

# **Multi-Sensor Environment Perception for Automated Vehicles with Semantic Evidential Grid Maps**

Zur Erlangung des akademischen Grades eines

**DOKTORS DER INGENIEURWISSENSCHAFTEN  
(Dr.-Ing.)**

von der KIT-Fakultät für Maschinenbau des  
Karlsruher Instituts für Technologie (KIT)  
angenommene

**DISSERTATION**

von

**Sven Richter, M.Sc.**

Tag der mündlichen Prüfung:

08.12.2022

Hauptreferent:  
Korreferent:

Prof. Dr.-Ing. Christoph Stiller  
Prof. Dr.-Ing. Klaus Dietmayer



# Danksagung

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Mess- und Regelungstechnik (MRT) des Karlsruher Instituts für Technologie (KIT). Ohne die vielfältige Unterstützung meiner Kollegen, Freunde und Familie wäre diese Arbeit nicht möglich gewesen.

Ein großer Dank geht an Herrn Prof. Christoph Stiller für die Betreuung meiner Arbeit. Das fachliche Feedback während der Promotion sowie die exzellenten Bedingungen am Institut die entwickelten Methoden selbst auf dem Testfahrzeug anwenden zu können, haben diese Arbeit getragen. Ebenfalls möchte ich mich bei Herrn Prof. Klaus Dietmayer für das Korreferat bedanken sowie bei Frau Prof. Jivka Ovtcharova für die Übernahme des Vorsitzes des Promotionsprüfungsausschusses.

Ein besonderer Dank gilt allen Kolleginnen und Kollegen des MRT, mit denen ich entweder in irgendeiner Form zusammengearbeitet oder abseits der Arbeit Zeit verbracht habe. Zu aller erst möchte ich hier Sascha und Johannes nennen, die mir als erfahrene Kollegen unglaublich viel mitgegeben haben. Durch euren Support in vielen wissenschaftlichen Gesprächen und eure Hilfe in programmiertechnischen Fragen habt ihr einen großen Anteil an dieser Arbeit. Vielen Dank auch an Carlos für deine Unterstützung als Gruppenleiter und an Jan für deinen wissenschaftlichen Input und dafür, dass du auf jede Frage mit drei Papertiteln antworten kannst. Ein weiterer Dank geht an alle, mit denen ich am Adenauerprojekt zusammengearbeitet habe, insbesondere an Eduardo, Fabian, Hao, Jannik, Pio, Christoph und Kevin. Ein besonderer Dank geht auch an die Probeleser dieser Arbeit Jan, Frank, Max und Sascha. Ebenfalls möchte ich mich bei dem Sekretariat des MRT, allen voran bei Erna bedanken, ohne die ich wohl entweder ein paar Dienstreisen selbst bezahlt hätte oder die Arbeit vorzeitig hätte abrechnen müssen, weil ich die Frist für meine Vertragsverlängerung verpasst hätte.

---

Abschließend möchte ich mich bei allen bedanken, die mich außerhalb der Arbeit kennen und mich während der Promotion ausgehalten und unterstützt haben. Das gilt vor allem für meine Familie und meine Frau Vanessa, die mir während dieser Zeit den Rücken freigehalten hat.

Vielen Dank!

Karlsruhe im Dezember 2022

*Sven Richter*



# Kurzfassung

Die automatische Navigation von Fahrzeugen hat das Potenzial, Leben zu retten und den Verkehr effizienter zu gestalten. Ein zentraler Bestandteil algorithmischer Lösungen ist die Wahrnehmung der relevanten Umgebung des automatisierten Fahrzeugs. Die gewünschte Repräsentation enthält Informationen über Verkehrsteilnehmer, deren semantischen Zustand und deren Bewegungszustand. Zusätzlich müssen Beobachtbarkeit, Freiraum und Fahrbarkeit bekannt sein, um das Fahrzeug sicher zu navigieren. Um Messungen zu sammeln, aus denen diese Informationen abgeleitet werden können, sind automatisierte Fahrzeuge mit heterogenen Sensoren wie LiDAR, Kameras und RaDAR ausgestattet.

In dieser Arbeit wird eine Methode zur Umgebungswahrnehmung vorgestellt, die zur Lösung dieser Aufgaben entwickelt wurde. Die erzeugte Repräsentation ist eine mehrschichtige Top-View Grid Map, bei der die Unsicherheit mittels Evidenztheorie modelliert wird. Im Vergleich zur bisherigen Forschung auf diesem Gebiet führt diese Arbeit die folgenden Neuerungen ein: Der Belegungszustandsraum, der traditionell aus den Hypothesen *belegt* und *frei* besteht, wird durch ein hybrides Evidenzmodell ersetzt, das den semantischen und den dynamischen Zustand der Belegung enthält. Außerdem wird der semantische Zustand des Bodens separat modelliert. Im ersten Verarbeitungsschritt wird die Grid Map-Darstellung für jeden Sensor separat geschätzt. Wir verarbeiten LiDAR- und Kamera-Messungen, die als Bilder vorliegen, und leiten daraus Belegungsdaten ab, indem wir die Oberflächenorientierung für jede Sensorreflexion analysieren. Dies macht die Schätzung eines parametrischen Bodenmodells, das für konkurrierende Methoden notwendig ist, überflüssig. Für die Sensordatenfusion wird eine gewichtete Kombination der sensorspezifischen Evidenzen vorgeschlagen, welche nachweislich zu einer besseren Auflösung von Konflikten führt als die in anderen Veröffentlichungen verwendeten Kombinationsregeln. Schließlich werden die fusionierten Gitterkarten in ein temporäres Fusionsmodul eingespeist, das rekursiv evidenzbasierte Grid Maps aktualisiert und dabei Informationen über die Bodensemantik, sowie den semantischen und dynamischen Zustand

---

der Zellbelegungen akkumuliert. Hier werden Evidential Networks für die Aktualisierung der Evidenzverteilungen auf den Hypothesenmengen genutzt. Dies ermöglicht es, Abhängigkeiten zwischen semantischen und dynamischen Zuständen explizit zu modellieren.

Die Vorteile der vorgeschlagenen Methodik werden in realen Verkehrsszenarien anhand von Messungen von einem LiDAR und einer Stereo Kamera demonstriert.

# Abstract

The automatic navigation of vehicles has the potential to save lives and make traffic more efficient. A central component of algorithmic solutions is the perception of the environment in the relevant surroundings of the automated vehicle. The desired representation contains information about road users, their semantic state and their state of motion. In addition, observability, free space and drivability must be known in order to safely navigate the vehicle. To collect measurements from which this information can be derived, automated vehicles are equipped with heterogeneous sensors such as LiDARs, Cameras and RaDARs.

In this thesis, we present an environment perception method developed to solve these tasks. The representation generated is a multi-layer top-view grid map where uncertainty is modeled using evidence theory. Compared to previous research in this area, this work introduces the following innovations: The occupancy state space, which traditionally consists of the hypotheses *occupied* and *free*, is replaced by a hybrid evidential model that includes the semantic and dynamic states of cell occupancy. Furthermore, the semantic state of the ground is modeled separately. In the first processing step, the grid map representation is estimated for each sensor separately. We process LiDAR and camera measurements, available as images, and derive occupancy evidence by analyzing the surface orientation for each sensor reflection. This eliminates the need to estimate a parametric ground model, which is necessary for competing methods. For sensor data fusion, weighted evidential reasoning is proposed, which is shown to resolve conflicts better than the combination rules used in other publications. Finally, the fused grid maps are fed into a temporal fusion module that recursively updates evidential grid maps accumulating information about occupancy semantics, ground semantics and the dynamics of occupied grid cells. Here, evidential networks are exploited for updating belief mass distributions on the hypotheses sets. This allows dependencies between semantic and dynamic states to be explicitly modeled.

---

The advantages of the proposed methodology are demonstrated in real traffic scenarios using LiDAR and stereo camera measurements.

# Table Of Contents

<b>Abbreviations and Symbols</b> . . . . .	<b>VII</b>
<b>1 Introduction</b> . . . . .	<b>1</b>
1.1 Environment Perception for Automated Vehicles . . . . .	2
1.2 Contributions of this Work . . . . .	5
<b>2 Evidential Grid Mapping</b> . . . . .	<b>9</b>
2.1 Fundamentals . . . . .	9
2.1.1 Evidence Theory . . . . .	9
2.1.2 Grid Maps . . . . .	12
2.2 Related Work . . . . .	14
2.3 Hybrid Evidential Representation . . . . .	16
2.4 Datasets and Evaluation Metrics . . . . .	21
<b>3 Sensor Measurement Mapping in Evidential Grid Maps</b> . . . . .	<b>27</b>
3.1 Fundamentals . . . . .	27
3.1.1 LiDARs . . . . .	27
3.1.2 Cameras . . . . .	29
3.1.3 Sensor Measurement Modeling on Grids . . . . .	29
3.2 Related Work . . . . .	30
3.2.1 Grid Mapping with Range Sensors . . . . .	30
3.2.2 Grid Mapping with Cameras . . . . .	31
3.3 Sensor Measurement Grid Mapping . . . . .	32
3.3.1 Grid Mapping with Point Sets . . . . .	35
3.3.2 Grid Mapping with Images . . . . .	36
3.3.3 Grid Mapping Considering Sensor Modalities . . . . .	46
3.4 Experiments . . . . .	48
3.4.1 Qualitative Results . . . . .	49

- 3.4.2 Quantitative Evaluation . . . . . 54
- 4 Sensor Data Fusion in Evidential Grid Maps . . . . . 61**
  - 4.1 Fundamentals . . . . . 62
  - 4.2 Related Work . . . . . 65
  - 4.3 Combining Evidential Grid Maps with Evidential Reasoning 66
    - 4.3.1 Conflict Adaptive Evidential Reasoning . . . . . 67
    - 4.3.2 Parameter Estimation . . . . . 67
  - 4.4 Experiments . . . . . 69
    - 4.4.1 Qualitative Results . . . . . 69
    - 4.4.2 Quantitative Evaluation . . . . . 71
- 5 Temporal Fusion in Evidential Grid Maps . . . . . 75**
  - 5.1 Fundamentals . . . . . 76
    - 5.1.1 Random Finite Sets . . . . . 76
    - 5.1.2 Evidential Networks . . . . . 78
  - 5.2 Related Work . . . . . 79
  - 5.3 Semantic Evidential Grid Mapping and Tracking . . . . . 81
    - 5.3.1 Particle Filter . . . . . 85
    - 5.3.2 Evidential Network Reasoning . . . . . 91
    - 5.3.3 Parameter Estimation . . . . . 98
  - 5.4 Experiments . . . . . 105
    - 5.4.1 Particle Filter . . . . . 105
    - 5.4.2 Semantic State . . . . . 109
    - 5.4.3 Dynamic State . . . . . 114
- 6 Conclusion . . . . . 119**
  - 6.1 Discussion . . . . . 119
  - 6.2 Outlook . . . . . 121
- References . . . . . 123**

# Abbreviations and Symbols

## Abbreviations

<b>2D</b>	two-dimensional
<b>3D</b>	three-dimensional
<b>BBA</b>	basic belief assignment
<b>DEVN</b>	directed evidential network with conditional belief functions
<b>eloU</b>	evidential intersection over union
<b>ENC</b>	evidential network with conditional belief functions
<b>ER</b>	evidential reasoning
<b>FoD</b>	frame of discernment
<b>GPU</b>	graphics processing unit
<b>LiDAR</b>	light detection and ranging
<b>PCR</b>	partial conflict redistribution
<b>PDF</b>	probability density function
<b>PHD</b>	probability hypothesis density
<b>PHD/MIB</b>	probability hypothesis density / multi-instance Bernoulli
<b>RaDAR</b>	radio detection and ranging
<b>RFS</b>	random finite set

## Operators and Notations

$v$	scalar variable
$\mathbf{v}$	vector variable
$\mathbf{V}$	matrix variable
$f$	scalar-valued function
$\mathcal{P}(X)$	the power set containing all subsets of the set $X$
$\inf(X)$	the infimum, i.e. the largest lower bound of an ordered set $X$
$\sup(X)$	the supremum, i.e. the smallest upper bound of an ordered set $X$
$\mathbb{E}(X)$	the expectation/first statistical moment of a random variable $X$
$\mu(\cdot)$	the two-dimensional (2D) Lebesgue-measure
$\mathbb{R}$	the set of real numbers
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	the normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$
$p$	often used for propability $0 \leq p \leq 1$
$\mathbf{p}$	position vector
$\mathbf{v}$	velocity vector
$v$	velocity absolute value, i.e. $\ \mathbf{v}\ $
$\mathbf{n}$	normal vector
$\mathcal{T}$	coordinate system transformation
$d_{\text{MHD}}(\cdot)$	the Mahalanobis distance (Equation (5.77))
$\mathcal{K}$	camera calibration containing intrinsic and extrinsic calibration parameters
$Z$	often used for measurements
$\rho$	the permeability (Equation (3.4))
$\mathcal{X}$	set of particles



---

$\chi$  one particle  $\chi \in \mathcal{X}$  (Equation (5.21))

## Evidence Theory

$m(\cdot)$  the basic belief assignment (BBA, Equation (2.1))

$\text{bel}(\cdot)$  the belief (Equation (2.6))

$\text{pl}(\cdot)$  the plausibility (Equation (2.6))

$e_D(\cdot)$  Deng's entropy (Equation (2.11))

$\text{ns}_D(\cdot)$  Deng's nonspecificity (Equation (2.12))

$d_D(\cdot)$  Deng's discord (Equation (2.13))

## Grid Mapping

$\mathcal{G}$  general regular grid

$\mathcal{G}_{xy}$  Cartesian grid

$C$  A grid cell  $C \in \mathcal{G}$

$h_Z$  the measurement grid map (Equation (3.11))

$g_Z$  the sensor measurement grid map (Equation (3.2))

IoU the intersection over union

TP the true positive rate

FP the false positive rate

FN the false negative rate

eIoU the evidential intersection over union (Equation (2.26))

eTP the evidential true positive rate (Equation (2.27))

eFP the evidential false positive rate (Equation (2.28))

eFN            the evidential false negative rate (Equation (2.29))

## Frames of Discernment

$\Omega_g$	the ground semantics frame of discernment (FoD)
$\Omega_d$	the occupancy dynamics FoD
$V_d$	the hypothesis “void” in the occupancy dynamics FoD
$F_d$	the hypothesis “free” in the occupancy dynamics FoD
$O_{du}$	the hypothesis “dynamically unclassified occupied”
$O_{mov}$	the hypothesis “occupied by a moving entity”
$O_{stat}$	the hypothesis “occupied by a stationary entity”
$P$	the hypothesis “passable”, i.e. free or occupied by a moving entity
$\Omega_s$	the occupancy semantics FoD
$V_s$	the hypothesis “void” in the occupancy semantics FoD
$F_s$	the hypothesis “free” in the occupancy semantics FoD
$O_{su}$	the hypothesis “semantically unclassified occupied”
$O_{car}$	the hypothesis “occupied by a car”
$O_{tw}$	the hypothesis “occupied by a two-wheeler”
$O_{ped}$	the hypothesis “occupied by a pedestrian”
$O_{om}$	the hypothesis “occupied by another mobile entity”
$O_{im}$	the hypothesis “occupied by an immobile entity”
$S$	set containing all singleton semantic hypotheses

# 1 Introduction

Considering the history of mobility where safety and efficiency have been continuously improved, the automation of transportation systems towards full autonomy is a logical next step. The successive automation of subtasks already started in the 1950s with the introduction of anti-lock braking systems (ABS). Over the last decades, driver assistance systems (ADAS) designed to support the driver have been installed in production vehicles, enabling automation levels 1 and 2 according to the SAE standard [On-21]. However, 38,824 traffic fatalities were counted in the United States alone in 2020, many of them involving human choice situations such as speeding and alcohol consumption, see [Ste22]. The full automation of all driving components has the potential to prevent a majority of those accidents. Besides a significant reduction of fatal traffic accidents, the automation of transportation systems also aims to make mobility more efficient. Robo-taxis could revolutionize mobility by improving accessibility in rural areas and reducing the space required for parking lots in cities. Furthermore, traffic congestion may be reduced by cooperative speed adaption and vehicle-to-everything (V2X) communication.

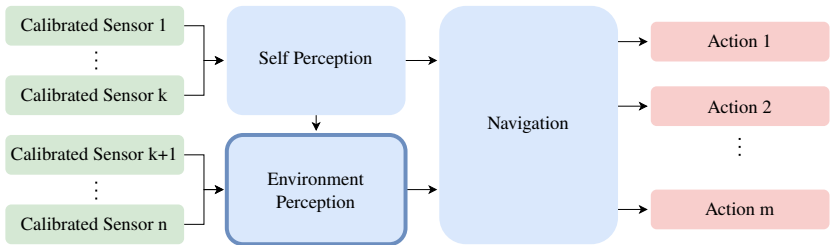


Figure 1.1: The environment perception module in context with a simplified software stack for automated vehicles. A more detailed consideration of a full software architecture can be found e.g. in [MM15].

A simplified software scheme for intelligent vehicles is shown in Figure 1.1. Input to the processing chain is a set of measurements recorded by  $n$  calibrated sensors. That means that the intrinsic sensor specifications and the extrinsic calibration are known. Extrinsic calibration refers to the pose estimation of a sensor with respect to a common coordinate system. The calibration can be conducted offline by solving an optimization problem based on sensor measurements recorded in a clearly defined scenario such as described in [BS18; Küm20]. As the mounting pose of the sensor may change slightly over time, online calibration methods as in [Xu+19] have also been proposed recently. The state of the ego vehicle is estimated in the self perception module using the sensor measurements from the calibrated sensors 1 to  $k$ . This state consists of at least the pose of the ego vehicle with respect to an ego motion compensated, or a map-fixed coordinate system. The ego motion compensation is usually done with inertial measurement units (IMU). Localization in a map can be obtained with GPS or by matching map features with features observed during the drive, as described in [Son20]. The estimated pose of the ego vehicle as well as measurements from sensors observing the environment, such as cameras, RaDARs and LiDARs, are fed into the environment perception module. The calculated environment model is sent to the scene understanding, behavior generation, planning and control which are combined here in the navigation module. Finally, the actions generated by the navigation module are sent to the vehicle's actuators.

## 1.1 Environment Perception for Automated Vehicles

The environment perception module calculates an environment model based on measurements recorded by sensors mounted on the vehicle. There are several ways how the collected information on the local environment may be represented. An overview of models with different levels of abstraction is given in [Sch18]. Here, they are grouped into two categories:

1. *Set of sparse features:* Independent of the level of abstraction, the environment model may consist of a list of entities indicating the presence of potentially interfering traffic participants. This information could be represented by unordered sets of point detections, bounding box

detections of objects, surface reconstructions or tracked extended object states. Note that this representation only provides indicators for the presence of entities in the environment but no explicit cues on the absence.

2. *Volumetric representations*: The region of interest is partitioned into sub volumes and a state is estimated for each of them. The state attached to the sub volumes may include information on the presence of obstacles but also on free space and occlusions. This representation enables deducing information on arbitrary subsets of the region of interest.

Schreier proposed in [Sch18] an environment model consisting of a list of extended object states for potentially moving traffic participants and a two-dimensions top-view grid. The latter is included to represent the static part of the environment in a volumetric fashion. In fact, when considering recent publications on motion planning both a highly abstracted list of tracked extended objects states representing traffic participants and volumetric information on sensor visibility, free space and occlusions are desired, see e.g. [Nau20].

In case multiple sensors are used to estimate the environment model, sensor data fusion may be applied to obtain one common representation. This potentially reduces uncertainty and makes the calculated environment representation more detailed and complete. Sensor data fusion methods can be grouped according

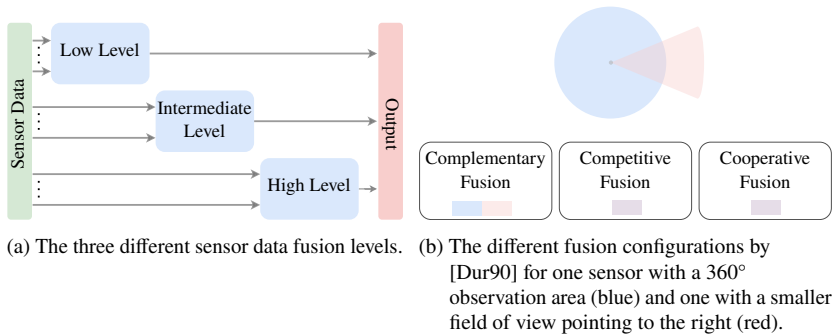


Figure 1.2: The groups of sensor data fusion concepts.

to the fusion level in the processing chain, see Figure 1.2a, as follows:

1. *Low Level Fusion*: The measurements from all sensors are combined at the beginning of the processing chain into a common representation which is subsequently abstracted into the output representation. An example is the accumulation of point detection sets from multiple range sensors.
2. *Intermediate Level Fusion*: For each sensor, the measurements are transformed into intermediate feature representations that are fed into a fusion operator. The fused state is then abstracted into the output representation. An example is grid mapping with subsequent object detection on a fused grid.
3. *High Level Fusion*: For each sensor, the measurements are processed individually until a late stage in the processing chain is reached. The state of each sensor is thereafter fused into the output representation. An example is sensor specific object detection with data fusion on object level.

Instead of considering the fusion level, Durrant-White defined in [Dur90] the following fusion configurations, see Figure 1.2b:

1. *Complementary Fusion*: Complementary sensors provide information about different aspects of the environment for instance by observing different areas. An example for complementary fusion is image stitching.
2. *Competitive Fusion*: Competitive sensors provide information about the same aspects of the environment. The collected information is redundant and can be combined to improve quality and confidence. An example is the combination of distance measurements of the same object from different sensors.
3. *Cooperative Fusion*: Cooperative sensors provide different information that are combined to infer new information about the environment. The information from all sources is required for doing the inference. An example for cooperative fusion is stereo disparity calculation based on two images in a stereo camera setup.

When working on multi-sensor environment perception tasks, developers face the following challenges:

- *Sensor redundancy*: To increase the area where meaningful information can be estimated, competitive sensor fusion approaches should be preferred whenever possible. Cooperative fusion approaches are limited to overlapping visibility areas and fail in case of sensor malfunction.
- *Conflict handling*: Combining measurements from competitive sensors does not necessarily improve quality. Applied fusion methods must be able to deal with highly conflicting information provided by the respective sensors. In order to handle conflicts reasonably, a meaningful uncertainty quantification is needed.
- *Computational complexity*: For online applications the processing time as well as memory consumption is limited and should be kept as low as possible to reduce the accumulated time delay. Therefore, an acceptable trade off between quality and efficiency must be found.

## 1.2 Contributions of this Work

The goal of this work is to develop a generic competitive sensor data fusion method. We expect the following prerequisites to be met regarding the information provided to the methods presented:

- All sensors are calibrated with respect to each other and to a vehicle-fixed rig coordinate system.
- The sensor measurements are assigned accurate time stamps so that they can be combined within a sliding time window. More advanced spatio-temporal measurement alignment techniques are not covered by this work.
- The pose of the ego vehicle consisting of a three-dimensional (3D) translation and a 3D rotation is known at all measurement time points.
- Pixel-wise semantic labeling, disparity and depth estimation are used, but the respective algorithms are outside the scope of this work.

The method proposed in this work is meant to calculate an intermediate grid-based top-view representation that contains extensive information on the presence of obstacles, semantics, dynamics, occlusion and free space

with uncertainties. An example for the estimated representation is shown in

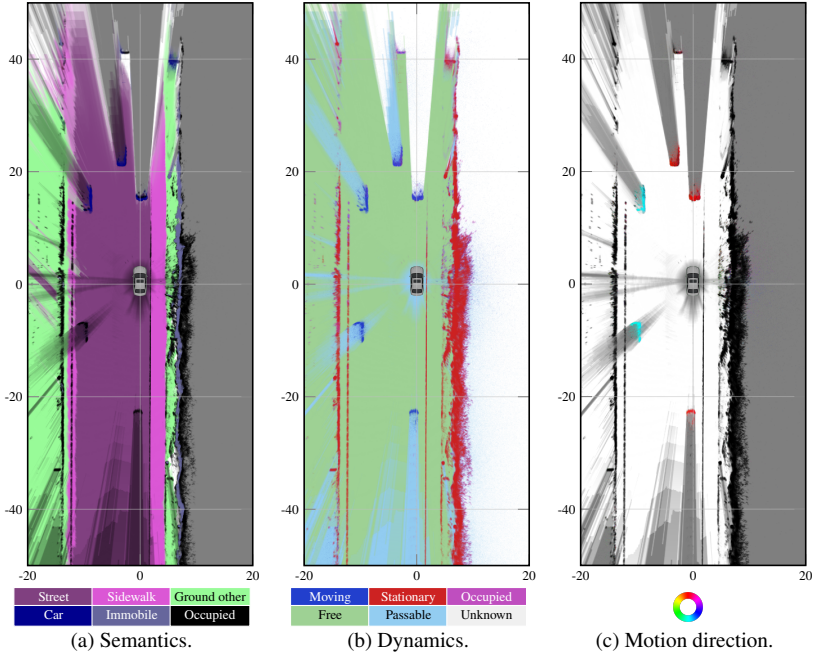


Figure 1.3: Visualizations of the grid-based environment representation estimated with the method presented in this work.

Figure 1.3. Uncertainty is modeled in an evidential framework by estimating belief assignments for three hypotheses sets that are introduced in Chapter 2. The first two are extensions of the classical occupancy frame consisting of the two hypotheses *occupied* and *free*. One divides the hypothesis *occupied* into semantic properties and the second into moving and stationary occupancy. The third hypotheses set models the semantic property of the ground. This hybrid grid map representation combines information on heterogeneous aspects of a traffic scene relevant for automated vehicles and enables further abstraction of object instances. Methods to detect and track extended object states based on such a grid-based representation have been proposed by Steyer in [STW17; Ste+20; Ste21] and Wirges in [Wir+18; Wir+19b; Wir+20].



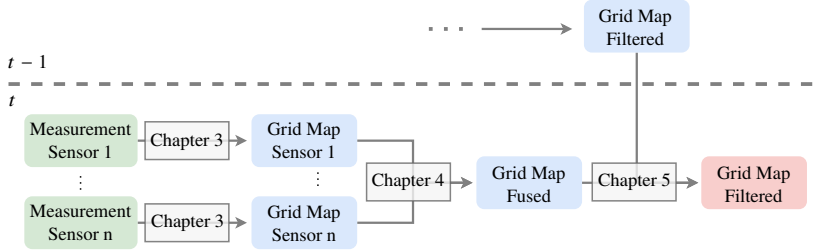


Figure 1.4: The grid mapping framework proposed in this thesis. It is separated into three submodules presented in the Chapters 3 to 5. First, grid maps are estimated based on sensor measurements for each sensor separately. Those grid maps are then fused in a common grid map thus containing measurements from all sensors. Finally, the fused grid map updates the filtered grid map recursively.

The presented grid mapping framework consists of the three submodules

1. sensor measurement grid mapping, where sensor measurements such as LiDAR range estimates and vision-based depth estimates are mapped into the evidential grid map representation (Chapter 3),
2. sensor measurement grid map fusion (Chapter 4), and
3. the temporal fusion of the fused measurement grid maps (Chapter 5).

The data flow is illustrated in Figure 1.4.

This work focuses on sensors providing measurements on a high resolution grid such as images from cameras and range images from LiDAR. The input image is projected into a sensor dependent measurement grid to enable a meaningful modeling of spatial uncertainty. Here, the neighborhood information in the scan image is used to infer information on the orientation of the reflecting surface which allows distinguishing blocking surfaces from passable areas. The proposed ray casting utilizes the 3D ray geometry to obtain an accurate and intuitive quantification of free space evidences based on the ray permeability.

The sensor measurement grid maps are transformed into a common Cartesian coordinate system so that evidential combination rules are applicable on cell level. This approach enables competitive and complementary data fusion where the belief assignments are combined to increase the visibility area, accuracy and confidence. Here, the semantic estimates that may be used in the sensor

measurement grid map estimation are redundant by design. In case no semantic estimates are provided to the method, the representation natively reduces to a classical occupancy grid map. We show that applying evidential reasoning (ER) instead of the combination rules used in past publications significantly improves the resolution of sensor measurement conflicts.

In the temporal fusion part, the evidential grid map is recursively updated by the fused measurement grid map. Classical temporal grid map fusion assumes the world to be fully static which results in occupied/free conflicts when observing moving entities, see e.g. [DRN14]. In order to mitigate this, the dynamic grid mapping framework based on a low level particle filter presented e.g. in [Nus+16; Rum+17; STW18] is adopted. In contrast to those publications, however, the method presented in this thesis enables the inclusion of semantic estimates and recursively updates belief assignments on three hypotheses sets, namely the ground semantics, the occupancy semantics and the occupancy dynamics in each grid cell. The presented temporal grid map fusion is based on evidential networks and the belief update depends on the vertical ray permeability for better conflict resolution. The data-driven parameter estimation presented in this thesis is demonstrated to significantly outperform competing methods in challenging scenarios.

## 2 Evidential Grid Mapping

In this chapter<sup>1</sup>, the hybrid evidential grid map model containing information on the ground semantics, the occupancy semantics and the occupancy dynamics is presented.

### 2.1 Fundamentals

We start with the fundamentals on evidence theory needed in this work and give a formal introduction to grid maps.

#### 2.1.1 Evidence Theory

Evidence theory, also referred to as Dempster–Shafer theory of evidence, was formally introduced by Shafer in [DS76]. As an extension of Bayes theory it provides a framework to model uncertainty and combine evidence degrees from independent sources. Let  $\Omega$  be a set consisting of mutually excluding hypotheses of interest called FoD and  $\mathcal{P}(\Omega)$  its power set containing all subsets  $A \subseteq \Omega$  including the empty set  $\emptyset$ . The mapping

$$m: \mathcal{P}(\Omega) \rightarrow [0, 1] , \quad m(\emptyset) = 0, \quad \sum_{A \in \mathcal{P}(\Omega)} m(A) = 1 \quad (2.1)$$

is called basic belief assignment (BBA) and assigns a degree of evidence to all possible combinations of hypotheses. The intention behind introducing another

---

<sup>1</sup> Parts of this chapter have been submitted to IEEE for possible publication and have been made available to the public via arXiv [Ric+22a; Ric+22b].

framework in addition to the well established probability theory is to explicitly handle ignorance. The additivity property

$$\Pr(A) = \sum_{X \in \mathcal{X}} \Pr(X), \quad A = \bigcup_{X \in \mathcal{X}} X \quad (2.2)$$

of a probability measure  $\Pr(\cdot)$  over a partition  $\mathcal{X}$  of  $A$  implies that  $\Pr(\cdot)$  provides redundant information for all hypotheses  $A$  with  $|A| > 1$ . This property is dropped for evidential BBAs in favor of assigning degrees of ignorance to those hypotheses. Degrees of evidence  $m(A)$  assigned to hypotheses  $A \neq \Omega$  with  $|A| > 1$  are called local ignorance and  $m(\Omega)$  is called global ignorance. Consider the FoD  $\Omega = \{A, B\}$  and the maximal entropy distributions shown in Table 2.1 indicating that there is no evidence pointing to either of the two hypotheses. In addition to the fact that both hypotheses are equally likely, the

	$A$	$B$	$\Omega$
$\Pr(\cdot)$	0.5	0.5	1
$m(\cdot)$	0	0	1

Table 2.1: Comparison of the BBA with the probability distribution assuming maximal entropy.

BBA  $m(\cdot)$  explicitly tells us that there is no evidence pointing to the hypotheses which is a fundamentally different statement than  $m(A) = m(B) = c \in (0, 0.5]$ . Naturally, every discrete probability distribution given by the probability measure  $\Pr(\cdot)$  induces a BBA  $m(\cdot)$  by setting

$$m(A) := \Pr(A) \text{ for all } A \in \Omega. \quad (2.3)$$

Vice-versa, every BBA without local and global ignorance induces the probability measure

$$\Pr(A) := m(A) \text{ for all } A \in \Omega. \quad (2.4)$$

In general, Shafer defined the following lower and upper bounds for the probability mass  $\Pr(\cdot)$  of the hypotheses  $A \in \mathcal{P}(\Omega)$ :

$$\text{bel}(A) \leq \Pr(A) \leq \text{pl}(A), \quad (2.5)$$

where

$$\text{bel}(A) = \sum_{B \subseteq A} m(B), \quad \text{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (2.6)$$

are denoted as belief and plausibility of  $A$  given the BBA  $m(\cdot)$ , respectively. In order to derive a probability measure from a general BBA, Smets [Sme90] proposed the pignistic transformation

$$\Pr(A) = \sum_{B \subseteq \Omega} \frac{|A \cap B|}{|B|} m(B). \quad (2.7)$$

When dealing with multiple FoDs, it might be necessary to model dependencies between the corresponding BBAs. Let  $\Omega$  and  $\Theta$  be two frames of discernment. The functions

$$\text{bel}(\cdot | \theta) : \mathcal{P}(\Omega) \rightarrow [0, 1], \quad (2.8)$$

$$\text{pl}(\cdot | \theta) : \mathcal{P}(\Omega) \rightarrow [0, 1] \quad (2.9)$$

are called *conditional belief function* and *conditional plausibility* on  $\Omega$  given  $\theta \subseteq \Theta$ . The conditional belief/plausibility represents the belief/plausibility under the assumption that  $\theta \subseteq \Theta$  is true similar to conditional probabilities. More information can be found e.g. in [XS96].

Analogously to Shannon's entropy measure for probabilistic random variables, similar measures in evidence theoretical contexts have been desired. Yager [Yag08] presented the two measures

$$e(m) = - \sum_{\emptyset \neq A \subseteq \Omega} m(A) \ln(\text{pl}(A)), \quad s(m) = \sum_{\emptyset \neq A \subseteq \Omega} \frac{m(A)}{|A|} \quad (2.10)$$

denoted as entropy  $e(m)$  and specificity  $s(m)$  of the BBA  $m$ . Deng investigates evidential uncertainty measures such as Yager's entropy and other concepts in [Den20] based on a set of five desired properties. He came up with the entropy measure

$$e_D(m) = - \sum_{\emptyset \neq A \subseteq \Omega} m(A) \log_2 \left( \frac{m(A)}{2^{|A|} - 1} \right), \quad (2.11)$$

that can be written as the sum of the nonspecificity

$$\text{ns}_D(m) = \sum_{\emptyset \neq A \subseteq \Omega} m(A) \log_2 \left( 2^{|A|} - 1 \right) \quad (2.12)$$

and the discord

$$d_D(m) = - \sum_{\emptyset \neq A \subseteq \Omega} m(A) \log_2(m(A)). \quad (2.13)$$

The nonspecificity  $ns_D(m)$  is a measure for the ignorance contained in the BBA and increases with BBA masses assigned to non singleton hypotheses  $\omega \subseteq \Omega$ ,  $|\omega| > 1$ . The discord  $d_D(m)$  is a measure for the indecision between several focal elements. The upper and lower bound of Deng's entropy are the BBA  $m_1$  assigning all the evidence mass  $m_1(\omega) = 1$  to one singleton hypothesis  $\omega \subset \Omega$ ,  $|\omega| = 1$  and the distribution of total ignorance  $m_2(\Omega) = 1$ :

$$e_D(m_1) = 0 \leq ns_D(m) \leq \log_2(2^{|\Omega|} - 1) = e_D(m_2) < |\Omega|. \quad (2.14)$$

	$ns_D(m)$	$d_D(m)$	$e_D(m)$
$m(A) = 0, m(B) = 0$	$\log_2(3)$	0	$\log_2(3)$
$m(A) = 1, m(B) = 0$	0	0	0
$m(A) = 0.5, m(B) = 0.5$	0	1	1

Table 2.2: Deng's entropy measures for different BBA distributions. In the first row, there is a full global ignorance  $m(\Omega) = 1$  leading to a maximal nonspecificity  $ns_D(m) = \log_2(3)$ . The BBA in the second row only supports the hypothesis  $A$  which yields the minimal entropy  $e_D(m) = 0$ . In the last row the BBA supports both  $A$  and  $B$  leading to the discord  $d_D(m) = 1$ .

In Table 2.2, three exemplary BBA distributions with corresponding Deng entropy measures are shown for the FoD  $\Omega = \{A, B\}$ .

## 2.1.2 Grid Maps

Grid mapping is the task of estimating a state  $x$  in the state space  $X$  for each cell in a regular grid. Therefore, a grid map  $g$  is a mapping

$$g: \mathcal{G} \rightarrow X \quad (2.15)$$

assigning an element in the state space  $x \in X$  to each grid cell  $C \in \mathcal{G}$ . The state space  $X$  may encode any formalizable information on the local environment.

A general regular grid  $\mathcal{G}$  is a tessellation of the  $n$ -dimensional Euclidean space  $\mathbb{R}^n$  in congruent disjoint subsets. In most of the literature the term grid mapping refers to the case  $n = 2$  where the 3-dimensional case is referred to as voxel grid mapping. Throughout this work, different kinds of tessellations in 2D spaces are introduced for intermediate representations depending on the measurement sources. For cell-wise fusing information from different sources a common grid representation is needed. The 2D Cartesian grid  $\mathcal{G}_{xy} = \mathcal{P}_1 \times \mathcal{P}_2$  on the rectangular region of interest  $\mathcal{R} = I_1 \times I_2 \subset \mathbb{R}^2$ , where

$$\begin{aligned} \mathcal{P}_i &= \{I_{i,k}, k \in \{0, \dots, s_i - 1\}\}, \\ I_{i,k} &= [o_i + k \delta_i, o_i + (k + 1) \delta_i], \quad i \in \{1, 2\} \end{aligned}$$

forms a partition of the interval  $I_i$  with equidistant length  $\delta_i \in \mathbb{R}$ , origin  $o_i \in \mathbb{R}$  and size  $s_i \in \mathbb{N}$ . Hence, each grid cell  $C \in \mathcal{G}_{xy}$  is a rectangle with side lengths  $(\delta_1, \delta_2)$ .

