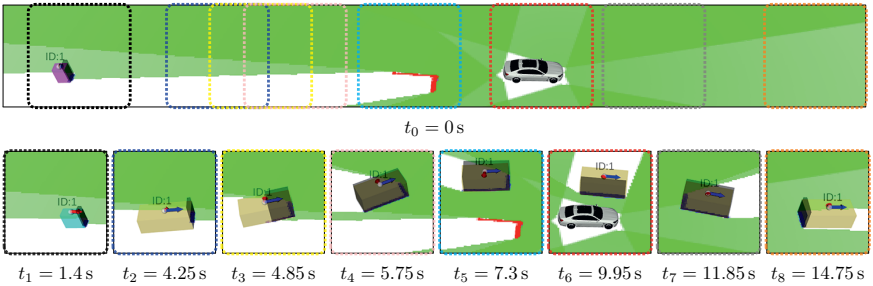
In the following, the geometry estimation and object classification are analyzed, primarily by a simulated sequence with two different variants, while the object classification is further quantitatively evaluated in a real scenario. This evaluation has been originally published in [1] in the context of this thesis, together with the proposed object state estimation.

Simulated Test Scenario

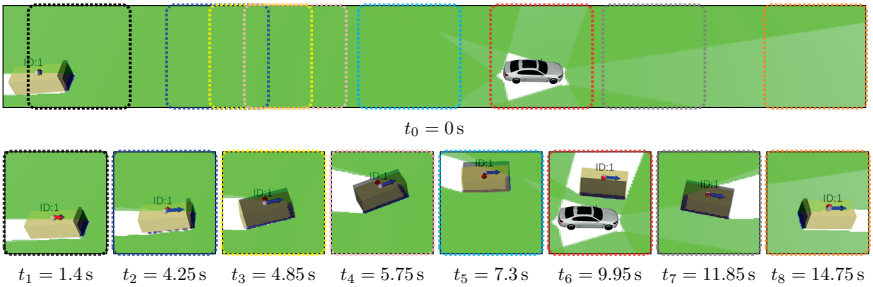
The evaluated simulation scenario consists of a non-moving ego vehicle, which is overtaken by another vehicle, as shown in Fig. 6.20. The first variant in Fig. 6.20a includes an additional parked vehicle that occludes parts of the overtaking vehicle in the beginning; the second variant in Fig. 6.20b is without occlusion. The trajectory of the overtaking vehicle is equal in both variants, the vehicle accelerates in the first 4 s of the sequence up to a velocity of 5 m/s and then moves with constant velocity. The simulated ground-truth object box has length $l = 3.9\,\mathrm{m}$ and 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. 6.21.

The first variant with occlusion, cf. Fig. 6.20a, illustrated by solid lines in Fig. 6.21, is simulated such that the measured width of the minimum bounding box in the initial acceleration phase is about 0.75 m, and no noteworthy length is 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 classes pedestrian and cyclist are selected higher than the prior 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 class pedestrian 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 m/s and thus the maximum estimated object velocity $v_\tau^{\max}$ also increases to that value. Due to that increasing velocity, the class pedestrian becomes less probable; the class cyclist achieves the highest probability. Around $t \approx 5\,\mathrm{s}$, the vehicle starts overtaking and larger parts of the width are observable, whereas also the class cyclist 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. 6.22 by the corresponding histogram probability distributions of the length and width over time for that sequence. Thereby also the inverse sensor model with the boundary visibility consideration is recognizable, as



(a) Scenario with a parking vehicle occluding parts of the approaching vehicle.



(b) Scenario without occlusion.

Fig. 6.20: 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 of the upper row are shown, which correspond to the respective dotted area. Figure similarly published in [1].

introduced in Fig. 5.8, 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. 6.20b and illustrated by dotted lines in Fig. 6.21, is simpler since the full width of the target is visible directly from the beginning. Due to that 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 regarding the likelihood of the class, resulting in an extracted length of approximately 4.4 m here, which is much more accurate than the maximum observed length that is only about 0.5 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.

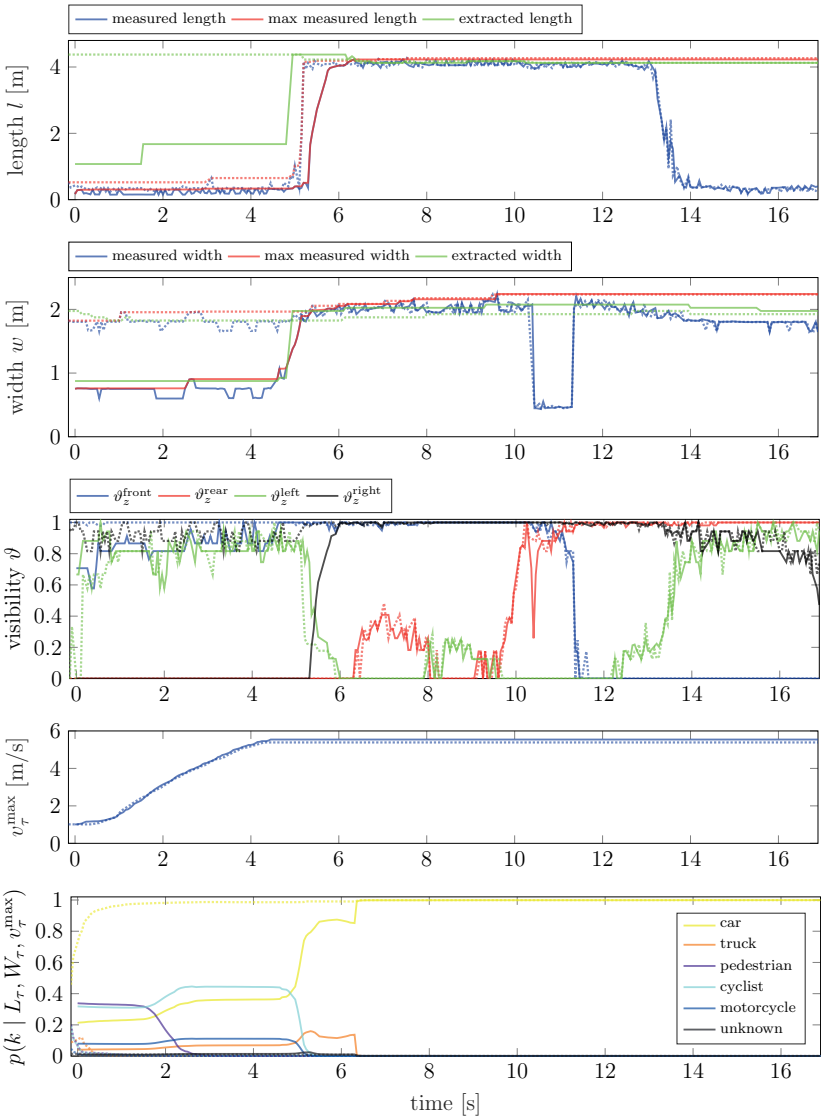
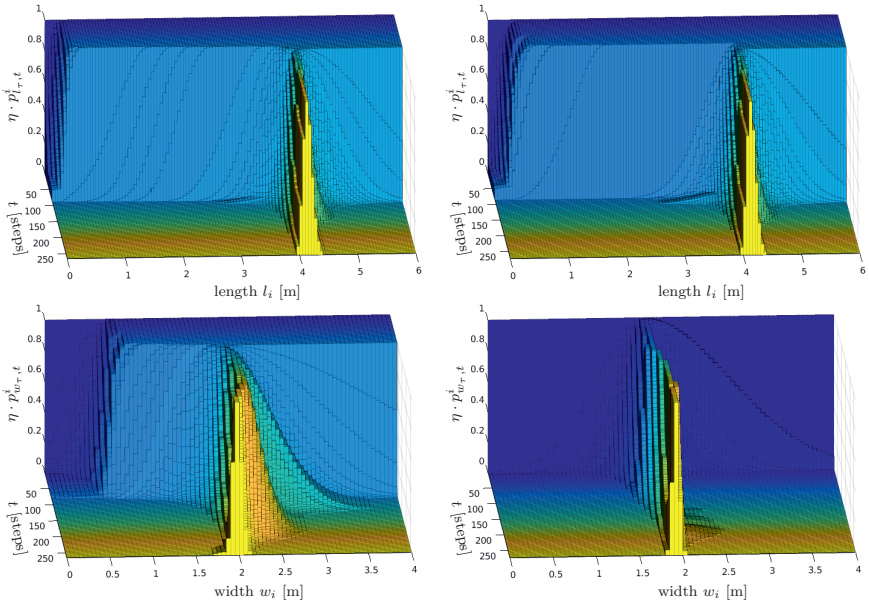


