

Johannes Pallauf

# Objektsensitive Verfolgung und Klassifikation von Fußgängern mit verteilten Multi-Sensor-Trägern



Johannes Pallauf

**Objektsensitive Verfolgung und Klassifikation  
von Fußgängern mit verteilten Multi-Sensor-Trägern**

**Forschungsberichte aus der Industriellen Informationstechnik**  
**Band 11**

Institut für Industrielle Informationstechnik  
Karlsruher Institut für Technologie  
Hrsg. Prof. Dr.-Ing. Fernando Puente León  
Prof. Dr.-Ing. habil. Klaus Dostert

Eine Übersicht aller bisher in dieser Schriftenreihe erschienenen Bände  
finden Sie am Ende des Buchs.

# **Objektsensitive Verfolgung und Klassifikation von Fußgängern mit verteilten Multi-Sensor-Trägern**

von  
Johannes Pallauf

Dissertation, Karlsruher Institut für Technologie (KIT)  
Fakultät für Elektrotechnik und Informationstechnik  
Tag der mündlichen Prüfung: 29. April 2016  
Referenten: Prof. Dr.-Ing. Fernando Puente León,  
Prof. Dr.-Ing. Christoph Stiller

#### Impressum



Karlsruher Institut für Technologie (KIT)  
KIT Scientific Publishing  
Straße am Forum 2  
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark of Karlsruhe  
Institute of Technology. Reprint using the book cover is not allowed.

[www.ksp.kit.edu](http://www.ksp.kit.edu)



*This document – excluding the cover, pictures and graphs – is licensed  
under the Creative Commons Attribution-Share Alike 3.0 DE License  
(CC BY-SA 3.0 DE): <http://creativecommons.org/licenses/by-sa/3.0/de/>*



*The cover page is licensed under the Creative Commons  
Attribution-No Derivatives 3.0 DE License (CC BY-ND 3.0 DE):  
<http://creativecommons.org/licenses/by-nd/3.0/de/>*

Print on Demand 2016

ISSN 2190-6629

ISBN 978-3-7315-0529-7

DOI 10.5445/KSP/1000054659





# **Objektsensitive Verfolgung und Klassifikation von Fußgängern mit verteilten Multi-Sensor-Trägern**

Zur Erlangung des akademischen Grades eines

**DOKTOR-INGENIEURS**

von der Fakultät für

Elektrotechnik und Informationstechnik

des Karlsruher Instituts für Technologie (KIT)

genehmigte

**DISSERTATION**

von

**Dipl.-Ing. Johannes Josef Anton Mac Dougall Pallauf**

geb. in Gießen

Tag der mündl. Prüfung: 29. April 2016  
Hauptreferent: Prof. Dr.-Ing. Fernando Puente León, KIT  
Korreferent: Prof. Dr.-Ing. Christoph Stiller, KIT



# Vorwort

Diese Arbeit entstand während meiner Zeit als akademischer Mitarbeiter am Institut für Industrielle Informationstechnik (IIIT) am Karlsruher Institut für Technologie (KIT). Für die Ermöglichung dieser sehr lehrreichen und spannenden Jahre möchte ich zuallererst besonders meinem Doktorvater Prof. Dr.-Ing. Puente León danken, der zum einen den nötigen Freiraum für eine solche Arbeit gewährte, aber auch im Zweifel motivierend und unterstützend den Rücken stärkte. Prof. Dr.-Ing. Stiller möchte ich herzlich für die Übernahme des Korreferats danken.

Den von mir mitbetreuten Bachelor-, Master- und Diplomstudenten möchte ich für ihr Engagement und ihre Begeisterung bei der Unterstützung von Teilen der Arbeit danken.

Ein besonderer Dank gilt allen Mitarbeitern des Instituts und insbesondere meiner Gruppe, die durch die gute Zusammenarbeit und die netten Arbeitstage und Feierabende sowohl fachlich als auch im Privaten zum Gelingen dieser Arbeit beigetragen haben. Für die hilfreichen Korrekturhinweise möchte ich darüber hinaus meinen Kollegen Sebastian Bauer, Pilar Hernández Mesa, Thomas Nürnberg und Sebastian Vater danken.

Ein letzter Dank richtet sich an meine Eltern Carol und Josef, meine drei Schwestern Kathrin, Anna und Antonia sowie an meine Freundin Catharina, die mir alle stets zur Seite gestanden haben und den nicht fachlichen Teil meiner Betreuung vervollständigt haben.