In this work, two coordinate systems are considered:

1. The *vehicle coordinate system* is located at a fixed position with respect to the ego-vehicle. The x-axis is pointing to the front of the vehicle, the y-axis to the left and the z-axis to the top.
2. The *reference coordinate system* is defined by the vehicle coordinate system at the first update time point  $t_0$  of the system. Hence, this coordinate system is independent of the ego-vehicle's motion.

For fusing grid maps temporally, the grid is defined with respect to the reference coordinate system. The region of interest  $\mathcal{R}_t$  is translated by whole grid cells in x- and y-direction to follow the ego vehicles movement. This avoids the interpolation of cell states when combining the BBAs from different time points. As opposed to a local grid in vehicle coordinates, the region of interest is not aligned with the ego motion direction. In practical applications, however, a region of interest covering a larger area in front of the vehicle may be desired. This can be achieved by dynamically localizing the ego vehicle on a circle centered in the current region of interest as proposed e.g. by Eraqi et al. [EHZ18]. The radius  $r_t$  at time  $t$  can be chosen based on different cues as the ego velocity or the current traffic scenario. The concept is sketched in Figure 2.1. In this work this method is applied to define the grid  $\mathcal{G}_t$  as the global  $\mathcal{G}$  restricted to the current region of interest  $\mathcal{R}_t$ .

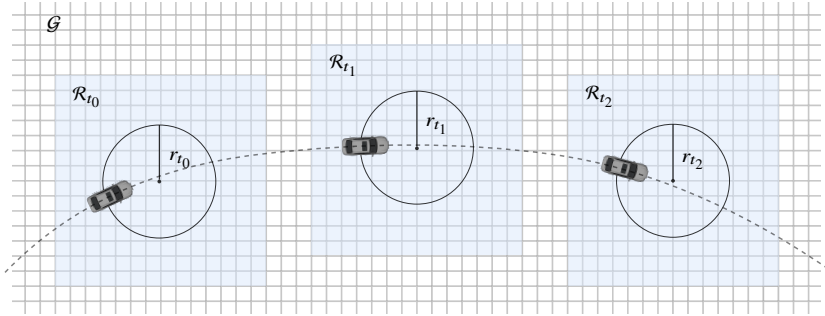


Figure 2.1: The location of the local grid  $\mathcal{G}_t$  at time  $t$  on the global grid  $\mathcal{G}$  at three not necessarily adjacent time points  $t_0$ ,  $t_1$  and  $t_2$ . The region of interest  $\mathcal{R}_t$  moves with the ego motion over the global grid by whole grid cells. The ego vehicle is located in the local grid with a radial offset  $r_t$  to the grid center.

## 2.2 Related Work

We give an overview on grid-based environment models that have been proposed in past publications.

Classical grid maps divide the environment into occupied and free cells and were introduced by Moravec et al. [ME85]. Whereas initially two grids were used to model free and occupied cells it was proposed in later publications to only use one grid for the occupancy probability [Mor89; Elf89]. Instead of a probabilistic model, Yi et al. [Yi+00] proposed using an evidential framework to reduce uncertainties resulting from specular reflections in a temporal filter. Yang et al. [YA06] further elaborated this showing that the evidential framework has advantages in dealing with uncertainties compared to the classical Bayesian framework.

Although occupancy grid maps are widely used, the term occupancy lacks a formal definition in literature. In a vast majority of the early publications, it was assumed that a range measurement supports the hypothesis occupied at its location and the hypothesis free along the ray between the sensors and the measurement's location. In the context of automated vehicles this assumption is not necessarily applicable. When considering range sensors such as RaDAR, Sonar and LiDAR with a low vertical field of view mounted in parallel with



a planar ground surface, reflections in fact are mostly caused by surfaces of obstacles. Using camera systems and LiDAR sensors with a higher vertical field of view, however, adds the challenge of classifying reflections as they may also occur on passable surfaces as the ground surface. The interpretation what parts of the environment are considered occupied ranges from “obstacles with a certain height” ([YCB14]) over “obstacles and curbs” ([HRL15]) to “everything except drivable areas” ([BVF15]) and seems to depend on the proposed estimator instead of being predefined. Depending on this consideration, the classification problem can be solved by either directly classifying each reflection based on semantic or geometric properties or by relating the detection coordinate to an estimated ground surface. Among others, Yu et al. [YCB14] did the latter by considering every reflection above a certain height above ground to be caused by obstacles. They applied their method to a Velodyne HDL64E LiDAR sensor and assumed a planar ground surface, but more complex models have been published, see [Rum+17; Wir+21]. Harms et al. [HRL15] used disparity and orientation images from Stereo Vision to build up a two-layer grid map where one layer contains occupancy probabilities for obstacles and one for curbs. Their orientation images contain the pixelwise deviation of the local normal vector from the global height axis and was used to determine the curb occupancy layer.

Besides occupancy, grid maps have also been used to model geometric and semantic properties. Geometric properties of interest are mostly limited to the height modeled in elevation maps, see e.g. [Sti+17; FBH18]. An evidential occupancy grid map with a refinement of the occupied hypothesis into the semantic hypotheses vehicle, building, vegetation and sidewalk was proposed by Bernardes Vitor et al. [BVF15]. Instead of including the semantic hypotheses in the evidential framework, they treat them as meta-knowledge of the occupied hypothesis. In recent years deep neural networks have been used to predict semantic top-view grid maps, see [Bie+20; Che+20a; Fei+21]. All those methods can only predict a single semantic class per grid cell thus neglecting uncertainties.

In [Ric+19], we introduced a novel evidential framework incorporating both occupancy and semantic estimates. As opposed to other publications, the semantic hypotheses were integrated in the evidential framework instead of assigning one semantic label to a grid cell only. The resulting semantic evidential grid mapping method was applied to LiDAR, RaDAR and Stereo

Vision measurements and further developed and applied to different Vision setups in [Ric+20] and [Ric+21].

## 2.3 Hybrid Evidential Representation

In this work, a hybrid semantic evidential state space that consists of the ground state and two occupancy states refining the classical occupancy space is proposed. The definition of the occupancy state is based on geometric constraints and two possibilities are proposed followed by a short comparison.

We define the driving corridor between the ground height and the maximal height above ground  $d_{z,\max}$  to be the height area that is relevant for the ego-vehicle. This excludes high obstacles such as bridges and tree branches that do not interfere with the automated vehicle. Let the surface of a traffic scene be implicitly given as  $f_S(x, y, z) = 0$  where  $x$ ,  $y$  and  $z$  are Cartesian coordinates in the vehicle coordinate system.

**Definition 2.1.** Let  $f_G: \mathbb{R}^2 \rightarrow \mathbb{R}$  be a function describing an approximation of the ground height and

$$D_C = \{z - f_G(x, y) \mid (x, y) \in C: f_S(x, y, z) = 0\} \quad (2.16)$$

be the set containing all distances between ground surface and traffic scene surface in grid cell  $C \in \mathcal{G}$ . The grid cell is called *occupied*, if the traffic scene surface intersects with the driving corridor, i.e.

$$\sup(D_C) > 0 \quad \wedge \quad \inf(D_C) < d_{z,\max}, \quad (2.17)$$

where the supremum  $\sup(A)$  is the smallest upper bound and the infimum  $\inf(A)$  is the largest lower bound of an ordered set  $A$ . If one of the two conditions in Equation (2.17) is not fulfilled the cell is called *free*.

The second definition of the term occupancy is based on the unit normal vector of the surface  $f_S(x, y, z) = 0$ . It can be shown that this normal vector  $n$  is given by the gradient of  $f_S$ :

$$(n_1, n_2, n_3)^T = \frac{\nabla f_S}{|\nabla f_S|}. \quad (2.18)$$

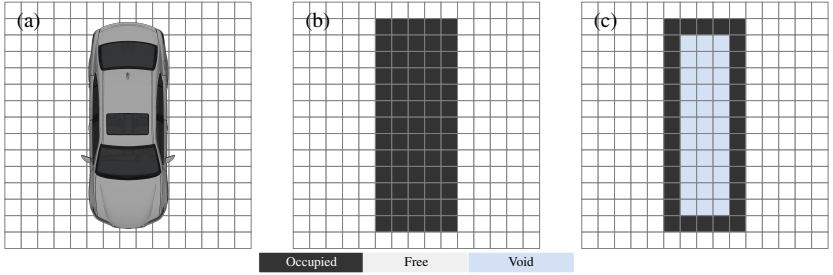


Figure 2.2: The two formalizations of the term *occupancy* presented in this work: (b) shows the occupancy pattern based on Definition 2.1 when observing the car shown in (a) and (c) shows the pattern when modeling occupancy based on Definition 2.2. In (b), all grid cells covered by the vehicle are considered occupied whereas in (c) only the cells at the car’s border are occupied.

**Definition 2.2.** A grid cell  $C \in \mathcal{G}$  is called

- *occupied*, if the angle difference between the surface normal and the z-axis in the vehicle coordinate system exceeds a given threshold  $b_T$ , i.e.

$$\arccos(n_3) > b_T, \quad (2.19)$$

and the traffic scene surface corresponding to the grid cell is not fully above the driving corridor, i.e.  $\inf(D_C) < d_{z,\max}$ .

- *free*, if the ego vehicle can enter the underlying area in space, i.e. the complete driving corridor is free, or
- *void*, if it is neither free nor occupied.

The cell state *void* is attained in grid cells covered by obstacles where the surface is non-blocking. This is mainly limited to areas that are separated from the ego vehicle by blocking surfaces such as the roof of a car. Figure 2.2 sketches the two definitions of the term occupancy.

In this work, the occupancy classification based on Definition 2.2 is used. The reason is that the condition “ $\sup(D) > 0$ ” that is only contained in Definition 2.1 is more critical than the condition “ $\inf(D) < d_{z,\max}$ ” contained in both Definitions 2.1 and 2.2. For the former, the ground surface estimation must be very accurate to exclude all the measurements reflected on the ground.

For the latter, however, a rough estimation is sufficient in most cases as there usually is a larger margin above the driving corridor that is free of obstacles. Hence, a small tolerance margin  $\delta_G > 0$  is added in some publications such as [Wir+21] so that the condition becomes “ $\text{sup}(D) > \delta_G$ ”. This, however increases the likelihood of missing low obstacles such as curb stones.

Furthermore, ground surface estimation is a hard task. When estimating the ground surface it is desirable to exclude any obstacle detections in the estimation process. However, this is not possible in cases where the ground estimation is performed in order to make this separation in the first place. This can be tackled by estimating the ground surface based on all detections and adding smoothness priors to the ground surface model as e.g. in [Wir+21]. Nevertheless, it cannot be guaranteed that the resulting ground surface model is not significantly influenced by obstacle detections.

Next, the hybrid semantic representation containing ground semantics, occupancy semantics and occupancy dynamics is introduced. We consider the semantic estimates occupied by “car”, “two-wheeler”, “pedestrian”, “other mobile entities” or “immobile entities” and the ground hypotheses “street”, “sidewalk” and “other ground”. Classical occupancy grid mapping is based on the general differentiation between occupied and free leading to a FoD consisting of those two excluding elementary hypotheses only. The considered semantic hypotheses, however, are not necessarily pairwise contradicting. For example both hypotheses “street” and “car” can hold for a grid cell as the car is just placed on top of the street. This fact violates the requirement that all elementary hypotheses in an evidential FoD must be contradicting. Therefore, ground semantics and occupancy semantics are modeled in two separate FoDs. In addition to the semantic state, the detection of moving parts of the environment is desired so that their motion can be estimated. Therefore, occupancy dynamics are estimated in a third layer completing our hybrid evidential representation.

**Ground Semantics.** The semantic state of the ground is classified by the hypotheses *street* ( $s$ ), *sidewalk* ( $sw$ ) and *other ground* ( $og$ ). This yields the ground semantics FoD

$$\Omega_g := \{s, sw, og\}. \quad (2.20)$$

**Occupancy Semantics.** The occupancy semantics describe the semantic state of cell occupancy in a rectangular cuboid on top of the considered grid cell. The size of the cuboids footprint is given by the grid cell and the cuboids height interval  $[0, d_{z,\max}]$  is defined by the driving corridor. This cuboid can either be free of obstacles or occupied by an entity with assigned semantic class. In particular, the occupancy semantics FoD

$$\Omega_s := \{c, cy, p, om, nm, f, v\} \quad (2.21)$$

consists of the hypotheses listed in Table 2.3. Here, the hypothesis *void*  $V_s$  is

Semantic class	Set	Letter
Occupied by a car	$\{c\}$	$O_{\text{car}}$
Occupied by a two-wheeler	$\{cy\}$	$O_{\text{tw}}$
Occupied by a pedestrian	$\{p\}$	$O_{\text{ped}}$
Occupied by another mobile object	$\{om\}$	$O_{\text{om}}$
Occupied by an immobile object	$\{nm\}$	$O_{\text{im}}$
Occupied by an object with unknown class	$\{c, cy, p, om, nm\}$	$O_{\text{su}}$
Free	$\{f\}$	$F_s$
Void, i.e. neither occupied nor free,	$\{v\}$	$V_s$

Table 2.3: Valid hypotheses induced by the occupancy semantics FoD  $\Omega_s$ .

needed to partition the whole environment into cell occupancy states based on Definition 2.2. In the remainder of this work, we refer to the hypothesis  $O_{\text{su}}$  as *semantically unclassified occupancy*. Note that this FoD can be partitioned into *semantically unclassified occupancy*  $O_{\text{su}}$  and *free*  $F_s$  as

$$\Omega_s \setminus V_s = O_{\text{su}} \dot{\cup} F_s \quad (2.22)$$

which is the separation used in classical occupancy grid mapping.

**Occupancy Dynamics.** In order to distinguish moving from stationary occupancy in the evidential context, this work follows Steyer et al. [STW18] and introduces the occupancy dynamics FoD

$$\Omega_d = \{m, nm, f, v\} \quad (2.23)$$

containing the valid hypotheses listed in Table 2.4. We refer to the hypothesis  $O_{\text{du}}$  as *dynamically unclassified occupancy*. Considering the hypotheses

Description	Set	Letter
Occupied by a moving object	$\{m\}$	$O_{\text{mov}}$
Occupied by a stationary object	$\{nm\}$	$O_{\text{stat}}$
Occupied by an object with unknown dynamic state	$\{m, nm\}$	$O_{\text{du}}$
Free	$\{f\}$	$F_d$
Passable, i.e. free or occupied by a moving object	$\{f, m\}$	$P$
Void, i.e. neither occupied nor free,	$\{v\}$	$V_s$

 Table 2.4: Valid hypotheses induced by the occupancy dynamics FoD  $\Omega_d$ .

*passable*  $P$  enables memorizing *free space* in cells that have been observed as *free* but are not observable anymore and thus might be occupied by a dynamic entity.

Note that some occupancy dynamics hypotheses  $\theta \subseteq \Omega_d$  depend on the semantic state  $\omega \subseteq \Omega_s$ . This will be formalized in Section 5.3.2 where the occupancy dynamics are inferred in a recursive temporal estimator under consideration of the occupancy semantics.

The hybrid evidential representation is sketched in Figure 2.3. It is based on the assumption that in traffic scenes each obstacle is placed on top of ground implying that every combination of occupancy semantics  $\theta_1 \in \Omega_s$  and ground semantics  $\theta_2 \in \Omega_g$  are non contradicting and can thus happen simultaneously. Note that the real world might be composed of several overlapping ground layers as e.g. at freeway exit ramps or bridges. However, even in those scenarios the region of interest can be limited to the ground layer as there is no direct interaction between traffic participants on different ground layers. The simplification of considering one occupancy layer only is justified by the fact that two objects placed on top of each other may be considered as one entity in the navigation module. The person sitting on a bicycle is considered as one entity represented by the hypotheses “occupied by a two-wheeler”.

The BBA  $m$  on  $\mathcal{P}(\Omega_i), i \in \{s, g, d\}$  is then represented by the multi-layer grid map

$$g_i : \mathcal{G}_{xy} \times \mathcal{P}(\Omega_i) \rightarrow [0, 1]. \quad (2.24)$$

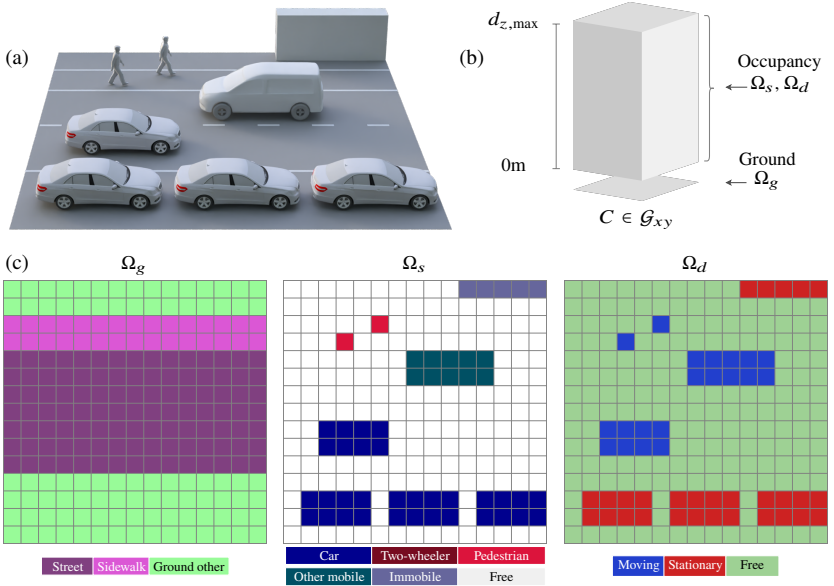
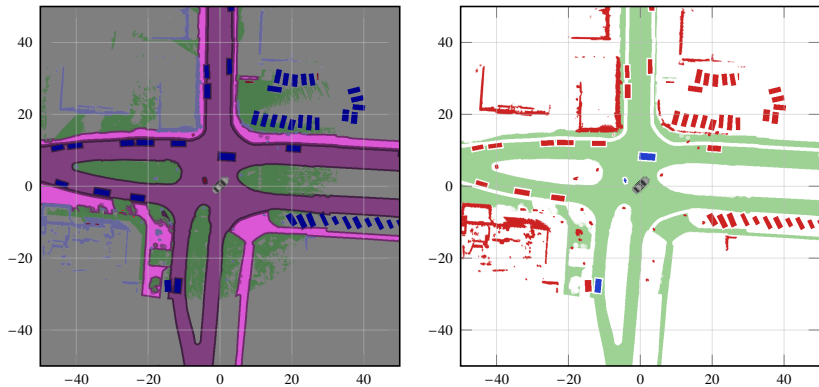


Figure 2.3: The hybrid evidential grid-based environment representation: (a) shows a traffic scene with three parking cars on the side of the road, two passing vehicles on the road and two walking pedestrians. (b) shows the subspace of the environment that is represented in a grid cell  $C \in \mathcal{G}_{xy}$ . Occupancy is modeled in the cuboid above the grid cell up to the maximal height  $d_{z,max}$ . In (c), the three grid layers for the ground semantics FoD  $\Omega_g$ , the occupancy semantics FoD  $\Omega_s$  and the occupancy dynamics FoD  $\Omega_d$  with corresponding color codings are depicted for the traffic scene in (a).

## 2.4 Datasets and Evaluation Metrics

We use the KITTI-360 dataset [LXG21] and the SemanticKITTI dataset [Beh+19] to estimate a reference representation. This reference representation is used later to evaluate the estimated representation.

**KITTI-360.** The 3D semantic bounding primitives and the semantic point cloud in the KITTI-360 dataset are used to generate a reference semantic evidential grid map, see Figure 2.4. The color map applied in Figure 2.4a shows both ground semantics and occupancy semantics and is a combination of



(a) For occupancy semantics  $\Omega_s$  and ground semantics  $\Omega_g$ .

(b) For occupancy dynamics  $\Omega_d$ .

Figure 2.4: The reference grid map  $g_{\text{ref}}$  for a frame in the first evaluation sequence in the KITTI-360 dataset.

the occupancy probability gray value map and a color coding of the semantic classes. It was generated based on the following rules:

- If both a BBA for ground semantics and an occupancy semantics have been assigned a high BBA, the object BBA is visualized.
- The color saturation scales with the assigned BBA.
- Above ground, a lower brightness indicates a low BBA for *free*.

KITTI-360 accumulates LiDAR measurement chunks containing over 300 frames. Each frame includes LiDAR reflections with a distance of at most 30m. The accumulated point clouds were semantically annotated and used to fit bounding primitives representing object instances and infrastructural entities such as buildings, poles and streets. The labeled bounding primitives and the accumulated point clouds were used to generate the reference evidential grid map  $g_{\text{ref}}$  based on the following rules:

- The BBA for traffic participants as *cars*, *two-wheelers* and *pedestrians* as well as for *street* and *sidewalk* is estimated by projecting the bounding



primitives into the Cartesian grid. We set  $g_{\text{ref}}(C, \omega) = 1$  in grid cell  $C$  located within the bounding primitive where  $\omega$  is the semantic hypotheses attached to the primitive.

- The BBA for *other ground* and *immobile occupancy* is estimated by mapping the accumulated LiDAR points with corresponding ground truth labels into the Cartesian grid.
- The BBA for the hypothesis *free* was set to one in grid cells belonging to street and sidewalk primitives where no object primitive is located. In order to avoid penalizing curbstone detections, those regions were removed from the *free* area using morphological erosion in the grid layers containing the BBA for *street* and *sidewalk*. In areas covered by other ground labels, no reference BBA for *free* is derived as those areas cannot be assumed to be non-occupying according to Definition 2.2.
- Moving and stationary bounding primitives are provided separately in the KITTI-360 dataset. However, it is not guaranteed that the moving primitives are moving during the whole time they are observed. In order to detect the frames where the primitives are actually moving the velocity  $v$  is estimated based on its location in two consecutive frames. In the reference BBA on the occupancy dynamics  $\Omega_d$ , we then set  $g_{\text{ref}}(C, \omega) = 1$ , where

$$\omega = \begin{cases} O_{\text{mov}}, & \text{if } v > 1.5 \text{ m/s} \\ O_{\text{stat}}, & \text{if } v < 0.01 \text{ m/s} \\ O_{\text{du}}, & \text{else.} \end{cases} \quad (2.25)$$

For stationary bounding primitives and immobile detections in the accumulated point cloud we directly set  $g_{\text{ref}}(C, O_{\text{stat}}) = 1$ .

- Furthermore, it was observed that the object primitives tend to overestimate the object dimensions significantly. This is probably due to the fact that they were generated based on accumulated sensor measurements where small errors in the ego pose accumulate over time. This would lead to a punishment of correctly estimated BBA for *free* in those regions in the evaluation process. Therefore, in the outermost 30cm of each object bounding box the BBA was assigned to the total ignorance  $m(\Omega) = 1$ .

After applying the above-mentioned modifications preventing wrong BBA assignments it is possible to do a cell-wise evaluation of the estimated evidential grid maps.

We separate the KITTI-360 dataset into subsequences used for evaluation and training as shown in Table 2.5. The evaluation sequences consist of 1000 frames





	Seq.	Frames	Scenario	Image
evaluation	0	2001 - 3000	Drive through a suburban area with buildings and parking cars on the side of the road.	
	2	4901 - 5900	Drive through a suburban area with ongoing and oncoming traffic.	
	3	31 - 1031	Drive on a country road with ongoing and oncoming traffic.	
training	10	501 - 1000	Drive through an urban area with heavy traffic.	

Table 2.5: The data sequences from the KITTI-360 dataset used in this work.

and the sequence used for training consists of 500 frames. This is enough data for the parameter tuning applied in this thesis. The sequences were selected so that diverse traffic scenarios are covered including as many moving traffic participants as possible.

**SemanticKITTI.** Besides the bounding primitives and accumulated point clouds from KITTI-360, we also use the semantic LiDAR point cloud labels from the SemanticKITTI dataset [Beh+19] for evaluation. It contains semantic annotations for each detection in the LiDAR point clouds contained in the Kitti odometry benchmark [GLU12]. Using this data, we estimate reference grid maps containing single semantically annotated LiDAR scans.

**Evaluation metrics.** In order to quantify the differences between the reference grid map and the estimated grid map, the evidential intersection over union (eIoU) is defined as

$$\text{eIoU}_\omega = \frac{\text{eTP}_\omega}{\text{eTP}_\omega + \text{eFP}_\omega + \text{eFN}_\omega}. \quad (2.26)$$

It uses the evidential true positive rate

$$\text{eTP}_\omega = \sum_{C \in \mathcal{G}_{xy}} \sum_{\psi \subseteq \omega} g_{\text{ref}}(C, \psi) g(C, \omega), \quad (2.27)$$

the false positive rate

$$\text{eFP}_\omega = \sum_{C \in \mathcal{G}_{xy}} \sum_{\psi \cap \omega = \emptyset} g_{\text{ref}}(C, \psi) g(C, \omega), \quad (2.28)$$

and false negative rate

$$\text{eFN}_\omega = \sum_{C \in \mathcal{G}_{xy}} \sum_{\psi \cap \omega = \emptyset} g_{\text{ref}}(C, \omega) g(C, \psi). \quad (2.29)$$

Note that the eIoU reduces to the classical intersection over union used for pixel wise semantic labeling if the BBAs are binary, i.e.  $g_{\text{ref}}(C, \psi) = 1, g(C, \omega) = 1$  for  $\psi, \omega \subset \Omega$ .



## 3 Sensor Measurement Mapping in Evidential Grid Maps

The first step of the grid-based environment model estimation presented in this work is the calculation of the sensor measurement grid maps.<sup>1</sup> A sensor measurement grid map contains the BBA on the occupancy semantics  $\Omega_s$  and the ground hypotheses  $\Omega_g$  estimated using measurements from one sensor. After giving a brief introduction to LiDAR and cameras, related work is presented in Section 3.2. Then, a generic sensor measurement grid map estimation pipeline is presented in Section 3.3. Subsequently, the framework is specified for mapping point set measurements in Section 3.3.1. The main focus of this work is to utilize the organized structure of measurements obtained by imaging sensors such as cameras and LiDAR scanners. The sensor measurement grid mapping with images is described in Section 3.3.2. It is subsequently applied to LiDAR and stereo camera sensory by modeling the sensor modalities in Section 3.3.3. Finally, the sensor measurement grid mapping pipeline is validated qualitatively and quantitatively in experiments using real sensor measurements in Section 3.4.

### 3.1 Fundamentals

#### 3.1.1 LiDARs

LiDAR sensors measure the distance to surrounding surfaces by firing laser beams and measuring the time until the reflected light returns. In order to obtain a map of the local environment different imaging strategies can be applied. Here, a brief overview is given, and the reader is referred to [RB19] for more details.

---

<sup>1</sup> A short version of this chapter has been submitted for publication in *Transactions on Intelligent Transportation Systems* and has been made available to the public via arXiv [Ric+22a].

In general, imaging strategies can be grouped into scanners and detector arrays. In scanners, one or multiple laser beams are incrementally repositioned to sweep over surfaces in the environment and measure distances. Detector arrays illuminate the whole scene and a detector matrix receives reflecting signals for subsections of the observed area. This work focuses on mechanical scanners as they are most commonly used in automotive applications. Although different scanning patterns exist, most scanners consist of multiple lasers that rotate in parallel resulting in parallel scan lines. Those scan lines can be combined vertically defining the sensor grid  $\mathcal{G}_{uv}$ . At each laser position a beam is fired

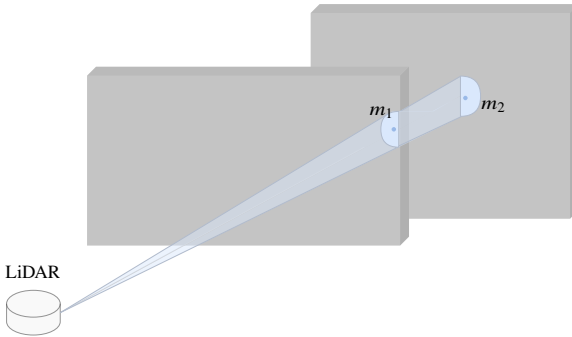


Figure 3.1: Measurement beam of a scanning LiDAR reflected at two distances. The footprint of the LiDAR beam covers two walls, each providing a range measurement according to its distance to the LiDAR sensor.

that diverges horizontally and vertically, see e.g. [Vel19]. Consequently, the laser beam footprint can be approximated by a circle with increasing diameter. The range measurement is the distance between the ray origin and a point within this circular laser beam footprint. Depending on how the driver software is implemented, LiDAR scanners not only provide the distance to the point with the highest reflectivity, but potentially further reflections, which is called multiple returns. Figure 3.1 shows the LiDAR scanner ray geometry in a scenario were a single LiDAR beam might return two reflections.

### 3.1.2 Cameras

Camera setups usually consist of one or multiple digital cameras where single cameras are called monocular cameras and two cameras with overlapping image planes are called stereo cameras. In this work, the recorded images are assumed to be rectified according to a suitable camera model and defined on the sensor grid  $\mathcal{G}_{uv}$  where  $u$  corresponds to the horizontal and  $v$  to the vertical pixel index. For each pixel, incoming light along rays intersecting with the pixel area is measured on the camera sensor. Therefore, one measurement element corresponds to an area on the reflecting surface instead of a point. In contrast to LiDAR scanners cameras do not measure the distance to reflecting surfaces directly. Instead, depth estimates must be provided based on the sensor readings by utilizing Computer Vision algorithms. In general, range estimates can be obtained with monocular cameras and stereo cameras. Stereo cameras enable the deduction of range estimates by finding corresponding pixels using epipolar geometry and calculated disparity values, see e.g. [Hir08; MR11]. The disparity is the pixel distance between matched pixels in the respective images. Those disparity values are then used to infer 3D pixel coordinates. In recent years, deep neural networks have also been trained to predict range information based on monocular [Qia+21; Yua+22] or stereo images [Zha+19; Che+20b].

### 3.1.3 Sensor Measurement Modeling on Grids

When modeling a sensor measurement  $m$  in a grid cell  $C$ , spatial uncertainty is modeled by the inverse sensor model given by the conditional probability

$$\Pr(x \in C | x_m), \quad (3.1)$$

where  $x_m$  is the measured position projected to the top-view grid and  $x$  is the random variable representing the real position. In the remainder of this thesis, we write  $\Pr(C|m)$  for short. While the final presentation is usually defined on a Cartesian grid, the inverse sensor model may be calculated in different coordinate systems such as Polar, u/distance or u/disparsity grids. The choice of the coordinate system depends on the sensor modalities and is made so that the grid can be aligned with the sensor measurement rays.

## 3.2 Related Work

We review past publications on how sensor measurements are interpreted to obtain probabilistic or evidential grid maps containing information about the presence and absence of obstacles.

### 3.2.1 Grid Mapping with Range Sensors

Elfes [Elf89] proposed to model a range measurement recorded by a Sound Navigation and Ranging (Sonar) sensor by a 2D Gaussian inverse sensor model where the two dimensions correspond to range and angle. Based on the inverse sensor model he derived an occupancy profile that is recursively fused in a Bayesian framework. Yguel et al. [YAL08] applied a simplified one-dimensional inverse sensor model to LiDAR range measurements in a Polar grid. They further focus on formulating the problem of switching coordinates from Polar to Cartesian mathematically and propose a suitable approximation that can be efficiently implemented on the graphics processing unit (GPU). Homm et al. [Hom+10] follow up on this and use a one-dimensional Gaussian sensor model in a Polar grid. Besides applying their method to LiDAR measurements, they further include RaDAR measurements for lane boundary detection and present an efficient GPU implementation. The early grid-based sensor models for range sensors were modeled for sensors providing measurements from one rotating laser scanner. In order to provide richer information on reflecting surfaces in the environment, LiDAR sensors used on automated vehicles consist of multiple lasers that may provide conflicting measurements. Yu et al. [YCB14] handle conflicting measurements by first collecting reflections above a given height threshold in a Polar grid and subsequently transform the measurement counts into a BBA. They treat ground detections as sources of evidence for the hypothesis *free* and apply backward extrapolation to propagate free space evidence to neighboring grid cells. Porębski [Por20] presented a customizable inverse sensor model to calculate occupancy grid maps. In order to be able to compute accurate probabilities in each grid cell, they proposed a cell selection process and apply either a Gaussian or an exponential distribution to compute the inverse sensor model. They also investigate the capability of handling sensor conflicts compared to past publications. Recently, Van Kempen et al. [Van+21] proposed an evidential occupancy grid mapping framework using end-to-end



learning. They generate synthetic LiDAR point clouds based on simulated scenarios and train a deep neural network that is able to generate BBA layers for the hypotheses *occupied* and *free* successfully modeling uncertainty.

### 3.2.2 Grid Mapping with Cameras

Badino et al. [BF07] computed an occupancy grid map based on stereo disparity images and used the resulting grid representation to infer free space areas. They compared Gaussian inverse sensor models on a Cartesian, Polar and u/disparity grid, respectively. Yu et al. [YCB15] modeled free space in a v/disparity grid and occupancy separately in a u/disparity grid based on stereo measurements and subsequently combine both in an evidential occupancy grid map. Valente et al. [VJF18] utilized a ground segmentation in a v/disparity grid and project obstacles into a u/disparity grid. They apply a 2D Gaussian inverse sensor model explicitly modeling errors in the stereo matching. The vast amount of semantic segmentation frameworks in Computer Vision suggests including semantic estimates in Vision-based grid maps. Bernardes Vitor et al. [BVF15] added an occupancy refinement value denoting the semantic state as meta information to their grid map representation. Thomas et al. [Tho+19] incorporated semantic hypotheses in an evidential framework in order to estimate a geometric road model. Their inverse sensor model considers confidences of the pixelwise semantic segmentation model and pixel location probabilities.

None of the above-mentioned publications models occupancy and semantic estimates in a joint evidential context such as the one introduced in Section 2.3. In [Ric+19; Ric+20; Ric+21] a sensor grid mapping pipeline was presented estimating a BBA on a FoD containing ground and occupancy hypotheses for range sensors and cameras. In this thesis, an advancement of this work for the evidential model in Section 2.3 is presented. Besides the modified evidential model, the main advancement is that the occupancy classification of the sensor measurement is incorporated in the estimation process instead of relying on an external ground surface estimation. This is achieved by utilizing the organized data structure of measurements obtained from imaging sensors.

### 3.3 Sensor Measurement Grid Mapping

Let  $Z$  be a sensor measurement consisting of individual measurement elements  $m \in Z$  with attached semantic label  $\omega_m \in \mathcal{P}(O_{\text{su}}) \cup \mathcal{P}(\Omega_g)$ . We present a generic framework for transforming the sensor measurement  $Z$  to a sensor measurement grid map  $g_Z$ . The outline of the methodology described in this chapter is sketched in Figure 3.2.

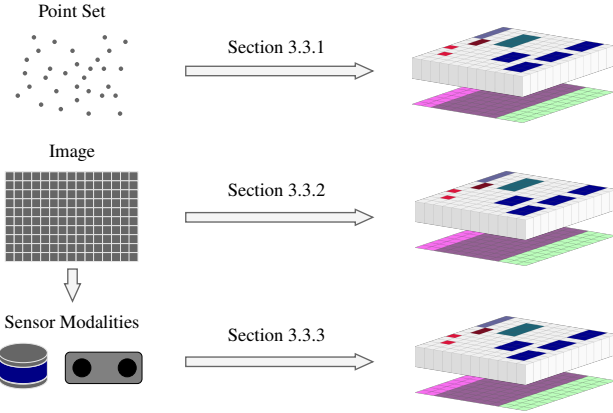


Figure 3.2: The methodology described in this chapter. The sensor measurement grid map estimation is presented for point sets and images. The latter is applied to specific sensors considering their modalities.

#### Occupancy estimation

Each measurement element  $m$  provides an evidence for occupancy  $\omega \subseteq O_{\text{su}}$  or ground  $\omega \subset \Omega_g$  depending on the attached semantic label  $\omega_m$ . The sensor measurement grid map for hypothesis  $\omega$  in the Cartesian grid cell  $C \in \mathcal{G}_{xy}$  is modeled as

$$g_Z(C, \omega) = 1 - \prod_{m \in Z} \Pr(m \nrightarrow \omega, C), \quad (3.2)$$

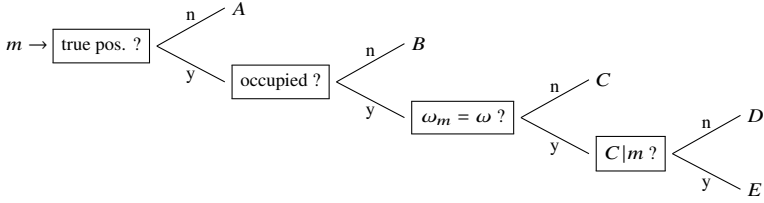


Figure 3.3: Tree visualizing the calculation of  $\Pr(m \rightarrow \omega, C)$  for  $\omega \subseteq \Omega_s$  which accumulates the probabilities of the events  $A, B, C$  and  $D$ .

The probability  $\Pr(m \rightarrow \omega, C)$  that the measurement element  $m$  is not relevant for hypothesis  $\omega$  in grid cell  $C$  is visualized in the tree diagram in Figure 3.3 for hypotheses  $\omega \subseteq O_{su}$  indicating that the grid cell is occupied.

It is calculated by considering the following four nested binary queries:

1. Is the measurement element  $m$  a true positive? The false positive rate  $p_{FP}$  is a sensor dependent design parameter quantifying the likelihood of obtaining ghost detections.
2. Was the measurement element  $m$  recorded on an occupying surface? Methods to calculate the according occupancy probability  $p_{occ}$  are discussed in Sections 3.3.1 and 3.3.2.
3. Does the semantic label  $\omega_m$  assigned to the measurement element  $m$  match with the considered hypothesis  $\omega$ ? The according probability  $p_\omega$  may be obtained from the confidences provided by the pixelwise semantic labeling algorithm. If no such information is available,  $p_\omega$  may be set to one if  $\omega_m = \omega$  and zero otherwise.
4. Does the measurement element  $m$  provide any evidence for grid cell  $C$  based on its spatial uncertainty? The according probability is given by the inverse sensor model  $\Pr(C|m)$ , presented in Section 3.3.3.