Fig. 6.21: Evaluation of the geometry estimation and classification of the overtaking vehicle scenario of Fig. 6.20 for both variants. Results for the scenario with occlusion (Fig. 6.20a) are shown by solid lines and for the scenario without occlusion (Fig. 6.20b) by dotted lines, respectively. The simulated ground-truth object box has length $l = 3.9\,\text{m}$ and width $w = 1.8\,\text{m}$. Figure initially published in [1], © 2019 IEEE.



(a) Scenario with occlusion.

(b) Scenario without occlusion.

Fig. 6.22: Histogram distributions over time of the length and width estimation for the simulated overtaking scenario of Fig. 6.20 for both variants with and without occlusion. 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. Figure similarly published in [1].

Object Classification in Real Urban Scenarios

The proposed object classification concept is finally evaluated by a real traffic scenario in an urban environment. This evaluation has also been published as part of [1], which further contains an attached video of the resulting object tracking in that sequence. The evaluated sequence has a duration of about 8 min with various vehicles, pedestrians, cyclists, etc. As originally presented in [1], neither camera information nor the roof-mounted lidar sensor with the higher vertical resolution are available in that scenario. Hence, as opposed to achieve the best possible classification results that also highly depends on the classification results of the camera or a dense lidar point cloud, the main focus of that evaluation is to demonstrate the general ability to classify objects based on limited measurement data using the 2-D occupancy grid representation without any height information.

The resulting confusion matrix of the object classification is presented in Table 6.2. Thereby only extracted moving objects are evaluated, since non-moving obstacles or parked vehicles remain in the static grid representation in this work. Hence, a simple ground

Tab. 6.2: 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. Results initially published in [1].

	Car	Truck	Pedestrian	(Motor-)Cyclist*1	Other	False-Positive
Car	23781	39	945	1035	114	882
Truck	45	437	0	0	567	324
Pedestrian	76	0	9530	1792	0	167
(Motor-)Cyclist*1	1800	0	834	9888	0	0
Other	294	0	0	0	403	0
# Occurrences*2	25996	476	11309	12715	1084	1373
Rate	0.915	0.918	0.842	0.778	0.372	0.026 (FPR)

*1 The classes cyclist and motorcycle are combined here due to occurring slow object speeds.
*2 Total number of evaluated occurrences of that actual class (sum of column).

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 (5.60) is considered. As noted before, those classification results are not directly comparable with other approaches, also since only moving objects are extracted here. But they indicate that the proposed simple Naive Bayes classification approach, solely based on the histogram filter geometry distribution estimation and the maximum observed track velocity over time, even without integrating additional camera information as presented in (5.62), achieves promising results with real sensor data without necessarily depending on a high vertical resolution sensor data or camera images.

6.4 Summary and Overall Approach Application

This chapter has presented a detailed evaluation of the different aspects of the proposed grid-based object tracking approach in various urban and highway environments.

First, a brief overview of the overall approach and evaluation setup has been given. Thereby the test vehicle setup with the different lidar, radar, and camera sensors has been presented, followed by a short illustrative recapitulation of the main processing steps and a discussion of the primary grid configuration as well as the algorithm runtime.

Second, different aspects of the proposed dynamic grid mapping have been evaluated. As demonstrated in the scenario with temporarily occluded areas of the static environment, the original approach of Tanzmeister et al. [120, 121] results in information loss in such scenarios due to the short-term filtering with static particles, whereas the proposed architecture retains that static occupancy information due to the persistent map accumulation. This increases not only the robustness of the occupancy filtering, but also the estimation efficiency, since the number of required particles is significantly reduced that way. Moreover,

it has been demonstrated how additional measurement data of radar and camera sensors as well as information of the subsequent object tracking further improve the static/dynamic occupancy classification. As a result, even complex highway scenarios with newly observed parts of straight guardrails and moving trucks with an observed long side are robustly represented, which otherwise cannot be fully solved from the cell-individual perspective. In addition, it has also been analyzed how the dynamic occupancy filtering can be adapted by the camera classification such that even non-moving objects can be extracted, which has been applied to pedestrians and cyclists, as those vulnerable road users are generally highly dynamic and therefore require an even more conservative safety distance than to stationary obstacles. Parked vehicles, in contrast, remain as static occupancy in the grid map without a subsequent object extraction in the application of this work, as this representation is spatially more accurate and more efficient, without requiring particles or object tracks or depending on the accuracy of the camera object estimation.

Third, the evaluation of the subsequent object tracking has been addressed. Thereby it has been demonstrated that generally arbitrary-shaped moving objects are extracted, even with sparse sensor data that do not fit a rectangular shape, also for large detection distances above 160 m. Moreover, the robustness of the association as well as the overall object tracking over time has been highlighted by complex urban and highway scenarios, even for scenarios with densely moving traffic, temporary occlusions, fast-moving vehicles, and unstructured environments of construction sites. Those qualitative results have been further extended by a detailed quantitative evaluation of the dynamic state estimation using a target vehicle with a reference measurement system, which has performed challenging highly dynamic full braking and "figure-eight" turning maneuvers. Those results have proven the accuracy and robustness of the object state estimation even for such extreme cases with high longitudinal or lateral accelerations, including a comparison of different strategies of the object state update. Finally, the combined object shape estimation and classification approach has been evaluated, where effects of the object shape visibility and occlusion have been discussed by different variations of a simulation scenario, whereas the robust object classification has been further proven for real traffic scenarios.