Hildesheim, im Juli 2016

Johannes Pallauf



# Inhaltsverzeichnis

<b>Vorwort</b> . . . . .	<b>i</b>
<b>Symbolverzeichnis</b> . . . . .	<b>vii</b>
<b>1 Einleitung</b> . . . . .	<b>1</b>
1.1 Problemstellung . . . . .	2
1.2 Eigener Beitrag . . . . .	4
1.3 Gliederung . . . . .	5
<b>2 Grundlagen</b> . . . . .	<b>7</b>
2.1 Bewegungsmodelle . . . . .	7
2.1.1 Zufällige Bewegung . . . . .	8
2.1.2 Konstante Geschwindigkeit . . . . .	9
2.2 Messmodelle . . . . .	10
2.3 Ein-Objekt-Bayes-Filter . . . . .	10
2.3.1 Modellierung des Schätzproblems . . . . .	11
2.3.2 Bayes-Filter . . . . .	11
2.3.3 Kalman-Filter . . . . .	15
2.3.4 Partikelfilter . . . . .	19
2.4 Detektionsverfahren . . . . .	22
2.4.1 Bildbasierte Verfahren . . . . .	23
2.4.2 Laserbasierte Verfahren . . . . .	26
2.5 Klassifikation . . . . .	27
2.5.1 Support Vector Machine . . . . .	27
2.5.2 Platt Scaling . . . . .	30
2.6 Lokalisierung . . . . .	31
<b>3 Multi-Objekt-Filterung</b> . . . . .	<b>33</b>
3.1 Endliche Mengen . . . . .	34
3.2 Multi-Objekt-Modelle . . . . .	38

3.3	Multi-Objekt-Bayes-Filter . . . . .	40
3.4	<i>Probability Hypothesis Density</i> -Filter . . . . .	41
3.5	<i>Cardinalized Probability Hypothesis Density</i> -Filter . . . . .	44
3.6	<i>Cardinality Balanced Multi-Target Multi-Bernoulli</i> -Filter . . . . .	50
3.7	Bewertung . . . . .	53
<b>4</b>	<b>Adaptive Detektionswahrscheinlichkeit . . . . .</b>	<b>57</b>
4.1	Sichtfelder . . . . .	59
4.2	Verdeckung . . . . .	61
4.2.1	Statische Verdeckungen . . . . .	62
4.2.2	Dynamische Verdeckungen . . . . .	64
4.3	Abstandsabhängige Detektionswahrscheinlichkeit . . . . .	65
4.3.1	Erfassung . . . . .	66
4.3.2	Auswertung . . . . .	70
4.4	Resultierende Detektionswahrscheinlichkeit . . . . .	72
4.5	Anwendung auf Statistik endlicher Mengen . . . . .	72
4.5.1	PHD . . . . .	73
4.5.2	CPHD . . . . .	79
4.5.3	CBMeMber . . . . .	81
<b>5</b>	<b>Richtungssensitive Verfolgung . . . . .</b>	<b>83</b>
5.1	Kandidatensuche und Personendetektion . . . . .	87
5.2	Orientierungsmessung . . . . .	89
5.2.1	Experten-Klassifikatoren . . . . .	90
5.2.2	Lernstrategie . . . . .	92
5.2.3	Anwendung des HOG-Merkmals . . . . .	94
5.3	Fusion für die Verfolgung . . . . .	96
5.3.1	Messmodell . . . . .	98
5.3.2	Korrekturschritt und Repräsentation der Unsicherheit . . . . .	99
5.3.3	Rolle der Unsicherheit . . . . .	102
<b>6</b>	<b>Ergebnisse . . . . .</b>	<b>105</b>
6.1	Systemaufbau . . . . .	105
6.1.1	Sensorträger . . . . .	106
6.1.2	Informationsaustausch . . . . .	107
6.1.3	Eingesetzte Detektionsverfahren . . . . .	108