Consequently, the calculation reads

$$\begin{aligned}
 \Pr(m \rightarrow \omega, C) &= \Pr(A) + \Pr(B) + \Pr(C) + \Pr(D) \\
 &= p_{\text{FP}} \\
 &\quad + (1 - p_{\text{FP}}) \cdot (1 - p_{\text{occ}}) \\
 &\quad + (1 - p_{\text{FP}}) \cdot p_{\text{occ}} \cdot (1 - p_{\omega}) \\
 &\quad + (1 - p_{\text{FP}}) \cdot p_{\text{occ}} \cdot p_{\omega} \cdot (1 - \Pr(C|m)). \quad (3.3)
 \end{aligned}$$

For ground semantics  $\omega \subseteq \Omega_g$  the calculations are done analogously where the second query is negated as ground surfaces are assumed to be non-occupying.

### Free space estimation

A grid cell is free, if no obstacles are present in a defined free space corridor. The free space corridor is limited by the values  $f_{z,\min}, f_{z,\max} \in \mathbb{R}$  denoting the distance to the ground, where  $0 \leq f_{z,\min} < f_{z,\max} \leq d_{z,\max}$ . This relation ensures that the free space corridor is part of the driving corridor. The reason for defining another corridor specifically for the free space estimation is to allow traversing measurement parts of the driving corridor without providing free space evidence. For instance, rays might traverse grid cells below cars. The free space corridor is the height interval where it is very unlikely to have traversing measurements, if the cell is not free. Evidence for the absence of obstacles is provided by measurement rays traversing the grid cell. The free space evidence deduced from each traversing measurement ray is quantified as the ray height relative to the height of the free space corridor. The ray permeability

$$\rho = \frac{d_z}{f_{z,\max} - f_{z,\min}}. \quad (3.4)$$

is then calculated as the ratio between the height portion  $d_z$  covered by traversing measurement rays and the overall height of the free space corridor. Furthermore, evidence for a cell not being free is obtained by any measurement that provides occupancy evidence. The BBA

$$m(F_s) = \rho \cdot \left( 1 - \sum_{\psi \neq F_s} m(\psi) \right) \quad (3.5)$$

for free space  $F_s$  is then calculated as the product of the ray permeability  $\rho$  and the part of the BBA mass that has not been assigned to any of the occupancy hypotheses  $\omega \subseteq O_{\text{su}}$ .

### 3.3.1 Grid Mapping with Point Sets

Let the sensor measurement  $Z$  be a point set, i.e. a measurement element  $m \in Z$  is a detection coordinate indicating the presence of a reflecting surface with attached semantic label  $\omega_m$ . Note that other information such as LiDAR intensities or RaDAR Doppler measurements are omitted here as they are not considered in Equation (3.3). Point set measurements may be obtained from range sensors such as RaDARs. In point sets, no neighborhood relations between the coordinates can be deduced from the data structure. The calculation of surface normal vectors used in Definition 2.2 requires finding neighboring elements which is computational expensive. Therefore, occupancy is modeled according to Definition 2.1 which means that the detection point set is segmented into obstacle and ground detections based on a ground surface model such as presented in [Wir+21]. In Equation (3.3), the occupancy probability  $p_{\text{occ}}$  is then set to one if  $m$  was classified as occupying and to zero otherwise.

The ray permeability  $\rho$  used for the BBA estimation for the hypothesis *free*  $F_s$  is approximated as

$$\rho \approx \frac{h_{\max} - h_{\min}}{f_{z,\max} - f_{z,\min}}, \quad (3.6)$$

where  $h_{\min}, h_{\max} \in \mathbb{R}$  are the minimal and maximal measured heights of traversing measurement rays within the driving corridor. Note that this approximation may differ from the real ray permeability significantly around obstacles not connected to the ground.

This grid mapping framework for point sets has the disadvantage that a ground surface estimation or an external ground segmentation module is required to estimate the occupancy probability  $p_{\text{occ}}$ . This occupies additional computational resources and may introduce errors.

### 3.3.2 Grid Mapping with Images

In this section, the generic grid mapping pipeline is put into concrete terms for measurements given as images. The measurement images may be provided by different sensor types in different forms such as range images from LiDARs or depth/disparity images from cameras.

#### Outline of the processing steps

Throughout the processing steps, the representation is transformed to grids defined in different coordinate systems. In particular, the following grids are considered:

- *Sensor Grid.* The sensor grid  $\mathcal{G}_{uv}$  represents the measurement pattern of the sensor. One sensor reading on  $\mathcal{G}_{uv}$

$$Z = \{f_{\text{range}} : \mathcal{G}_{uv} \rightarrow \mathbb{R} \cup \{\text{unknown}\}, \\ f_{\text{sem}} : \mathcal{G}_{uv} \rightarrow \mathcal{S} \cup \{\text{unknown}\}\},$$

consists of a range measurement given by the mapping  $f_{\text{range}}$  and potentially semantic estimates given by the mapping  $f_{\text{sem}}$ . Here,  $\mathcal{S} = \mathcal{O}_{\text{su}} \cup \Omega_{\text{g}}$  is the set containing all singleton semantic hypotheses. One sensor element  $m \in Z$  is identified by the 3-tuple  $(C, r_m, \omega_m)$  consisting of the sensor grid cell  $C \in \mathcal{G}_{uv}$ , the range measurement  $r_m \in \mathbb{R}_{>0}$  and the semantic measurement  $\omega_m \in \mathcal{S}$ . The meaning of the range measurement  $r_m$  depends on the sensor and may be the measured distance for LiDAR sensors or the pixel disparity for stereo cameras. Note that the sensor reading  $Z$  marks the entry point of the estimation pipeline presented in this work and that there might be preprocessing steps required to obtain that information from the raw sensor measurements such as disparity calculation or pixelwise semantic labeling.

- *Measurement Grid.* The measurement grid  $\mathcal{G}_{ur}$  consists of the horizontal sensor grid index  $u$  in the first dimension and discretizes the range measurements interval of interest in the second dimension. When collecting measurements from the sensor grid in the measurement grid, an orthographic projection along the upright Cartesian coordinate axis in

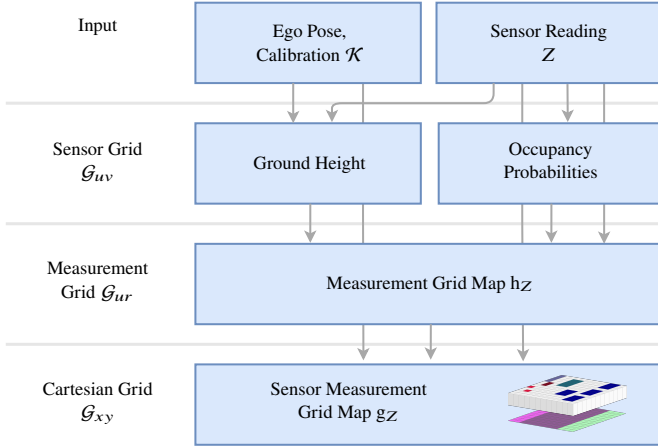


Figure 3.4: The processing blocks for calculating the BBA on the Cartesian grid based on input images.

the sensor coordinate system is performed. An example for a measurement grid is a grid in Polar coordinates.

- *Cartesian Grid*. The final sensor measurement grid map is defined on a Cartesian top-view grid  $\mathcal{G}_{xy}$ . It is defined on the reference coordinate system.

To indicate the corresponding coordinate system, the considered region of interest is subscripted analogously as  $\mathcal{R}_{uv}$ ,  $\mathcal{R}_{ur}$  and  $\mathcal{R}_{xy}$ , respectively. Furthermore, the mappings

$$\mathcal{T}_{uv}^{ur} : \mathcal{R}_{uv} \rightarrow \mathcal{R}_{ur}, \quad \mathcal{T}_{ur}^{xy} : \mathcal{R}_{ur} \rightarrow \mathcal{R}_{xy}$$

are introduced for transforming coordinates from one system to another.

The individual processing blocks for calculating the sensor measurement grid map  $g_z$  are depicted in Figure 3.4. The extrinsic and intrinsic sensor calibrations as well as the 6-dimensional ego pose consisting of the 3D position and orientation are assumed to be known. The first processing layer contains image processing steps on the sensor grid  $\mathcal{G}_{uv}$ . First, the surface normal vector is calculated for each measurement element based on the sensor calibration

and the input range image. Given the surface normal vectors, the occupancy probability  $p_{\text{occ}}$  from Equation (3.3) can be computed. Subsequently, each pixel is assigned a height above ground by propagating the heights of ground detections along each image column. This information is later used for the free space estimation based on the vertical sensor ray coverage. In the second processing layer a change of coordinates is applied from the sensor grid  $\mathcal{G}_{uv}$  to the top-view measurement grid  $\mathcal{G}_{ur}$ . This coordinate system is chosen according to the sensor characteristics, so that noise can be handled reasonably, and individual rays can be traced efficiently. Here, individual sensor reflections are mapped into the measurement grid map  $h_Z$  where evidence for occupancy and free space is accumulated. Finally, the sensor measurement grid map  $g_Z$  is calculated based on the measurement grid map  $h_Z$  in a common Cartesian grid  $\mathcal{G}_{xy}$ . In the following, the calculation steps on the three grids are explained in detail.

## Sensor Grid

On the sensor grid  $\mathcal{G}_{uv}$ , a surface analysis determining pixelwise occupancy probabilities  $p_{\text{occ}}$  and a ground analysis approximating the height above ground in each pixel are performed.

The surface at the reflection locations is analyzed to identify measurement elements that stem from occupying surfaces according to Definition 2.2. The decision if a measurement element stems from an occupying surface is made based on the surface normal vector at that location. Instead of providing a binary classification of each measurement element into occupying and non-occupying, we calculate the probability  $p_{\text{occ}} \in [0, 1]$  that the surface reflecting the measurement element is occupying. This occupancy probability  $p_{\text{occ}}$  can scale the resulting BBA whereas a binary classification results in a loss of information.

Following [New+11], a bilateral filter is applied to the measurement elements to eliminate noise while preserving edges. Here, the geometric context plays an important role. Averaging applied in the bilateral filter provides desired results if the measurement elements used for calculating the average have similar sources with zero-mean disturbances. In order to increase the likelihood for this, the range image  $f_{\text{range}}$  is transformed to images  $f_{\text{height}}$  and  $f_{\text{distXY}}$  representing



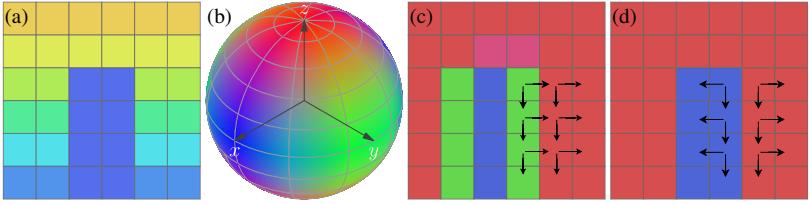


Figure 3.5: Pixel-wise surface normal vector calculation based on neighboring pixels: (a) an except of a range image with foreground (blue) and background (blue-orange), (b) the color map on the unit sphere used to visualize surface normal vectors, (c) naive normal vector calculation, (d) normal vector calculation using nearest neighboring pixels.

the height of the measurement element relative to the sensor origin and the distance to the sensor origin projected to the XY-plane. This selection is based on the assumption that environments in traffic scenes can be separated into sub-planes that are mostly oriented along the XY-plane or perpendicular to it. This holds for the ground surface as well as many objects as buildings and traffic participants. The bilateral filter is then applied to  $f_{\text{height}}$  and  $f_{\text{distXY}}$  with parameters  $\sigma_{r,\text{height}}$  and  $\sigma_{r,\text{distXY}}$  denoting the range standard deviation and  $\sigma_{d,\text{height}}$  and  $\sigma_{d,\text{distXY}}$  denoting the spatial standard deviation.

Based on the filtered images  $\tilde{f}_{\text{height}}$  and  $\tilde{f}_{\text{distXY}}$ , the surface normal vector can be approximated for each measurement element. Here, the image representation has strong advantages over other representations such as unordered point sets as it implicitly defines a neighborhood. The selection of neighboring pixels used for approximating the surface normal can be crucial. When considering the same neighborhood in each measurement element, the considered values might represent surfaces of different entities and the surface normal calculation will be erroneous, see Figure 3.5c. To minimize this effect, the considered neighborhood is adapted based on the range measurements. To find the horizontally adjacent pixel  $C_h$ , the next pixel to the left and to the right are considered and the one that has the smaller Euclidean distance in the 3D Cartesian space to the measurement elements is chosen. For the vertically adjacent pixel  $C_v$ , the next pixel below and above are considered. In case no direct neighbors can be found, the second next pixels are considered and so forth until a maximal neighborhood size is reached. This might be necessary in case no reliable measurements were recorded at neighboring locations. The selection

of adjacent pixels used for the normal vector calculation is demonstrated in Figure 3.5d.

Given the filtered images  $\tilde{f}_{\text{height}}$  and  $\tilde{f}_{\text{distXY}}$ , the sensor calibration  $\mathcal{K}$  and the pixels  $C, C_h, C_v \in \mathcal{G}_{uv}$ , the Cartesian coordinates  $\mathbf{p}, \mathbf{p}_h, \mathbf{p}_v \in \mathbb{R}^3$  can be calculated and the normal vector  $\mathbf{n}$  can be determined by computing and normalizing the corresponding cross product

$$\mathbf{n} = (n_1, n_2, n_3)^T = \frac{(\mathbf{p}_h - \mathbf{p}) \times (\mathbf{p}_v - \mathbf{p})}{\|(\mathbf{p}_h - \mathbf{p}) \times (\mathbf{p}_v - \mathbf{p})\|}. \quad (3.7)$$

Based on the surface normal vector  $\mathbf{n}$ , an occupancy weight  $w_{\text{occ}}$  is calculated. A logistic function centered around  $\frac{\pi}{4}$  is applied to  $\arccos(n_3)$ , i.e. the angle between  $\mathbf{n}$  and the North Pole  $(0, 0, 1)^T$  as

$$w_{\text{occ}} = \frac{1}{1 + \exp(-k(\arccos(n_3) - \frac{\pi}{4}))}, \quad (3.8)$$

where  $k \in \mathbb{R}_{>0}$  is a scaling factor. The logistic function is parametrized so that we have  $w_{\text{occ}} = 0.5$  for  $\arccos(n_3) = \frac{\pi}{4}$  which is based on the geometric consideration that a surface is considered blocking if its slope exceeds  $45^\circ$ .

The credibility of the calculated surface normal vector depends on the distance between the measurement element and its neighbors. If the distance is small it might be dominated by inaccurate range measurements and the resulting normal vector might be disturbed. Therefore, we apply another logistic function to model a normal vector confidence value

$$\text{conf}_{\mathbf{n}} = \frac{1}{1 + \exp(-k'(\min(\|\mathbf{p}_h - \mathbf{p}\|, \|\mathbf{p}_v - \mathbf{p}\|) - \sigma_{\text{range}}))} \quad (3.9)$$

quantifying if the minimal distance to the neighboring pixel coordinates is smaller or larger than the standard deviation  $\sigma_{\text{range}}$  of the range measurement. The constant  $k'$  is another scaling factor. Finally, the occupancy probability is calculated as

$$p_{\text{occ}} = \text{conf}_{\mathbf{n}} \cdot w_{\text{occ}}. \quad (3.10)$$

The algorithm transforming the input range image  $f_{\text{range}}$  to an image  $f_{\text{occ}}$  containing the occupancy probabilities for each measurement element is summarized in Algorithm 3.1.

**ALGORITHM 3.1:** Calculate occupancy cell weights.

---

```

Input :  $f_{\text{range}}, \mathcal{K}$  /* Range image and calibration */
Output:  $f_{\text{occ}}$  /* Image with occupancy cell weights */
1  $\tilde{f}_{\text{height}} \leftarrow \text{computeHeight}(f_{\text{range}}, \mathcal{K})$  /* Height relative to sensor */
2  $\tilde{f}_{\text{distXY}} \leftarrow \text{computeDistXY}(f_{\text{range}}, \mathcal{K})$  /* Distance projected to XY-plane */
3  $\tilde{\tilde{f}}_{\text{height}} \leftarrow \text{bilateralFilter}(\tilde{f}_{\text{height}})$  /* Denoising */
4  $\tilde{\tilde{f}}_{\text{distXY}} \leftarrow \text{bilateralFilter}(\tilde{f}_{\text{distXY}})$  /* Denoising */
5 for  $C$  in  $\mathcal{G}_{uv}$  do
   /* Find adjacent pixel indices likely to be on same plane */
6    $C_h \leftarrow \text{adjazentHorizontal}(f_{\text{range}}, \mathcal{K}, C)$ 
7    $C_v \leftarrow \text{adjazentVertical}(f_{\text{range}}, \mathcal{K}, C)$ 
   /* Calculate 3D coordinates */
8    $\mathbf{p} \leftarrow \text{toCoordinate}(\tilde{\tilde{f}}_{\text{height}}, \tilde{\tilde{f}}_{\text{distXY}}, \mathcal{K}, C)$ 
9    $\mathbf{p}_h \leftarrow \text{toCoordinate}(\tilde{\tilde{f}}_{\text{height}}, \tilde{\tilde{f}}_{\text{distXY}}, \mathcal{K}, C_h)$ 
10   $\mathbf{p}_v \leftarrow \text{toCoordinate}(\tilde{\tilde{f}}_{\text{height}}, \tilde{\tilde{f}}_{\text{distXY}}, \mathcal{K}, C_v)$ 
   /* Compute normal vector (Equation (3.7)) */
11   $\mathbf{n} \leftarrow \text{surfaceNormal}(\mathbf{p}, \mathbf{p}_h, \mathbf{p}_v)$ 
   /* Compute occupancy weight (Equation (3.8)) */
12   $w_{\text{occ}} \leftarrow \text{occupancyWeight}(\mathbf{n}, k)$ 
   /* Compute normal vector confidence (Equation (3.9)) */
13   $\text{conf}_{\mathbf{n}} \leftarrow \text{surfaceNormalConfidence}(\mathbf{p}, \mathbf{p}_h, \mathbf{p}_v, \sigma_{\text{range}})$ 
   /* Compute occupancy probability (Equation (3.10)) */
14   $f_{\text{occ}}(C) \leftarrow \text{conf}_{\mathbf{n}} \cdot w_{\text{occ}}$ 
15 return  $f_{\text{occ}}$ 

```

---

Besides the image  $f_{\text{occ}}$  containing the occupancy probabilities, the distance to ground is estimated for each measurement element. This information is needed to exclude measurements outside the considered driving corridor and to clip measurement rays that intersect with the free space corridor boundaries. One option is to explicitly estimate a ground surface using a parametric model such as a plane or a 2D B-spline in the Cartesian space as in [Wir+21]. However, estimating a parametric model often comes along with solving an optimization problem with computationally time-consuming numerical solvers. Instead, we estimate the distance to ground directly for each measurement element. Therefore, measurement elements that belong to measurement rays hitting the ground surface are classified, and the measured height is assigned directly. Then the missing regions, i.e. rays that have been reflected by objects, have to be filled. One best guess for this is to traverse the measurement image column-wise and propagate the last known ground height. The corresponding algorithm is

**ALGORITHM 3.2:** Calculate pixel wise ground height.

---

```

Input :  $f_{\text{normals}}, f_{\text{height}}$  /* Blocking probability and detection height */
Output:  $f_{\text{ground}}$  /* Image containing pixel wise ground height */

1 hit ← False
2 for  $C$  in  $\mathcal{G}_{uv}$  do
3    $\mathbf{n}$  ←  $f_{\text{normals}}(C)$ 
4   if isGround( $\mathbf{n}$ ) AND (hit == False) then
5      $f_{\text{ground}}(C)$  ←  $f_{\text{height}}(C)$ 
6      $h_{\text{last}}$  ←  $f_{\text{height}}(C)$ 
7   else if isGround( $\mathbf{n}$ ) AND (hit == True) AND ( $f_{\text{height}}(C) < h_{\text{last}}$ ) then
8      $f_{\text{ground}}(C)$  ←  $f_{\text{height}}(C)$ 
9      $h_{\text{last}}$  ←  $f_{\text{height}}(C)$ 
10  else
11     $f_{\text{ground}}(C)$  ←  $h_{\text{last}}$ 
12    hit ← True
13 return  $f_{\text{ground}}$ 

```

---

formulated in Algorithm 3.2. The function `isGround` classifies measurement elements into ground and obstacle detections based on the following heuristic considerations: Let  $C_{uv} \in \mathcal{G}_{uv}$  be the current measurement element. In case  $C_{uv}$  is not located in the bottom row, let  $C_{uv}^{-1} \in \mathcal{G}_{uv}$  be the element located below  $C_{uv}$ . Based on similar geometric considerations as in the calculations of the occupancy weight  $w_{\text{occ}}$  (Equation (3.8)), the measurement element is classified as obstacle, if

- the angle between the surface normal vector  $\mathbf{n}$  and the North Pole  $(0, 0, 1)^T$  exceeds  $45^\circ$ , i.e.  $\arccos(n_3) > \frac{\pi}{4}$ , or
- $C_{uv}$  is located in the bottom row and the vertical component of the detection coordinate minus the sensor height exceeds a given threshold, or
- $C_{uv}$  is not located in the bottom row and the Euclidean distance in the 3D Cartesian space to the measurement in  $C_{uv} \in \mathcal{G}_{uv}$  is smaller than the distance to the measurement in  $C_{uv}^{-1} \in \mathcal{G}_{uv}$ .

After the first obstacle detection was found in a column, subsequent elements are only classified as ground, if the upper conditions are not fulfilled and the

measured height has decreased compared to the measurement in  $C_{uv}^{-1}$ . This is crucial to prevent classifying horizontal surfaces on obstacles as ground.

## Measurement Grid

On the measurement grid  $\mathcal{G}_{ur}$ , measurement elements are assigned to the occupancy hypotheses  $\omega \subseteq \Omega_s$  and ground hypotheses  $\omega \subseteq \Omega_g$  and spatial uncertainty is modeled by applying the inverse sensor model. The multi-layer grid map

$$h_Z: \mathcal{G}_{ur} \times (\mathcal{P}(\Omega_g) \cup \mathcal{P}(\Omega_s)) \rightarrow \mathbb{R} \quad (3.11)$$

accumulates measurement elements for each hypothesis  $\omega \subseteq \Omega_s$  and  $\omega \subseteq \Omega_g$ , respectively, in the corresponding grid layer  $h_Z(\cdot, \omega)$ . Based on Equations (3.2) and (3.3), the logarithms of the probabilities  $\Pr(m \rightarrow \omega, C)$  are accumulated as

$$h_Z(C, \omega) = \sum_{C_{uv} \in \mathcal{G}_{uv}} \log(\Pr(m \rightarrow \omega, C)) \quad (3.12)$$

in grid cell  $C \in \mathcal{G}_{ur}$  for the hypotheses  $\omega \in \mathcal{P}(\Omega_g) \cup \mathcal{P}(\Omega_s) \setminus F_s$ .

For the free space hypothesis  $F_s \subset \Omega_s$ , the ray permeability  $\rho$  defined in Equation (3.4) is estimated. Therefore, we calculate

$$h_Z(C_{ur}, F_s) = \rho_{C_{ur}}. \quad (3.13)$$

For this purpose, a 3D ray casting is applied based on the sensor intrinsics. The rationale behind this is that the portion of space covered by a measurement ray between the reflecting surface and the sensor origin provides free space evidence. Each measurement ray contributes to the observed height where the contribution is defined by the ray divergence as sketched in Figure 3.6. For this purpose, we approximately model the vertical ray coverage to be dense, i.e. it is assumed that there are no gaps between the rays. This is justified by the high vertical resolution of the sensors used in this work due to which only insignificant entities may be missed at full ray coverage. Consequently, the ray permeability  $\rho_{C_{ur}}$  in grid cell  $C_{ur} \in \mathcal{G}_{ur}$  is computed as

$$\rho_{C_{ur}} = \frac{d_z}{f_{z,\max} - f_{z,\min}}, \quad d_z = \sum_{C_{uv} \in \mathcal{G}_{uv}} z(r_m), \quad (3.14)$$

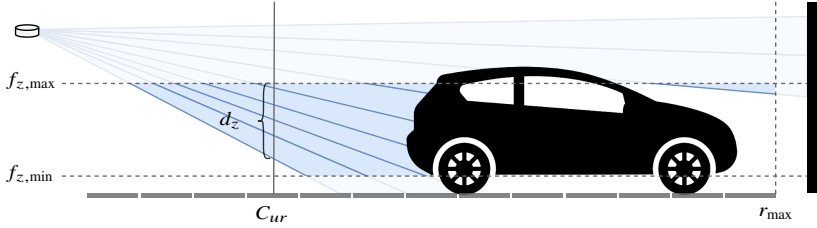


Figure 3.6: Calculation of the ray permeability  $\rho_{C_{ur}}$  in a measurement grid cell  $C_{ur} \in \mathcal{G}_{ur}$ . The rays are clipped according to the minimal height  $f_{z,\min}$ , the maximal height  $f_{z,\max}$  and maximal range  $r_{\max}$ .

where  $z$  is a function calculating the height of a measurement ray at range  $r_m$ . Each measurement ray is clipped according to  $f_{z,\min}$ ,  $f_{z,\max}$  as only parts of the ray that traverse the defined free space corridor contribute to the free space estimation. Positive and negative ray heights are mapped into the underlying grid cells to mark end and start points of the clipped measurement rays. The missing gaps can then be filled by simply computing the running sum and correcting the ray heights to account for the ray divergence. Algorithm 3.3 summarizes the steps to fill  $h_Z(C, F_s)$  with the accumulated height intervals observed by measurement rays.

## Cartesian Grid

After calculating the measurement grid map  $h_Z$ , the second change of coordinates from the measurement grid  $\mathcal{G}_{ur}$  to the Cartesian grid  $\mathcal{G}_{xy}$  is applied. Therefore, cell values  $h_Z(C, \omega)$  must be transformed properly into the Cartesian representation  $h_{xy}(C, \omega)$ . For all hypotheses  $\omega \neq F_s$  except free space, this is done by integrating  $h_Z$  over  $\mathcal{T}_{xy}^{ur}(C) \subset \mathcal{R}_{ur}$  as

$$h_{xy}(C, \omega) = \int_{\mathcal{T}_{xy}^{ur}(C)} h_Z(x, \omega) dx. \quad (3.15)$$

**ALGORITHM 3.3:** Calculate accumulated observed height  $h_Z(C, F_s)$ 


---

```

Input :  $f_{\text{range}}, f_{\text{ground}}, \mathcal{K}$ 
Output:  $h_Z(C, F_s)$ 

/* Map measurement elements */
1 for  $C$  in  $\mathcal{G}_{uv}$  do
2    $C_{\text{cur}}$   $\leftarrow$  computeSensorGridCell( $f_{\text{range}}, \mathcal{K}, C$ )
3   [ $C_{\text{start}}, C_{\text{end}}$ ]  $\leftarrow$  clipRay( $\mathcal{K}, C, f_{\text{ground}}, f_{z,\text{min}}, f_{z,\text{max}}$ )
   /* Compute ray heights */
4    $h_{\text{start}}$   $\leftarrow$   $z(f_{\text{range}}, \mathcal{K}, \text{cellStart})$ 
5    $h_{\text{end}}$   $\leftarrow$   $z(f_{\text{range}}, \mathcal{K}, \text{cellEnd})$ 
6    $h_Z(C_{\text{start}}, F_s)$   $\leftarrow$   $h_Z(C_{\text{start}}, F_s) - h_{\text{start}}$ 
7    $h_Z(C_{\text{end}}, F_s)$   $\leftarrow$   $h_Z(C_{\text{end}}, F_s) + h_{\text{end}}$ 

/* Cast rays */
8 for  $u$  in [ $u_{\text{Min}}, \dots, u_{\text{Max}}$ ] do
9    $\text{acc}$   $\leftarrow$  0 /* Running sum */
10   $h_{\text{last}}$   $\leftarrow$   $z(f_{\text{range}}, \mathcal{K}, (u, r_{\text{last}}))$ 
11  for  $r$  in [ $r_{\text{Max}}, \dots, r_{\text{Min}}$ ] do
12     $h$   $\leftarrow$   $z(f_{\text{range}}, \mathcal{K}, (u, r))$ 
13     $\text{acc}$   $\leftarrow$   $h/h_{\text{last}} * \text{acc} + h_{\text{Obs}}((u, r))$ 
14     $h_{\text{Obs}}((u, r))$   $\leftarrow$   $\text{acc}$ 
15     $h_{\text{last}}$   $\leftarrow$   $h$ 
16 return  $h_{\text{Obs}}$ 

```

---

For free space  $\omega = F_s$ , the average over  $\mathcal{T}_{xy}^{ur}(C)$  is calculated as

$$h_{xy}(C, F_s) = \frac{1}{\mu(\mathcal{T}_{xy}^{ur}(C))} \int_{\mathcal{T}_{xy}^{ur}(C)} h_Z(x, \omega) dx. \quad (3.16)$$

Here,  $\mu(\cdot)$  denotes the 2D Lebesgue-measure which calculates the area of  $\mathcal{T}_{xy}^{ur}(C)$ .

The Cartesian grid map  $h_{xy}$  is subsequently transformed to a consistent BBA represented by the sensor measurement grid map  $g_Z$ . Following Equations (3.2), (3.3) and (3.5), it is computed as

$$g_Z(C, \omega) = \begin{cases} k(1 - \exp(h_{xy}(C, \omega))), & \text{if } \omega \subseteq \Omega_s \text{ or } \omega \in \Omega_g, \\ \left(1 - \sum_{\psi \neq F_s} g_Z(C, \psi)\right) h_{xy}(C, \omega), & \text{if } \omega = F_s, \\ 0, & \text{else,} \end{cases}$$

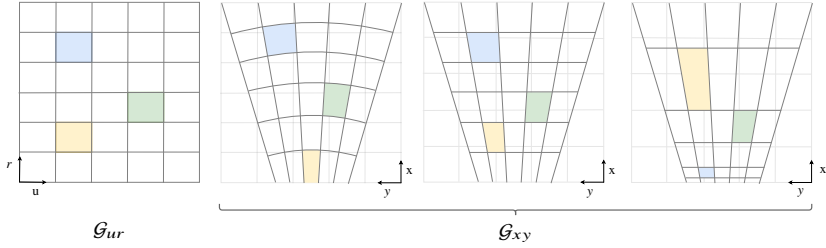


Figure 3.7: The measurement grid  $\mathcal{G}_{ur}$  on the left and the transformed Polar, u/distance and u/disparsity grid  $\mathcal{T}_{ur}^{xy}(\mathcal{G}_{ur})$ . Matching grid cell colors indicate the same grid cell in the source grid and the warped target grid.

where

$$k = \frac{1 - \exp\left(-\sum_{\omega \subseteq \Omega_s} h_Z(C, \omega)\right)}{\sum_{\omega \subseteq \Omega_s} 1 - \exp(h_Z(C, \omega))} \quad (3.17)$$

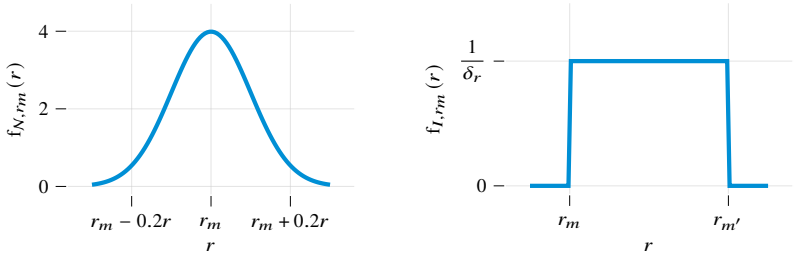
is a normalization factor distributing BBA masses equally to conflicting hypotheses.

### 3.3.3 Grid Mapping Considering Sensor Modalities

Up to this point, the inverse sensor model  $\Pr(C|m)$  and the measurement grid  $\mathcal{G}_{ur}$  were not specified as they depend on the sensor modalities. In this section, the measurement grids designed for LiDARs, monocular cameras and stereo cameras and two inverse sensor models are presented.

The measurement grid  $\mathcal{G}_{ur}$  used for LiDAR is defined in Polar coordinates, i.e. the horizontal index  $u$  corresponds to the angle location of the spinning laser and the range measurement is the measured distance projected to the  $xy$ -plane. For measurements obtained from a monocular camera where the distance was estimated directly in each pixel, an u/distance grid is used. In case disparity estimates obtained from a stereo camera are mapped, an u/disparsity grid is used. Figure 3.7 shows the relation between the measurement grid  $\mathcal{G}_{ur}$  and the Cartesian grid  $\mathcal{G}_{xy}$  for all three measurement grids used in this work. The cell values in the measurement grid map  $h_Z$  are transformed to Cartesian coordinates using Equations (3.15) and (3.16).





(a) The PDF for the Gaussian inverse sensor model.

(b) The pseudo PDF for the interval inverse sensor model.

Figure 3.8: The (pseudo) PDFs of the range  $r$  used in the two presented inverse sensor models.

The inverse sensor model is calculated in the measurement grid  $\mathcal{G}_{ur}$ . Let a measurement grid cell  $C = I_u \times I_r \in \mathcal{G}_{ur}$  be divided into its  $u$  and  $r$  components. We model the measurement  $m$  to be uniformly distributed in  $I_u$ . Hence, the inverse sensor model simplifies to

$$\Pr(C|m) = \Pr(r|r_m) = \int_{r \in I_r} f_{r_m}(r) dr, \quad (3.18)$$

where  $f_{r_m}$  is the probability density function (PDF) of the range  $r$  given the range measurement  $r_m$ . In this work, two options for  $f_{r_m}$  are presented:

1. *The Gaussian Model:* We model the range to be normally distributed with mean  $r_m$  and standard deviation  $\sigma_r$ , i.e.

$$f_{N,r_m}(r) = \frac{1}{\sqrt{2\pi} \sigma_r} \exp\left(-\frac{1}{2} \left(\frac{r - r_m}{\sigma_r}\right)^2\right). \quad (3.19)$$

In the remainder of this thesis, we refer to this model as Gaussian inverse sensor model  $\Pr_N(C|m)$ . It is sketched in Figure 3.8a with standard deviation  $\sigma_r = 0.1r$ .

2. *The Interval Model:* The measurement element  $m$  covers the whole interval  $[r_m, r_{m'}]$ , i.e.

$$f_{I,r_m}(r) = \begin{cases} \frac{1}{\delta_r}, & \text{if } r \in [r_m, r_{m'}] \\ 0, & \text{else,} \end{cases} \quad (3.20)$$

where  $\delta_r$  is the length of the range interval  $I_r$  of one grid cell  $C \in \mathcal{G}_{uv}$  and  $r_{m'}$  is the range measurement in the sensor grid cell  $C' = (u, v + 1)$  that is vertically adjacent to the considered measurement element  $m$  in  $C$ . Note that  $f_{I,r_m}(r)$  is normalized so that it integrates to one over the range interval  $I_r$  of a fully supported grid cell. This model is based on the assumption that the measurement  $Z$  partitions the measured surface which means that there are no gaps between areas on the world surface covered by rows in the sensor grid  $\mathcal{G}_{uv}$ . In the remainder of this thesis, we refer to this model as interval inverse sensor model  $\text{Pr}_I(C|m)$ . It is sketched in Figure 3.8b.

We apply  $f_{I,r_m}$  only in the calculation of the BBA on the ground hypotheses  $\Omega_g$  as the model assumption is violated for occupying surfaces. Furthermore, we propose using it only for camera measurements and not for LiDAR measurements as LiDAR scan lines are usually not adjacent. In comparison, the gaps between pixel rows in camera sensor used in automotive applications are negligible.

## 3.4 Experiments

Our proposed grid mapping framework is validated using the Kitti Vision Benchmark [GLU12] and the semantic LiDAR point cloud labels from the SemanticKITTI dataset extension [Beh+19]. They contain measurements from a Velodyne HDL-64E LiDAR scanner with surround view, one RGB and one grayscale stereo camera setup pointing to the front, the sensor calibrations and 6D ego pose annotations. In this work, the measurements from the Velodyne HDL-64E and the stereo RGB camera setup are processed with the presented grid mapping pipeline. The sensor coverage of this sensor setup is shown in Figure 3.9.

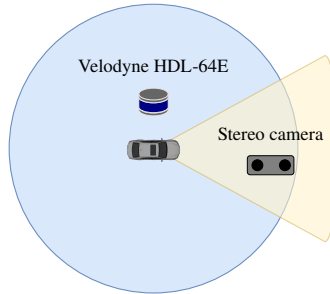


Figure 3.9: The sensor setup consisting of a 360° LiDAR scanner and a stereo camera pointing to the front.

### 3.4.1 Qualitative Results

In the sensor grid, surface normal vectors are calculated based on the range estimates and occupancy probabilities are deduced as described in Algorithm 3.1.

Figure 3.10a shows the image processing steps for a range image of a KITTI-360 measurement taken from the Kitti dataset. Note that the intrinsic calibration is already applied here meaning that rotational position and distance of each reflection was corrected according to the intrinsic calibration parameters. The gaps in the range image result from shadows caused by sensors and antennas on top of the test vehicle or missing reflections due to surfaces with low reflectivity. The latter tends to occur on surfaces at high distances larger than 100m or on dark surfaces such as black cars or windows. In order to obtain one single image containing the range measurements, multiple returns are omitted in this work meaning that only the lowest range measurement is stored in the range image. Figure 3.10b contains the semantic labels assigned to each LiDAR detection in the SemanticKITTI dataset extension. Figure 3.10c shows the image containing the surface normal vectors calculated in Algorithm 3.1 colorcoded according to Figure 3.5b and Figure 3.10d shows the resulting image  $f_{\text{occ}}$  containing the occupancy probabilities. Recall that the occupancy probabilities in  $f_{\text{occ}}$  scale with the orientation of the reflecting surface and thus do not segment the environment in objects and ground. Consequently, there are areas on objects with low occupancy probability on horizontal surfaces such as the hood or the rooftop of cars.