Altogether, this chapter has demonstrated that the proposed grid-based object tracking approach of this thesis results in a generic and robust real-time environment estimation with a full surrounding 360° detection area based on multiple lidar, radar, and camera sensors. Moving objects as well as static obstacles and passable freespace are consistently estimated in various unstructured environments, even in scenarios with densely moving traffic participants, partial occlusions, or highly dynamic maneuvers. The implemented algorithm is actively used within autonomous driving test vehicle prototypes for both the multi-sensor object tracking and the obstacle detection of the surrounding static environment. Moreover, the resulting object tracking is further used as real sensor data input to other related research projects in the field of autonomous driving that are based on high-level object instances, e.g., for learning interaction-aware probabilistic driver behavior models [107], for the set-based prediction of pedestrians [66], for fail-safe trajectory planning [92, 93], or for a semantic grid-based road model estimation [124].

7 Conclusion

Modeling and estimating the current local environment remain major challenges of mobile robots, as also the requirements on the accuracy, the robustness, and the detection range are permanently increasing. This is particularly the case for the ongoing development of autonomous driving with the generally high scenario complexity of unstructured dynamic environments. Therefore, an accurate environment model is required that properly represents moving objects, static obstacles, and passable freespace, which has to be robustly estimated by processing measurement data of multiple sensors – those aspects represented the overall goal and the scope of this work.

This thesis presented the *Grid-Based Object Tracking* – a novel environment perception strategy that combines a multi-sensor multi-object tracking with a consistent estimation of static obstacles as well as passable freespace. Measurement data are thereby fused and preprocessed by a low-level environment estimation based on the generic concept of dynamic occupancy grids, whereas the object-level estimation is performed subsequently based on that grid representation. Hence, as opposed to popular high-level object fusion approaches that are based on an error-prone early-stage sensor-individual object abstraction, the proposed strategy corresponds to an early-stage sensor fusion and dynamic estimation with a late-stage object abstraction. Objects are thus estimated based on the combined measurement data of all sensors in this work; the additional low-level dynamic grid estimation further enables a generic detection of arbitrary-shaped moving objects, while the static environment is retained in the grid representation without abstracting object instances.

Overall, the proposed strategy results in a highly robust and accurate multi-sensor environment estimation, even with many occurring objects, unstructured surroundings, sparse sensor data, or partial occlusions. Thereby measurement data of multiple lidar, radar, and camera sensors are fused, obtaining a full surrounding 360° detection area of heterogeneous sensors. Moreover, the grid-based framework of this work is optimized to combine the advantages of the different sensor principles – spatially accurate lidar sensors represent the main sensors for the generic occupancy/freespace grid derivation and thus also for the object pose and shape estimation; radar sensors primarily contribute to the dynamic estimation on both the grid-level and the object-level by their Doppler velocity measurements; and the semantic information of camera sensors further improves the static/dynamic occupancy differentiation as well as the object extraction and classification.

Another important benefit of the presented concept is its high versatility for different environments and applications. First, the approach is solely sensor-based, i.e., the environment is fully estimated by the current sensor measurement data without requiring or relying on any prior map knowledge or localization, which enables an application in various unknown environments. Second, the algorithm is able to be processed online using a conventional modern computer, i.e., no offline post-processing is required, due to the efficient implementation with highly parallelized GPU computation. Third, the generic

grid-based estimation framework is directly adaptable to modifications and extensions of the sensor setup, as it only requires that the corresponding measurement data are modeled in the uniform measurement grid representation. This also applies to the subsequent object tracking, as it is based on the abstracted dynamic grid representation rather than sensor-specific characteristics.

Altogether, the proposed approach is well suited for real-time multi-sensor environment perception applications. The implemented algorithm is actively used for the multi-sensor object tracking and obstacle detection within test vehicle prototypes for autonomous driving in urban environments, demonstrating the successful application with promising and robust results in challenging real-world scenarios. Thereby the output is also used as real sensor data input to other related research projects in the field of autonomous driving.

As part of the grid-based object tracking strategy, the different aspects of sensor data fusion, temporal filtering of the dynamic grid mapping, object extraction and association, and object state estimation have been addressed in this work. Novel concepts have been proposed for all these fields, primarily based on the characteristics of the dynamic grid estimation, to fully utilize the potential of the overall approach. In the following, the main contributions of each of those processing steps are briefly summarized.

Measurement Grid Representation and Fusion: The measurement data of all sensors are modeled and fused using a uniform evidential Dempster–Shafer occupancy grid representation. Thereby all measurements have been considered in this work, as opposed to common grid-based approaches that model only the static environment and try to filter out moving parts. To handle asynchronous sensor data with different measurement and latency time instances, a new fusion strategy has been presented such that all fused data approximately represent the same time instance and thus the same status of the dynamic environment. Moreover, to retain important additional measurement data of radar and camera sensors, separate velocity and classification grid layers have been introduced.

Dynamic Grid Mapping and Particle Tracking: A novel evidential dynamic grid mapping strategy has been proposed that combines the benefits of a persistent grid map accumulation with a short-term particle filter tracking of the dynamic state of the grid. To differentiate static and dynamic occupancy over time, solely based on unclassified occupancy measurements, the occupancy filtering of the evidence masses has been adapted within the Dempster–Shafer framework, also regarding implicit temporal transitions between the individual hypotheses and their supersets as well as conflicts between contradicting hypotheses. Dynamic occupancy of the grid map is thereby directly coupled with the particle tracking, overall resulting in a robust velocity distribution estimation of individual grid cells and thus also a robust prediction of their dynamic occupancy masses. Static occupancy, in contrast, is directly accumulated in the persistent map representation without requiring particles. Compared to state-of-the-art approaches, especially in scenarios with temporarily occluded areas of the static environment, false dynamic classification has been significantly reduced due to the persistent static occupancy accumulation of the grid map, as opposed to a short-term estimation with static particles that results in information loss in such scenarios. Not only the occupancy classification is improved that way, but also the computational efficiency due to the drastically reduced number of required particles. Moreover, the proposed approach has been further enhanced by additionally

evaluating radar velocity measurements and camera classification information as well as using feedback of the subsequent object tracking. As a result, even challenging scenarios with newly observed parts of straight guardrails and moving trucks with an observed long side are robustly estimated, which otherwise generally cannot be fully solved from the cell-individual perspective.