6.1.4	Fusion . . . . .	108
6.1.5	Implementierung . . . . .	108
6.2	Experimentelle Evaluierung . . . . .	109
6.2.1	Motion-Capture-Systeme . . . . .	110
6.2.2	Aufbau . . . . .	111
6.3	Evaluierung der adaptiven Detektionswahrscheinlichkeit . . . . .	112
6.3.1	Versuchsbeschreibung . . . . .	112
6.3.2	Simulation . . . . .	114
6.3.3	Realdaten . . . . .	119
6.3.4	Wahl der Detektionswahrscheinlichkeit pro Gauß-Komponente . . . . .	131
6.4	Evaluierung der richtungssensitiven Verfolgung . . . . .	133
6.4.1	Orientierungsmessung . . . . .	134
6.4.2	Fusion von Position und Orientierung im Multi-Objekt-Fall . . . . .	142
<b>7</b>	<b>Zusammenfassung . . . . .</b>	<b>151</b>
7.1	Adaptive Detektionswahrscheinlichkeit . . . . .	151
7.2	Richtungssensitive Verfolgung . . . . .	153
7.3	Ausblick . . . . .	154
<b>A</b>	<b>Bewegungsmodell konstanter polarer Geschwindigkeit . . . . .</b>	<b>157</b>
<b>B</b>	<b>Klassen endlicher Zufallsmengen . . . . .</b>	<b>159</b>
B.1	Endliche <i>IID-Cluster</i> -Zufallsmenge . . . . .	159
B.2	Endliche Poisson-Zufallsmenge . . . . .	160
B.3	Endliche Bernoulli-Zufallsmenge . . . . .	160
B.4	Endliche Multi-Bernoulli-Zufallsmenge . . . . .	160
<b>Literaturverzeichnis . . . . .</b>		<b>161</b>
	Eigene Veröffentlichungen . . . . .	176
	Betreute studentische Arbeiten . . . . .	177



# Symbolverzeichnis

## Allgemeine Abkürzungen

Abkürzung	Bedeutung
Ada-Boost	<i>Adaptive Boosting</i>
adapt.	adaptiv
CBMeMber	<i>Cardinality Balanced Multi-Target Multi-Bernoulli(-Filter)</i>
CPHD	<i>Cardinalized Probability Hypothesis Density</i>
Dist.	Distanz
dyn.	dynamisch
EAP	<i>Expected A Posteriori</i>
EKF	Erweitertes Kalman-Filter
FiStt	<i>Finite Set Statistics</i>
FoV	<i>Field of View</i>
GM	Gauß-Mixtur
HOG	<i>Histogram of Oriented Gradients</i>
iid	<i>independent identically distributed</i>
JIPDA	<i>Joint Integrated Probability Data Association</i>
KF	Kalman-Filter
LBP	<i>Local Binary Pattern</i>
LIDAR	<i>Light detection and ranging</i>
MAE	<i>Mean Absolute Error</i>
MAP	<i>Maximum-a-posteriori-Schätzer</i>
MC	Monte-Carlo
NCV	<i>Nearly Constant Velocity model</i>
Occ	<i>Occlusion</i>
OpenCV	<i>Open Computer Vision library</i>
Ori	Orientierung

<b>Abkürzung</b>	<b>Bedeutung</b>
OSPA	<i>Optimal Sub-Pattern Assignment-Metrik</i>
PCL	<i>Point Cloud Library</i>
PHD	<i>Probability Hypothesis Density</i>
RFS	<i>Random Finite Set</i>
RFS	<i>Random Finite Set</i>
RGB-D	<i>Rot, Grün, Blau und Tiefe (Depth)</i>
ROI	<i>Region Of Interest</i>
ROS	<i>Robot Operating System</i>
SDE	<i>Stochastic Differential Equation</i>
SIFT	<i>Scale-Invariant Feature Transform</i>
SIS	<i>Sequential Importance Sampling</i>
SLAM	<i>Simultaneous Localization And Mapping</i>
SMC	<i>sequentielle Monte-Carlo-Methoden</i>
stat.	<i>statisch</i>
STIP	<i>Space-Time Interest Points</i>
SURF	<i>Speeded Up Robust Features</i>
SVM	<i>Support Vector Machine</i>
UKF	<i>Unscented-Kalman-Filter</i>
UK	<i>Unscented-Kalman-Transformation</i>
W-LAN	<i>Wireless Local Area Network</i>

## Symbole

### Lateinische Buchstaben

<b>Symbol</b>	<b>Bedeutung</b>
$A$	<i>lineares Bewegungsmodell</i>
$a$	<i>Bewegungsmodell</i>
$B$	<i>Menge neugeborener Objekte</i>
$c$	<i>Cut-off-Parameter der OSPA-Metrik</i>
$D$	<i>Menge detektierter Objekte</i>
$D(\bullet)$	<i>Intensitätsfunktion (PHD)</i>
$d_p^{(c)}(\bullet)$	<i>Distanz der Ordnung <math>p</math> mit cut-off <math>c</math></i>
$d_{\text{jump}}$	<i>Jumping distance</i>