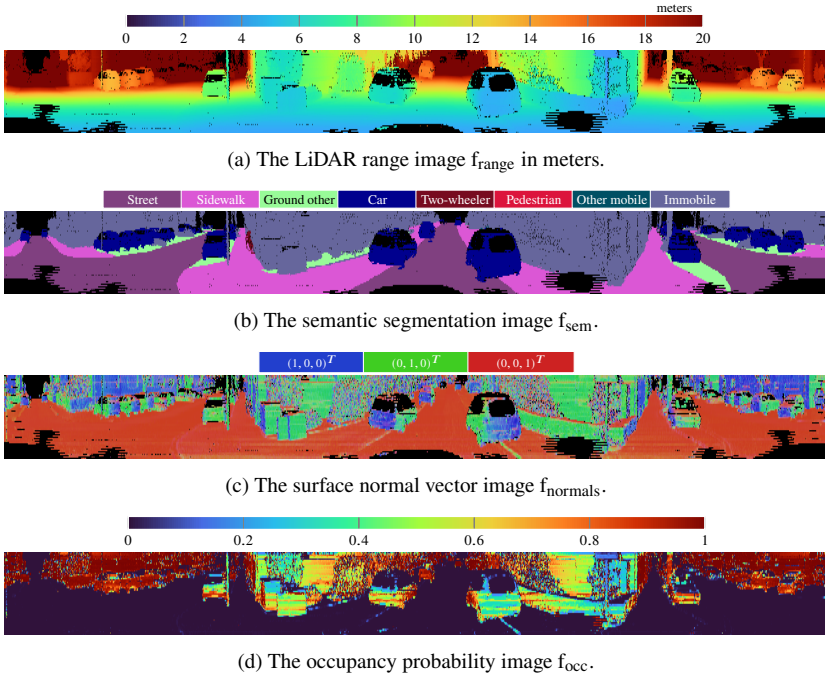


Figure 3.10: Impressions of the sensor grid processing chain for a LiDAR sensor reading.

Figure 3.11 shows the image processing results for an image pair of the stereo camera in the Kitti Tracking dataset. The image view, see Figure 3.11a, shows the center part of the LiDAR range image shown in Figure 3.10. The disparity image depicted in Figure 3.11b was estimated using the guided aggregation net for stereo matching presented by Zhang et al. [Zha+19]. This is a well performing stereo disparity estimator generating dense disparity maps that comes along with a real-time capable implementation running at 15-20 frames per seconds. The pixelwise semantically labelled image in Figure 3.11c was obtained by feeding the RGB image recorded by the left camera into the network presented by Zhu et al. [Zhu+19]. The inference on this network is not real-time capable, but similarly performing, real-time capable alternatives haven been proposed e.g. in [Hon+21]. Figure 3.11d visualizes the surface normal vectors  $f_{\text{normals}}$  and Figure 3.11e the corresponding occupancy probability image  $f_{\text{occ}}$ .

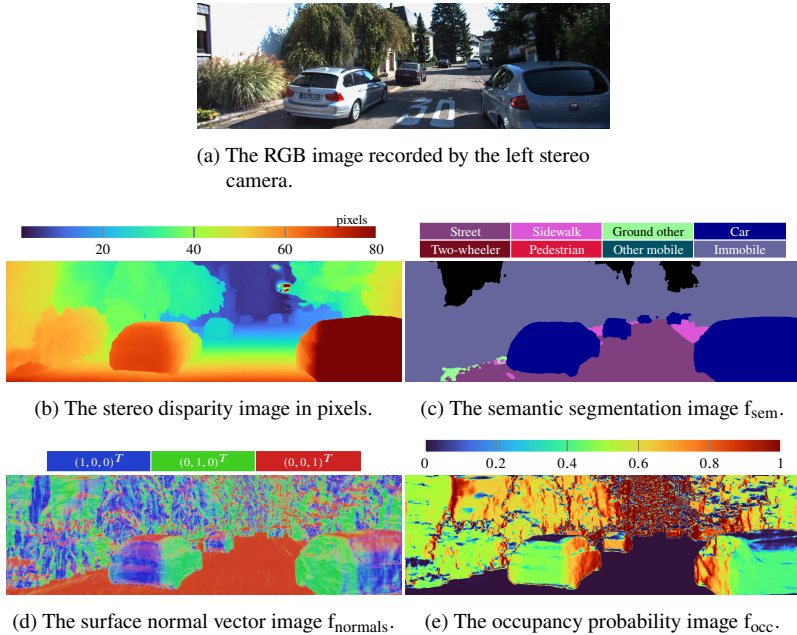
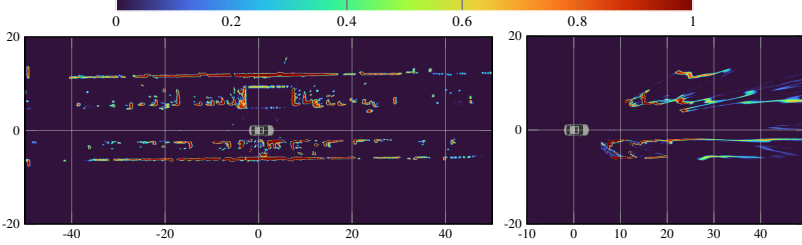


Figure 3.11: Impressions of the sensor grid processing chain for a stereo camera sensor reading.

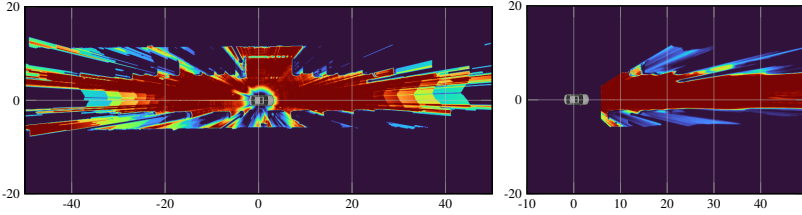
Figure 3.12 shows the final BBA for the hypotheses *occupied* and *free* based on a LiDAR scan and stereo camera images, respectively. It can be seen that the proposed estimation process accounts for the different sensory characteristics and their effect to the measurement uncertainty. This becomes visible in the BBAs for *occupied* shown in Figure 3.12b. The uncertainty of the range estimate leads to blurry occupancy patterns. The sensor measurement grid map of the stereo camera shows increasing blurriness and thus increasing range estimate uncertainty with higher distances to the sensor origin. In the LiDAR sensor measurement grid map on the other hand, this uncertainty is independent of the distance. This is due to the fact that stereo measurement elements are mapped in an  $u/\text{disparity}$  grid where the discretization becomes coarser at larger distances. Recall that the BBA of the hypothesis *free* in Figure 3.12c models the 3D ray geometry as depicted in Figure 3.1. That means that it shows the percentage of the height interval of interest that can be observed. It can be seen



(a) Image taken by the front left color camera.



(b) The BBA of the hypothesis *occupied*  $m(O_{su})$ .

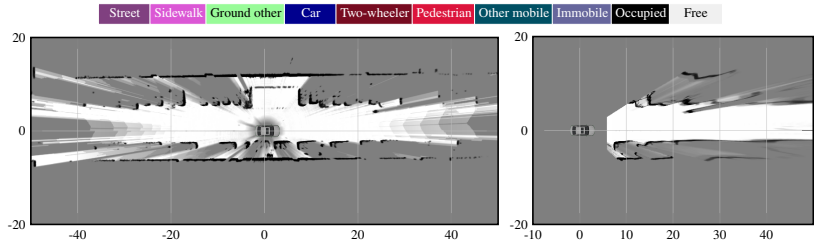


(c) The BBA of the hypothesis *free*  $m(F_s)$ .

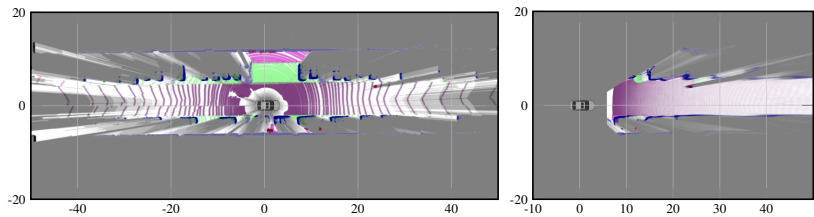
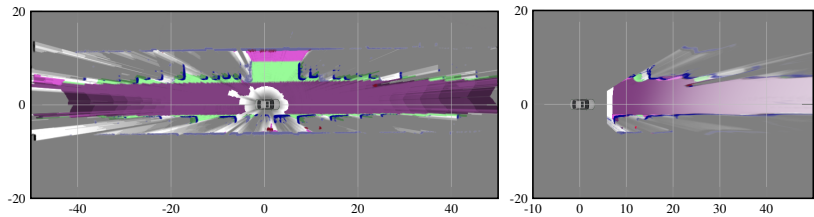
Figure 3.12: Resulting BBA for the hypotheses *occupied*  $O_{su}$  and *free*  $F_s$  in the Cartesian grid  $\mathcal{G}_{xy}$  using LiDAR and stereo camera measurements without semantic estimates.

that no-returns caused by low reflectivity or rays reflected by the sensor setup on the ego vehicles rooftop lower the free space evidence masses. Additionally, it is visible that cells very close to the sensor origin cannot be observed.

Figure 3.13 visualizes the sensor measurement grid maps estimated with the same measurements. First, no semantic estimates are included, i.e. only the  $360^\circ$  LiDAR scan and the stereo disparity map are processed. Figure 3.13a shows the occupancy probability after applying the pignistic transformation (Equation (2.7)). Here, the gray values transition from white for zero to black denoting an occupancy probability of one. In Figures 3.13b and 3.13c the resulting sensor measurement grid map is visualized when additionally



(a) Sensor measurement grid maps without semantic estimates.

(b) Sensor measurement grid maps with semantic estimates using the Gaussian inverse sensor model  $\Pr_{\mathcal{N}}(C|m)$  for both ground semantics  $\Omega_g$  and occupancy semantics  $\Omega_s$ .(c) Sensor measurement grid maps with semantic estimates using the Gaussian inverse sensor model  $\Pr_{\mathcal{N}}(C|m)$  for occupancy semantics  $\Omega_s$  and the interval inverse sensor model  $\Pr_I(C|m)$  for ground semantics  $\Omega_g$ .Figure 3.13: Resulting BBA visualizations in the Cartesian grid  $\mathcal{G}_{xy}$  using LiDAR (left) and stereo camera measurements (right) with and without semantic estimates.

processing the semantic estimates shown in Figures 3.10b and 3.11c. The results clearly show the differences in modelling ground detections with the Gaussian inverse sensor model  $\Pr_{\mathcal{N}}(C|m)$  in Figure 3.13b and with the interval inverse sensor model  $\Pr_I(C|m)$  in Figure 3.13c. Whereas the BBA estimation for the hypothesis street is sparse using the Gaussian model, evidence for the whole area of the street covered by the sensor grid is obtained when using the interval model.

### 3.4.2 Quantitative Evaluation

One of the key tasks in the sensor measurement grid map estimation is the deduction of occupancy evidence based on the measurements. We demonstrate the differences when applying the presented framework to point sets and images. Recall that two ways of defining the term occupied in a geometric manner were presented where Definition 2.1 is used for point sets and Definition 2.2 for images. To make the evaluation representable, a subsequence of the Kitti odometry benchmark is chosen that was recorded on challenging terrain with altering height. The BBA for the hypothesis *occupied* is calculated with one of the following three methods:

1. *Flat world model.* Derive evidence for occupancy according to Definition 2.1 as described in Section 3.3.1. The ground surface is modeled as a xy-plane  $\{(x, y, z) \mid z = 0\}$  in vehicle coordinates and the tolerance margin is set to  $\delta_G = 0.3\text{m}$ . The resulting BBA is denoted as  $m_{\text{flat}}$ .
2. *B-spline model.* Derive evidence for occupancy according to Definition 2.1 as described in Section 3.3.1. The ground surface is represented by the uniform B-spline model proposed by Wirges et al. [Wir+21] and the tolerance margin is set to  $\delta_G = 0.3\text{m}$ . The resulting BBA is denoted as  $m_{\text{spline}}$ .
3. *Surface normals.* Derive evidence for occupancy according to Definition 2.2 as described in Section 3.3.2. The resulting BBA is denoted as  $m_{\text{normals}}$ .

To evaluate the resulting BBA for a cell being occupied the following BBAs are calculated:



- The reference BBA  $m_{\text{ref}}$  calculated using Equation (3.2) where the occupancy probability is set to

$$p_{\text{occ}} = \begin{cases} 1, & \text{if } \omega_{\text{ref}} \subseteq O_{\text{su}}, \\ 0, & \text{else,} \end{cases}$$

where  $\omega_{\text{ref}}$  is the semantic label added in the SemanticKITTI dataset extension.

- The BBA  $m_{\text{all}}$  containing all detections, i.e. the occupancy weight is  $p_{\text{occ}} = 1$ .
- The three estimated BBAs  $m_{\text{flat}}$ ,  $m_{\text{spline}}$  and  $m_{\text{normals}}$  calculated as described above.

Note that the reference BBA  $m_{\text{ref}}$  is not a ground truth classification based on the same geometric cues as used in the estimation. As opposed to defining occupancy based on geometric constraints as in Definitions 2.1 and 2.2, the reference BBA deduces occupancy based on semantic constraints. In order to create a ground truth BBA based on geometric constraints, a complete 3D surface model of the environment would be required. However, the comparison considered here still yields interpretable information on the performance of the BBA estimation. Based on  $m_{\text{ref}}$ ,  $m_{\text{all}}$  and  $m_i$ ,  $i \in \{\text{flat}, \text{spline}, \text{normals}\}$ , the confusion metrics

$$\begin{aligned} \xi_{\text{TP},i} &= \tilde{m}_i(O_{\text{su}}) \tilde{m}_{\text{ref}}(O_{\text{su}}) m_{\text{all}}(O_{\text{su}}), \\ \xi_{\text{FP},i} &= \tilde{m}_i(O_{\text{su}}) (1 - \tilde{m}_{\text{ref}}(O_{\text{su}})) m_{\text{all}}(O_{\text{su}}), \\ \xi_{\text{FN},i} &= (1 - \tilde{m}_i(O_{\text{su}})) \tilde{m}_{\text{ref}}(O_{\text{su}}) m_{\text{all}}(O_{\text{su}}), \\ \xi_{\text{TN},i} &= (1 - \tilde{m}_i(O_{\text{su}})) (1 - \tilde{m}_{\text{ref}}(O_{\text{su}})) m_{\text{all}}(O_{\text{su}}) \end{aligned}$$

are defined per grid cell  $C \in \mathcal{G}_{xy}$ , where

$$\tilde{m}_i(O_{\text{su}}) = \frac{m_i(O_{\text{su}})}{m_{\text{all}}(O_{\text{su}})}, \quad i \in \{\text{flat}, \text{spline}, \text{normals}, \text{ref}\} \quad (3.21)$$

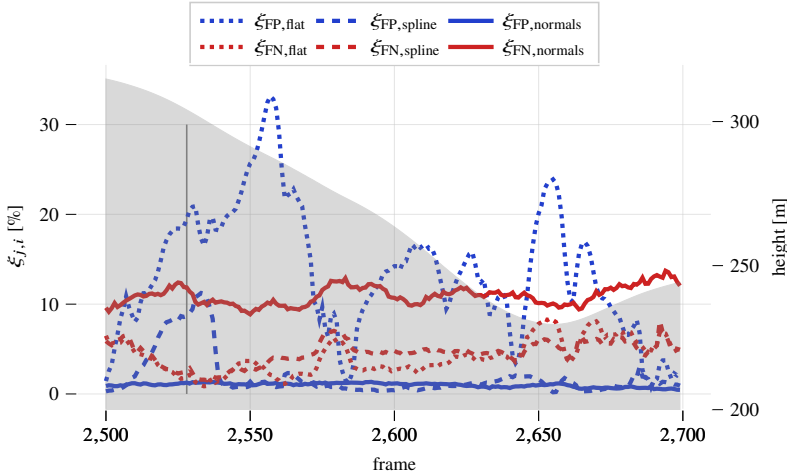


Figure 3.14: Confusion rates  $\xi_{j,i}$  for the occupancy BBAs  $m_{\text{flat}}$ ,  $m_{\text{spline}}$  and  $m_{\text{normals}}$ . The height of the road surface with respect to a global reference coordinate system is visualized in gray behind the plots for the six metrics.

is the part of  $m_{\text{all}}(O_{\text{su}})$  that was classified as occupied. The confusion rates on the whole grid  $\mathcal{G}_{xy}$  are then calculated as

$$\xi_{j,i} = \frac{\sum_{C \in \mathcal{G}_{xy}} \xi_{j,i}(C)}{\sum_{j \in J} \sum_{C \in \mathcal{G}_{xy}} \xi_{\text{TP},i}(C)}, \quad (3.22)$$

where  $j \in J = \{\text{TP}, \text{FP}, \text{FN}, \text{TN}\}$ . A high false positive rate  $\xi_{\text{FP},i}$  indicates that measurement elements  $m$  with attached semantic label  $\omega_{\text{ref}} \in \Omega_g$  contributed to a high BBA  $m(O_{\text{su}})$  whereas a high false negative rate  $\xi_{\text{FN},i}$  indicates that measurement elements with semantic label  $\omega_{\text{ref}} \in O_{\text{su}}$  had little contribution to  $m(O_{\text{su}})$ . Figure 3.14 shows the confusion metrics  $\xi_{j,i}$  for one traffic scene consisting of 200 consecutive frames in the SemanticKITTI dataset. Here, all grid cells with assigned ground label *other ground*, i.e. everything but street and sidewalk are excluded as occupancy derived from semantic properties might differ significantly from the geometric occupancy on terrain like meadows and other vegetation. Including those areas would distort the evaluation results. It can be seen that the false positive rate  $\xi_{\text{FP},\text{flat}}$  of the flat world model is heavily

influenced by uneven terrain violating the flat world assumption. In most of the frames, the false positive rate  $\xi_{FP,spline}$  is reduced to approximately zero for the B-spline model. However, between frame 2510 and 2540, a significant rise of  $\xi_{FP,spline}$  can be observed. The false positive rate  $\xi_{FP,normal}$  for the proposed surface normal vector-based method, on the other hand, stays almost constant close to zero in the whole test sequence. As in the proposed method, occupancy evidence is not deduced for horizontal surfaces on objects such as car roofs, the false negative rate  $\xi_{FN,normal}$  is the highest almost throughout the whole sequence. The false negative rate  $\xi_{FN,spline}$  indicates that the largest number of detections reflected on object surfaces were not missed in the B-spline model. However, it should be emphasized that this is due to the differences between the two underlying occupancy concepts in Definition 2.1 and Definition 2.2. The qualitative results presented in the remainder of this section show that missing detections in  $m_{normal}$  are in fact almost entirely located within objects and thus are negligible in top-view object shape estimation.

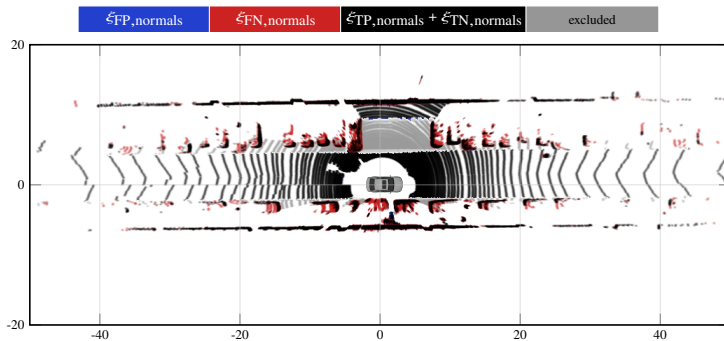


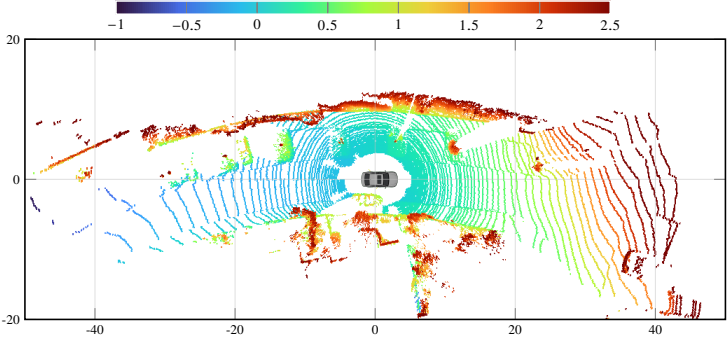
Figure 3.15: Visualization of the confusion rates  $\xi_{j,normal}$  for the occupancy BBAs  $m_{normal}$ .

Figure 3.15 visualizes the confusion metrics for one frame in the Kitti odometry benchmark based on the proposed mapping algorithm using surface normal vectors. In each grid cell, the false positive rate  $\xi_{FP,normal}$  is visualized in blue, the false negative rate  $\xi_{FN,normal}$  in red and the true positive rate  $\xi_{TP,normal}$  in black. Black grid cells visualize the sum of the two rates  $\xi_{TP,normal}$  and  $\xi_{TN,normal}$ , i.e. the estimated classification coincides with the reference. Grid cells without any detection are white and grid cells with detections that have been excluded from the evaluation are shown in gray. As mentioned, it can be

seen that red grid cells mainly occur on horizontal surfaces on objects such as the rooftop of cars whereas the object boundaries are estimated as occupied.



(a) Image taken by the front left color camera.



(b) The maximal detected height of all LiDAR detections located in a specific grid cell.

Figure 3.16: The frame that is indicated by the vertical line in Figure 3.14.

Figure 3.16 shows a specific frame in the same sequence as shown in Figure 3.14. The ego vehicle enters a street with a significant incline that leads to a crossing on rather flat terrain. The frame is marked by a gray vertical line in Figure 3.14. The camera image of the front left camera in Figure 3.16a shows that there is a significant change in the gradient of the ground surface in the vicinity of the ego vehicle. This is further highlighted in Figure 3.16 that shows the maximal detected height of all LiDAR detections located in a grid cell. Here, the LiDAR detections were transformed into the vehicle coordinate system.

In Figure 3.17, the advantages of the surface normal-based occupancy classification compared to a classification based on a ground model are demonstrated. Figure 3.17a depicts the mapping result when applying grid mapping with point sets using the flat world model. In the area around the junction there are many ground detections that are classified as obstacles contributing to a high occupancy mass in those grid cells. On challenging terrain as in this scenario

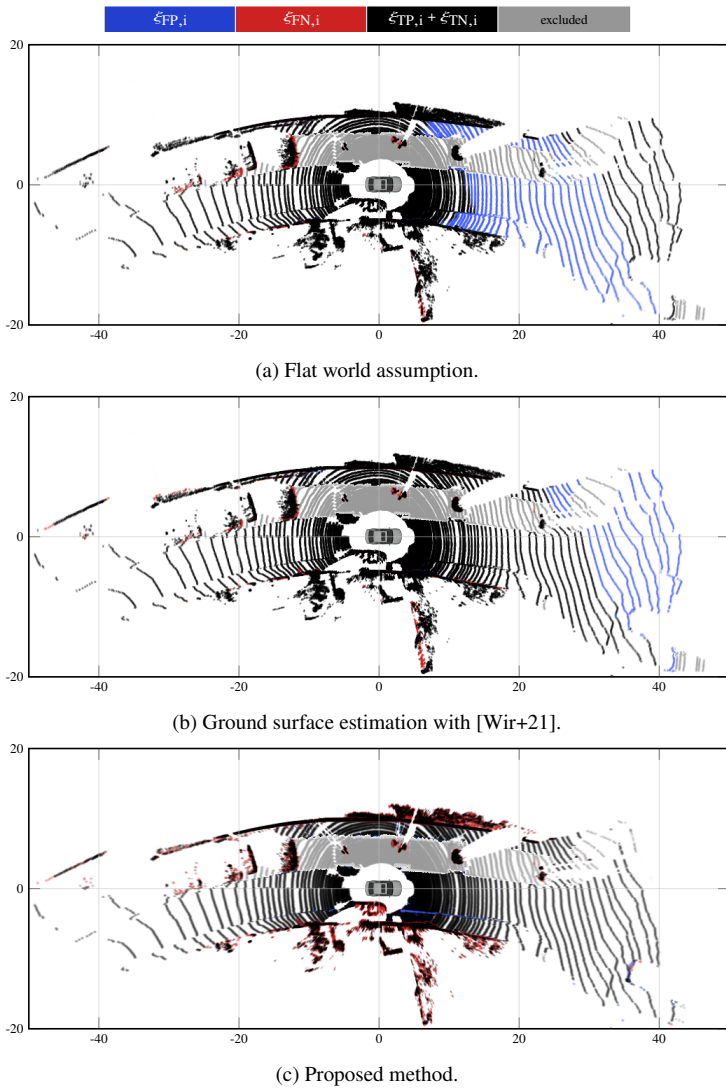


Figure 3.17: Comparison of the occupancy BBA estimation on challenging terrain.

this is expected as the ground model cannot capture the actual ground geometry. Figure 3.17b shows the results when applying grid mapping with point sets using the B-spline model. The fact that the uniform B-spline is able to capture the real surface significantly better leads to fewer grid cells falsely classified as occupied. However, there are still some grid cells on the ground that are classified as occupied. This can be resolved when applying the proposed surface normal vector model. The results are shown in Figure 3.17c where a high BBA for the hypothesis occupied is mostly obtained in areas where obstacles are assumed to be present. One exception are curb stones. As opposed to the other two models, the surface normals model classifies grid cells located at curb stones as *occupied*. They are visualized in blue in Figure 3.17c as curb stones are labeled as sidewalk in the SemanticKITTI labels and thus no occupancy evidence is deduced in the calculation of the reference BBA  $m_{\text{ref}}$ .

## 4 Sensor Data Fusion in Evidential Grid Maps

After estimating the sensor measurement grid map based on measurements in a defined time window, they are combined in the fused measurement grid map. Similarly as the sensor measurement grid map, the fused measurement grid map contains the BBAs on the occupancy semantics  $\Omega_s$  and the ground hypotheses  $\Omega_g$ . Combining estimates from all sensors in a joint model simplifies the further processing steps as they only have to be done once instead of applying the calculations for each sensor. Moreover, it may help to reduce measurement noise and missed detections by resolving measurement conflicts. Recall that a competitive sensor data fusion is desired in this work which means that information from more sensors may improve the accuracy, but is not required for estimating the output representation. This is achieved here as the input representation is the same as the output representation.

Let  $g_1, \dots, g_n$  be sensor measurement grid maps computed based on measurements from  $n$  independent sensor sources. Mathematically, all grid maps are combined in the sensor data fusion as

$$g_{1,\dots,n} = f(g_1, \dots, g_n), \quad (4.1)$$

where the fusion operator  $f$  is to be specified. The sensor data fusion is sketched in Figure 4.1. In the remainder of this chapter<sup>1</sup>, we restrict this to the fusion of two sensor measurement grid maps assuming that this can be generalized to an arbitrary number of sensor measurement grid maps e.g. by evaluating the fusion operator recursively. Recall that all sensor measurement grid maps

---

<sup>1</sup> A short version of this chapter has been submitted for publication on the *25th International Conference on Information Fusion (FUSION)* and has been made available to the public via arXiv [Ric+22b].

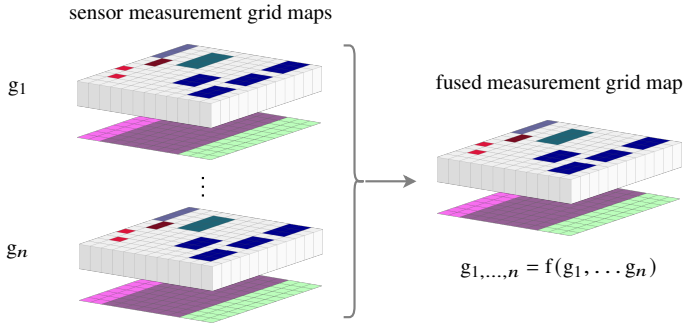


Figure 4.1: In the sensor data fusion, sensor measurement grid maps  $g_1, \dots, g_n$  from  $n$  sensors are combined in one fused measurement grid map  $g_{1,\dots,n}$ .

are defined on the same Cartesian grid  $\mathcal{G}_{xy}$ , so that the fusion can be applied cell-wise.

## 4.1 Fundamentals

Conditional probabilities  $\Pr(o|z_1)$  and  $\Pr(o|z_2)$  can be combined with the binary Bayes filter as

$$\Pr(o|z_1, z_2) = \frac{\Pr(o|z_1)\Pr(o|z_2)}{\Pr(o|z_1)\Pr(o|z_2) + (1 - \Pr(o|z_1))(1 - \Pr(o|z_2))}, \quad (4.2)$$

see [Bon08].

In the evidential context, Dempster's rule of combination was introduced in [Dem67]. Here, two BBAs  $m_1$  and  $m_2$  from independent sources are combined as

$$(m_1 \oplus m_2)(A) = \frac{1}{1 - K} \sum_{X \cap Y = A} m_1(X) m_2(Y). \quad (4.3)$$

The normalization constant  $\frac{1}{1-K}$  distributes the conflicts

$$K = \sum_{X \cap Y = \emptyset} m_1(X) m_2(Y) \quad (4.4)$$



equally to all focal elements. Dempster’s rule satisfies desired properties as commutativity and associativity. Furthermore, Nuss derived in [Nus17] the following relation to the binary Bayes filter: Let the FoD  $\Omega = \{A, B\}$  and two independent BBAs  $m_1$  and  $m_2$  be given where

$$m_1(\Omega) = m_2(\Omega) = 0, \quad m_i(X) > 0 \text{ for all } i \in \{1, 2\}, X \in \Omega. \quad (4.5)$$

Then combining the pignistic transformations (Equation (2.7)) of  $m_1$  and  $m_2$  with the binary Bayes filter leads to the same result as applying the pignistic transformation to  $m_1 \oplus m_2$ . Although Dempster’s rule is frequently used in literature, it has also been criticized. Zadeh showed in [Zad79] that combining highly conflicting BBAs with Dempster’s rule leads to counterintuitive results. This effect is also known as Zadeh’s paradox. Table 4.1 shows such an example. Although both  $m_1$  and  $m_2$  hold high evidences masses against hypothesis  $B$ , Dempster’s rule yields  $(m_1 \oplus m_2)(B) = 1$  ignoring the conflict mass  $K = 0.99$ .

	$A$	$B$	$C$	$\Omega$
$m_1$	0.9	0.1	0	0
$m_2$	0	0.1	0.9	0
$m_1 \oplus m_2$	0	1	0	0

Table 4.1: Two BBAs  $m_1$  and  $m_2$  on  $\Omega = \{A, B, C\}$  combined with Dempster’s rule (Equation (4.3)).

Different combination rules have been proposed aiming at resolving this counterintuitivity that all address different ways of dealing with conflicts. Yager [Yag87] defined the conjunctive rule of combination given as

$$(m_1 \odot m_2)(A) = \begin{cases} \sum_{X \cap Y = A} m_1(X) m_2(Y), & \text{if } A \neq \Omega \\ m_1(\Omega) m_2(\Omega) + \sum_{X \cap Y = \emptyset} m_1(X) m_2(Y), & \text{if } A = \Omega. \end{cases} \quad (4.6)$$

Note that it merely drops the normalization constant and assigns the conflict mass  $K$  to  $\Omega$  compared to Dempster’s rule in Equation (4.3) which is why it is also referred to as unnormalized Demster’s rule. Table 4.2 shows the results when applying the conjunctive rule to the example introduced in Table 4.1. Compared to Dempster’s rule the conjunctive rule assigns the conflict mass  $K$  to  $(m_1 \odot m_2)(\Omega)$  indicating a high degree of uncertainty. Although the conjunctive

	$A$	$B$	$C$	$\Omega$
$m_1$	0.9	0.1	0	0
$m_2$	0	0.1	0.9	0
$m_1 \odot m_2$	0	0.01	0	0.99

Table 4.2: Two BBAs  $m_1$  and  $m_2$  on  $\Omega = \{A, B, C\}$  combined with the conjunctive rule (Equation (4.6)).

rule gives a more intuitive result in this example it discards a significant amount of information by assigning the whole conflict mass to  $\Omega$ . Other examples for modified combination rules are Duboi's and Prade's rule presented in [DP88] and the partial conflict redistribution (PCR) rules introduced in [SD06]. All of them are based on Yager's rule (Equation (4.6)) and assign conflict masses in different ways.

Yang et al. [YX13] propose another approach to dealing with conflicts when combining BBAs. They discuss the mathematical properties of different evidential combination rules. All the above-mentioned modified rules lose some of the mathematical interpretability that holds for Dempster's original rule. Dempster's rule is the only one that is a probabilistic reasoning process meaning that the combination of two BBAs that can be represented as a probability distribution can again be represented as a probability distribution. Yang et al. therefore presented an alternative approach to evidential reasoning. Given the FoD  $\Omega$ , they consider sources of evidence  $\{e_i, i = 1, \dots, n\}$  with weight  $0 \leq w_i \leq 1$  and reliability  $0 \leq r_i \leq 1$  each providing a BBA  $m_i$ . The weight models the relative importance of the source of evidence, whereas the reliability models the information quality. The BBA  $m_i$  of the source of evidence  $e_i$  is modified based on the weight  $w_i$  and the reliability  $r_i$  as

$$\tilde{m}_i(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ \frac{1}{1+w_i-r_i} m_i(A), & \text{if } A \in \mathcal{P}(\Omega) \setminus \emptyset. \end{cases} \quad (4.7)$$

Two independent sources of evidence  $e_1$  and  $e_2$  with reliabilities  $0 \leq r_1, r_2 \leq 1$  and modified BBAs  $\tilde{m}_1$  and  $\tilde{m}_2$  are then combined as

$$m_{1,2}(A) = \begin{cases} 0, & \text{if } A = \emptyset, \\ \frac{\tilde{m}_{1,2}(A)}{\sum_{B \in \mathcal{P}(\Omega)} \tilde{m}_{1,2}(B)}, & \text{if } A \in \mathcal{P}(\Omega) \setminus \emptyset, \end{cases} \quad (4.8)$$

where

$$\begin{aligned} \tilde{m}_{1,2}(A) &= (1 - r_2) \tilde{m}_1(A) + (1 - r_1) \tilde{m}_2(A) \\ &+ \sum_{B \cap C = A} \tilde{m}_1(B) \tilde{m}_2(C). \end{aligned} \quad (4.9)$$

Note that for  $w_1 = w_2 = r_1 = r_2 = 1$ , Equation (4.8) reduces to Dempster's rule. In the remainder of this dissertation, Equation (4.8) will be referred to as evidential reasoning (ER) rule. Yang et al. further stated the following properties of the ER rule:

- The combination of  $n > 2$  sources of evidence can be evaluated recursively, see [YX13, Corollary 4].
- Similar as Dempster's rule, it forms a probabilistic reasoning process.

The two reliability parameters  $r_1, r_2$  influence the combination results significantly. Table 4.3 shows the results when combining the two BBAs from

	$A$	$B$	$C$	$\Omega$
$m_1$ with $r_1 = 0.7$	0.9	0.1	0	0
$m_2$ with $r_2 = 0.3$	0	0.1	0.9	0
$m_1 \odot m_2$	0.67	0.11	0.22	0

Table 4.3: Two BBAs  $m_1$  and  $m_2$  on  $\Omega = \{A, B, C\}$  combined with the ER rule (Equation (4.8)).

Tables 4.1 and 4.2 with the ER combination rule where the reliabilities were set to  $r_1 = 0.7$  and  $r_2 = 0.3$  and  $w_1 = w_2 = 1$ . Due to the higher reliability assigned to  $m_1$ , the conflict between  $A$  and  $C$  is mostly assigned to  $A$ .

In summary, evidential reasoning with the ER combination rule (Equation (4.8)) provides a mathematical framework for dealing with differently credible sources of evidence without losing the mathematical properties of Dempster's original combination rule.

## 4.2 Related Work

When combining BBAs from independent sensor sources in grid maps, Dempster's rule is usually applied [Nus+14; TW17]. However, because of the

above-mentioned shortcomings of this rule, other combination operators haven been explored as well: Moras et al. [MDP15] applied the PCR6 to the fusion of evidential occupancy grid maps. They tested their method with simulated LiDAR data that they fuse over time and showed an improved conflict resolution compared to Dempster’s rule. Li et al. [LLZ20] adapt the BBA obtained from LiDAR and stereo cameras by adding a similarity factor and apply Dempster’s rule to the adapted BBAs. Compared to applying Dempster’s rule to the original BBAs, they obtain a reduced entropy and a higher specificity in their fused BBA. Ullah et al. [UYH21] proposed a new uncertainty measure based on Deng’s entropy and combine their entropy measures using Dempster’s rule. By doing so, they could increase the accuracy of the fusion results compared to Dempster’s rule and apply to using Deng’s original entropy.

The critical cases when fusing heterogeneous sensor data in top-view grid maps occur if the BBAs obtained from the individual sensors are highly conflicting similar to the example demonstrated in Tables 4.1 and 4.2. The inclusion of information on the credibility of the individual sensors may help to resolve those conflicts correctly. This is not covered in all the above-mentioned publications.

### 4.3 Combining Evidential Grid Maps with Evidential Reasoning

We apply the ER combination rule presented by Yang et al. [YX13] to sensor measurement grid maps and model the reliability  $r_i$  of sensor sources to improve conflict resolution. The importance weights  $w_i$  are set to one modeling all sources to be equally important. Given two sensor measurement grid maps  $g_1$  and  $g_2$  calculated as described in Chapter 3 using measurements from two independent sensors  $s_1$  and  $s_2$ , the combination  $g_{1,2}$  is computed. The two sensors  $s_1$  and  $s_2$  are interpreted as sources of evidence with reliabilities  $0 \leq r_1, r_2 \leq 1$ . In a fixed grid cell  $C \in \mathcal{G}$ , the BBAs  $g_1(C, \cdot)$  and  $g_2(C, \cdot)$  are then combined to the BBA  $g_{1,2}(C, \cdot)$  using the ER rule.

### 4.3.1 Conflict Adaptive Evidential Reasoning

The reliability of a source of evidence should be chosen carefully. Let  $m_1$  and  $m_2$  be two BBAs to be combined. We propose to model the reliabilities  $r_i$  as a function of the conflict mass

$$K = \sum_{X \cap Y = \emptyset} m_1(X) m_2(Y) \quad (4.10)$$

and the credibility coefficient  $b_i$  as

$$r_i = f_r(K, b_i) = 1 - (1 - b_i) K. \quad (4.11)$$

The coefficient  $0 \leq b_i \leq 1$  models the credibility of a source of evidence in light of a conflict. If  $K = 0$ , then  $r_1 = r_2 = 1$  and the ER rule reduces to Dempster's rule. The higher the conflict mass  $K$ , the more unintuitive the combination result with Dempster's rule becomes as shown in Table 4.1 and the combination rule is adapted. If  $K = 1$ , we have  $r_i = b_i$  and the credibility coefficient fully serves as reliability value.