Object Extraction and Association: Moving objects are robustly detected based on the dynamic grid estimation with the low-level particle tracking, which corresponds to a generic movement-based track-before-detect approach. That strategy enables detecting arbitrary moving objects by their occupancy motion, without depending on specific features or training data, in contrast to common appearance-based detection algorithms. The robustness of the object extraction has been presented for objects with unstructured shapes, unknown classes, and sparse sensor data in far distances as well as partial occlusions. In addition to that generic movement-based object detection, the proposed approach has been further extended to also incorporate additional input information of object instances. Hence, non-moving objects can be extracted as well by adapting the dynamic occupancy filtering within the areas of such input objects. This has been demonstrated for camera-detected pedestrians and cyclists due to their generally high dynamic and vulnerability, but also for the feedback of already tracked objects in order to continue tracking them when they temporarily stop moving. Newly occurring objects are extracted by a novel combination of a density-based and connectivity-based clustering, where dynamic occupied cells with similar velocities are compared with the local neighborhood regarding the cell velocity variance to minimize false positives. For tracks that have been extracted in the past, individual grid cells are directly associated on the grid cell-level rather than on the object-level. A new association approach has been proposed, where particles of the underlying low-level particle tracking are linked to object tracks by an attached label. The occupancy likelihood of a track, i.e., the expected areas of measured occupancy, including the current visibility and free-form shape, is thus accurately represented by the predicted particle population linked to that track. Overall, this results in a robust cell-wise association, even for scenarios with clutter measurements, densely moving objects, or partial occlusions with gaps in-between, where clustering-based approaches without any information of the predicted tracks commonly fail.

Object State Estimation: The associated occupied grid cells of each object track are abstracted to a popular oriented bounding box model of the measurement. However, as opposed to other object tracking approaches, the presented object state estimation has been further optimized for the characteristics of the underlying dynamic grid. The object pose is thereby updated by selecting the best visible reference point regarding the current visibility of the bounding box edges, determined by the surrounding freespace of the grid, and a generic orientation extraction by combining the particle-based movement direction with a freespace-based optimization approach. In contrast to popular L-shape box fitting approaches that require dense and accurate sensor data, the proposed strategy ensures a robust dynamic state estimation even in challenging scenarios with sparse sensor data, occlusions, and changing sensor fields of view. The approach has been further extended by directly processing the radar velocity measurements for the object state update, where a new Doppler-based UKF update has been proposed that robustly utilizes the radial ve-

locity measurement space. 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, as demonstrated for highly dynamic full braking and turning maneuvers. Finally, a new strategy for the object size estimation has been proposed by combining a histogram filter geometry distribution estimation with a Naive Bayes object classification concept, which in turn is based on the filtered geometry distribution, the maximum observed velocity, and obtained camera classification information. This enables to model non-Gaussian distributions of the length and width by distinguishing lower and upper bounds of the measurement, also evaluated by the freespace-based visibility, and to use prior class knowledge. As a result, the extracted object box size implicitly adapts to the assumed size of the object class and the actually observed lower and upper bounds of the measurement data.

In conclusion, this thesis has presented new methods for sensor data fusion, temporal filtering and mapping, data association, and object tracking, which overall significantly improve the perception capabilities and thus the safety of mobile robots, especially in the context of future autonomous driving applications. Still, fully autonomous driving remains a very challenging task that requires further advancements of the respective algorithms and sensors to achieve the ultimate goal of being completely driverless.

As final remarks, some conceivable similar applications and future research fields are briefly discussed. The presented methods could possibly be further adapted, for instance, for other types of mobile robots, also as part of an offline post-processing ground-truth generation or reference algorithm for perception systems with low online computation power, as well as for removing artifacts in terms of filtering out dynamic parts for offline map generation or localization of the static environment, or for tracking moving objects of premises or indoor applications, possibly also with distributed or stationary sensors. Potential algorithmic improvements can be achieved, for example, by integrating the height information into the environment model by a 2.5-D or 3-D model, e.g., for an enhanced estimation of small obstacles like curbs or also a more detailed object size estimation and classification. Another possibility is to extend the object tracking by a temporally filtered existence probability estimation for improving both the extraction as well as the deletion of object tracks. Finally, there is also great potential in a more extensive use and integration of latest machine learning approaches for the object detection and semantic segmentation into this approach – while still also retaining the robustness of the generic movement-based dynamic grid estimation with the ability of detecting and tracking even unknown moving object types and the consistent estimation along with the static environment.

Own Publications

[1] S. Steyer, C. Lenk, D. Kellner, G. Tanzmeister, and D. Wollherr. Grid-Based Object Tracking with Nonlinear Dynamic State and Shape Estimation. *IEEE Trans. Intell. Transp. Syst.*, 21(7):2874–2893, July 2020.

[2] 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.*, pages 3036–3043, 2018.

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

[4] S. Steyer, G. Tanzmeister, and D. Wollherr. Grid-Based Environment Estimation Using Evidential Mapping and Particle Tracking. *IEEE Trans. Intell. Veh.*, 3(3):384–396, Sept. 2018.

[5] 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.*, pages 231–237, 2016.

Bibliography

[6] R. Adams and L. Bischof. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 16(6):641–647, 1994.

[7] M. Aeberhard. *Object-Level Fusion for Surround Environment Perception in Automated Driving Applications*. PhD thesis, Dept. Elect. Eng. Inf. Tech., Technische Universität Dortmund, Dortmund, Germany, 2017.

[8] 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.*, 13(4):1717–1726, Dec. 2012.

[9] Alphabet Inc. Alphabet Project Wing. `https://x.company/projects/wing/`. Online; retrieved 2019-05-28.

[10] Amazon.com Inc. Amazon Prime Air. `https://amazon.com/Amazon-Prime-Air/b?node=8037720011`. Online; retrieved 2019-05-28.

[11] A. Asvadi, P. Peixoto, and U. Nunes. Detection and Tracking of Moving Objects Using 2.5D Motion Grids. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 788–793, 2015.

[12] 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.*, pages 330–335, 2013.

[13] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Hoboken, NJ, USA, 2004.

[14] R. Barea, C. Perez, L. M. Bergasa, E. Lopez-Guillen, E. Romera, E. Molinos, M. Ocana, and J. Lopez. Vehicle Detection and Localization using 3D LIDAR Point Cloud and Image Semantic Segmentation. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 3481–3486, 2018.

[15] M. Beard, S. Reuter, K. Granström, B.-T. Vo, B.-N. Vo, and A. Scheel. Multiple Extended Target Tracking with Labelled Random Finite Sets. *IEEE Trans. Signal Process.*, 64(7):1638–1653, Apr. 2016.

[16] 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*, pages 850–865, 2016.

[17] S. S. Blackman. Multiple hypothesis tracking for multiple target tracking. *IEEE Aerosp. Electron. Syst. Mag.*, 19(1):5–18, Jan. 2004.