### 4.3.2 Parameter Estimation

When combining LiDAR and stereo camera sensor measurement grid maps with the ER rule, the sensor credibility coefficients  $b_l$  for the LiDAR and  $b_s$  for the stereo camera need to be specified. In this work, a data-driven approach assigning the values  $b_l, b_s$  resulting in the best fusion performance in terms of a quantitative evaluation is presented. The performance is quantified by the eIoU (Equation (2.26)) for the hypotheses *occupied*  $O_{su} \subseteq \Omega_s$ . More specificity, we apply the ER rule to LiDAR and stereo camera sensor measurement grid maps based on measurements in the KITTI-360 [LXG21] training sequence (see Table 2.5) for credibility values

$$\{(b_l, b_s) \mid b_l, b_s \in 0.1 \cdot \mathbb{N}_{\leq 10}\}.$$

In this context, all the occupancy evidence is assigned to the superset  $O_{su}$  as the individual semantic hypotheses are not considered in the analysis. The results are shown in the table in Figure 4.2. Each entry contains the eIoU averaged over all cell states in every tenth frame of the training sequence. Therefore,

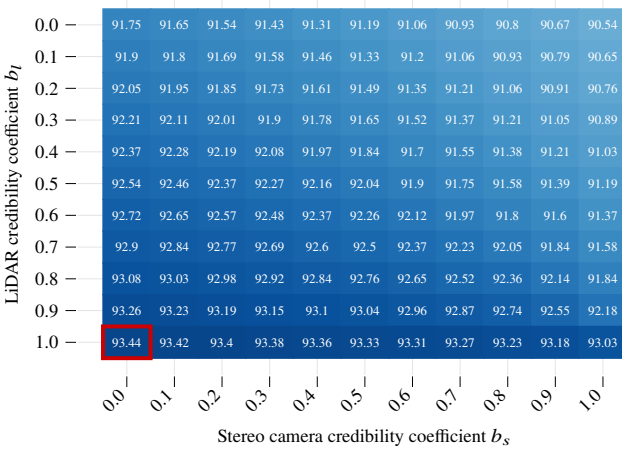


Figure 4.2: The eIoUs of the hypotheses *occupied* for different combinations of credibility coefficients ( $b_l$ ,  $b_s$ ). The highest eIoU is achieved for  $b_l = 1$  and  $b_s = 0$ .

a total of 50 frames were used for each configuration. It can be seen that the eIoU increases with increasing LiDAR credibility  $b_l$ , independent of the stereo camera credibility. Furthermore, the eIoUs increases for decreasing stereo camera credibilities. The highest eIoU is measured for the LiDAR credibility  $b_l = 1$  and the stereo camera credibility  $b_s = 0$ . For this combination of credibility values, the eIoU is 0.41 percentage points higher compared to applying Dempster’s rule and 2.9 percentage points higher than with  $b_l = b_s = 0$ . This difference is quite significant considering that the sensor estimates are not conflicting in the majority of the grid cells and both rules coincide in those cases. Recall that  $b_s = 0$  does not mean that the BBA estimated with the stereo camera is not regarded at all in the combination. It merely means that in cases where the stereo camera provides a measurement that disagrees with the LiDAR measurement, the LiDAR measurement shall be considered more reliable. The eIoU-based analysis shows how close the fusion result is to the reference grid map  $g_{ref}$  based on the semantic labels and bounding box primitives in the KITTI-360 dataset. As those labels were annotated using the LiDAR measurements this result does not really come as a surprise. However, this demonstrates that the sensor fusion results can be tuned as desired by adapting the credibility coefficients in the ER rule. As we consider the

annotations in the KITTI-360 dataset to be accurate in this work, we therefore set  $b_l = 1$  and  $b_s = 0$  for the remainder of this dissertation. This is sensible as in fact, LiDAR scanners provide more accurate depth estimates compared to the disparity-based depth estimation with stereo cameras.

## 4.4 Experiments

We evaluate the fusion results qualitatively and quantitatively based on Velodyne HDL-64E LiDAR and stereo camera measurements in the KITTI-360 evaluation sequences (see Table 2.5).

### 4.4.1 Qualitative Results

We show qualitative results for the fusion of sensor measurement grid maps from Velodyne HDL-64E LiDAR scans without semantic estimates with sensor measurement grid maps from disparity maps from stereo images with semantic estimates.

Figure 4.3 shows the grid maps for one frame. In the sensor grid map based on LiDAR measurements that is depicted in Figure 4.3a, BBA estimates are only available for the hypotheses *free* and *occupied by unknown object type*. The grid map based on stereo camera measurements depicted in Figure 4.3b contains BBA estimates for the individual semantic hypotheses based on the semantic labeling provided by the neural network. It can be seen that the spatial uncertainty is higher in the stereo camera grid map indicated by a more blurry occupancy pattern. The result of combining the two sensor grid maps with the ER rule is shown in Figure 4.3c. The semantic estimates provided by the stereo camera is successfully included in the occupancy pattern obtained from the LiDAR scanner.

In order to demonstrate the effect of using different evidential combination rules in case of conflicting sensor BBAs, the same sensor grid maps are combined using Dempster’s rule (Equation (4.3)), Yager’s rule (Equation (4.6)) and the ER rule (Equation (4.8)) in Figure 4.4. In this example, the distance of an observed car was underestimated by the stereo pipeline leading to an occupied-free conflict in front of the car. With Dempster’s rule, conflicting BBAs are

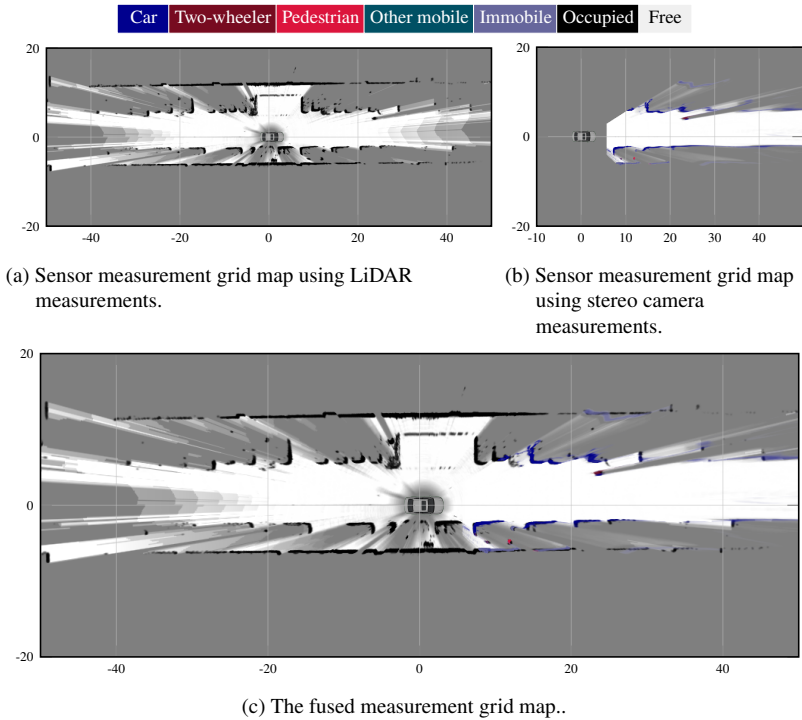


Figure 4.3: Resulting evidential grid map for the fusion of LiDAR measurements without and stereo camera measurements with semantic estimates using the ER rule.

distributed equally over all focal elements. Hence, large parts of the conflict mass are assigned to the hypothesis car. Yager’s rule assigns all conflict masses to  $\Omega$ , thus leading to a low BBA for both hypotheses free and car. The ER rule on the other hand is able to correctly assign large parts of the conflict masses to the hypothesis free due to the lower credibility coefficient  $r_s$  assigned to the stereo camera.



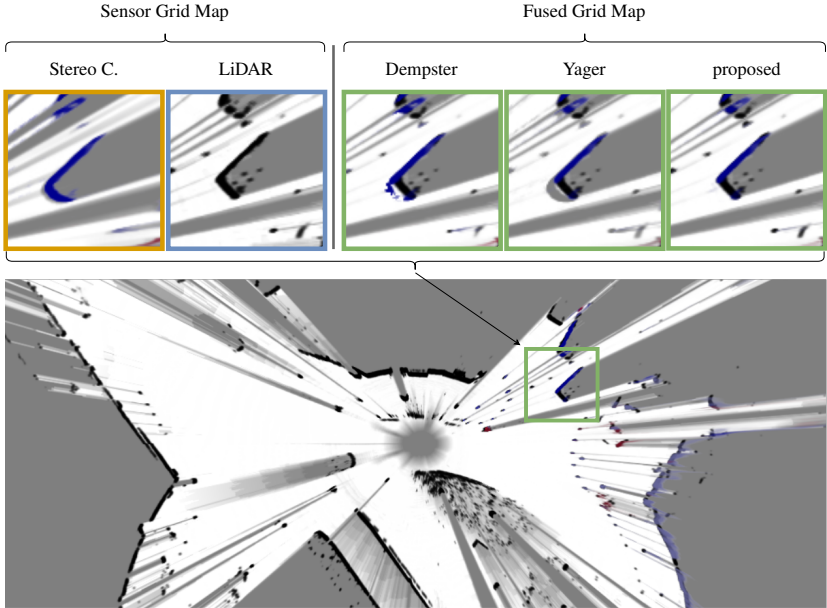


Figure 4.4: Results for fusing LiDAR grid maps with stereo camera grid maps using different combination rules. A subarea is highlighted where a vehicle is located. The measurement conflict between the BBAs provided by the two sensors is resolved differently. Only the proposed method resolves the conflict correctly in most of the grid cells.

#### 4.4.2 Quantitative Evaluation

We evaluate the accuracy of the BBA by calculating the eIoU for the fused grid maps and comparing it to the results obtained for the sensor measurement grid maps. The hypotheses *occupied by immobile object*  $O_{im}$  and *occupied by other mobile object*  $O_{om}$  are not considered here as the reference BBA for the former was found to be not credible, and the latter is barely observed in the evaluation dataset.

Figure 4.5 shows the results for the FoD  $\Omega_s$  in the described sensor setup. Because no semantic estimates are included in the LiDAR-based estimation chain, the eIoU values are zero for all  $\omega \subseteq O_{su}$ . As expected, it can be seen

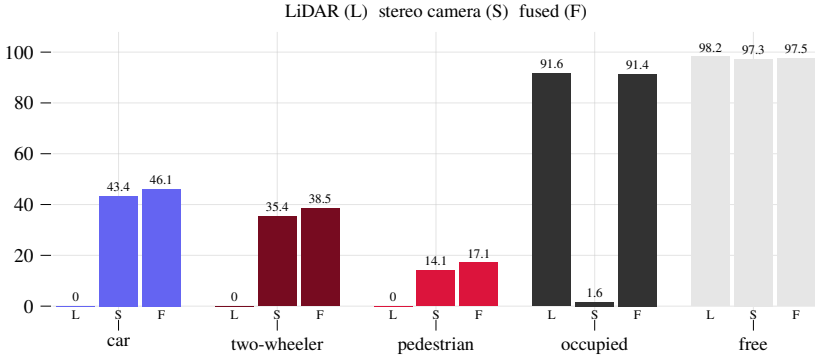


Figure 4.5: The eIoUs in % for five hypotheses in the FoD  $\Omega_s$ . For each hypothesis, the bar plots show the eIoUs measured in the grid map based on LiDAR measurements, stereo camera measurements and for the fused measurement grid map. The BBA from the stereo camera contains estimates for the semantic occupancy hypotheses whereas the LiDAR-based BBA only contains evidence for the hypotheses *occupied by unknown object type*  $O_{su}$  and *free*  $F_s$ .

that the LiDAR scanner leads to a more accurate BBA estimation than the stereo camera for the hypotheses estimated by both sensors. Furthermore, the metrics show that the accuracy for the singleton hypotheses *car*, *two-wheeler* and *pedestrian* can be significantly improved by fusing the occupancy semantics estimation obtained from the stereo camera with the occupancy information obtained from the LiDAR. Although no further semantic estimates are added in the fusion process the eIoU can be relatively improved by 6.2% for *cars*, by 8.8% for *two-wheelers* and by 21.3% for *pedestrians*. This is due to the improved conflict resolution when applying the ER combination rule as verified in Figure 4.4.

In addition to the comparison with reference BBA maps, the uncertainty incorporated in the estimated BBA is evaluated based on Deng's entropy measures defined in Equations (2.11) to (2.13). Figure 4.6 shows Deng's nonspecificity, discord and entropy averaged over all frames in the evaluation sequences for LiDAR, stereo camera and fused measurements. For each grid map, either all grid cells within a distance of 30 m to the ego vehicle (360° view) or the grid cells that are additionally within the viewing area of the stereo camera (camera view), respectively, are taken into account. In the 360°

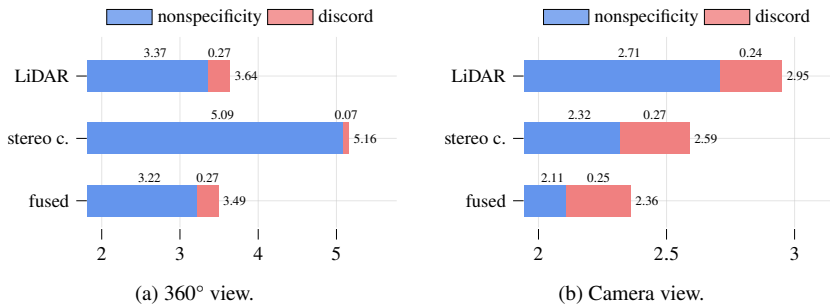


Figure 4.6: Nonspecificity and discord for LiDAR, stereo camera and fused grid maps.

view shown in Figure 4.6a, we observe that the average entropy is highest in the stereo camera grid maps. This is expected as the stereo camera does not provide measurements outside the stereo camera view. After applying the proposed sensor data fusion, the entropy is reduced by 4.1% compared to the LiDAR grid map. The competitive part of the sensor data fusion is evaluated in the overlapping viewing areas of the sensors. The corresponding entropy measures are plotted in Figure 4.6b. Here, the entropy in the LiDAR grid map is higher than the entropy in the stereo camera. Again the best result is obtained after applying the proposed sensor data fusion. Due to a significant reduction of the nonspecificity, the entropy in the fused grid map is reduced by 8.9% compared to the stereo camera grid map. This demonstrates that the introduced fusion operator successfully aggregates information from different sources while keeping the discord at a constant level.



## 5 Temporal Fusion in Evidential Grid Maps

When perceiving the environment for automated vehicles, it is desired to not only depend on the sensor measurements acquired at a moment but to incorporate all past measurements into the estimation. By doing so, the estimated model is more complete and the system may be able to detect and resolve erroneous measurements. Furthermore, the motion state of other traffic participants can be estimated by setting consecutive measurements into context.

However, this task is non-trivial. As the ego vehicle moves relatively to the static part of the environment, the ego motion needs to be compensated. Next, all dynamic parts whose motion state is to be estimated using temporally consecutive measurements move independently themselves. Finally, the dynamic parts need to be separated from the static parts of the environment as not all sensor modalities offer inherent information about the dynamics of a measurement.

In this chapter, a temporal fusion framework in evidential grid maps is proposed. After presenting a grid mapping framework for heterogeneous sensor data in Chapter 3 estimating the BBA on the ground semantics (Equation (2.20)) and occupancy semantics (Equation (2.21)), and the subsequent sensor data fusion in Chapter 4, it is the third and last processing block presented in this thesis. Besides accumulating estimates on the ground and occupancy semantics, we infer the dynamics in the environment and estimate the BBA on the occupancy dynamics (Equation (2.23)) as well.

Formally, temporal fusion in grid maps is the estimation of a state in the grid map  $\mathbf{g}^{(t)}$  at time  $t$  by combining two or more grid maps estimated at different time points. This can be done by either combining a batch of grid map measurements  $\mathbf{g}_Z^{(t_0)}, \dots, \mathbf{g}_Z^{(t_k)}$  in a potentially noncausal fashion

$$\mathbf{g}^{(t)} = f_b \left( \mathbf{g}_Z^{(t_0)}, \dots, \mathbf{g}_Z^{(t_k)} \right), \quad (5.1)$$

or by recursively updating based on the current measurement grid map  $\mathbf{g}_Z^{(t)}$

$$\mathbf{g}^{(t)} = \mathbf{f}_r \left( \mathbf{g}^{(t-1)}, \mathbf{g}_Z^{(t)} \right). \quad (5.2)$$

As it has advantages in computational efficiency and lower memory consumption, this work focuses on the recursive estimation.

After giving the fundamentals needed for the proposed methodology and a summary of related work, our framework combining measurement grid maps recursively to estimate ground semantics, occupancy semantics and occupancy dynamics in a joint model is presented in Section 5.3. In contrast to competitive methods, our proposal contains a data-driven parameter estimation which leads to a significant performance boost. This is demonstrated by presenting qualitative results and a detailed quantitative evaluation in Section 5.4.

## 5.1 Fundamentals

The proposed temporal grid mapping pipeline is based on the concepts of random finite set statistics and evidential networks which are introduced in this section.

### 5.1.1 Random Finite Sets

A random finite set (RFS)  $X$ , formally introduced e.g. by Mahler [Mah14], is a set of real-valued random variables

$$X = \{X_1, \dots, X_n\}, \quad (5.3)$$

where the cardinality  $n$  is random and finite. The distribution of a RFS is given by the cardinality distribution  $\rho(n)$ ,  $n \in \mathbb{N}$  and a set of PDFs

$$\{\mathbf{f}_1(x_1), \mathbf{f}_2(x_1, x_2), \dots, \mathbf{f}_n(x_1, \dots, x_n), n \in \mathbb{N} \mid \rho(n) > 0\}, \quad (5.4)$$

each representing the PDF given a fixed cardinality  $n \in \mathbb{N}$ . The PDF  $f_X$  of a RFS  $X$  is defined as

$$f_X(X = \{x_1, \dots, x_n\}) = \begin{cases} \rho(0), & \text{if } X = \emptyset \\ n! \cdot \rho(n) \cdot f_n(x_1, \dots, x_n), & \text{else.} \end{cases} \quad (5.5)$$

In object tracking applications, multi-Bernoulli RFSs are frequently considered. A multi-Bernoulli RFS is defined by the parameter set  $\{(p_1, f_1), \dots, (p_N, f_N)\}$  where  $N$  is the maximal possible cardinality,  $p_i$  is the existence probability and  $f_i$  is the PDF over the state of the  $i$ -th entity. For  $1 \leq n \leq N$ , the cardinality distribution is then given as

$$\rho(n) = \prod_{i=1}^N (1 - p_i) \sum_{\pi \in \Pi_n} \prod_{j=1}^n \frac{p_{\pi(j)}}{1 - p_{\pi(j)}}, \quad (5.6)$$

where  $\Pi_n$  is the set containing all permutations  $\pi: \{1, \dots, n\} \rightarrow \{1, \dots, n\}$  of indices  $1, \dots, n$ . For  $n = 0$  this reduces to

$$\rho(0) = \prod_{i=1}^N (1 - p_i). \quad (5.7)$$

The probability hypothesis density (PHD) of a RFS is defined as the first statistical moment

$$f_{\text{PHD}}(x) = \mathbb{E} \left( \sum_{w \in X} \delta(x - w) \right) \quad (5.8)$$

where  $\delta(\cdot)$  is the Dirac delta measure defined by the property

$$\int_{-\infty}^{\infty} f(x) \delta(x) dx = f(0) \text{ for all real-valued continuous functions } f. \quad (5.9)$$

Integrating over a PHD yields the expected number of objects of an RFS

$$\hat{n} = \int_{x \in X} f_{\text{PHD}}(x) dx. \quad (5.10)$$

### 5.1.2 Evidential Networks

Shafer et al. [SSM87] were the first who generalized Bayesian networks [KN09] to the evidential context and proposed to propagate beliefs in directed networks. As opposed to specifying the dependencies in the network by joint belief functions as in the early publications of Shafer et al., Xu et al. [XS96] proposed to give the dependencies in the form of conditional belief functions:

**Definition 5.1.** An *evidential network with conditional belief functions (ENC)* is a directed acyclic graph  $(V, E)$  where each node  $X_i \in V$  represents a variable with domain  $\Omega_{X_i}$  and each edge  $(X_i, X_j) \in E, i \neq j$  represents a conditional dependency between the variables  $X_i$  and  $X_j$ . The dependency denoted by the edge  $(X_i, X_j)$  is defined by the conditional belief function  $\text{bel}(X_j|X_i)$ .

ENCs are closely related to Bayesian Networks. As opposed to their probabilistic counterpart, however, each edge in an ENC models the dependency between source and target node independently of the dependency to other adjacent nodes. The ENC depicted in Figure 5.1a for instance has two dependencies defined by the conditional belief functions  $\text{bel}(X_3|X_1)$  and  $\text{bel}(X_3|X_2)$ . A Bayesian Network represented by the same graph on the other hand would be defined by the conditional probability  $\text{Pr}(X_3|X_1, X_2)$ . In order to represent conditional

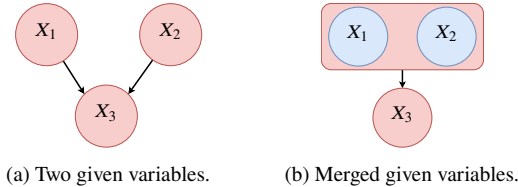


Figure 5.1: Example of an ENC with three variables.

beliefs  $\text{bel}(X|X_1, \dots, X_n)$  given multiple variables  $X_1, \dots, X_n$ , Xu proposed to merge the nodes of the given variables as shown in Figure 5.1b. The merging step is concluded by calculating the so-called *ballooning extension* for each child node and combining them by applying the conjunctive rule of combination, see [XS96] for details.



Yaghlane et al. [YSM03] proposed to extend the concept of ENCs to allow relations between any number of nodes. Let the set of parent nodes of a node  $X$  in a directed graph be denoted as  $P_X$ .

**Definition 5.2.** A directed evidential network with conditional belief functions (DEVN) is a directed acyclic graph  $(V, E)$  where each node  $X_i \in V$  represents a variable with domain  $\Omega_{X_i}$ . For each root node  $X_{r_i}$ , an a priori belief function  $\text{bel}(x), x \subseteq X_{r_i}$  is given and for the other nodes  $X_i \in V, \emptyset \neq P_{X_i} \subset V$ , dependencies are modeled by the conditional belief function  $\text{bel}(x | P_{X_i})$ .

In [YM08], Yaghlane et al. demonstrate in detail how beliefs of latent variables in DEVNs are inferred. The conditional belief functions modeling the dependencies between states are either determined based on expert knowledge or are learned from data. Ben Hariz et al. present in [BB14] a method to learn the conditional beliefs in DEVNs. They generalize the maximum likelihood estimation to the evidential context and learn the parameters based on evidential databases.

## 5.2 Related Work

Related work on temporal grid map fusion can be separated into methods assuming a static world and methods explicitly modeling the dynamics of moving entities in the environment.

In static grid mapping all entities occupying grid cells are assumed to be stationary. Hence, the prediction of the grid map  $g^{(t-1)}$  from the last time point is limited to the compensation of the ego motion. The evidential or probability masses either stay constant or are discounted based on a time-dependent fading factor to model possible state switches.

Dietmayer et al. [DRN14] applied the binary Bayes filter (Equation (4.2)) to the hypothesis *occupied* and *free* in the grid map update step. The grid map is updated by a probabilistic measurement occupancy grid map that can be calculated using range measurements from LiDAR and RaDAR.

Rummelhard et al. [RNL15] presented the Conditional Monte Carlo Dense Occupancy Tracker (CMCDOT) where they estimate the static part in an

occupancy grid map and model the dynamic part by moving particles. The separation into stationary and moving areas is done by introducing stationary and moving occupancy states in the filtering process. Instead of using an evidential framework, they use a Bayesian network to model state dependencies. They model the transition between stationary and moving occupancy by assigning a fixed percentage of one state to the other in the prediction process.

Nuss et al. [Nus+16] proposed to model the grid cell states as a RFS. Based on this formulation, they motivate a probabilistic filter and deduce a real-time capable approximation based on evidence theory. Instead of separating occupancy into static and dynamic parts, they model all occupancy masses by a particle filter. By fusing LiDAR and RaDAR measurements they achieve accurate cell velocity estimates. The modeling of static occupancy by particles, however, leads to a fast decay of occupancy masses in unobserved cells.

Tanzmeister et al. [TW17] were the next who adapted the idea to base a dynamic grid mapping framework on a particle filter. They model their evidential grid map with a FoD containing the elementary hypotheses *free*, *occupied by a stationary entity* and *occupied by a moving entity*. This prevents the fast decay of unobserved static occupancy as it is modeled explicitly and not solely deduced from the particle weights. In collaboration with the authors, Steyer et al. [STW18] presented an advancement of their approach. They got rid of static particles and only simulate particles in dynamic cells which significantly improves the computational efficiency. By doing so, they could further reduce the rate of falsely classified grid cells as dynamic in partially occluded static areas.

Vatavu et al. [Vat+20] focused on improving the particle-based velocity estimation. Instead of managing one set of particles for the whole grid map, they spawn several independent particle filters called tracklets. Particle weights are updated based on a two-layer measurement grid where the first layer contains the BBA on a classical occupancy frame and the second holds the most likely semantic class. They further pass on the cell independence assumption and update particle states based on their distance to tracklet landmarks on object boundaries. Especially in grid cells within large objects, they achieve improved velocity estimations and could reduce both false positive and false negative rates.

All the above-mentioned publications have two disadvantages:

1. Semantic estimates are either not handled at all or are not contained in the evidential context but treated separately ([Vat+20]). Incorporating this information in a joint framework has the potential to stabilize the dynamic state estimation.
2. Only the occupancy information projected to the top-view is used. Incorporating 3D information in the 2D grid representation might help to better resolve temporal conflicts while keeping the computational complexity at a minimum.

In consequence, they are not able to handle occupancy provided by very low obstacles as curb stones as the cells might easily be estimated as being free in the measurement grid and newly observed free space tends to wipe out historically observed occupancy. The method proposed in this work tackles those shortcomings.

In recent years, researchers started to apply learning of optical flow to top-view grid maps. Wirges et al. presented a self-supervised approach using a fully convolutional neural network in [Wir+19a]. By applying motion and spatial consistency regularization, they are able to both estimate the odometry and per-cell object velocities. Lee et al. combined in [Lee+20] a Pillar Feature Network with a flow estimation network to estimate the dynamics in LiDAR-based top-view grid maps. Input to their network are two consecutive LiDAR point clouds. They showed an improved performance both in computational performance and quality when feeding their velocity estimates into an object tracking module.

As apposed to the presented recursive estimators, however, those networks only estimate the flow of currently observed detections. Hence, no filtering is applied, and no uncertainty is given in the output representation. Furthermore, no semantic estimates are considered in the estimation process.

### 5.3 Semantic Evidential Grid Mapping and Tracking

We propose a recursive temporal grid map fusion that estimates the current grid map state  $g^{(t)}$  based on the measurement grid map  $g_Z^{(t)}$  and the grid map  $g^{(t-1)}$

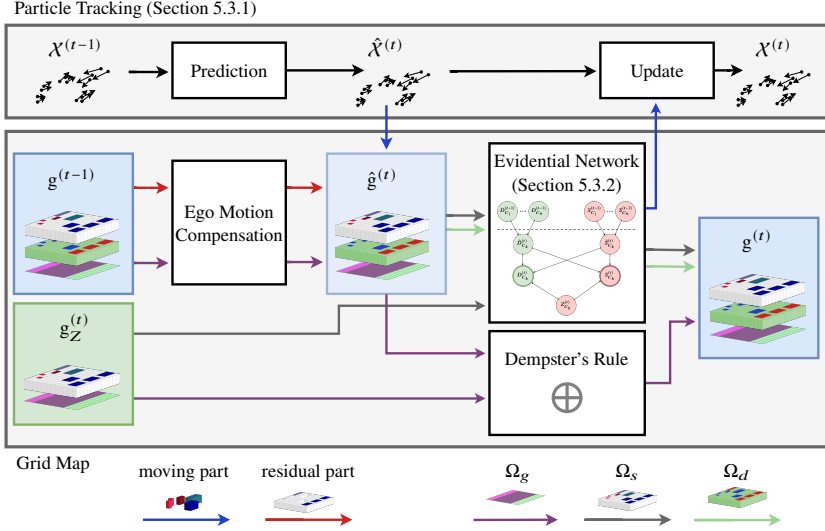


Figure 5.2: The temporal grid map fusion proposed in this chapter: Input to the recursive estimator is the internal grid map state  $g^{(t-1)}$  from the last time point and the current measurement grid map  $g_Z^{(t)}$ . Moving occupancy is modeled by the RFS  $X_{O_{\text{mov}}}^{(t)}$  whose PHD is estimated by the particle population  $\chi^{(t)}$ , see Section 5.3.1. The BBA on the ground semantics  $\Omega_g$  is updated with the measurement grid map by applying Dempster's rule whereas the occupancy semantics  $\Omega_s$  and the occupancy dynamics  $\Omega_d$  are updated in the evidential network presented in Section 5.3.2.

from the last update time point  $t - 1$ . Here, the grid map state  $g^{(t)}$  represents the BBA on the ground semantics  $\Omega_g$ , the occupancy semantics  $\Omega_s$  and the occupancy dynamics  $\Omega_d$  introduced in Section 2.3. The measurement grid map  $g_Z^{(t)}$  represents the BBA on the ground semantics  $\Omega_g$  and the occupancy semantics  $\Omega_s$  and is estimated using the sensor measurement grid mapping presented in Chapter 3 and the sensor data fusion from Chapter 4. The temporal fusion framework is sketched in Figure 5.2. It can be separated into a grid map prediction step estimating the predicted grid map  $\hat{g}^{(t)}$  by propagating the last grid map state to the current measurement time point  $t$  and an update step combining the predicted grid map with the measurement grid map  $g_Z^{(t)}$ . In the following, we also use the notation  $m_C^{(t)} = g^{(t)}(C, \cdot)$  referring to the BBA in the grid cell  $C \in \mathcal{G}_{xy}$  at time  $t$ .

**Grid map prediction.** The BBA for the ground semantics  $\omega \subseteq \Omega_g$  remains unchanged in the prediction process. Hence, for predicting the grid map, only the ego motion is compensated. Recall that the grid is translated by whole grid cells  $\Delta C^{(t)}$  between two update time points so that the grid cell  $C = C' + \Delta C^{(t)}$  is the grid cell at time  $t$  representing the same area in space as  $C'$  at time  $t - 1$ . Then we have

$$\hat{m}_C^{(t)}(\omega) = g^{(t-1)}(C', \omega), \quad \omega \subseteq \Omega_g, \quad (5.11)$$

i.e. the BBA remains unchanged.

Because the prediction of moving occupancy is handled differently than the prediction of stationary occupancy, the prediction is modeled jointly on the occupancy dynamics  $\Omega_d$  and occupancy semantics  $\Omega_s$ . Let

$$\hat{m}_C^{(t)}(\omega, \theta): \mathcal{P}(\Omega_s) \times \mathcal{P}(\Omega_d) \rightarrow [0, 1] \quad (5.12)$$

be the predicted BBA for occupancy dynamics and occupancy semantics. By estimating this predicted BBA, the occupancy semantics  $\omega \subseteq O_{su}$  can be predicted separately for moving and stationary parts.

The BBA for stationary occupancy  $O_{stat}$  and dynamically unclassified occupancy  $O_{du}$  remains unchanged during prediction. We compensate for ego motion and set for occupancy semantics  $\omega \subseteq O_{su}$

$$\hat{m}_C^{(t)}(\omega, \theta) = k m_{C'}^{(t-1)}(\omega), \quad \theta \in \{O_{stat}, O_{du}\}, \quad (5.13)$$

where the additional factor

$$k = \frac{m_{C'}^{(t-1)}(\theta)}{\text{bel}_{C'}^{(t-1)}(O_{du})} \quad (5.14)$$

ensures that only the part of the occupancy belief  $\text{bel}_{C'}^{(t-1)}(O_{du})$  that was assigned to the BBA of the considered occupancy dynamics, i.e.  $m_{C'}^{(t-1)}(\theta)$ , is assigned here. Following a conservative design, no free space BBA is propagated in the prediction step. Instead, the BBA of the hypotheses that a

grid cell is either free  $F_d$  or occupied by a moving entity  $O_{\text{mov}}$ , represented by the hypotheses passable  $P$  is predicted. It is calculated as

$$\hat{m}_C^{(t)}(\Omega_s, P) = \left( 1 - \sum_{\omega \subseteq O_{\text{du}}} \hat{m}_C^{(t)}(\omega, O_{\text{mov}}) \right) m_P, \quad (5.15)$$

with

$$m_P = \left( g^{(t-1)}(C', P) + g^{(t-1)}(C', F_d) + g^{(t-1)}(C', O_{\text{mov}}) \right), \quad (5.16)$$

i.e. the part that was not predicted as moving  $O_{\text{mov}}$  that was estimated passable  $P$ , free  $F_d$  or moving  $O_{\text{mov}}$  at the last update time point.

For moving occupancy  $O_{\text{mov}} \subset \Omega_d$ , the BBAs must be propagated according to the motion of the occupying objects. This is not possible with the information contained in the hybrid evidential grid map representation. Therefore, we follow past publications [Nus+16; STW18] and link a particle filter to the grid map representation. This is explained in detail in Section 5.3.1.

**Grid map update.** For the ground semantics  $\Omega_g$ , the grid map is updated by applying Dempster's rule (Equation (4.3)) to the predicted grid map  $\hat{g}^{(t)}$  and the measurement grid map  $g_Z^{(t)}$ .

For the occupancy semantics  $\Omega_s$  and the occupancy dynamics  $\Omega_d$ , regular evidential combination rule are not applicable as the measurement is only available for  $\Omega_s$ . Hence, a novel updating method using evidential networks is presented in Section 5.3.2 that explicitly regards dependencies between the occupancy semantics and occupancy dynamics FoDs. Past publications either apply Dempster's rule in a simpler occupancy model [Nus+16] or assign conflicts manually based on hand selected parameters [STW18]. As opposed to that, a data-driven estimation of the parameters introduced by the evidential network is proposed in Section 5.3.3.

### 5.3.1 Particle Filter

The particle filter used in this dissertation is a low-level cell velocity estimator that is utilized to propagate moving cell occupancy. In this work, we adopt the probability hypothesis density / multi-instance Bernoulli (PHD/MIB) filter and its evidential approximation derived in detail in [Nus+16]. In their work, Nuss et. al estimated the PHD of a RFSs representing all occupied grid cells. Here, we adopt this filter to estimate the PHD of a RFS representing moving occupancy similar as the particle representation presented by Steyer et al. [STW18]. This way, fewer particles are required as they are only spawn in a subset of occupied grid cells. Furthermore, the separation of occupancy into moving and stationary parts based on evidential reasoning can make the estimator more robust against falsely converging particles and thus falsely detected cell dynamics. The occupancy semantics  $\Omega_s$  introduced in this work further support this separation.

Mathematically, we model the grid map at time  $t$  by the multi-Bernoulli RFSs  $X_{O_{\text{mov}}}^{(t)}$  defined by the parameter set

$$\left\{ \left( p_X^{(C,t)}, f_X^{(C,t)} \right) \mid C \in \mathcal{G}_{xy} \right\}, \quad (5.17)$$

where  $p_X^{(C,t)}$  is the existence probability of the Bernoulli RFS in grid cell  $C$  and  $f_X^{(C,t)}$  its PDF. This way, grid cells occupied by objects that start moving can be modeled by increasing the existence probability  $p_X^{(C,t)}$  and objects that stop moving by decreasing the existence probability. The state of one Bernoulli RFS is given by

$$x_{O_{\text{mov}}} = (\mathbf{p}, \mathbf{v}, l), \quad (5.18)$$

where  $\mathbf{p} \in \mathbb{R}^2$  is the position in the reference coordinate system,  $\mathbf{v} \in \mathbb{R}^2$  is the velocity vector and  $l \subseteq O_{\text{su}}$  is the semantic label. The Bernoulli RFS in the grid cell  $C \in \mathcal{G}_{xy}$  is linked to the grid map representation via its existence probability  $p_X^{(C,t)}$  as

$$g^{(t)}(C, O_{\text{mov}}) = p_X^{(C,t)}, \quad (5.19)$$

i.e. the BBA of moving occupancy  $O_{\text{mov}}$  represents the existence probability of the RFS in the corresponding grid cell. The semantic label  $l$  in the RFS state further links the RFS to the BBA on the occupancy semantics  $\omega \subseteq \Omega_s$  as

$$m_C(\omega, \phi = O_{\text{mov}}) = p_X^{(C,t)} \Pr_{O_{\text{mov}}}^{(C,t)}(\omega), \quad (5.20)$$

where  $\Pr_{O_{\text{mov}}}^{(C,t)}(\omega)$  is the probability that the Bernoulli RFS instance  $X_{O_{\text{mov}}}^{(C,t)}$  has the semantic label  $\omega$ . The task is to jointly estimate the multi-object state of the multi-Bernoulli RFS  $X_{O_{\text{mov}}}^{(t)}$  and the BBAs on the FoDs  $\Omega_s$  and  $\Omega_d$ .

The PHD of  $X_{O_{\text{mov}}}^{(t)}$  is estimated with a particle filter based on the PHD/MIB filter presented in [Nus+16]. Let a particle

$$\chi = (\mathbf{p}_\chi, \mathbf{v}_\chi, l_\chi, w_\chi) \in \mathcal{X} \quad (5.21)$$

consist of the particle weight  $w_\chi \in [0, 1]$  and the RFS state, i.e. the position  $\mathbf{p}_\chi \in \mathbb{R}^2$ , the velocity vector  $\mathbf{v}_\chi \in \mathbb{R}^2$  and the semantic label  $l_\chi \subseteq O_{\text{su}}$ . The set of particles located in the grid cell  $C \in \mathcal{G}_{xy}$ , i.e.  $\mathbf{p}_\chi \in C$ , is denoted as  $\mathcal{X}_C$ . Figure 5.3 sketches the linkage between the evidential grid map representation and the particle population. The existence probability of a RFS in a grid cell is approximated by the sum of particle weights in that cell. Following Equation (5.19), the BBA of moving occupancy  $O_{\text{mov}}$  in grid cell  $C \in \mathcal{G}_{xy}$  is thus approximated as

$$m_C(O_{\text{mov}}) = \sum_{\chi \in \mathcal{X}_C} w_\chi. \quad (5.22)$$

The BBA of the occupancy semantic hypothesis  $\omega \subseteq O_{\text{su}}$  and the moving occupancy hypothesis  $O_{\text{mov}}$  is computed by approximating Equation (5.20) as

$$\begin{aligned} m_C(\omega, O_{\text{mov}}) &= \text{bel}_C(\omega | O_{\text{mov}}) m_C(O_{\text{mov}}), \\ &= \sum_{\chi \in \mathcal{X}_C: l_\chi = \omega} w_\chi, \end{aligned} \quad (5.23)$$

i.e. the accumulated particle weights in the grid cell with matching semantic label.

The filter follows the typical particle filter scheme consisting of the three steps particle prediction, particle weight update and particle resampling. In the



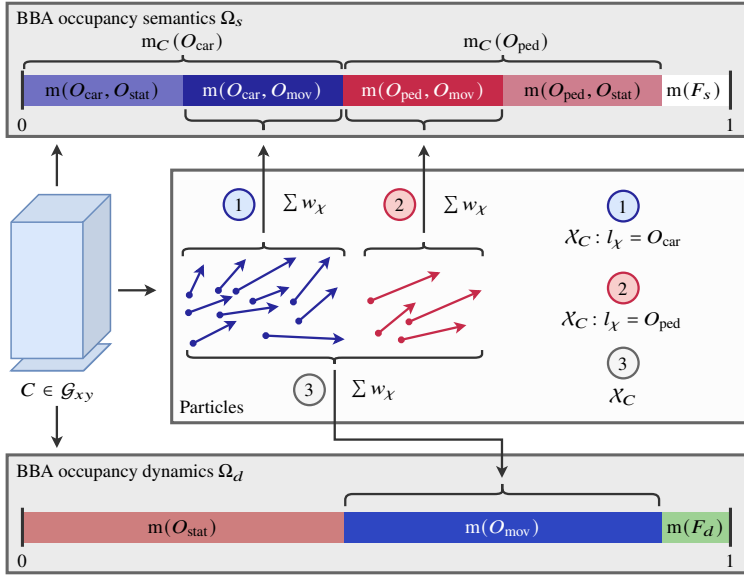


Figure 5.3: The linkage between the evidential grid map representation and the particle set  $\mathcal{X}_C$ : The sketch shows a grid cell  $C \in \mathcal{G}_{xy}$  where the BBA on the occupancy semantics is visualized on the top and the BBA on the occupancy dynamics is shown on the bottom. The weights of all particles  $\chi \in \mathcal{X}_C$  accumulate to the BBA of moving occupancy  $m_C(O_{mov})$  and the weights of particles with a specific semantic label  $\omega \in \Omega_s$  accumulate to  $m_C(\omega, O_{mov})$ , here visualized for the hypotheses  $\omega = O_{car}$  and pedestrian  $\omega = O_{ped}$ .

following, the calculations steps of the particle filter and the interaction steps with the grid-based BBA representation are explained.

## Particle prediction

Each particle state is predicted to the current time  $t$  according to a constant velocity motion model while adding 2D Gaussian noise  $\sigma_p \sim \mathcal{N}(\mu_p, \Sigma_p)$  to the position and  $\sigma_v \sim \mathcal{N}(\mu_v, \Sigma_v)$  to the velocity components as

$$\hat{\mathbf{p}}_{\chi}^{(t)} = \mathbf{p}_{\chi}^{(t-1)} + \Delta t \mathbf{v}_{\chi}^{(t-1)} + \sigma_p, \quad (5.24)$$

$$\hat{\mathbf{v}}_{\chi}^{(t)} = \mathbf{v}_{\chi}^{(t-1)} + \sigma_v. \quad (5.25)$$

The particle weight is predicted as

$$\hat{w}_{\chi}^{(t)} = p_{\text{pers}} w_{\chi}^{(t-1)}, \quad (5.26)$$

where the persistence probability  $p_{\text{pers}}$  models the likelihood that a Bernoulli RFS stays active until the next update time point. The semantic label  $\chi_l$  stays constant during prediction.

## Link predicted particles to grid map

The predicted BBA for the moving part of the occupancy semantics is calculated based on the predicted particle population  $\hat{\mathcal{X}}^{(t)}$  as

$$\hat{m}_C^{(t)}(\omega, O_{\text{mov}}) = \min \left( \sum_{\chi \in \hat{\mathcal{X}}_C^{(t)} : l_{\chi} = \omega} \hat{w}_{\chi}^{(t)}, p_{\text{pers}} \right), \quad (5.27)$$

where  $\hat{\mathcal{X}}_C^{(t)}$  is the set of particles predicted into grid cell  $C \in \mathcal{G}_{xy}$ . Note that the predicted BBA cannot exceed the persistence probability  $p_{\text{pers}}$ , because  $p_{\text{pers}}$  is an upper bound for the existence probability  $p_X^{(C,t)}$  of the Bernoulli RFS  $X_{O_{\text{mov}}}^{(C,t)}$ .

### Particle weight update

Analogously to Nuss et al. [Nus+16], the updated moving BBA is separated into persistent moving occupancy mass  $m_{C,\text{pers}}^{(t)}$  and new moving occupancy mass  $m_{C,\text{new}}^{(t)}$ :

$$m_C^{(t)}(O_{\text{mov}}) = m_{C,\text{pers}}^{(t)} + m_{C,\text{new}}^{(t)}. \quad (5.28)$$

Here, we apply [Nus+16, Equation (67)] to moving occupancy only, i.e.

$$\frac{m_{C,\text{new}}^{(t)}}{m_{C,\text{pers}}^{(t)}} = \frac{p_{\text{new}} \left(1 - \hat{m}_C^{(t)}(O_{\text{mov}})\right)}{\hat{m}_C^{(t)}(O_{\text{mov}})} \quad (5.29)$$

between the updated persistent and new moving occupancy masses. Combining Equations (5.28) and (5.29) results in the following two formulas for the updated persistent and new moving occupancy mass:

$$m_{C,\text{new}}^{(t)} = \frac{m_C^{(t)}(O_{\text{mov}}) \cdot p_{\text{new}} \cdot \left(1 - \hat{m}_C^{(t)}(O_{\text{mov}})\right)}{\hat{m}_C^{(t)}(O_{\text{mov}}) + p_{\text{new}} \cdot \left(1 - \hat{m}_C^{(t)}(O_{\text{mov}})\right)} \quad (5.30)$$

$$m_{C,\text{pers}}^{(t)} = m_C^{(t)}(O_{\text{mov}}) - m_{C,\text{new}}^{(t)}. \quad (5.31)$$

The particle weights are updated proportionally to the persistent moving occupancy mass as

$$w_\chi^{(t)} = \frac{m_{C,\text{pers}}^{(t)}}{|\mathcal{X}_C|}. \quad (5.32)$$

In the following, the set of persistent particles, i.e. updated particles predicted at least once, is denoted by  $\mathcal{X}_{\text{pers}}$ .

### Particle resampling

Two sets of new particles are initialized to prevent particle deprivation, each containing a fixed number of particles. A number of  $n_{\text{new}}$  particles in the first particle set  $\mathcal{X}_{\text{new}}$  are initialized in grid cells  $C$  with new moving occupancy

mass  $m_{C,\text{new}}^{(t)} > 0$ . More specifically, the number of particles drawn in grid cell  $C$  is set to

$$n_{C,\text{new}} = \left\lfloor \frac{n_{\text{new}}}{m_{\text{new}}^{(t)}} m_{C,\text{new}}^{(t)} \right\rfloor, \quad (5.33)$$

where

$$m_{\text{new}}^{(t)} = \sum_{C \in \mathcal{G}_{xy}} m_{C,\text{new}}^{(t)} \quad (5.34)$$

is the new moving occupancy mass accumulated over all grid cells  $C \in \mathcal{G}_{xy}$ . Consequently, the weight of a new particle is calculated as

$$w_{\chi,\text{new}} = \frac{m_{C,\text{new}}^{(t)}}{n_{C,\text{new}}}. \quad (5.35)$$

Up to this point, the update of persistent particles and the initialization of new particles is based on the assumption that  $m_C^{(t)}(O_{\text{mov}}) > 0$ . Hence, the existence of moving occupancy can only be determined in the BBA update step. In order to additionally derive moving occupancy based on converging particle states a second set of  $n_0$  new particles  $\mathcal{X}_0$  is initialized. The weight is calculated analogously to Equation (5.35) but proportionally to the newly gained BBA of dynamically unclassified occupancy  $\Delta^+ m_C^{(t)}(O_{\text{du}})$ . Note that the calculation of  $\Delta^+ m_C^{(t)}(O_{\text{du}})$  depends on the BBA update. It will be formalized in Equation (5.65).

For each new particle, the position  $\mathbf{p}_\chi$  is drawn from a uniform distribution defined on the current grid cell and the velocity  $\chi_v$  is drawn from a uniform distribution defined between a minimal and maximal velocity. The semantic label  $\chi_l$  is sampled from the measured BBA on  $\Omega_s$ . More specifically, label  $\omega$  is drawn with the probability

$$p_\omega = \frac{m_{C,Z}^{(t)}(\omega)}{\sum_{\theta \subseteq O_{\text{su}}} m_{C,Z}^{(t)}(\theta)}. \quad (5.36)$$

Finally,  $n$  particles are drawn out of the  $n + n_{\text{new}} + n_0$  updated and newly initialized particles according to their particle weights. For particles sampled from the sets  $\mathcal{X}_{\text{pers}}$  and  $\mathcal{X}_{\text{new}}$ , the new weight is set to

$$w_{\chi} = \frac{1}{n} \sum_{C \in \mathcal{G}_{xy}} m_C^{(t)}(O_{\text{mov}}). \quad (5.37)$$

For particles drawn from the set  $\mathcal{X}_0$ , the weight is set to zero. This way, the resampled particle set still approximates the PHD of the multi-Bernoulli RFS representing moving occupancy  $O_{\text{mov}}$  while at the same time zero-weight particles have been added for potentially moving occupancy.

### Statistical moments of the RFS

The mean velocity in grid cell  $C \in \mathcal{G}_{xy}$  can be calculated as

$$v_C = \sum_{\chi \in \mathcal{X}_C} w_{\chi} v_{\chi} \in \mathbb{R}^2, \quad (5.38)$$

i.e. the sum of the particles' velocity component weighted by the updated particle weights. Analogously, the velocity covariance matrix reads as

$$\Sigma_C = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_{yy} \end{pmatrix}, \quad (5.39)$$

where

$$\Sigma_{ab} = \frac{1}{W_C} \sum_{\chi \in \mathcal{X}_C} w_{\chi} (v_{a,\chi} - v_C) (v_{b,\chi} - v_C), \quad (5.40)$$

for  $a, b \in \{x, y\}$  and

$$W_C = \sum_{\chi \in \mathcal{X}_C} w_{\chi}. \quad (5.41)$$

### 5.3.2 Evidential Network Reasoning

As mentioned above, measurements are only available as BBAs on the occupancy semantics  $\Omega_s$  whereas both the BBA on the occupancy semantics  $\Omega_s$  and

the BBA on the occupancy dynamics  $\Omega_d$  is to be updated. Hence, basic combinations rules such as Dempster's rule are not applicable in the BBA update step of this work. Therefore, the concept of DEVNs is utilized to model the update of the BBAs on  $\Omega_s$  and  $\Omega_d$  jointly. This further enables modeling dependencies between the two FoDs.

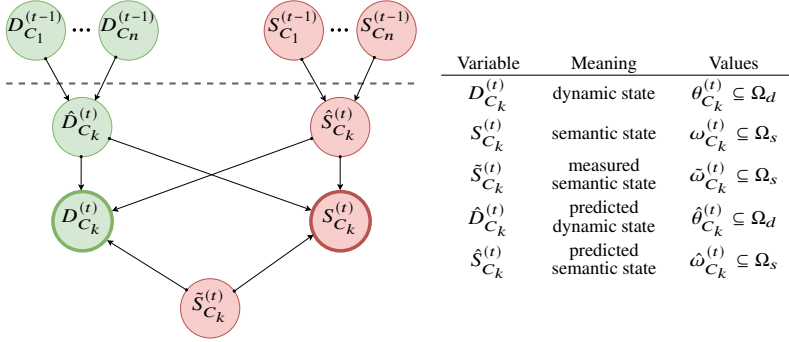


Figure 5.4: The DEVN for the temporal fusion of the BBA on the occupancy semantics  $\Omega_s$  and the occupancy dynamics  $\Omega_d$ . The color of the nodes indicates the underlying FoDs  $\Omega_d$  (green) and  $\Omega_s$  (red). The circled nodes  $D_{C_k}^{(t)}$  and  $S_{C_k}^{(t)}$  are the states linked to the updated BBAs.

The evidential network modeling the dependencies in the temporal fusion on the FoDs  $\Omega_s$  and  $\Omega_d$  is shown in Figure 5.4. The objective of the dynamic grid mapping pipeline is to estimate the belief distributions of the dynamic state  $D_{C_k}^{(t)}$  and the semantic state  $S_{C_k}^{(t)}$  based on the predicted occupancy semantics  $\hat{S}_{C_k}^{(t)}$ , the predicted occupancy dynamics  $\hat{D}_{C_k}^{(t)}$  and the measured occupancy semantics  $\tilde{S}_{C_k}^{(t)}$ . The conditional belief functions induced by the evidential network are

- the prediction of the dynamic and semantic state given the states in neighboring grid cells  $C_1, \dots, C_n \in \mathcal{G}_{xy}$  at time  $t - 1$  represented by the conditional belief

$$\text{bel}_{\hat{D}_{C_k}^{(t)}} \left( \hat{\theta}_{C_k}^{(t)} \mid \theta_{C_1}^{(t-1)}, \dots, \theta_{C_n}^{(t-1)} \right), \quad (5.42)$$

$$\text{bel}_{\hat{S}_{C_k}^{(t)}} \left( \hat{\omega}_{C_k}^{(t)} \mid \omega_{C_1}^{(t-1)}, \dots, \omega_{C_n}^{(t-1)} \right), \quad (5.43)$$

- the conditional belief functions of the semantic state given the predicted occupancy semantics  $\hat{\omega}_{C_k}^{(t)}$ , the predicted occupancy dynamics  $\hat{\theta}_{C_k}^{(t)}$  and the measured occupancy semantics  $\tilde{\omega}_{C_k}^{(t)}$

$$\text{bel}_{\hat{D}_{C_k}^{(t)}} \left( \omega_{C_k}^{(t)} \mid \hat{\omega}_{C_k}^{(t)}, \hat{\theta}_{C_k}^{(t)}, \tilde{\omega}_{C_k}^{(t)} \right), \quad (5.44)$$

- the conditional belief functions of the dynamic state given the predicted occupancy semantics  $\hat{\omega}_{C_k}^{(t)}$ , the predicted occupancy dynamics  $\hat{\theta}_{C_k}^{(t)}$  and the measured occupancy semantics  $\tilde{\omega}_{C_k}^{(t)}$

$$\text{bel}_{\hat{D}_{C_k}^{(t)}} \left( \theta_{C_k}^{(t)} \mid \hat{\omega}_{C_k}^{(t)}, \hat{\theta}_{C_k}^{(t)}, \tilde{\omega}_{C_k}^{(t)} \right). \quad (5.45)$$

In the following, the super- and subscripts denoting the time  $t$  and grid cell  $C_k$  are omitted as they stay constant throughout the remainder of this section. Furthermore, the abbreviation

$$\hat{S}\hat{D}\tilde{S} = \left( \hat{S}_{C_k}^{(t)}, \hat{D}_{C_k}^{(t)}, \tilde{S}_{C_k} \right)$$

is introduced. The inference in the DEVN is done in the following two steps:

### 1. *Belief Prediction*

Calculate the predicted semantic and dynamic state beliefs. This is estimated in the grid map (Equations (5.13) and (5.15)) and using the predicted particle population (Equation (5.27)), respectively.

### 2. *Belief Update*

Calculate the updated semantic and dynamic state beliefs:

$$\text{bel}_S(\omega) = \sum_{\hat{\omega}, \hat{\theta}, \tilde{\omega}} \text{bel}_S(\omega \mid \hat{\omega}, \hat{\theta}, \tilde{\omega}) m_{\hat{S}\hat{D}\tilde{S}}(\hat{\omega}, \hat{\theta}, \tilde{\omega}), \quad (5.46)$$

$$\text{bel}_D(\theta) = \sum_{\hat{\omega}, \hat{\theta}, \tilde{\omega}} \text{bel}_D(\theta \mid \hat{\omega}, \hat{\theta}, \tilde{\omega}) m_{\hat{S}\hat{D}\tilde{S}}(\hat{\omega}, \hat{\theta}, \tilde{\omega}). \quad (5.47)$$

Note that the measured occupancy semantics  $\tilde{\omega}_{C_k}^{(t)}$  is independent of the predicted states  $\hat{\omega}_{C_k}^{(t)}$  and  $\hat{\theta}_{C_k}^{(t)}$ . Thus, the joint BBA can be written as the product

$$m_{\hat{S}_D \hat{S}}(\hat{\omega}, \hat{\theta}, \tilde{\omega}) = m_{\hat{S}_D}(\hat{\omega}, \hat{\theta}) m_{\hat{S}}(\tilde{\omega}), \quad (5.48)$$

where  $m_{\hat{S}_D}$  is known from the prediction step and  $m_{\hat{S}}$  is contained in the measurement grid map.

The main objective of the belief update step is to estimate the conditional belief functions for the semantic and the dynamic state given its ancestors in the DEVN. This work proposes to model the conditional belief functions based on three parameters  $0 \leq \lambda_i \leq 1$ . To avoid the need for heuristic expert knowledge and re-tuning for various domains, it is also shown how to learn these parameters from data.

Some occupancy dynamics hypotheses  $\theta \subseteq \Omega_d$  depend on the occupancy semantics  $\omega \subseteq \Omega_s$ . For example, only the mobile types  $O_{\text{car}}$ ,  $O_{\text{tw}}$ ,  $O_{\text{ped}}$ ,  $O_{\text{om}}$  and semantically unclassified occupancy  $O_{\text{su}}$  might be moving. The dependency

	$O_{\text{car}}$	$O_{\text{tw}}$	$O_{\text{ped}}$	$O_{\text{om}}$	$O_{\text{im}}$	$O_{\text{su}}$	$F_s$	$\Omega_s$
$O_{\text{mov}}$	0/1	0/1	0/1	0/1	0/0	0/1	0/0	0/1
$O_{\text{stat}}$	0/1	0/1	0/1	0/1	1/1	0/1	0/0	0/1
$O_{\text{du}}$	1/1	1/1	1/1	1/1	1/1	1/1	0/0	0/1
$F_d$	0/0	0/0	0/0	0/0	0/0	0/0	1/1	0/1
$\Omega_d$	1/1	1/1	1/1	1/1	1/1	1/1	1/1	1/1

Table 5.1: The conditional beliefs and plausibilities  $\text{bel}_D(\theta|\omega)/\text{pl}_D(\theta|\omega)$  of the occupancy dynamics  $\theta \subseteq \Omega_d$  given the occupancy semantics  $\omega \subseteq \Omega_s$ .

between the semantic and the dynamic states are formalized in Table 5.1. The table shows the conditional beliefs and plausibilities  $\text{bel}_D(\theta|\omega)$  and  $\text{pl}_D(\theta|\omega)$  of the occupancy dynamics  $\theta \subseteq \Omega_d$  given the occupancy semantics  $\omega \subseteq \Omega_s$ .

In the following, we state the conditional BBAs

$$m_X(x | \hat{\omega}, \hat{\theta}, \tilde{\omega}) = \sum_{y \subseteq x} (-1)^{|x|-|y|} \text{bel}_X(x | \hat{\omega}, \hat{\theta}, \tilde{\omega}), \quad X \in \{D, S\}, \quad (5.49)$$

and assume

$$m_X(x | \hat{\omega}, \hat{\theta}, \tilde{\omega}) = 0, \quad (5.50)$$



if not stated otherwise. We distinguish between consenting dependent states, occupancy/free switches and semantic conflicts.

**Consenting states.** Let the predicted occupancy semantics  $\hat{\omega} \subseteq \Omega_s$  and the measured occupancy semantics  $\tilde{\omega} \subseteq \Omega_s$  be consenting, i.e.  $\hat{\omega} \cap \tilde{\omega} \neq \emptyset$ . For occupancy semantics, the BBA is assigned to the intersecting hypothesis by setting

$$m_S(\hat{\omega} \cap \tilde{\omega} \mid \hat{\omega}, \hat{\theta}, \tilde{\omega}) = 1. \quad (5.51)$$

For occupancy dynamics, we aim to separate moving from stationary occupancy. Here, we consider three categories of consenting states, namely

1. newly observed occupancy, i.e. unknown predicted states  $\hat{\omega} = \Omega_s$ ,  $\hat{\theta} = \Omega_d$  and measured occupancy semantics  $\tilde{\omega} \subseteq O_{su}$ ,
2. newly observed free space, i.e. unknown predicted states  $\hat{\omega} = \Omega_s$ ,  $\hat{\theta} = \Omega_d$  and measured free space  $\tilde{\omega} = F_s$  and
3. repeatedly observing occupancy, i.e. predicted occupancy  $\hat{\theta} \subseteq O_{du}$  and  $\hat{\omega} \subseteq O_{su}$  and measured occupancy  $\tilde{\omega} \subseteq O_{su}$ .

In the first category, newly observed occupancy, moving and dynamically unclassified occupancy can be deduced. For moving occupancy  $O_{mov} \subset \Omega_d$ , the conditional belief depends on the mobility of the observed entity that is represented by the conditional plausibility  $pl_D(O_{mov} \mid \tilde{\omega})$  of moving occupancy given the measured occupancy semantics  $\tilde{\omega}$ . We parametrize the conditional belief for moving occupancy given that occupancy  $\tilde{\omega} \subseteq O_{su}$  was newly observed as

$$m_D(O_{mov} \mid \Omega_s, \Omega_d, \tilde{\omega}) = \lambda_{\Omega_d \rightarrow O_{mov}} pl_D(O_{mov} \mid \tilde{\omega}), \quad (5.52)$$

where  $0 \leq \lambda_{\Omega_d \rightarrow O_{mov}} \leq 1$  is a parameter specifying the amount of newly observed occupancy BBA assigned to moving occupancy  $O_{mov}$ . As evidence for moving occupancy  $O_{mov}$  is gained, the belief of dynamically unclassified occupancy  $O_{du} \subset \Omega_d$  is reduced accordingly as

$$m_D(O_{du} \mid \Omega_s, \Omega_d, \tilde{\omega}) = 1 - m_D(O_{mov} \mid \Omega_s, \Omega_d, \tilde{\omega}). \quad (5.53)$$

As opposed to occupied grid cell hypotheses that may be moving or stationary, the free space hypothesis is not subdivided. Thus, in the second category of

consenting states, newly observed free space is fully assigned to the free space hypotheses as

$$m_D(F_d | \Omega_s, \Omega_d, F_s) = 1. \quad (5.54)$$

The third category of consenting states, i.e. repeatedly observing occupancy, implies a stationary dynamic state by setting

$$\hat{\omega} \cap \tilde{\omega} \subseteq O_{\text{su}}, \hat{\theta} \in \{O_{\text{stat}}, O_{\text{du}}\} \Rightarrow m_D(O_{\text{stat}} | \hat{\omega}, \hat{\theta}, \tilde{\omega}) = 1, \quad (5.55)$$

if the predicted occupancy dynamics is stationary occupied  $O_{\text{stat}}$  or dynamically unclassified occupied  $O_{\text{du}}$ . Note that this does not consider that moving entities might occupy the same grid cell at several consecutive update steps. However, this is handled by assigning parts of newly observed occupancy to the moving occupancy hypothesis  $O_{\text{mov}}$  in Equation (5.52) so that it is not predicted to dynamically unclassified occupancy  $\hat{\theta} = O_{\text{du}}$  in the first place. For predicted moving occupancy  $O_{\text{mov}}$ , the belief remains at moving occupancy, i.e.

$$\hat{\omega} \cap \tilde{\omega} \subseteq O_{\text{su}}, \hat{\theta} \in \{O_{\text{mov}}\} \Rightarrow m_D(O_{\text{mov}} | \hat{\omega}, \hat{\theta}, \tilde{\omega}) = 1. \quad (5.56)$$

**Occupied/free switches.** For occupied/free switches, we distinguish between

1. grid cells previously observed free, i.e. predicted as passable  $\hat{\theta} = P$  and now observed occupied  $\tilde{\omega} \subseteq O_{\text{su}}$ ,
2. grid cells previously observed moving  $\hat{\theta} = O_{\text{mov}}$  or dynamically unclassified occupied  $\hat{\theta} = O_{\text{du}}$  and now observed free  $\tilde{\omega} = F_s$ , and
3. grid cells previously observed stationary occupied  $\hat{\theta} = O_{\text{stat}}$  and now observed free  $\tilde{\omega} = F_s$ .

The case that a grid cell is predicted as passable  $\hat{\theta} = P$  and currently observed as occupied  $\tilde{\omega} \subseteq O_{\text{su}}$  is technically not a conflict in the presented evidential model. It intersects to the moving occupancy hypotheses  $O_{\text{mov}} \subset \Omega_d$ . However, to prevent falsely deducing moving occupancy in case of noisy measurements

the occupancy evidence is not fully assigned to moving occupancy  $O_{\text{mov}}$  but only partly based on the parameter  $0 \leq \lambda_{P \rightarrow O_{\text{mov}}} \leq 1$  as

$$m_D(O_{\text{mov}} | \hat{\omega}, P, \tilde{\omega}) = \lambda_{P \rightarrow O_{\text{mov}}} \text{pl}_D(O_{\text{mov}} | \tilde{\omega}) + (1 - \lambda_{P \rightarrow O_{\text{mov}}}) \lambda_{\Omega_d \rightarrow O_{\text{mov}}} \text{pl}_D(O_{\text{mov}} | \tilde{\omega}). \quad (5.57)$$

Here, the part  $\lambda_{\Omega_d \rightarrow O_{\text{mov}}}$  of  $1 - \lambda_{P \rightarrow O_{\text{mov}}}$  is assigned to moving occupancy  $O_{\text{mov}}$  as well instead of assigning it to dynamically unclassified occupancy  $O_{\text{du}}$ . The residual part of the observed occupancy semantics  $\tilde{\omega} \subseteq O_{\text{su}}$  is assigned to dynamically unclassified occupancy  $O_{\text{du}}$  by setting

$$m_D(O_{\text{du}} | \hat{\omega}, P, \tilde{\omega}) = 1 - m_D(O_{\text{mov}} | \hat{\omega}, P, \tilde{\omega}). \quad (5.58)$$

For occupancy semantics, the BBA is fully assigned to the measured occupancy  $\tilde{\omega} \subseteq O_{\text{su}}$ , i.e.

$$m_S(\tilde{\omega} | \hat{\omega}, P, \tilde{\omega}) = 1. \quad (5.59)$$

Cells previously observed moving occupied  $\hat{\theta} = O_{\text{mov}}$  or dynamically unclassified  $\hat{\theta} = O_{\text{du}}$  and now observed free  $\tilde{\omega} = F_s$  are assumed to be free, that means for  $\hat{\omega} \subseteq O_{\text{su}}$ , the conditional beliefs read

$$m_S(F_s | \hat{\omega}, O_{\text{du}}, F_s) = 1, \quad m_D(F_d | \hat{\omega}, O_{\text{du}}, F_s) = 1, \quad (5.60)$$

$$m_S(F_s | \hat{\omega}, O_{\text{mov}}, F_s) = 1, \quad m_D(F_d | \hat{\omega}, O_{\text{mov}}, F_s) = 1. \quad (5.61)$$

Note that this overwrites potentially stationary occupancy contained in the superset  $O_{\text{du}} \supset O_{\text{stat}}$ . The above modeling of the conditional belief  $m_D(F_d | \hat{\omega}, O_{\text{du}}, F_s)$  ensures that occupied/free conflicts are only resolved in favor of the occupancy hypothesis, if stationary occupancy has been confirmed even for immobile predicted occupancy semantics  $\hat{\omega} = O_{\text{im}}$ .

Finally, we consider cells previously observed as stationary occupied  $\hat{\theta} = O_{\text{stat}}$  and now observed as free  $\tilde{\omega} = F_s$ . Here, we distinguish two cases:

1. The measured occupancy semantics is free due to missed detections. In this case, the conflict should be resolved in favor of stationary occupancy  $O_{\text{stat}}$ .
2. The predicted occupancy dynamics is stationary occupied due to false detections in past measurements or because a stationary, mobile object

starts moving. In those cases, the conflict should be resolved in favor of the free space hypothesis  $F_d$ .

Therefore, the belief is partially assigned to free space and stationary occupancy based on the parameter  $0 \leq \lambda_{O_{\text{stat}} \rightarrow F_d} \leq 1$  by setting

$$m_S(F_s | \hat{\omega}, O_{\text{stat}}, F_s) = m_D(F_d | \hat{\omega}, O_{\text{stat}}, F_s) = \lambda_{O_{\text{stat}} \rightarrow F_d}, \quad (5.62)$$

$$m_S(\hat{\omega} | \hat{\omega}, O_{\text{stat}}, F_s) = m_D(O_{\text{stat}} | \hat{\omega}, O_{\text{stat}}, F_s) = 1 - \lambda_{O_{\text{stat}} \rightarrow F_d}. \quad (5.63)$$

**Semantic conflicts.** In case of conflicting occupancy semantics  $\hat{\omega}, \tilde{\omega} \subseteq O_{\text{su}}$  with  $\hat{\omega} \cap \tilde{\omega} = \emptyset$ , the BBA can be assigned to semantically unclassified occupancy  $O_{\text{su}} \subset \Omega_s$  as

$$m_S(O_{\text{su}} | \hat{\omega}, \hat{\theta}, \tilde{\omega}) = m_D(O_{\text{du}} | \hat{\omega}, \hat{\theta}, \tilde{\omega}) = 1. \quad (5.64)$$

This ensures that even in case of conflicting semantic estimates, the evidence on cell occupancy is preserved.

### 5.3.3 Parameter Estimation

For modeling the conditional beliefs induced by the DEVN (Figure 5.4) the parameters  $\lambda_{\Omega_d \rightarrow O_{\text{mov}}}$  for the initialization of moving occupancy, and  $\lambda_{P \rightarrow O_{\text{mov}}}, \lambda_{O_{\text{stat}} \rightarrow F_d}$  for resolving occupied/free switches are introduced. We propose ways to estimate those parameters.

#### Initialization of moving occupancy

The initialization of moving occupancy in previously unobserved grid cells represented by the conditional belief  $\text{bel}_D(O_{\text{mov}} | \hat{\omega}, \Omega_d, \tilde{\omega})$  is parametrized by  $\lambda_{\Omega_d \rightarrow O_{\text{mov}}} \in [0, 1]$  in Equations (5.52) and (5.53). We propose to estimate this

parameter based on the predicted particle population  $\hat{\chi}^{(t)}$ . Recall that particles initialized in newly observed occupancy  $\tilde{\omega} \subseteq O_{\text{su}}$

$$\begin{aligned} \Delta^+ m_C^{(t)}(O_{\text{du}}) = & \sum_{\tilde{\omega} \subseteq O_{\text{su}}} m_D(O_{\text{du}} | \Omega_s, \Omega_d, \tilde{\omega}) m_{\hat{S}\hat{D}}(\Omega_s, \Omega_d) m_{\hat{S}}(\tilde{\omega}) \\ & + m_D(O_{\text{du}} | \Omega_s, P, \tilde{\omega}) m_{\hat{S}\hat{D}}(\Omega_s, P) m_{\hat{S}}(\tilde{\omega}) \end{aligned} \quad (5.65)$$

have weight zero after the resampling step. Therefore, not all particles predicted into a grid cell contribute to the predicted BBA for moving occupancy  $O_{\text{mov}}$ . The idea for estimating  $\lambda_{\Omega_d \rightarrow O_{\text{mov}}}$  is that all particles including those with weight zero can be used to confirm measured occupancy as moving. Let

$$m_\chi^{(t)} = \frac{1}{n} \sum_{C \in \mathcal{G}_{xy}} m_C^{(t)}(O_{\text{mov}}) + \Delta^+ m_C^{(t)}(O_{\text{du}}) \quad (5.66)$$

be the BBA for moving and the newly gained BBA for dynamically unclassified occupancy  $\Delta^+ m_C^{(t)}(O_{\text{du}})$  per particle at time  $t$ . The sum of weights  $m_\chi^{(t-1)}$  of the predicted particle population in a grid cell represents moving and potentially moving occupancy. Therefore,  $\lambda_{\Omega_d \rightarrow O_{\text{mov}}}$  is estimated as

$$\lambda_{\Omega_d \rightarrow O_{\text{mov}}} = \left| \hat{\chi}_C^{(t)} \right| m_\chi^{(t-1)}, \quad (5.67)$$

so that moving occupancy is initialized in previously unobserved grid cells if it contains predicted particles and is observed as occupied in the current measurement.

### Occupied/free switches

Consider the parameters  $\lambda_{P \rightarrow O_{\text{mov}}}, \lambda_{O_{\text{stat}} \rightarrow F_d} \in [0, 1]$  modeling occupied/free switches.

**Baseline parameterization.** Note that the described DEVN is a generalization of the manual conflict assignment based on Yager's rule on the FoD  $\Omega_d$  presented by Steyer et al. [STW18]. This will serve as baseline for the now following parameter estimation. A comparative evaluation of their parameterization and our parameterization described in the next section will be presented

in Section 5.4. For reference, we state the corresponding parameterization used by Steyer et al. [STW18] in Table 5.2.

$\lambda_{O_{\text{stat}} \rightarrow F_d}$	0.5
$\lambda_{P \rightarrow O_{\text{mov}}}$	0.4

Table 5.2: The parameterization proposed by Steyer et al. [STW18].

**Data-driven parameterization.** Instead of assigning the two parameters  $\lambda_{O_{\text{stat}} \rightarrow F_d}$  and  $\lambda_{P \rightarrow O_{\text{mov}}}$  a constant value as in the baseline parameterization, a data driven approach is introduced that estimates them based on observed cues and statistical considerations. This might help to detect the causes for the occupied/free switches and could thus lead to a more robust and accurate BBA estimation.

Consider the conditional belief functions

1.  $\text{bel}_D(\cdot \mid \Omega_s, P, \tilde{\omega})$  for the case that a grid cell is predicted as passable  $P$  and measured occupied, i.e.  $\tilde{\omega} \subseteq O_{\text{su}}$ , and
2.  $\text{bel}_D(\cdot \mid \hat{\omega}, O_{\text{stat}}, F_s)$  for the conflict that a grid cell predicted as stationary occupied  $O_{\text{stat}}$  with occupancy semantics  $\hat{\omega} \subseteq O_{\text{su}}$  was observed free  $F_s$ .

Both conditional beliefs represent occupied/free switches as they either depend on the grid cell to be free at an earlier point in time and observed occupied at the current time point or the other way around. Those occupied/free switches may be due to moving entities entering or leaving the grid cell. We model this by defining binary classifiers for a moving entity to be present and estimate the probability for this by applying a logistic regression.

For the conditional belief  $\text{bel}_D(\cdot \mid \hat{\omega}, P, \tilde{\omega})$  with  $\tilde{\omega} \subseteq O_{\text{su}}$ , the grid cell may be moving occupied in case the switch from free to occupied was caused by a moving entity entering the grid cell. Hence, we define the binary classifier

$$y_{P \rightarrow O_{\text{mov}}} = \begin{cases} 1, & \text{if a moving entity enters the cell,} \\ 0, & \text{else.} \end{cases} \quad (5.68)$$

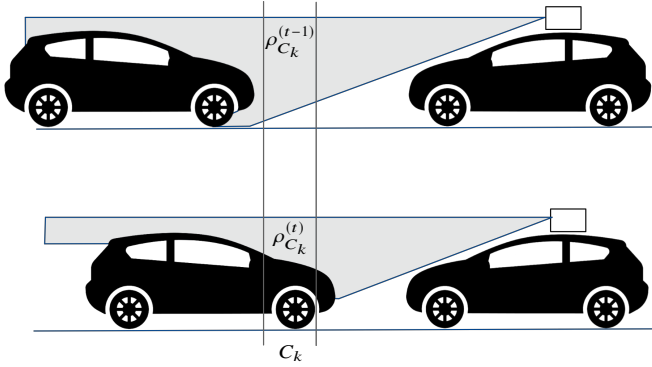


Figure 5.5: Visualization of the permeability change for moving objects. The ego vehicle is on the right and a passing vehicle is coming from the left. When the vehicle enters the considered grid cell  $C_k$ , parts of the driving corridor are blocked leading to a permeability drop.

Other reasons for the state switch may be that the grid cell is still free and false detections caused the current measurement or the grid cell was falsely estimated as free in past update steps. Recall that in Equation (5.58) the part that is not assigned to moving occupancy  $O_{\text{mov}}$  is assigned to dynamically unclassified occupancy  $O_{\text{du}}$  following a conservative policy.

For the conditional belief  $\text{bel}_D(\cdot | \hat{\omega}, O_{\text{stat}}, F_s)$ , the grid cell may either be free in case a moving entity has left the grid cell or still be stationary occupied in case the occupying entity was missed by the sensor. Therefore, the binary classifier

$$y_{O_{\text{stat}} \rightarrow F_d} = \begin{cases} 1, & \text{if a moving entity leaves the cell,} \\ 0, & \text{if cell remains occupied.} \end{cases} \quad (5.69)$$

covers the two possible causes for the occupied/free conflict.