[18] T. Bonkenburg. Robotics in Logistics. Technical report, DHL Customer Solutions & Innovation, Mar. 2016. `https://www.dhl.com/content/dam/downloads/g0/about_us/logistics_insights/dhl_trendreport_robotics.pdf`.

[19] N. Bose. Wal-Mart's new robots scan shelves to restock items faster. Technical report, Reuters, Oct. 2017. `https://www.reuters.com/article/us-usa-walmart-robots/wal-marts-new-robots-scan-shelves-to-restock-items-faster-idUSKBN1CV1N4`.

[20] 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.*, pages 294–300, 2010.

[21] J. E. Bresenham. Algorithm for Computer Control of a Digital Plotter. *IBM Syst. J.*, 4(1):25–30, Mar. 1965.

[22] K. C. Chang, R. K. Saha, and Y. Bar-Shalom. On optimal track-to-track fusion. *IEEE Trans. Aerosp. Electron. Syst*, 33(4):1271–1276, Oct. 1997.

[23] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 77–85, 2017.

[24] R. O. Chavez-Garcia and O. Aycard. Multiple Sensor Fusion and Classification for Moving Object Detection and Tracking. *IEEE Trans. Intell. Transp. Syst.*, 17(2):525–534, Feb. 2016.

[25] V. Chen, Fayin Li, Shen-Shyang Ho, and H. Wechsler. Micro-doppler effect in radar: Phenomenon, model, and simulation study. *IEEE Trans. Aerosp. Electron. Syst.*, 42(1):2–21, Jan. 2006.

[26] C. Coué, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere. Bayesian Occupancy Filtering for Multitarget Tracking: An Automotive Application. *Int. J. Robot. Res.*, 25(1):19–30, Jan. 2006.

[27] F. C. Crow. Summed-area Tables for Texture Mapping. In *Proc. Comput. Graph. and Interactive Techniques Conf.*, SIGGRAPH '84, pages 207–212, 1984.

[28] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 886–893, 2005.

[29] R. Danescu, F. Oniga, and S. Nedevschi. Modeling and Tracking the Driving Environment With a Particle-Based Occupancy Grid. *IEEE Trans. Intell. Transp. Syst.*, 12(4):1331–1342, Dec. 2011.

[30] R. Danescu, C. Pantilie, F. Oniga, and S. Nedevschi. Particle grid tracking system stereovision based obstacle perception in driving environments. *IEEE Intell. Transp. Syst. Mag.*, 4(1):6–20, Jan. 2012.

[31] M. H. Daraei, A. Vu, and R. Manduchi. Velocity and shape from tightly-coupled LiDAR and camera. In *Proc. IEEE Intell. Veh. Symp.*, pages 60–67, 2017.

[32] F. E. Daum. Extended kalman filters. In J. Baillieul and T. Samad, editors, *Encyclopedia of Systems and Control*, pages 411–413. Springer London, London, 2015.

[33] P. Del Moral. Non-linear filtering: Interacting particle resolution. *Markov Process. Relat. Fields*, 2(4):555–581, 1996.

[34] A. P. Dempster. Upper and Lower Probabilities Induced by a Multivalued Mapping. *Ann. Math. Statist.*, 38(2):325–339, Apr. 1967.

[35] A. P. Dempster. A Generalization of Bayesian Inference. *J. Roy. Statist. Soc. Ser. B (Methodol.)*, 30(2):205–247, 1968.

[36] K. C. J. Dietmayer, S. Reuter, and D. Nuss. Representation of Fused Environment Data. In H. Winner, S. Hakuli, F. Lotz, and C. Singer, editors, *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort*, pages 567–603. Springer Int. Publishing, Cham, Switzerland, 2016.

[37] A. Elfes. Using occupancy grids for mobile robot perception and navigation. *Computer*, 22(6):46–57, June 1989.

[38] 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.*, pages 3852–3858, 2018.

[39] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov. Scalable Object Detection Using Deep Neural Networks. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 2155–2162, 2014.

[40] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Kdd*, volume 96, pages 226–231, 1996.

[41] T. E. Fortmann, Y. Bar-Shalom, and M. Scheffe. Multi-target tracking using joint probabilistic data association. In *Proc. IEEE Conf. on Decision and Control Including the Symp. on Adaptive Processes*, pages 807–812, 1980.

[42] GNU. GNU Scientific Library – Reference Manual: Weighted Samples. `https://www.gnu.org/software/gsl/doc/html/statistics.html#weighted-samples`. Online; retrieved 2018-10-02.

[43] K. Granström, M. Baum, and S. Reuter. Extended Object Tracking: Introduction, Overview, and Applications. *J. Adv. Inf. Fusion*, 12(2):138–174, Dec. 2017.

[44] P. S. Heckbert. Survey of Texture Mapping. *IEEE Comput. Graph. Appl.*, 6(11):56–67, Nov. 1986.

[45] M. Himmelsbach, A. Muller, and T. Luttel. LIDAR-based 3D Object Perception. In *Proc. Int. Workshop on Cognition for Technical Systems*, volume 1, page 7, 2008.

[46] M. Himmelsbach, F. v. Hundelshausen, and H.-J. Wuensche. Fast segmentation of 3D point clouds for ground vehicles. In *Proc. IEEE Intell. Veh. Symp.*, pages 560–565, 2010.

[47] P. Hitczenko and S. Kwapień. On the Rademacher Series. In J. Hoffmann-Jørgensen, J. Kuelbs, and M. B. Marcus, editors, *Probability in Banach Spaces, 9*, pages 31–36. Birkhäuser Boston, 1994.

[48] S. Hoermann, P. Henzler, M. Bach, and K. Dietmayer. Object Detection on Dynamic Occupancy Grid Maps Using Deep Learning and Automatic Label Generation. In *Proc. IEEE Intell. Veh. Symp.*, pages 826–833, 2018.

[49] F. Homm. *Fahrzeugeigenlokalisation Im Kontext Hochautomatisierter Fahrfunktionen*. PhD thesis, Technische Universität München. Lehrstuhl für Echtzeitsysteme und Robotik, 2012.

[50] F. Homm, N. Kaempchen, J. Ota, and D. Burschka. Efficient occupancy grid computation on the GPU with lidar and radar for road boundary detection. In *Proc. IEEE Intell. Veh. Symp.*, pages 1006–1013, 2010.

[51] J. Honer and H. Hettmann. Motion State Classification for Automotive LIDAR Based on Evidential Grid Maps and Transferable Belief Model. In *Proc. Int. Conf. Inform. Fusion*, pages 1069–1076, 2018.

[52] M. Hutson. Watch just a few self-driving cars stop traffic jams. Technical report, Science Magazine, Nov. 2018. `https://www.sciencemag.org/news/2018/11/watch-just-few-self-driving-cars-stop-traffic-jams`.

[53] S. Hyken. Walmart To Use Drones In Stores To Provide Better Customer Service. Technical report, Forbes, June 2018. `https://www.forbes.com/sites/shephyken/2018/06/17/walmart-to-use-drones-in-stores-to-provide-better-customer-service/#146d5dfd4e84`.