The higher the vertical portion of the driving corridor that is covered by the obstacle the less likely it is to miss it. False and missed detections, however, usually cover a small vertical portion of the driving corridor as they usually occur for single measurement rays. The progression of the ray permeability defined in Equation (3.4) over time provides a way to separate the presence of

moving entities from missed or false detections as indicated in Figure 5.5. The idea is that moving entities have a certain height, thus covering a large portion of the driving corridor leading to large changes in the ray permeability when entering or leaving a grid cell. On the other hand, low entities such as curb stones that are likely to be missed detections and measurement noise can be detected by small changes in the ray permeability.

In order to visualize the dependency between the ray permeabilities  $\rho^{(t-1)}$  and  $\rho^{(t)}$  at two consecutive update time points and the binary classifiers  $y_{P \rightarrow O_{\text{mov}}}$  and  $y_{O_{\text{stat}} \rightarrow F_d}$ , we collect reference classifications in permeability histograms. The reference classification is generated based on the reference grid map  $\mathbf{g}_{\text{ref}}$  representing the BBA on the occupancy dynamics  $\Omega_d$  estimated using the KITTI-360 dataset as described in Section 2.4. The histogram

$$h(y_{\text{ref}, P \rightarrow O_{\text{mov}}} | F_s, O_{\text{su}}): \mathcal{G}_{[0,1]} \rightarrow \mathbb{R}_{\geq 0} \quad (5.70)$$

counts the number of grid cells with reference classification  $y_{\text{ref}, P \rightarrow O_{\text{mov}}}$  given that the measured occupancy semantics BBA  $m_C^{(t-1)}(F_s)$  at time  $t - 1$  and  $m_C^{(t)}(O_{\text{su}})$  at time  $t$  exceeded the threshold  $\tau = 0.7$ . Analogously,

$$h(y_{\text{ref}, O_{\text{stat}} \rightarrow F_d} | O_{\text{su}}, F_s): \mathcal{G}_{[0,1]} \rightarrow \mathbb{R}_{\geq 0} \quad (5.71)$$

counts the number of grid cells with reference classification  $y_{\text{ref}, O_{\text{stat}} \rightarrow F_d}$  given that the measured occupancy semantics BBA  $m_C^{(t-1)}(O_{\text{su}}), m_C^{(t)}(F_s) > 0.7$ . The histogram bins are defined by a grid  $\mathcal{G}_{[0,1]}$  discretizing the unit square where each cell  $C \in \mathcal{G}_{[0,1]}$  represents pairs of ray permeabilities  $(\rho^{(t-1)}, \rho^{(t)}) \in C$ . The statistics were collected in the KITTI-360 training sequence, see Table 2.5.

Figure 5.6 depicts the histograms representing moving entities entering or leaving a grid cell, false and missed detections. Figures 5.6a and 5.6b show the cases that a moving entity is either entering or leaving a grid cell. In both cases, large changes in the ray permeability can be observed indicated by high histogram counts either in the lower right (Figure 5.6a) or the upper left corner (Figure 5.6b). In the case that a moving entity is entering a grid cell, the permeability variance is higher than for entities leaving a grid cell as indicated by  $\rho^{(t)}$  in Figure 5.6a and  $\rho^{(t-1)}$  in Figure 5.6b. This can be explained by the fact that if the entity is a car, the front is observed in such cases that is lower and thus leads to a higher ray permeability. For false detections visualized in



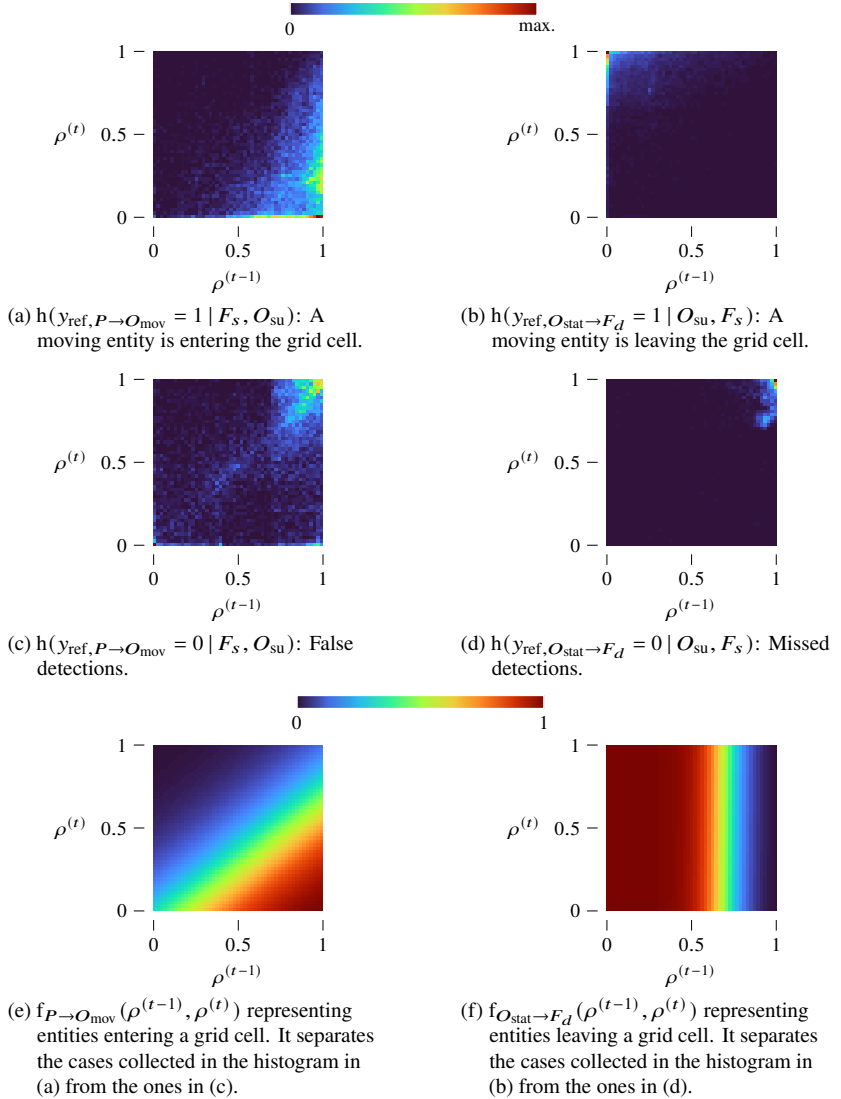


Figure 5.6: Histograms on  $\mathcal{G}_{[0,1]}$  representing the ray permeabilities  $\rho^{(t-1)}, \rho^{(t)} \in [0, 1]$  at two consecutive frames. The values in (a) - (d) are normalized between zero and the maximal cell count.

Figure 5.6c, the ray permeability remains unchanged mostly at a high level indicated by high histogram counts in the upper right corner of  $\mathcal{G}_{[0,1]}$ . Missed detections also lead to small changes in the ray permeability, see Figure 5.6d. The average is reached for  $\rho^{(t-1)} \approx 1$  and  $\rho^{(t)}$  slightly below one. This is due to the fact that missed detections are mostly observed at curb stones leading to a small permeability drop according to their height.

Based on the statistics contained in the histograms, we can fit a logistic function

$$f\left(\rho^{(t-1)}, \rho^{(t)}\right) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 \rho^{(t-1)} + \beta_2 \rho^{(t)}\right)\right)} \quad (5.72)$$

to quantify the probability for the binary classifiers based on the ray permeabilities  $\rho^{(t-1)}$  and  $\rho^{(t)}$ . The logistic regression problem is solved using the Python library scikit-learn [Ped+11] using the coordinate descent algorithm presented by Friedman et al. [FHT10] with  $l_2$  regularization in the cost function. To account for an unbalanced number of samples for  $y = 0$  and  $y = 1$ , the sample weights are set inversely proportional to the sample frequencies. By solving the logistic regression, we obtain the parameters for the logistic function  $f_{P \rightarrow O_{\text{mov}}}$  quantifying the probability for  $y_{P \rightarrow O_{\text{mov}}} = 1$  depicted in Figure 5.6e. The probability for  $y_{O_{\text{stat}} \rightarrow F_d}$  is modeled by the logistic function  $f_{O_{\text{stat}} \rightarrow F_d}$  visualized in Figure 5.6f.

Since the predicted BBAs for passable space  $P$  and stationary occupancy  $O_{\text{stat}}$  might not only contain the measurement from the last time point  $t - 1$  but also earlier measurements, we accumulate the ray permeability over time. As the BBA is recursively updated for all hypotheses simultaneously, we do the accumulation separately for the assumption that the grid cell is stationary occupied and free, respectively. The ray permeability  $\rho_{O_{\text{stat}}}^{(t)}$  at time  $t$  given that the grid cell is occupied by a stationary entity is recursively updated as the weighted average between the value from the last time point  $t - 1$  and the current measured permeability. More specifically, the weight is set to the BBA for stationary occupancy  $O_{\text{stat}} \subseteq \Omega_d$ , so that

$$\rho_{O_{\text{stat}}}^{(t)} = m_C^{(t)}(O_{\text{stat}}) \rho^{(t)} + \left(1 - m_C^{(t)}(O_{\text{stat}})\right) \rho_{O_{\text{stat}}}^{(t-1)}. \quad (5.73)$$

Analogously, the ray permeability given that the grid cell is passable is updated as the weighted average

$$\rho_P^{(t)} = m_C^{(t)}(F_d) \rho^{(t)} + \left(1 - m_C^{(t)}(F_d)\right) \rho_P^{(t-1)} \quad (5.74)$$

based on the free space BBA.

Finally, the parameters  $\lambda_{P \rightarrow O_{\text{mov}}}$  and  $\lambda_{O_{\text{stat}} \rightarrow F_d}$  are set based on the ray permeability as

$$\lambda_{P \rightarrow O_{\text{mov}}} = f_{P \rightarrow O_{\text{mov}}}(\rho_P^{(t-1)}, \rho^{(t)}), \quad (5.75)$$

$$\lambda_{O_{\text{stat}} \rightarrow F_d} = f_{O_{\text{stat}} \rightarrow F_d}(\rho_{O_{\text{stat}}}^{(t-1)}, \rho^{(t)}) \quad (5.76)$$

using the logistic functions  $f_{P \rightarrow O_{\text{mov}}}$  and  $f_{O_{\text{stat}} \rightarrow F_d}$ .

## 5.4 Experiments

In the temporal fusion of the fused measurement grid map, the BBA on the ground semantics  $\Omega_g$ , the occupancy semantics  $\Omega_s$  and the occupancy dynamics  $\Omega_d$  are estimated. In this section the BBAs are evaluated visually and qualitatively. Furthermore, the quality of the particle-based cell velocity estimation is investigated.

### 5.4.1 Particle Filter

The particle filter estimates the motion state of moving occupancy. We first test the capability of the particle filter to converge towards occupancy motion in simulated scenarios. For this purpose, extended object states are spawn following motion patterns with a constant acceleration and constant turn rate motion model. Based on those object states, a simple range sensor is simulated by casting rays from a defined sensor location to the first intersection with the object's hull in Polar coordinates. Then the measurement BBA is calculated by applying an inverse sensor model similar as described in Section 3.3.3. We process the simulated measurements on a 60 m by 40 m large grid with a grid

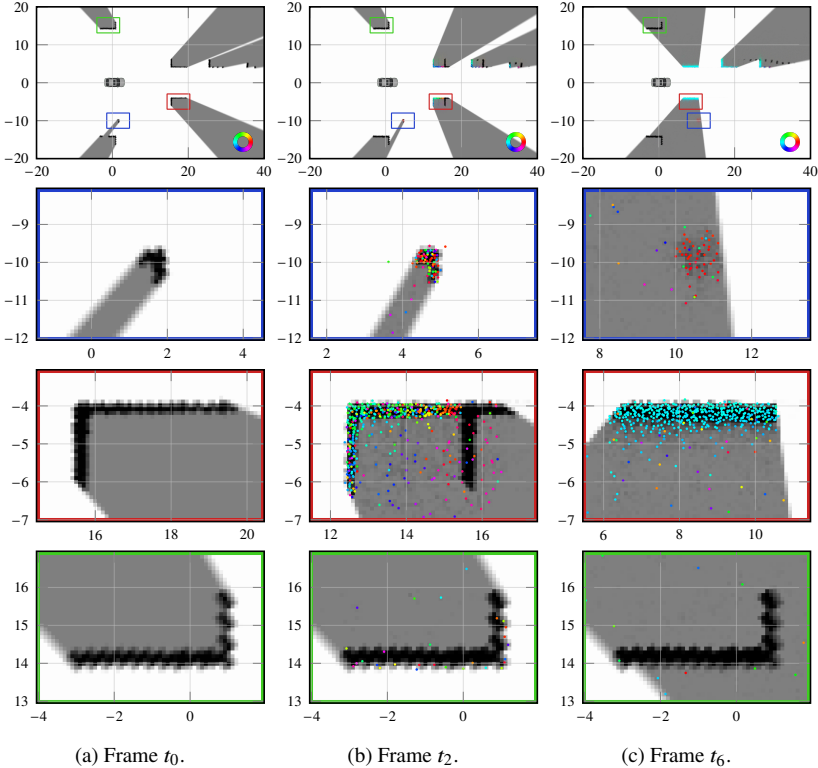


Figure 5.7: Results of the particle filter after the resampling step in a simulated scenario. The first row shows the whole grid map visualizing the occupancy probability and color-coded velocity orientation. The other rows magnify one specific object of interest and visualizes every 50th particles with non-zero weight with the same color-coding. Each column corresponds to one of the three update time points  $t_0$ ,  $t_2$  and  $t_6$ .

cell size of 10 cm by 10 cm. The number of particles  $n$  is set to 100 000 and the numbers of new particles  $n_{\text{new}}, n_0$  is set to 10 000 each.

Figure 5.7 shows a simulated scenario where the ego vehicle is stopping, and several other vehicles are passing. Furthermore, two stationary vehicles are located on the left and on the right of the ego vehicle. In the first row, the occupancy probability is shown where the direction of motion is additionally

visualized in each grid cell  $C \in \mathcal{G}_{xy}$  by the hue component in the HSV color space. Here the brightness is determined by the Mahalanobis distance

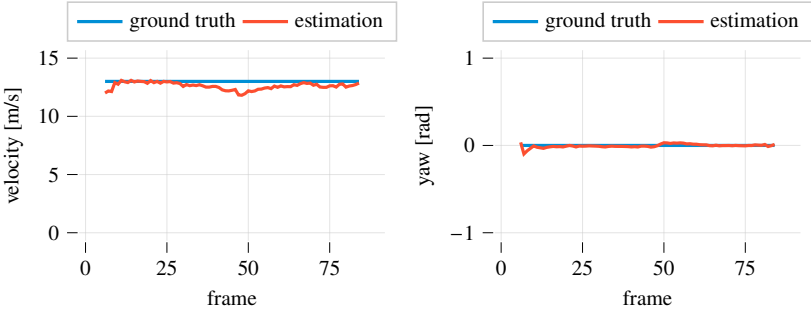
$$d_{\text{MHD}}(v_C) = v_C^T \Sigma_C^{-1} v_C \in \mathbb{R}_{>0}, \quad (5.77)$$

where the maximal brightness is achieved for  $d_{\text{MHD}}(v_C) > 4$  m/s. In the second, third, and fourth row, every 50th particle is visualized with the hue component of the HSV color selected based on the direction of motion. Figure 5.7a shows the state after receiving the first measurement where the view is magnified around three objects. As the particles have only been sampled from the new particle set  $X_0$ , all particles have weight zero. After two further measurement updates, more particles are sampled in cells occupied by moving objects, but they have not yet converged to the correct cell velocities, see Figure 5.7b. Following the survival-of-the-fittest principle, the vast majority of the samples eventually have the correct velocity after six measurement updates, see Figure 5.7c. As the particle filter only models moving occupancy, no particles are spawned in stationary cell occupancy. This can be seen in the bottom row showing one of the stationary objects where only a few particles are sampled in the occupied grid cells.

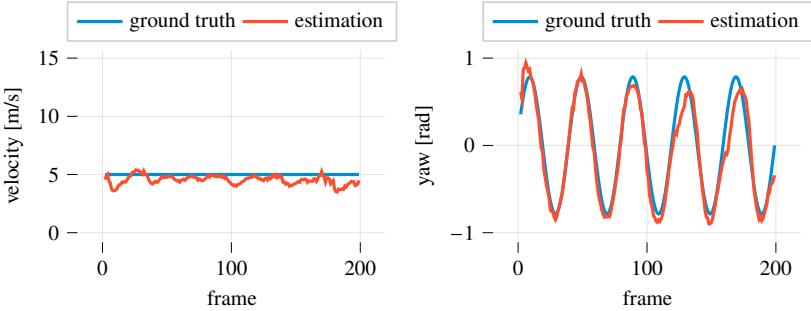
For a scenario-based velocity evaluation, one observed vehicle is simulated in the following scenarios:

1. *Constant velocity scenario*: The observed vehicle drives with a constant velocity of 13 m/s with orientation of 0 rad.
2. *Slalom scenario*: The observed vehicle drives a slalom course with orientation varying between -1 rad and 1 rad and a constant velocity of 5 m/s.
3. *Start-stop scenario*: The observed vehicle decelerates from 10 m/s to a full stop and accelerates back to 10 m/s with orientation of 0 rad.

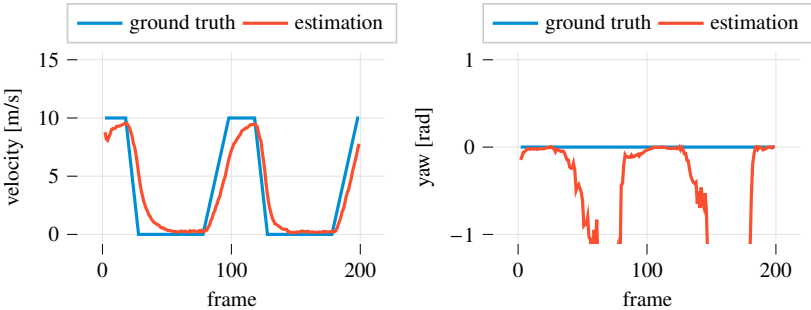
The evaluation of the velocity and orientation estimates of the particle filter in the three simulated scenarios is shown in Figure 5.8. The results for the constant velocity scenario is plotted in Figure 5.8a. Both the simulated velocity and the orientation are estimated accurately for the whole simulated sequence. The same applies for the highly dynamic slalom scenario depicted in Figure 5.8b. Here, small orientation offsets can be seen at the turning points, i.e. at  $\pm 1$  rad. In the start-stop scenario shown in Figure 5.8c the low-pass behavior



(a) Constant velocity scenario.



(b) Slalom scenario.



(c) Start-stop scenario.

Figure 5.8: Velocity and orientation estimates for one observed vehicle in simulated scenarios.

of the particle filter can be observed. When the motion state changes from acceleration or deceleration phases to a constant velocity phase, the particle filter slowly converges towards the object’s velocity. Note that for stopping cell occupancy, no cell orientation can be estimated as the orientation is exclusively estimated based on the moving direction. This effect can be observed in the yaw plot in Figure 5.8c for frames where the simulated velocity drops to 0 m/s.

## 5.4.2 Semantic State

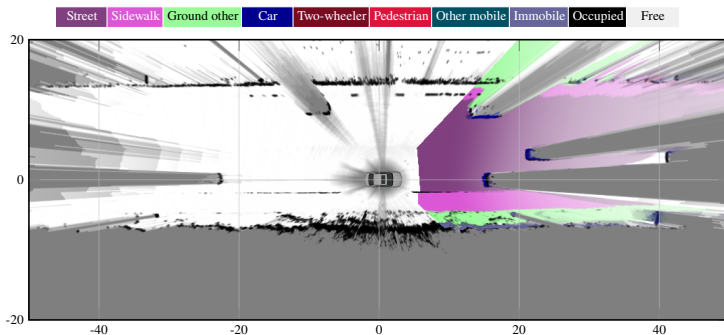
Next, we consider the BBA on the occupancy semantics  $\Omega_s$  in the filtered grid map. Here, a 100 m by 100 m large grid with a grid cell size of 10 cm by 10 cm is chosen. The number of particles  $n$  is set to 500 000 and the numbers of new particles  $n_{\text{new}}, n_0$  is set to 50 000 each. With this setup and the GPU implementation used in this work, the processing time for one cycle of the temporal fusion pipeline is 20 ms on a NVIDIA GeForce RTX 2080 Ti.

Figure 5.9 shows the BBA on the occupancy semantics  $\Omega_s$  and the ground semantics  $\Omega_g$  after applying temporal fusion in comparison to the sensor fusion results in the current frame. In the current fused grid map, semantic estimates are only available in the viewing area of the stereo camera, see Figure 5.9b. In the filtered grid map semantic estimates are successfully accumulated and fused with occupancy from LiDAR over time in the area behind the ego vehicle as can be seen e.g. in areas labeled as street and sidewalk, see Figure 5.9c. The vegetation on the right of the ego lane is correctly classified as immobile occupancy and other ground. The advantage of dividing occupancy semantics and ground semantics into two not necessarily excluding frames can be seen behind the passing vehicle on the left. The evidence mass accumulated for the hypothesis *street* is not overwritten by the passing car and thus stays in the grid map. If ground semantics and occupancy semantics was modeled in a common FoD, the BBAs for *street* would be moved to *occupied* and eventually to *unknown* once the car has passed.

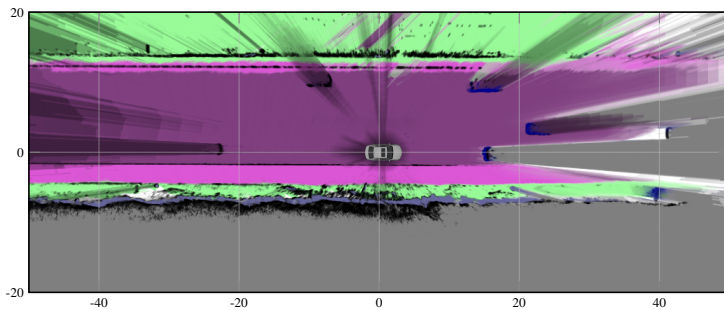
The eIoUs (Equation (2.26)) for the occupancy semantics  $\Omega_s$  are shown in Figure 5.10. The statistics were computed for all frames in the KITTI-360 evaluation sequences, see Table 2.5. To demonstrate the performance evolution at different stages of the pipeline the metrics are presented for the stereo sensor measurement grid map, the grid map after additionally fusing LiDAR



(a) Image of the traffic scene recorded by the left front camera.



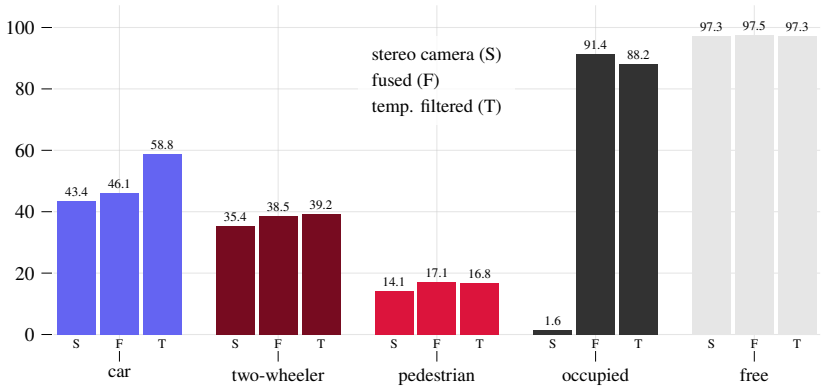
(b) The BBA on the occupancy semantics  $\Omega_s$  and the ground semantics  $\Omega_g$  after fusing LiDAR and stereo camera measurements.



(c) The BBA on the occupancy semantics  $\Omega_s$  and the ground semantics  $\Omega_g$  after fusing LiDAR and stereo camera measurements and temporal fusion.

Figure 5.9: Visualization of the BBA on the occupancy semantics  $\Omega_s$  and the ground semantics  $\Omega_g$ .



Figure 5.10: eIoU for the occupancy semantics  $\Omega_s$ .

measurements and the filtered grid map. In addition to the improvements in the fused grid map compared to the stereo grid map, the eIoUs are once again improved in the filtered grid map. For the hypothesis *car*, the eIoU improved from 46.1% in the fused grid map to 58.8% in the filtered grid map. This is a relative enhancement of 27.5% compared to the fused grid map and 35.5% compared to the stereo grid map. For *two-wheelers*, *pedestrians* and *free space*, the eIoU are slightly improved or stay at a similar level, respectively. In the filtered grid map, the eIoU for *semantically unclassified occupancy* is reduced from 91.4% in the fused grid map to 88.2%. This effect can be explained by the fact that in many grid cells the BBA masses assigned to *semantically unclassified occupancy* in the fused measurement grid map can be assigned to one of the semantic hypotheses at some point in the temporal fusion. This is the opposite effect to the increase of the eIoU for the other hypotheses such as *car*.

Figure 5.11 shows Deng's non specificity and discord, see Equations (2.12) and (2.13), for the occupancy semantics  $\Omega_s$  in the fused measurement grid map and the filtered grid map. Recall that the nonspecificity is a measure for the ignorance contained in the BBA whereas the discord is a measure for the indecision between several hypotheses with non-zero BBA. In Figure 5.11a, all grid cells within a distance of 30 m ( $360^\circ$  view) are considered. In Figure 5.11b, only grid cells are considered that are additionally within the visible area of the stereo camera (camera view). In both cases, nonspecificity is slightly reduced,

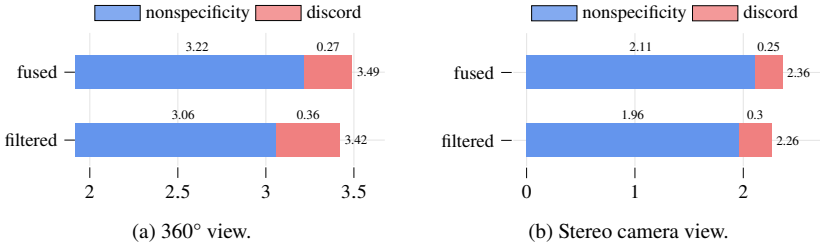


Figure 5.11: Nonspecificity and discord for fused and filtered grid maps on the occupancy semantics  $\Omega_g$ .

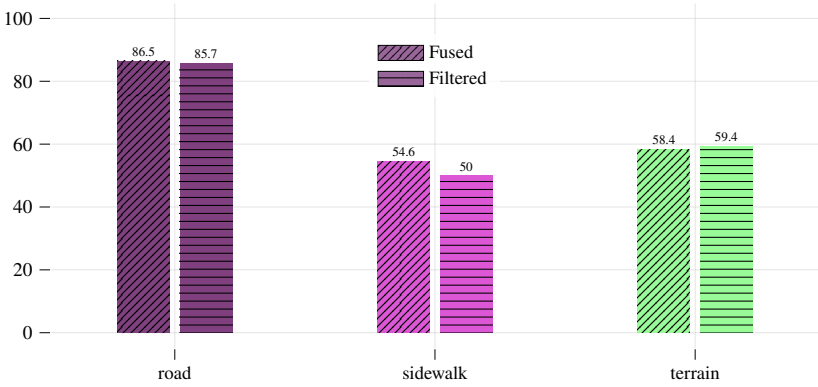


Figure 5.12: eIoUs for the ground semantics  $\Omega_g$ .

and the discord is slightly increased. This is expected when accumulating information over time as more knowledge is gained that, however, might lead to evidential conflicts. Overall, it can be stated that the temporal fusion proposed in those work successfully reduces uncertainty which is demonstrated by a reduced entropy, i.e. the sum of nonspecificity and discord.

Next, the BBA on the ground semantics  $\Omega_g$  after applying temporal filtering is compared to the BBA in the fused grid map. Note that the latter coincides with the BBA obtained from stereo measurements only as no evidences for the ground hypotheses are obtained from LiDAR in this work’s evaluation setup.

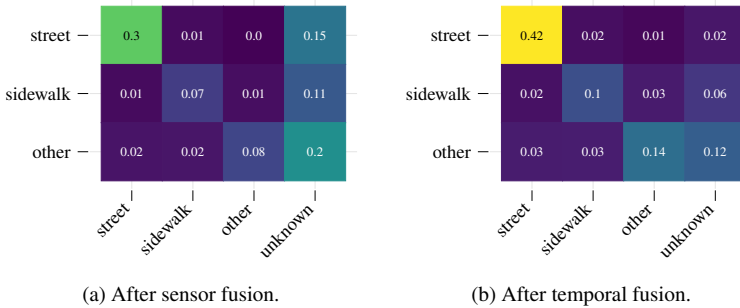


Figure 5.13: Confusion matrices for the ground state estimation. Rows correspond to the reference state and columns to the estimation.

The eIoUs for the ground semantics  $\Omega_g$  are depicted in Figure 5.12. Compared to the BBA in the fused grid map, the eIoU is slightly reduced for the hypotheses *street* and *sidewalk* and slightly improved for *other ground* after applying temporal fusion. The reason behind this is that the true positive rate can be improved by the temporal filter, however, the false positive rate is enhanced as well.

This becomes visible when considering the confusion matrices in Figure 5.13. Here, each entry  $e_{ij}$  contains the percentage of the BBA with reference label  $i$  assigned to the hypotheses  $j$ . Both matrices are normalized so that the sum of all entries is one. It can be seen that the relative false positive rate is higher for all three hypotheses after applying temporal fusion. At the same time however, the off-diagonal entries are higher as well indicating higher confusion between the individual hypotheses. Those numbers are increased as a significantly lower amount of the BBA is assigned to the hypothesis *unknown*. The overall amount of the BBA assigned to *unknown* is reduced from 0.46 to 0.2. This reduction of uncertainty can also be seen in Deng's entropy.

Figure 5.14 shows nonspecificity and discord for the ground semantics  $\Omega_g$  in the fused and the filtered grid map for both the  $360^\circ$  and the camera view. As opposed to the numbers for the occupancy semantics  $\Omega_s$ , both values are greatly reduced here after applying temporal filtering. Overall, Deng's entropy decreases by 39.4% in the  $360^\circ$  view and by 41.1% in the camera view.

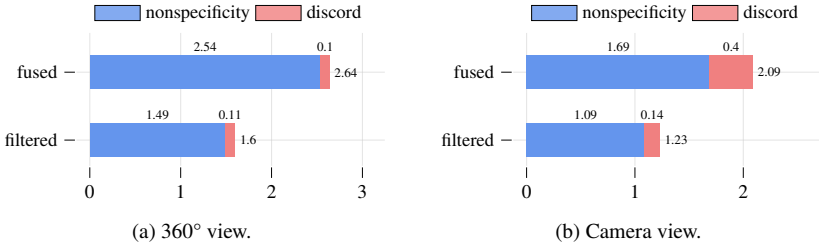


Figure 5.14: Nonspecificity and discord for fused and filtered grid maps on the ground semantics  $\Omega_g$ .

### 5.4.3 Dynamic State

Finally, we consider the BBA on the occupancy dynamics  $\Omega_d$ .

Figures 5.15a and 5.15b show visualizations of the BBA on the occupancy dynamics  $\Omega_d$  as well as the estimated cell velocities in the same traffic scene that is shown in Figure 5.9. It can be seen that the curb stones on the sides of the street behind the ego vehicle remain in the grid map as stationary occupancy. This is achieved due to the advanced conflict resolution in the evidential network using the data-driven parameterization. There are two cars driving on the oncoming lane and four cars on the ongoing lanes indicated by a high BBA for the hypothesis moving. The movement direction is correctly detected as indicated by the color coded grid cells in Figure 5.15b.

Next we compare the data-driven parameterization presented in Section 5.3.3 with the baseline parameterization stated in Table 5.2. Recall that the baseline parameterization is based on the manual conflict assignment proposed by Steyer et al. [STW18] whereas the data-driven approach incorporates the temporal progression of the ray permeability.

Figure 5.16 shows the results for the baseline parameterization compared to the proposed data-driven parameterization. Figure 5.16a, in the left column, shows the grid map containing the BBA on the occupancy dynamics  $\Omega_d$  and two magnified subareas for the baseline parameterization. In Figure 5.16b in the right column, the same results are shown for the proposed data-driven parameterization. It can be observed that *stationary* occupancy stemming from low obstacles such as curb stones cannot be tracked in the baseline configuration.

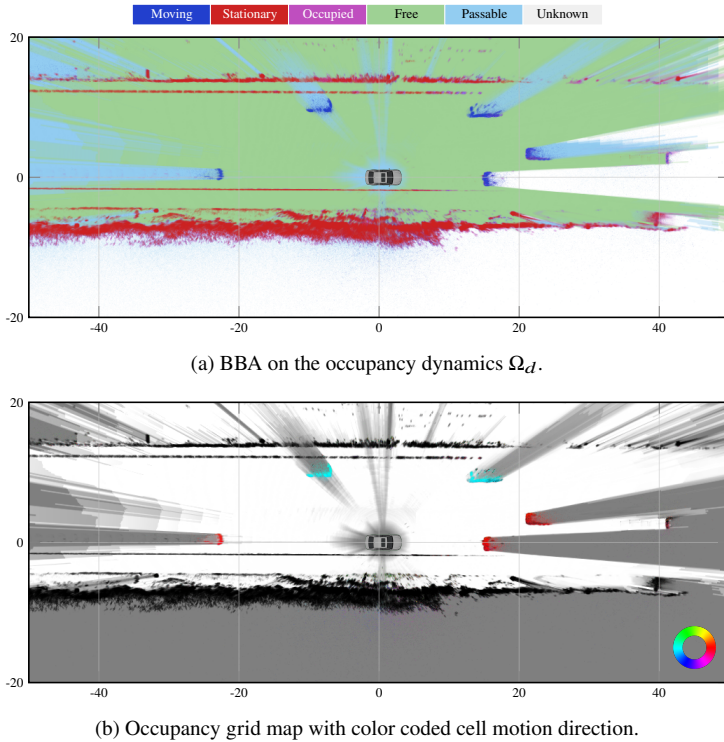


Figure 5.15: Visualizations of the estimated dynamic grid cell states on the Cartesian grid  $\mathcal{G}_{xy}$ .

This is due to the fact that the *occupied/free* conflicts in those cells can only be resolved correctly after adding the ray permeability to the evidential network. This is especially visible in the lower arm of the T-crossing which is highlighted in the middle column. It is crucial that the improved stationary occupancy detection for low obstacle is not at the expense of a slower detection of moving occupancy. In the right column, one of the passing cars that is moving from the left to the right is magnified. Both configurations show similar results in this example where moving occupancy is detected slightly better using the proposed parameterization. Furthermore, the proposed parameterization fully assigns the remaining occupancy masses behind the car on the left to the *free space* hypothesis. When using the baseline method, however, small amounts

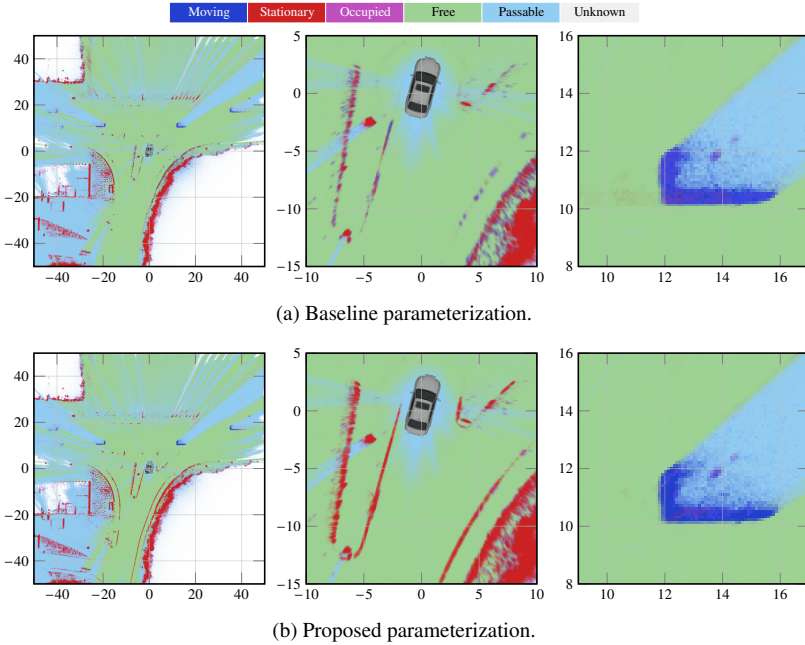
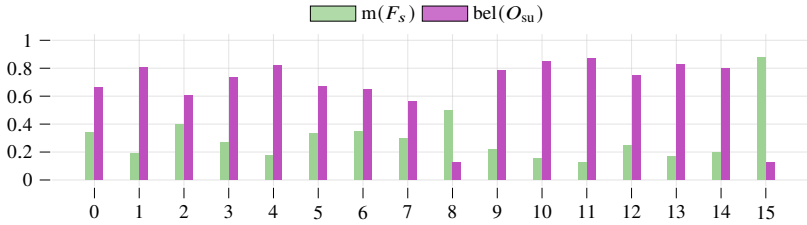


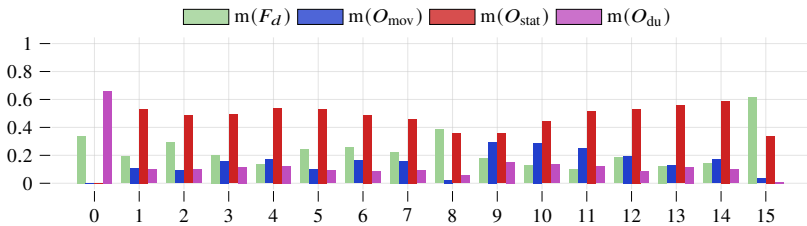
Figure 5.16: The occupancy grid map with color coded BBA on the occupancy dynamics  $\Omega_d$  using the baseline parameterization and our proposed parameterization.

of the BBA are still assigned to *occupied* in some grid cells to the left of the car. This is due to the parameter  $\lambda_{O_{\text{stat}} \rightarrow F_d}$  which regulates the amount of BBA assigned to *free*, if the cell was observed as stationary occupied and is observed free in the current measurement. Whereas in the baseline  $\lambda_{O_{\text{stat}} \rightarrow F_d}$  is set to 0.5, it is estimated based on the ray permeability progression in the proposed parameterization.

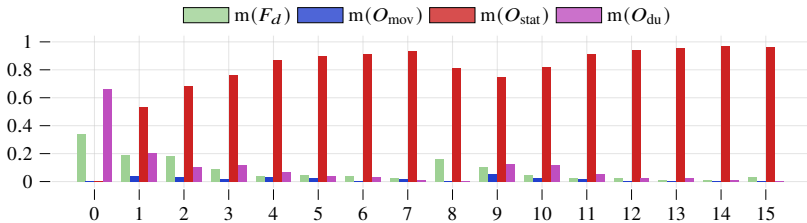
For a deeper investigation of the improved conflict resolution at low obstacles such as curb stones and guard rails, the BBA progression is plotted in Figure 5.17 over 16 frames for a fixed location in the grid. This fixed location is at the curb stone on the right side of the t-crossing from Figure 5.16. In Figure 5.17a, the BBA in the fused measurement grid map is plotted. The BBA assigns significant amounts to both the free space hypothesis  $F_s$  and the occupancy hypotheses  $O_{\text{su}}$  where the separation varies over time. The challenge is to correctly resolve



(a) The BBA in the fused measurement grid map.



(b) The BBA in the filtered grid map using the baseline parameterization.



(c) The BBA in the filtered grid map using the proposed data-driven parameterization.

Figure 5.17: The progression of the estimated BBA on the occupancy dynamics  $\Omega_d$  over 16 frames in a grid cell located on a curb stone. The BBA in the fused measurement grid map shown in (a) assigns significant amounts to both the free space hypothesis  $F_s$  and the occupancy hypotheses  $O_{su}$ . The baseline method depicted in (b) is not able to resolve this in favor of stationary occupancy  $O_{stat}$ . The proposed method on the other hand correctly assigns the majority of the BBA mass to stationary occupancy  $O_{stat}$ .

the resulting conflicts in the temporal fusion in favor of stationary occupancy. The baseline configuration, depicted in Figure 5.17b, generates a high discord between the hypotheses *free*  $F_d$ , moving occupancy  $O_{mov}$ , stationary occupancy  $O_{stat}$  and dynamically unclassified occupancy  $O_{du}$ . In contrast, the proposed configuration based on the progression of the ray permeability is able to robustly

resolve those conflicts in favor of stationary occupancy as demonstrated in the BBA plotted in Figure 5.17c.

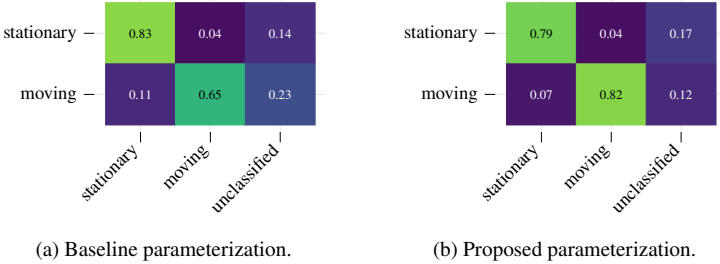


Figure 5.18: Confusion matrices for the dynamic state estimation in occupied cells comparing our proposed method with the baseline approach. Rows correspond to the reference state and columns to the estimation.

The visual impression that moving occupancy can be better detected using the proposed parameterization can be confirmed by the confusion matrix of all occupancy masses in the dynamic state estimation. The confusion matrices for the dynamic classification of occupancy are shown in Figure 5.18. The two rows are normalized and represent the reference states *stationary* and *moving* occupied. Each column corresponds to the estimated state and each entry on the diagonal shows the recall of the corresponding hypothesis. For *moving* occupancy the recall can be increased by 26.2%. This improvement can be achieved due to assigning measured occupancy in grid cells previously observed as *free* to *moving* occupancy based on the change of the ray permeability. The recall for *stationary* occupancy is slightly lower for the proposed parameterization. However, note that the reference grid map generated based on the labeled bounding primitives contained in the KITTI-360 dataset (cf. Section 2.4) does not contain occupancy BBA masses for curb stones. Hence, the improvements made in these areas demonstrated in Figure 5.17 are not included in the confusion matrices in Figure 5.18.



## 6 Conclusion

In this thesis, a generic evidential grid mapping framework for multi-sensor environment perception in the context of automated driving was presented. A novel hybrid evidential model was introduced in order to incorporate detailed semantic information on occupancy and ground level. Sensor models for cameras and LiDAR scanners were presented where spatial uncertainty is modeled on sensor-specific grids. The proposed sensor data fusion uses ER combination rules to explicitly model sensor credibility. Finally, an algorithm for temporal fusion was presented where the update step of the BBA is modeled in an evidential network. The resulting grid map representation contains evidence masses on the ground semantics, the occupancy semantics and the occupancy dynamics. The performance of the sensor measurement grid mapping, the sensor data fusion and the temporal fusion was evaluated in challenging traffic scenarios with measurements from a stereo camera and a LiDAR scanner.

### 6.1 Discussion

The whole grid mapping pipeline presented in this thesis including sensor measurement grid mapping, the sensor data fusion and temporal fusion was implemented in CUDA [NVF20] to utilize the massive parallelization capabilities of GPUs. Hence, the presented framework can be processed with an update frequency of at least 10 Hz on high performance GPUs such as the NVIDIA GeForce RTX 2080 Ti.

To the author's best knowledge, the proposed method is the first that estimates a BBA on occupancy semantics in top-view grid maps with the presented level of detail. The competitive sensor data fusion approach enables a redundant inclusion of the semantic estimates reducing it to a classical occupancy mapper in case no semantic estimates are available. Besides this unique characteristic, the sensor grid mapping is more efficient as no parametric ground model is

needed to differ between ground and obstacle detections. The temporal fusion shows advantages over competitive approaches in detecting the movement of entities and tracking low obstacles such as curb stones. As demonstrated in the experiments, both the sensor data fusion and the temporal fusion steps further show significant improvements in resolving evidential conflicts.

Note that the proposed framework generalizes traditional methods in two ways: First, the sensor data fusion based on ER reduces to Dempster's rule if both sensor reliabilities are set to one. Second, BBA combination rules used in past publications can be formulated as evidential networks. In evidential networks, however, more complex dependencies can be modeled which was utilized in this work.

The grid mapping with images, proposed in Section 3.3.2 as one of the main contributions, requires the measurement to be organized on a high-resolution sensor grid. If this requirement is not fulfilled, grid mapping with point sets, summarized in Section 3.3.1, may be applied. This is necessary for sensors providing sparse detections such as RaDAR sensors. When applying the proposed framework to camera data, the depth estimation quality is crucial. Although uncertainty in depth or disparity estimation is incorporated in the inverse sensor models, an appropriate quality level is required to infer reasonable velocities in the occupancy tracker. The same holds for the pixelwise semantic segmentation in the sensor grid. However, note that label confidences may be incorporated in the BBA estimation if provided by the neural network. Besides depth estimation for cameras and pixelwise semantic segmentation, sensor calibration and ego motion estimation heavily influence the quality of the mapping result.

As mentioned in the introduction of this thesis, the evidential multi-layer top view representation is meant to be an intermediate representation in the environment perception module and is subject to further abstraction. For most applications, an object detection and tracking module will be attached to represent traffic participants in a more compact and informative way. The grid map layers representing the BBA for the hypotheses *free* and *unknown* may form the basis for calculating confidence values for drivability and visibility in different regions of interest obtained from the navigation module. The evidential context makes this framework also well suited for self diagnostics. An increasing average entropy in the estimated BBA might give a hint on inaccurate sensor calibration or errors in the ego state estimated by the self

perception module. A comparison of the estimated semantics with a high resolution map could indicate localization errors or even facilitate localization by matching the observations with the corresponding information in the map.

## 6.2 Outlook

Depending on the sensor setup used, the quality of the estimated evidential grid map depends on the quality of pixel-wise semantic labeling, depth or disparity estimation. Therefore, the presented pipeline profits from future improvements in the respective Computer Vision tasks.

The main limiting factor of the described method is the assumption that the BBAs can be estimated cell-wise independently of neighboring grid cells. Especially for the temporal fusion, context information might be helpful for a better detection of moving entities. This could be mitigated by adding a dependency to the state in neighboring grid cells in the evidential network. However, this makes an explicit parameter estimation impracticable so that it is desirable to apply deep learning techniques with convolutional operators.

Last but not least, future research could address algorithms transforming the presented semantic evidential grid map representation to a more abstract and compact model. It would be interesting to evaluate how the object detection and tracking pipeline presented by Steyer in [Ste21] benefits from the BBA on the extended hypotheses set including semantic information. In this context, it is desirable not to overfit object detection and tracking to specific sensor setups so that the competitive fusion property can be retained.



## References

- [BB14] N. Ben Hariz and B. Ben Yaghlane. “Learning Parameters in Directed Evidential Networks with Conditional Belief Functions”. In: *Belief Functions: Theory and Applications*. Ed. by F. Cuzzolin. Cham: Springer International Publishing, 2014, pp. 294–303. ISBN: 978-3-319-11191-9 (cit. on p. 79).
- [Beh+19] J. Behley, M. Garbade, A. Milioto, J. Quenzel, S. Behnke, C. Stachniss, and J. Gall. “SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences”. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South), 2019, pp. 9296–9306. DOI: 10.1109/ICCV.2019.00939 (cit. on pp. 21, 24, 48).
- [BF07] H. Badino and U. Franke. *Free Space Computation Using Stochastic Occupancy Grids and Dynamic Programming*. Tech. rep. 2007, pp. 1–12. URL: <http://vision.jhu.edu/iccv2007-wdv/WDV07-badino.pdf> (visited on 04/19/2022) (cit. on p. 31).
- [Bie+20] F. Bieder, S. Wirges, J. Janosovits, S. Richter, Z. Wang, and C. Stiller. “Exploiting Multi-Layer Grid Maps for Surround-View Semantic Segmentation of Sparse LiDAR Data”. In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. Las Vegas, NV, USA, 2020, pp. 1892–1898. DOI: 10.1109/IV47402.2020.9304848 (cit. on p. 15).
- [Bon08] J. Bongard. “Probabilistic Robotics. Sebastian Thrun, Wolfram Burgard, and Dieter Fox. (2005, MIT Press.) 647 pages”. In: *Artificial Life* 14.2 (2008), pp. 227–229. DOI: 10.1162/artl.2008.14.2.227 (cit. on p. 62).
- [BS18] J. Beck and C. Stiller. “Generalized B-spline Camera Model”. In: *2018 IEEE Intelligent Vehicles Symposium (IV)*. Changshu, China, 2018, pp. 2137–2142. DOI: 10.1109/IVS.2018.8500466 (cit. on p. 2).

- [BVF15] G. Bernardes Vitor, A. C. Victorino, and J. V. Ferreira. “Stereo Vision for Dynamic Urban Environment Perception Using Semantic Context in Evidential Grid”. In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Gran Canaria, Spain, 2015, pp. 2471–2476. DOI: 10.1109/ITSC.2015.398 (cit. on pp. 15, 31).
- [Che+20a] K. Chen, R. Oldja, N. Smolyanskiy, S. Birchfield, A. Popov, D. Wehr, I. Eden, and J. Pehserl. “MVLidarNet: Real-Time Multi-Class Scene Understanding for Autonomous Driving Using Multiple Views”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, NV, USA, 2020, pp. 2288–2294. DOI: 10.1109/IROS45743.2020.9341450 (cit. on p. 15).
- [Che+20b] X. Cheng, Y. Zhong, M. Harandi, Y. Dai, X. Chang, H. Li, T. Drummond, and Z. Ge. “Hierarchical neural architecture search for deep stereo matching”. In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 22158–22169 (cit. on p. 29).
- [Dem67] A. P. Dempster. “Upper and Lower Probabilities Induced by a Multivalued Mapping”. In: *The Annals of Mathematical Statistics* 38.2 (1967), pp. 325–339. DOI: 10.1214/aoms/1177698950 (cit. on p. 62).
- [Den20] Y. Deng. “Uncertainty measure in evidence theory”. In: *Science China Information Sciences* 63.11 (2020), p. 210201. DOI: 10.1007/s11432-020-3006-9 (cit. on p. 11).
- [DP88] D. Dubois and H. Prade. “Representation and combination of uncertainty with belief functions and possibility measures”. In: *Computational Intelligence* 4.3 (1988), pp. 244–264. DOI: <https://doi.org/10.1111/j.1467-8640.1988.tb00279.x> (cit. on p. 64).
- [DRN14] K. C. J. Dietmayer, S. Reuter, and D. Nuss. “Representation of Fused Environment Data”. In: *Handbook of Driver Assistance Systems: Basic Information, Components and Systems for Active Safety and Comfort*. Ed. by H. Winner, S. Hakuli, F. Lotz, and C. Singer. Cham: Springer International Publishing, 2014, pp. 1–30. ISBN: 978-3-319-09840-1. DOI: 10.1007/978-3-319-09840-1\_25-1 (cit. on pp. 8, 79).

- 
- [DS76] A. Dempster and G. Shafer. *A Mathematical Theory of Evidence*. Limited paperback editions. Princeton, NJ, USA: Princeton University Press, 1976. ISBN: 9780691100425 (cit. on p. 9).
- [Dur90] H. F. Durrant-Whyte. “Sensor Models and Multisensor Integration”. In: *Autonomous Robot Vehicles*. Ed. by I. J. Cox and G. T. Wilfong. New York, NY: Springer New York, 1990, pp. 73–89. ISBN: 978-1-4613-8997-2. DOI: 10.1007/978-1-4613-8997-2\_7 (cit. on pp. 3, 4).
- [EHZ18] H. Eraqi, J. Honer, and S. Zuther. “Static Free Space Detection with Laser Scanner using Occupancy Grid Maps”. In: *Conference Record - IEEE Conference on Intelligent Transportation Systems* (Jan. 2018) (cit. on p. 13).
- [Elf89] A. Elfes. “Using Occupancy Grids for Mobile Robot Perception and Navigation”. In: *Computer* 22.6 (1989), pp. 46–57. DOI: 10.1109/2.30720 (cit. on pp. 14, 30).
- [FBH18] P. Fankhauser, M. Bloesch, and M. Hutter. “Probabilistic Terrain Mapping for Mobile Robots with Uncertain Localization”. In: *IEEE Robotics and Automation Letters* 3.4 (2018), pp. 3019–3026. DOI: 10.1109/LRA.2018.2849506 (cit. on p. 15).
- [Fei+21] J. Fei, K. Peng, P. Heidenreich, F. Bieder, and C. Stiller. “PillarSeg-Net: Pillar-based Semantic Grid Map Estimation using Sparse LiDAR Data”. In: *2021 IEEE Intelligent Vehicles Symposium (IV)*. Nagoya, Japan, 2021, pp. 838–844. DOI: 10.1109/IV48863.2021.9575694 (cit. on p. 15).
- [FHT10] J. H. Friedman, T. Hastie, and R. Tibshirani. “Regularization Paths for Generalized Linear Models via Coordinate Descent”. In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22. DOI: 10.18637/jss.v033.i01 (cit. on p. 104).
- [GLU12] A. Geiger, P. Lenz, and R. Urtasun. “Are we ready for autonomous driving? The KITTI vision benchmark suite”. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA, 2012, pp. 3354–3361. DOI: 10.1109/CVPR.2012.6248074 (cit. on pp. 24, 48).

- [Hir08] H. Hirschmuller. “Stereo Processing by Semiglobal Matching and Mutual Information”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2 (2008), pp. 328–341. doi: 10.1109/TPAMI.2007.1166 (cit. on p. 29).
- [Hom+10] 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: *2010 IEEE Intelligent Vehicles Symposium*. La Jolla, CA, USA, 2010, pp. 1006–1013. doi: 10.1109/IVS.2010.5548091 (cit. on p. 30).
- [Hon+21] Y. Hong, H. Pan, W. Sun, and Y. Jia. *Deep Dual-resolution Networks for Real-time and Accurate Semantic Segmentation of Road Scenes*. 2021. doi: 10.48550/ARXIV.2101.06085. URL: <https://arxiv.org/abs/2101.06085> (visited on 04/19/2022) (cit. on p. 50).
- [HRL15] H. Harms, E. Rehder, and M. Lauer. “Grid map based free space estimation using stereo vision”. In: *1st Workshop on Environment Perception for Automated On-road Vehicles, IEEE Intelligent Vehicles Symposium, 2015*. IEEE. Seoul, Korea (South), 2015 (cit. on p. 15).
- [KN09] T. Koski and J. Noble. *Bayesian Networks: An Introduction*. Wiley Series in Probability and Statistics. John Wiley & Sons, Sept. 2009. ISBN: 9780470743041. doi: 10.1002/9780470684023 (cit. on p. 78).
- [Küm20] J. V. Kümmerle. “Multimodal Sensor Calibration with a Spherical Calibration Target”. Doctoral Dissertation. Karlsruher Institut für Technologie (KIT), 2020. 185 pp. doi: 10.5445/IR/1000124721 (cit. on p. 2).
- [Lee+20] K.-H. Lee, M. Kliemann, A. Gaidon, J. Li, C. Fang, S. Pillai, and W. Burgard. “PillarFlow: End-to-end Birds-eye-view Flow Estimation for Autonomous Driving”. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Las Vegas, NV, USA, 2020, pp. 2007–2013. doi: 10.1109/IROS45743.2020.9340931 (cit. on p. 81).



- 
- [LLZ20] J. Li, H. Li, and H. Zeng. “SEGM: A Novel Semantic Evidential Grid Map by Fusing Multiple Sensors”. In: *Pattern Recognition and Computer Vision: Third Chinese Conference, PRCV 2020, Nanjing, China, October 16–18, 2020, Proceedings, Part I*. Nanjing, China: Springer-Verlag, 2020, pp. 155–166. ISBN: 978-3-030-60632-9. DOI: 10.1007/978-3-030-60633-6\_13 (cit. on p. 66).
- [LXG21] Y. Liao, J. Xie, and A. Geiger. *KITTI-360: A Novel Dataset and Benchmarks for Urban Scene Understanding in 2D and 3D*. 2021. DOI: 10.48550/ARXIV.2109.13410. URL: <https://arxiv.org/abs/2109.13410> (visited on 04/19/2022) (cit. on p. 21, 67).
- [Mah14] R. Mahler. *Advances in Statistical Multisource-Multitarget Information Fusion*. Electronic Warfare. Artech House, 2014. ISBN: 9781608077984 (cit. on p. 76).
- [MDP15] J. Moras, J. Dezert, and B. Pannetier. “Grid occupancy estimation for environment perception based on belief functions and PCR6”. In: *Signal Processing, Sensor/Information Fusion, and Target Recognition XXIV*. Ed. by I. Kadar. Vol. 9474. International Society for Optics and Photonics. Baltimore, MD, United States: SPIE, 2015, pp. 208–220. DOI: 10.1117/12.2177653 (cit. on p. 66).
- [ME85] H. Moravec and A. Elfes. “High resolution maps from wide angle sonar”. In: *Proceedings. 1985 IEEE International Conference on Robotics and Automation*. Vol. 2. St. Louis, MO, USA, 1985, pp. 116–121. DOI: 10.1109/ROBOT.1985.1087316 (cit. on p. 14).
- [MM15] R. Matthaei and M. Maurer. “Autonomous driving – a top-down-approach”. In: *at - Automatisierungstechnik* 63.3 (2015), pp. 155–167. DOI: doi:10.1515/auto-2014-1136 (cit. on p. 1).
- [Mor89] H. P. Moravec. “Sensor Fusion in Certainty Grids for Mobile Robots”. In: *Sensor Devices and Systems for Robotics* 9.2 (1989), pp. 253–276. DOI: 10.1007/978-3-642-74567-6\_19 (cit. on p. 14).

- [MR11] C. R. Michael Bleyer and C. Rother. “PatchMatch Stereo - Stereo Matching with Slanted Support Windows”. In: *Proceedings of the British Machine Vision Conference*. Dundee, United Kingdom: BMVA Press, 2011, pp. 14.1–14.11. ISBN: 1-901725-43-X. DOI: <http://dx.doi.org/10.5244/C.25.14> (cit. on p. 29).
- [Nau20] M. Naumann. “Probabilistic Motion Planning for Automated Vehicles”. Doctoral Dissertation. Karlsruher Institut für Technologie (KIT), 2020. 161 pp. DOI: [10.5445/IR/1000123725](https://doi.org/10.5445/IR/1000123725) (cit. on p. 3).
- [New+11] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon. “KinectFusion: Real-time dense surface mapping and tracking”. In: *2011 10th IEEE International Symposium on Mixed and Augmented Reality*. Basel, Switzerland, 2011, pp. 127–136. DOI: [10.1109/ISMAR.2011.6092378](https://doi.org/10.1109/ISMAR.2011.6092378) (cit. on p. 38).
- [Nus+14] 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: *17th International Conference on Information Fusion (FUSION)*. Salamanca, Spain, 2014, pp. 1–8 (cit. on p. 65).
- [Nus+16] 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”. In: *The International Journal of Robotics Research* 37 (May 2016). DOI: [10.1177/0278364918775523](https://doi.org/10.1177/0278364918775523) (cit. on pp. 8, 80, 84–86, 89).
- [Nus17] D. Nuss. “A random finite set approach for dynamic occupancy grid maps”. Doctoral Dissertation. Universität Ulm, 2017. DOI: [10.18725/OPARU-4361](https://doi.org/10.18725/OPARU-4361) (cit. on p. 63).
- [NVF20] NVIDIA, P. Vingelmann, and F. H. Fitzek. *CUDA, release: 10.2.89*. 2020. URL: <https://developer.nvidia.com/cuda-toolkit> (visited on 04/19/2022) (cit. on p. 119).
- [On-21] On-Road Automated Driving (ORAD) committee. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. SAE International, Apr. 2021. DOI:

- 
- 10.4271/J3016\_202104. URL: [https://doi.org/10.4271/J3016\\_202104](https://doi.org/10.4271/J3016_202104) (visited on 04/27/2022) (cit. on p. 1).
- [Ped+11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on p. 104).
- [Por20] J. Porębski. “Customizable Inverse Sensor Model for Bayesian and Dempster-Shafer Occupancy Grid Frameworks”. In: *Advanced, Contemporary Control*. Ed. by A. Bartoszewicz, J. Kabziński, and J. Kacprzyk. Cham: Springer International Publishing, 2020, pp. 1225–1236. ISBN: 978-3-030-50936-1 (cit. on p. 30).
- [Qia+21] S. Qiao, Y. Zhu, H. Adam, A. Yuille, and L.-C. Chen. “ViP-DeepLab: Learning Visual Perception with Depth-aware Video Panoptic Segmentation”. In: *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA, 2021, pp. 3996–4007. DOI: 10.1109/CVPR46437.2021.00399 (cit. on p. 29).
- [RB19] S. Royo and M. Ballesta-Garcia. “An Overview of Lidar Imaging Systems for Autonomous Vehicles”. In: *Applied Sciences* 9.19 (2019). ISSN: 2076-3417. DOI: 10.3390/app9194093 (cit. on p. 27).
- [Ric+19] S. Richter, S. Wirges, H. Königshof, and C. Stiller. “Fusion of Range Measurements and Semantic Estimates in an Evidential Framework”. In: *tm - Technisches Messen* 86.s1 (2019), pp. 102–106. DOI: 10.1515/teme-2019-0052 (cit. on pp. 15, 31).
- [Ric+20] S. Richter, J. Beck, S. Wirges, and C. Stiller. “Semantic Evidential Grid Mapping Based on Stereo Vision”. In: *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. Karlsruhe, Germany, 2020, pp. 179–184. DOI: 10.1109/MFI49285.2020.9235217 (cit. on pp. 16, 31).
- [Ric+21] S. Richter, Y. Wang, J. Beck, S. Wirges, and C. Stiller. “Semantic Evidential Grid Mapping Using Monocular and Stereo Cameras”. In: *Sensors* 21.10 (2021). ISSN: 1424-8220. DOI: 10.3390/s21103380 (cit. on pp. 16, 31).

- [Ric+22a] S. Richter, F. Bieder, S. Wirges, and C. Stiller. *Mapping LiDAR and Camera Measurements in a Dual Top-View Grid Representation Tailored for Automated Vehicles*. 2022. DOI: 10.48550/ARXIV.2204.07887. URL: <https://arxiv.org/abs/2204.07887> (visited on 04/21/2022) (cit. on pp. 9, 27).
- [Ric+22b] S. Richter, F. Bieder, S. Wirges, and C. Stiller. *Sensor Data Fusion in Top-View Grid Maps using Evidential Reasoning with Advanced Conflict Resolution*. 2022. DOI: 10.48550/ARXIV.2204.08780. URL: <https://arxiv.org/abs/2204.08780> (visited on 04/21/2022) (cit. on pp. 9, 61).
- [RNL15] L. Rummelhard, A. Nègre, and C. Laugier. “Conditional Monte Carlo Dense Occupancy Tracker”. In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems*. Gran Canaria, Spain, 2015, pp. 2485–2490. DOI: 10.1109/ITSC.2015.400 (cit. on p. 79).
- [Rum+17] L. Rummelhard, A. Paigwar, A. Nègre, and C. Laugier. “Ground estimation and point cloud segmentation using SpatioTemporal Conditional Random Field”. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. Los Angeles, CA, USA, 2017, pp. 1105–1110. DOI: 10.1109/IVS.2017.7995861 (cit. on pp. 8, 15).
- [Sch18] M. Schreier. “Environment representations for automated on-road vehicles”. In: *at - Automatisierungstechnik* 66.2 (2018), pp. 107–118. DOI: doi:10.1515/auto-2017-0104 (cit. on pp. 2, 3).
- [SD06] F. Smarandache and J. Dezert. “Proportional conflict redistribution rules for information fusion”. In: *Advances and Applications of DSMT for Information Fusion (Collected Works) 2* (2006), pp. 3–68 (cit. on p. 64).
- [Sme90] P. Smets. “The combination of evidence in the transferable belief model”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12.5 (1990), pp. 447–458. DOI: 10.1109/34.55104 (cit. on p. 11).
- [Son20] M. Sons. “Automatische Erzeugung langzeitverfügbarer Punktmehrkarten zur robusten Lokalisierung mit Multi-Kamera-Systemen für automatisierte Fahrzeuge”. German. Doctoral Dissertation. Karlsruher Institut für Technologie (KIT), 2020. 156 pp. DOI: 10.5445/IR/1000117845 (cit. on p. 2).

- 
- [SSM87] G. Shafer, P. P. Shenoy, and K. Mellouli. “Propagating belief functions in qualitative Markov trees”. In: *Int. J. Approx. Reason.* 1 (1987), pp. 349–400 (cit. on p. 78).
- [Ste+20] S. Steyer, C. Lenk, D. Kellner, G. Tanzmeister, and D. Wollherr. “Grid-Based Object Tracking With Nonlinear Dynamic State and Shape Estimation”. In: *IEEE Transactions on Intelligent Transportation Systems* 21.7 (2020), pp. 2874–2893. doi: 10.1109/TITS.2019.2921248 (cit. on p. 6).
- [Ste21] S. J. Steyer. “Grid-based object tracking”. Doctoral Dissertation. München: Technische Universität München, 2021 (cit. on pp. 6, 121).
- [Ste22] T. Stewart. “Overview of motor vehicle crashes in 2020”. In: *National Highway Traffic Safety Administration March* (2022) (cit. on p. 1).
- [Sti+17] K. Stiens, J. Keilhacker, G. Tanzmeister, and D. Wollherr. “Local elevation mapping for automated vehicles using lidar ray geometry and particle filters”. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. Los Angeles, CA, USA, 2017, pp. 481–486. doi: 10.1109/IVS.2017.7995764 (cit. on p. 15).
- [STW17] S. Steyer, G. Tanzmeister, and D. Wollherr. “Object tracking based on evidential dynamic occupancy grids in urban environments”. In: *2017 IEEE Intelligent Vehicles Symposium (IV)*. Los Angeles, CA, USA, 2017, pp. 1064–1070. doi: 10.1109/IVS.2017.7995855 (cit. on p. 6).
- [STW18] S. Steyer, G. Tanzmeister, and D. Wollherr. “Grid-based Environment Estimation Using Evidential Mapping and Particle Tracking”. In: *IEEE Transactions on Intelligent Vehicles* 3 (3 2018), pp. 384–396. doi: 10.1109/TIV.2018.2843130 (cit. on pp. 8, 19, 80, 84, 85, 99, 100, 114).
- [Tho+19] J. Thomas, J. Tatsch, W. van Ekeren, R. Rojas, and A. Knoll. “Semantic Grid-Based Road Model Estimation for Autonomous Driving”. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. Paris, France, 2019, pp. 2329–2336. doi: 10.1109/IVS.2019.8813790 (cit. on p. 31).

- [TW17] G. Tanzmeister and D. Wollherr. “Evidential Grid-Based Tracking and Mapping”. In: *IEEE Transactions on Intelligent Transportation Systems* 18.6 (2017), pp. 1454–1467. DOI: 10.1109/TITS.2016.2608919 (cit. on pp. 65, 80).
- [UYH21] I. Ullah, J. Youn, and Y.-H. Han. “Multisensor Data Fusion Based on Modified Belief Entropy in Dempster–Shafer Theory for Smart Environment”. In: *IEEE Access* 9 (2021), pp. 37813–37822. DOI: 10.1109/ACCESS.2021.3063242 (cit. on p. 66).
- [Van+21] R. Van Kempen, B. Lampe, T. Woopen, and L. Eckstein. “A Simulation-based End-to-End Learning Framework for Evidential Occupancy Grid Mapping”. In: *2021 IEEE Intelligent Vehicles Symposium (IV)*. Nagoya, Japan, 2021, pp. 934–939. DOI: 10.1109/IV48863.2021.9575715 (cit. on p. 30).
- [Vat+20] A. Vatavu, M. Rahm, S. Govindachar, G. Krehl, A. Mantha, S. R. Bhavsar, M. R. Schier, J. Peukert, and M. Maile. “From Particles to Self-Localizing Tracklets: A Multilayer Particle Filter-Based Estimation for Dynamic Grid Maps”. In: *IEEE Intelligent Transportation Systems Magazine* 12.4 (2020), pp. 149–168. DOI: 10.1109/MITS.2020.3014428 (cit. on pp. 80, 81).
- [Vel19] Velodyne Lidar, Inc. *Velodyne VLP-16 User Manual*. Ed. by Velodyne Lidar, Inc. VLP-16. Rev. E. 2019. URL: <https://velodynelidar.com/wp-content/uploads/2019/12/63-9243-Rev-E-VLP-16-User-Manual.pdf> (visited on 04/07/2022) (cit. on p. 28).
- [VJF18] M. Valente, C. Joly, and A. de la Fortelle. “Fusing Laser Scanner and Stereo Camera in Evidential Grid Maps”. In: *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. Singapore, Singapore, 2018, pp. 990–997. DOI: 10.1109/ICARCV.2018.8580635 (cit. on p. 31).
- [Wir+18] S. Wirges, T. Fischer, C. Stiller, and J. B. Frias. “Object Detection and Classification in Occupancy Grid Maps Using Deep Convolutional Networks”. In: *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. Maui, HI, USA, 2018, pp. 3530–3535. DOI: 10.1109/ITSC.2018.8569433 (cit. on p. 6).

- 
- [Wir+19a] S. Wirges, J. Gräter, Q. Zhang, and C. Stiller. “Self-Supervised Flow Estimation using Geometric Regularization with Applications to Camera Image and Grid Map Sequences”. In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand, 2019, pp. 1782–1787. DOI: 10.1109/ITSC.2019.8916989 (cit. on p. 81).
- [Wir+19b] S. Wirges, M. Reith-Braun, M. Lauer, and C. Stiller. “Capturing Object Detection Uncertainty in Multi-Layer Grid Maps”. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*. Paris, France, 2019, pp. 1520–1526. DOI: 10.1109/IVS.2019.8814073 (cit. on p. 6).
- [Wir+20] S. Wirges, Y. Yang, S. Richter, H. Hu, and C. Stiller. “Learned Enrichment of Top-View Grid Maps Improves Object Detection”. In: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. Rhodes, Greece, 2020, pp. 1–6. DOI: 10.1109/ITSC45102.2020.9294330 (cit. on p. 6).
- [Wir+21] S. Wirges, K. Rösch, F. Bieder, and C. Stiller. “Fast and Robust Ground Surface Estimation from LiDAR Measurements using Uniform B-Splines”. In: *2021 IEEE 24th International Conference on Information Fusion (FUSION)*. Sun City, South Africa, 2021, pp. 1–7 (cit. on pp. 15, 18, 35, 41, 54, 59).
- [XS96] H. Xu and P. Smets. “Reasoning in evidential networks with conditional belief functions”. In: *International Journal of Approximate Reasoning* 14.2 (1996), pp. 155–185. ISSN: 0888-613X. DOI: [https://doi.org/10.1016/0888-613X\(96\)00113-2](https://doi.org/10.1016/0888-613X(96)00113-2) (cit. on pp. 11, 78).
- [Xu+19] H. Xu, G. Lan, S. Wu, and Q. Hao. “Online Intelligent Calibration of Cameras and LiDARs for Autonomous Driving Systems”. In: *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*. Auckland, New Zealand, 2019, pp. 3913–3920. DOI: 10.1109/ITSC.2019.8916872 (cit. on p. 2).
- [YA06] T. Yang and V. Aitken. “Evidential Mapping for Mobile Robots with Range Sensors”. In: *IEEE Transactions on Instrumentation and Measurement* 55 (4 2006), pp. 1422–1429. DOI: 10.1109/TIM.2006.876399 (cit. on p. 14).

- [Yag08] R. R. Yager. “Entropy and Specificity in a Mathematical Theory of Evidence”. In: *Classic Works of the Dempster-Shafer Theory of Belief Functions*. Ed. by R. R. Yager and L. Liu. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 291–310. ISBN: 978-3-540-44792-4. DOI: 10.1007/978-3-540-44792-4\_11 (cit. on p. 11).
- [Yag87] R. R. Yager. “On the dempster-shafer framework and new combination rules”. In: *Information Sciences* 41.2 (1987), pp. 93–137. ISSN: 0020-0255. DOI: [https://doi.org/10.1016/0020-0255\(87\)90007-7](https://doi.org/10.1016/0020-0255(87)90007-7) (cit. on p. 63).
- [YAL08] M. Yguel, O. Aycard, and C. Laugier. “Efficient GPU-based construction of occupancy grids using several laser range-finders”. In: *International Journal of Vehicle Autonomous Systems* 6.1-2 (2008), pp. 48–83. DOI: 10.1504/IJVAS.2008.016478 (cit. on p. 30).
- [YCB14] C. Yu, V. Cherfaoui, and P. Bonnifait. “An evidential sensor model for Velodyne scan grids”. In: *2014 13th International Conference on Control Automation Robotics Vision (ICARCV)*. Singapore, Singapore, 2014, pp. 583–588. DOI: 10.1109/ICARCV.2014.7064369 (cit. on pp. 15, 30).
- [YCB15] C. Yu, V. Cherfaoui, and P. Bonnifait. “Evidential occupancy grid mapping with stereo-vision”. In: *2015 IEEE Intelligent Vehicles Symposium (IV)*. Seoul, Korea (South), 2015, pp. 712–717. DOI: 10.1109/IVS.2015.7225768 (cit. on p. 31).
- [Yi+00] Z. Yi, Y. K. Ho, C. S. Chua, and X. W. Zhou. “Multi-Ultrasonic Sensor Fusion for Autonomous Mobile Robots”. In: *Sensor Fusion: Architectures, Algorithms, and Applications IV*. Ed. by B. V. Dasarathy. Vol. 4051. IV. Orlando, FL, USA, 2000, pp. 314–321 (cit. on p. 14).
- [YM08] B. B. Yaghlane and K. Mellouli. “Inference in directed evidential networks based on the transferable belief model”. In: *International Journal of Approximate Reasoning* 48.2 (2008). In Memory of Philippe Smets (1938–2005), pp. 399–418. ISSN: 0888-613X. DOI: <https://doi.org/10.1016/j.ijar.2008.01.002> (cit. on p. 79).



- 
- [YSM03] B. B. Yaghlane, P. Smets, and K. Mellouli. “Directed Evidential Networks with Conditional Belief Functions”. In: *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Ed. by T. D. Nielsen and N. L. Zhang. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 291–305. ISBN: 978-3-540-45062-7 (cit. on p. 79).
- [Yua+22] W. Yuan, X. Gu, Z. Dai, S. Zhu, and P. Tan. *NeW CRFs: Neural Window Fully-connected CRFs for Monocular Depth Estimation*. 2022. DOI: 10.48550/ARXIV.2203.01502. URL: <https://arxiv.org/abs/2203.01502> (visited on 04/19/2022) (cit. on p. 29).
- [YX13] J.-B. Yang and D.-L. Xu. “Evidential reasoning rule for evidence combination”. In: *Artificial Intelligence* 205 (2013), pp. 1–29. ISSN: 0004-3702. DOI: <https://doi.org/10.1016/j.artint.2013.09.003> (cit. on pp. 64–66).
- [Zad79] L. Zadeh. *On the Validity of Dempster’s Rule of Combination of Evidence*. Tech. rep. UCB/ERL M79/24. EECS Department, University of California, Berkeley, Mar. 1979. URL: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/1979/28427.html> (visited on 04/27/2022) (cit. on p. 63).
- [Zha+19] F. Zhang, V. Prisacariu, R. Yang, and P. H. S. Torr. “GA-Net: Guided Aggregation Net for End-To-End Stereo Matching”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA, 2019, pp. 185–194. DOI: 10.1109/CVPR.2019.00027 (cit. on pp. 29, 50).
- [Zhu+19] Y. Zhu, K. Sapra, F. A. Reda, K. J. Shih, S. Newsam, A. Tao, and B. Catanzaro. “Improving Semantic Segmentation via Video Propagation and Label Relaxation”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA, 2019, pp. 8848–8857. DOI: 10.1109/CVPR.2019.00906 (cit. on p. 50).