[54] IFR International Federation of Robotics. Executive Summary World Robotics 2018 Service Robots. Technical report, IFR International Federation of Robotics, 2018. `https://ifr.org/downloads/press2018/Executive_Summary_WR_Service_Robots_2018.pdf`.

[55] S. J. Julier and J. K. Uhlmann. Unscented filtering and nonlinear estimation. *Proc. IEEE*, 92(3):401–422, 2004.

[56] R. Jungnickel and F. Korf. Object tracking and dynamic estimation on evidential grids. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 2310–2316, 2014.

[57] R. E. Kalman. A new approach to linear filtering and prediction problems. *J. Basic Eng.*, 82(1):35–45, 1960.

[58] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei. Large-Scale Video Classification with Convolutional Neural Networks. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 1725–1732, 2014.

[59] D. Kellner. *Verfahren zur Bestimmung von Objekt- und Eigenbewegung auf Basis der Dopplerinformation hochauflösender Radarsensoren*. PhD thesis, Universität Ulm, 2017.

[60] 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.*, pages 324–329, 2014.

[61] 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.*, 17(5):1341–1353, May 2016.

[62] D. Kellner, J. Klappstein, and K. Dietmayer. Grid-based DBSCAN for clustering extended objects in radar data. In *Proc. IEEE Intell. Veh. Symp.*, pages 365–370, June 2012.

[63] K. Klasing, D. Wollherr, and M. Buss. A clustering method for efficient segmentation of 3D laser data. In *Proc. IEEE Int. Conf. Robot. Autom.*, pages 4043–4048, May 2008.

[64] E. F. Knott, J. F. Schaeffer, and M. T. Tulley. *Radar Cross Section*. SciTech Publishing, 2004.

[65] P. Konstantinova, A. Udvarev, and T. Semerdjiev. A study of a target tracking algorithm using global nearest neighbor approach. In *Proc. Int. Conf. on Comput. Syst. and Technol.*, pages 290–295, 2003.

[66] M. Koschi, C. Pek, M. Beikirch, and M. Althoff. Set-Based Prediction of Pedestrians in Urban Environments Considering Formalized Traffic Rules. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 2704–2711, 2018.

[67] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, May 2017.

[68] R. Kumar and Richa. Autonomous Vehicle Market Size, Share and Growth Analysis — Forecast. Technical report, Allied Market Research, May 2018. `https://www.alliedmarketresearch.com/press-release/autonomous-vehicle-market.html`.

[69] M. Kurdej, J. Moras, V. Cherfaoui, and P. Bonnifait. Map-Aided Evidential Grids for Driving Scene Understanding. *IEEE Intell. Transp. Syst. Mag.*, 7(1):30–41, Jan. 2015.

[70] E. Lefevre, O. Colot, and P. Vannoorenberghe. Belief function combination and conflict management. *Inf. Fusion*, 3(2):149–162, 2002.

[71] X. R. Li and V. Jilkov. Survey of Maneuvering Target Tracking. Part I: Dynamic Models. *IEEE Trans. Aerosp. Electron. Syst.*, 39(4):1333–1364, Oct. 2003.

[72] D. Lowe. Object recognition from local scale-invariant features. In *Proc. IEEE Int. Conf. Comput. Vision*, pages 1150–1157 vol.2, 1999.

[73] P. C. Mahalanobis. On the generalized distance in statistics. *Proc. Natl. Inst. Sci. India*, 2:49–55, 1936.

[74] R. Mahler. A brief survey of advances in random-set fusion. In *Proc. Int. Conf. on Control, Automation and Information Sciences*, pages 62–67, 2015.

[75] A. Manjunath, Y. Liu, B. Henriques, and A. Engstle. Radar Based Object Detection and Tracking for Autonomous Driving. In *Proc. IEEE MTT-S Int. Conf. Microwaves for Intell. Mobility*, pages 1–4, 2018.

[76] 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*, pages 69–81. Springer, Berlin, Germany, 2009.

[77] K. Mekhnacha, Y. Mao, D. Raulo, and C. Laugier. Bayesian occupancy filter based "Fast Clustering-Tracking" algorithm. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sept. 2008.

[78] A. Mendes, L. Bento, and U. Nunes. Multi-target detection and tracking with a laserscanner. In *Proc. IEEE Intell. Veh. Symp.*, pages 796–801, Parma, Italy, 2004.

[79] C. Mertz, L. E. Navarro-Serment, R. MacLachlan, P. Rybski, A. Steinfeld, A. Suppé, C. Urmson, N. Vandapel, M. Hebert, C. Thorpe, D. Duggins, and J. Gowdy. Moving object detection with laser scanners: Moving Object Detection with Laser Scanners. *J. Field Robot.*, 30(1):17–43, Jan. 2013.

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

[81] J. Moras, V. Cherfaoui, and P. Bonnifait. Moving Objects Detection by Conflict Analysis in Evidential Grids. In *Proc. IEEE Intell. Veh. Symp.*, pages 1122–1127, 2011.

[82] H. Nam and B. Han. Learning Multi-domain Convolutional Neural Networks for Visual Tracking. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 4293–4302, 2016.

[83] A. Nègre, L. Rummelhard, and C. Laugier. Hybrid Sampling Bayesian Occupancy Filter. In *Proc. IEEE Intell. Veh. Symp.*, pages 1307–1312, 2014.

[84] D. Nuss. *A Random Finite Set Approach for Dynamic Occupancy Grid Maps*. Ph.D. dissertation, Faculty of Engineering, Computer Science and Psychology, Ulm University, Ulm, Germany, 2017.

[85] D. Nuss, S. Reuter, M. Thom, T. Yuan, G. Krehl, M. Maile, A. Gern, and K. Dietmayer. A random finite set approach for dynamic occupancy grid maps with real-time application. *Int. J. Robot. Res.*, 37(8):841–866, July 2018.

[86] 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.*, pages 1371–1377, 2014.

[87] D. Nuss, M. Thom, A. Danzer, and K. Dietmayer. Fusion of laser and monocular camera data in object grid maps for vehicle environment perception. In *Proc. Int. Conf. Inform. Fusion*, pages 1–8, 2014.

[88] 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.*, pages 1074–1081, 2015.

[89] NVIDIA Corporation. CUDA Zone Developer Webpage. `https://developer.nvidia.com/cuda-zone`. Online; retrieved 2020-05-12.

[90] A. Ohnsman. The End Of Parking Lots As We Know Them: Designing For A Driverless Future. Technical report, Forbes, May 2018. `https://www.forbes.com/sites/alanohnsman/2018/05/18/end-of-parking-lot-autonomous-cars/`.

[91] D. Pagac, E. Nebot, and H. Durrant-Whyte. An evidential approach to map-building for autonomous vehicles. *IEEE Trans. Robot. Automat.*, 14(4):623–629, Aug. 1998.

[92] C. Pek and M. Althoff. Computationally Efficient Fail-safe Trajectory Planning for Self-driving Vehicles Using Convex Optimization. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 1447–1454, 2018.

[93] C. Pek and M. Althoff. Efficient Computation of Invariably Safe States for Motion Planning of Self-Driving Vehicles. In *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, pages 3523–3530, 2018.

[94] A. Petrovskaya and S. Thrun. Model based vehicle detection and tracking for autonomous urban driving. *Auton. Robots*, 26(2-3):123–139, Apr. 2009.

[95] 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.*, pages 392–398, 2017.

[96] D. Reid. An algorithm for tracking multiple targets. *IEEE Trans. Autom. Control*, 24(6):843–854, Dec. 1979.

[97] 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.*, 39(6):1137–1149, June 2017.

[98] S. Reuter, A. Danzer, M. Stübler, A. Scheel, and K. Granström. A fast implementation of the Labeled Multi-Bernoulli filter using gibbs sampling. In *Proc. IEEE Intell. Veh. Symp.*, pages 765–772, 2017.

[99] S. Reuter, B.-T. Vo, B.-N. Vo, and K. Dietmayer. Multi-Object Tracking Using Labeled Multi-Bernoulli Random Finite Sets. In *Annual Conf. Inf. Fusion*, page 8, 2014.

[100] S. Reuter, B. Wilking, J. Wiest, M. Munz, and K. Dietmayer. Real-time multi-object tracking using random finite sets. *IEEE Trans. Aerosp. Electron. Syst.*, 49(4):2666–2678, Oct. 2013.

[101] L. Rummelhard, A. Nègre, and C. Laugier. Conditional Monte Carlo Dense Occupancy Tracker. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 2485–2490, 2015.

[102] M. Saval-Calvo, L. Medina-Valdés, J. M. Castillo-Secilla, S. Cuenca-Asensi, A. Martínez-Álvarez, and J. Villagrá. A Review of the Bayesian Occupancy Filter. *Sensors*, 17(2):344, 2017.

[103] 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*, pages 1–4, 2018.

[104] M. Schreier, V. Willert, and J. Adamy. Grid mapping in dynamic road environments: Classification of dynamic cell hypothesis via tracking. In *Proc. IEEE Int. Conf. Robot. Autom.*, pages 3995–4002, 2014.

[105] R. Schubert, E. Richter, and G. Wanielik. Comparison and evaluation of advanced motion models for vehicle tracking. In *Proc. Int. Conf. Inform. Fusion*, pages 1–6, 2008.

[106] 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.*, pages 692–697, 2012.

[107] J. Schulz, C. Hubmann, N. Morin, J. Lochner, and D. Burschka. Learning Interaction-Aware Probabilistic Driver Behavior Models from Urban Scenarios. In *Proc. IEEE Intell. Veh. Symp.*, pages 1326–1333, 2019.

[108] M. Schütz, N. Appenrodt, J. Dickmann, and K. Dietmayer. Occupancy grid map-based extended object tracking. In *Proc. IEEE Intell. Veh. Symp.*, pages 1205–1210, 2014.

[109] C. Seeger, M. Manz, P. Matters, and J. Hornegger. 2-D Evidential Grid Mapping with Narrow Vertical Field of View Sensors Using Multiple Hypotheses and Spatial Neighborhoods. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 1843–1848, 2015.

[110] C. Seeger, M. Manz, P. Matters, and J. Hornegger. Locally adaptive discounting in multi sensor occupancy grid fusion. In *Proc. IEEE Intell. Veh. Symp.*, pages 266–271, 2016.

[111] G. Shafer. *A Mathematical Theory of Evidence*. Princeton Univ. Press, Princeton, NJ, USA, 1976.

[112] M. Skolnik. *Radar Handbook, Third Edition*. Electronics Electrical Engineering. McGraw-Hill Education, 2008.

[113] P. Smets and R. Kennes. The transferable belief model. *Artif. Intell.*, 66(2):191–234, 1994.

[114] Starship Technologies. `https://starship.xyz/news/`. Online; retrieved 2019-05-28.

[115] P. Stepan, M. Kulich, and L. Preucil. Robust data fusion with occupancy grid. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.*, 35(1):106–115, Feb. 2005.

[116] S. Steyer. *Multisensorielle Grid-basierte Umfeldmodellierung im Bereich des automatisierten Fahrens*. Masterthesis, Karlsruhe Institute of Technology, Institute of Systems Optimization, Karlsruhe, Germany, 2015.

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

[118] 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.

[119] G. Tanzmeister and S. Steyer. Verfahren zur Fusion von Kamera-Sensordaten mit Abstandsmessender Sensorik. German Patent DE102018127990A1, Assignee: Bayerische Motoren Werke AG, Filing Date: 2018-08-11.

[120] 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.*, pages 6090–6095, 2014.

[121] G. Tanzmeister and D. Wollherr. Evidential Grid-Based Tracking and Mapping. *IEEE Trans. Intell. Transp. Syst.*, 18(6):1454–1467, June 2017.

[122] M. K. Tay, K. Mekhnacha, M. Yguel, C. Coue, C. Pradalier, C. Laugier, T. Fraichard, and P. Bessiere. The Bayesian occupation filter. In *Probabilistic Reasoning and Decision Making in Sensory-Motor Systems*, pages 77–98. Springer, Berlin, Germany, 2008.

[123] A. Tchamova and J. Dezert. On the behavior of Dempster's rule of combination and the foundations of Dempster-Shafer Theory. In *Proc. IEEE Int. Conf. Intell. Sys.*, pages 108–113, Sofia, Bulgaria, 2012.

[124] J. Thomas, J. Tatsch, W. van Ekeren, R. Rojas, and A. Knoll. Semantic Grid-Based Road Model Estimation for Autonomous Driving. In *Proc. IEEE Intell. Veh. Symp.*, pages 2329–2336, 2019.

[125] K. Thormann, J. Honer, and M. Baum. Extended Target Tracking Using Gaussian Processes with High-Resolution Automotive Radar. In *Proc. Int. Conf. Inform. Fusion*, pages 1764–1770, 2018.

[126] S. Thrun. Learning occupancy grids with forward models. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 1676–1681 vol.3, 2001.

[127] S. Thrun. Simultaneous Localization and Mapping. In M. E. Jefferies and W.-K. Yeap, editors, *Robotics and Cognitive Approaches to Spatial Mapping*, Springer Tracts in Advanced Robotics, pages 13–41. Springer, Berlin, Heidelberg, 2008.

[128] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, Cambridge, MA, USA, 2005.

[129] T. N. Tran, K. Drab, and M. Daszykowski. Revised DBSCAN algorithm to cluster data with dense adjacent clusters. *Chemom. Intell. Lab. Syst.*, 120:92–96, Jan. 2013.

[130] M. Valente, C. Joly, and A. de la Fortelle. Fusing Laser Scanner and Stereo Camera in Evidential Grid Maps. In *Proc. IEEE International Conference on Control, Automation, Robotics and Vision*, pages 990–997, 2018.

[131] A. Vatavu, N. Rexin, S. Appel, T. Berling, S. Govindachar, G. Krehl, J. Peukert, M. Schier, O. Schwindt, J. Siegel, C. Zalidis, T. Rehfeld, D. Nuss, M. Maile, S. Zimmermann, K. Dietmayer, and A. Gern. Environment Estimation with Dynamic Grid Maps and Self-Localizing Tracklets. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 3370–3377, 2018.

[132] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages I–511–I–518, 2001.

[133] B.-N. Vo, B.-T. Vo, and D. Phung. Labeled Random Finite Sets and the Bayes Multi-Target Tracking Filter. *IEEE Trans. Signal Process.*, 62(24):6554–6567, Dec. 2014.

[134] T. D. Vu and O. Aycard. Laser-based detection and tracking moving objects using data-driven Markov chain Monte Carlo. In *Proc. IEEE Int. Conf. Robot. Autom.*, pages 3800–3806, 2009.

[135] 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.*, pages 190–195, 2007.

[136] 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.*, pages 684–689, 2008.

[137] C.-C. Wang and C. Thorpe. Simultaneous localization and mapping with detection and tracking of moving objects. In *Proc. IEEE Int. Conf. Robot. Autom.*, volume 3, pages 2918–2924, 2002.

[138] L. Wang, W. Ouyang, X. Wang, and H. Lu. Visual Tracking with Fully Convolutional Networks. In *Proc. IEEE Int. Conf. Comput. Vision*, pages 3119–3127, 2015.

[139] L. Wang, J. Tang, and Q. Liao. A study on radar target detection based on deep neural networks. *IEEE Sens. Lett.*, 3(3):1–4, Mar. 2019.

[140] J. W. Wedel, B. Schünemann, and I. Radusch. V2X-based traffic congestion recognition and avoidance. In *Proc. Int. Symp. Pervasive Systems, Algorithms, and Networks*, pages 637–641, 2009.

[141] E. W. Weisstein. Trigonometric Addition Formulas. From MathWorld – A Wolfram Web Resource. `https://mathworld.wolfram.com/TrigonometricAdditionFormulas.html`. Online; retrieved 2020-04-04.

[142] H. Winner. Radarsensorik. In H. Winner, S. Hakuli, F. Lotz, and C. Singer, editors, *Handbuch Fahrerassistenzsysteme*, pages 259–316. Springer Fachmedien Wiesbaden, Wiesbaden, 2015.

[143] 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.*, pages 3530–3535, 2018.

[144] World Health Organization. *Global Status Report on Road Safety 2018*. World Health Organization, S.l., 2019.

[145] C. Wu, A. Kreidieh, K. Parvate, E. Vinitsky, and A. M. Bayen. Flow: Architecture and Benchmarking for Reinforcement Learning in Traffic Control. *ArXiv171005465 Cs*, Oct. 2017. `http://arxiv.org/abs/1710.05465`.

[146] X Rong Li and V. Jilkov. Survey of maneuvering target tracking. part v: Multiple-model methods. *IEEE Trans. Aerosp. Electron. Syst.*, 41(4):1255–1321, Oct. 2005.

[147] M. Yguel, O. Aycard, and C. Laugier. Efficient GPU-based construction of occupancy grids using several laser range-finders. *Int. J. Veh. Auton. Syst.*, 6(1/2):48, 2008.

[148] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comput. Surv.*, 38(4):13–es, Dec. 2006.

[149] 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.*, pages 603–608, 2016.

[150] F. Zhang, D. Clarke, and A. Knoll. Vehicle detection based on LiDAR and camera fusion. In *Proc. IEEE Intell. Transp. Syst. Conf.*, pages 1620–1625, 2014.

[151] G. Zhang, H. Li, and F. Wenger. Object Detection and 3D Estimation via an FMCW Radar Using a Fully Convolutional Network. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, pages 4487–4491, 2020.

[152] X. Zhang, W. Xu, C. Dong, and J. M. Dolan. Efficient L-shape fitting for vehicle detection using laser scanners. In *Proc. IEEE Intell. Veh. Symp.*, pages 54–59, 2017.

[153] Y. Zhou and O. Tuzel. VoxelNet: End-to-End Learning for Point Cloud Based 3D Object Detection. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 4490–4499, 2018.

VDI verlag

Alle 23 Reihen der „Fortschritt-Berichte VDI“ in der Übersicht. Bequem recherchieren und bestellen unter:

Gebundene Ausgabe:
www.vdi-nachrichten.com/shop

Digitale Ausgabe:
elibrary.vdi-verlag.de

Reihe 01 Konstruktionstechnik / Maschinenelemente
Reihe 02 Fertigungstechnik
Reihe 03 Verfahrenstechnik
Reihe 04 Bauingenieurwesen
Reihe 05 Grund- und Werkstoffe/Kunststoffe
Reihe 06 Energietechnik
Reihe 07 Strömungstechnik
Reihe 08 Mess-, Steuerungs- und Regelungstechnik
Reihe 09 Elektronik/Mikro- und Nanotechnik
Reihe 10 Informatik/Kommunikation
Reihe 11 Schwingungstechnik
Reihe 12 Verkehrstechnik/Fahrzeugtechnik
Reihe 13 Fördertechnik/Logistik
Reihe 14 Landtechnik/Lebensmitteltechnik
Reihe 15 Umwelttechnik
Reihe 16 Technik und Wirtschaft
Reihe 17 Biotechnik/Medizintechnik
Reihe 18 Mechanik/Bruchmechanik
Reihe 19 Wärmetechnik/Kältetechnik
Reihe 20 Rechnergestützte Verfahren
Reihe 21 Elektrotechnik
Reihe 22 Mensch-Maschine-Systeme
Reihe 23 Technische Gebäudeausrüstung

INGENIEUR.de
TECHNIK - KARRIERE - NEWS

powered by VDI Verlag

Starten Sie durch – auf INGENIEUR.de!

Das TechnikKarriereNews-Portal für Ingenieure und IT-Ingenieure.

Was immer Sie für Ihre Karriere brauchen – Sie finden es auf ingenieur.de: Auf Sie zugeschnittene Infos und Services, Stellenangebote in der Jobbörse, Firmenprofile, Fachartikel, Gehaltstest, Bewerbungstipps, Newsletter und alles zu den VDI nachrichten Recruiting Tagen.

VDI

REIHE 08
MESS-,
STEUERUNGS-
UND REGELUNGS-
TECHNIK

NR. 1272

ISBN 978-3-18-527208-0

BAND
1|1

VOLUME
1|1