<b>Symbol</b>	<b>Bedeutung</b>
$f(\bullet)$	kontinuierliche Dichtefunktion
$f(h((\bullet)))$	Funktional
$g(\bullet)$	Likelihood-Funktion
$G[h(s)]$	wahrscheinlichkeitserzeugendes Funktional
$G(s)$	wahrscheinlichkeitserzeugende Funktion
<b>H</b>	lineares Messmodell
$h(\bullet)$	Messmodell
<b>I</b>	Kalman-Innovation
$J$	Anzahl an Gauß-Komponenten
$\mathcal{K}$	Menge an Falschalarman
<b>K</b>	Kalman-Verstärkung
$l^{(i)}$	Label des $i$ -ten Objekts
<b>m</b>	Mittelwertvektor
$M$	Anzahl der Messungen
$N$	Anzahl von Objekten oder Klassen
<b>P</b>	Kovarianzmatrix des Zustands
$p$	Ordnung der OSPA-Metrik
$p_D(\bullet)$	Detektionswahrscheinlichkeit
$p_D^{\text{Dist}}(\bullet)$	Detektionswahrscheinlichkeit aufgrund von Sensor-Objekt-Distanz
$p_D^{\text{Exp}}$	experimentell ermittelte Detektionswahrscheinlichkeit
$p_D^{\text{FoV}}(\bullet)$	Detektionswahrscheinlichkeit aufgrund des Sichtfelds
$p_D^{\text{Occ}}(\bullet)$	Detektionswahrscheinlichkeit aufgrund Verdeckung
$p^{(i)}(\bullet)$	Verteilung des $i$ -ten Objekts im Zustandsraum
$p_S(\bullet)$	Überlebenswahrscheinlichkeit
<b>Q</b>	Kovarianzmatrix des Systemmodells
$q_r(\bullet)$	Modell eines Experten
<b>R</b>	Kovarianzmatrix des Messmodells
$r^{(i)}$	Existenzwahrscheinlichkeit des $i$ -ten Objekts
$\mathcal{S}$	Menge weiterbestehender Objekte
<b>S</b>	Region im Zustandsraum

Symbol	Bedeutung
$t(x, y)$	Tiefenbild
$t$	Zeit
$V$	Punkt einer Punktwolke
$WB$	Winkelbereich
$\mathcal{X}$	Menge der Objektzustände
$\mathbf{x}$	Zustandsvektor
$\mathbf{x}_i$	Merkmalsvektor
$\mathcal{Z}$	Menge der Messungen

## Griechische Buchstaben

Symbol	Bedeutung
$\alpha$	Gradientenrichtung
$\theta$	Orientierung einer Person
$\kappa(\bullet)$	Falschalarmintensität
$\mu_r$	Mittelwert der Klasse $r$
$\rho(n)$	Kardinalitätsverteilung
$\sigma^2$	Varianz
$\sigma_{\text{Exp}}$	Parameter des Expertenmodells
$\sigma_{\text{SVM}}$	Gewichtungsparameter beim Training einer SVM
$\omega$	Gewicht

## Indizes

Index	Bedeutung
$(\bullet)_{\text{B}}$	<i>Birth</i>
$(\bullet)_{k+1}$	ausgewertet am diskreten Zeitpunkt $t_{k+1}$
$(\bullet)_{1:k}$	von Zeitpunkt 1 bis $k$
$(\bullet)_{\text{L}}$	<i>Legacy</i>
$(\bullet)_{\text{P}}$	<i>Persistent</i>
$(\bullet)_r$	Richtungsklasse $r$
$(\bullet)_{\text{T}}$	Trainingsbeispiel
$(\bullet)_{\text{U}}$	<i>Update</i>

## Mathematische Operatoren

Operator	Bedeutung
$\lfloor \bullet \rfloor$	Abrunden
$ \mathcal{X} $	Anzahl der Elemente in Menge $\mathcal{X}$
$\delta\mathcal{X}$	Differential in einem Mengenintegral
$E\{\bullet\}$	Erwartungswert
mod	Modulo

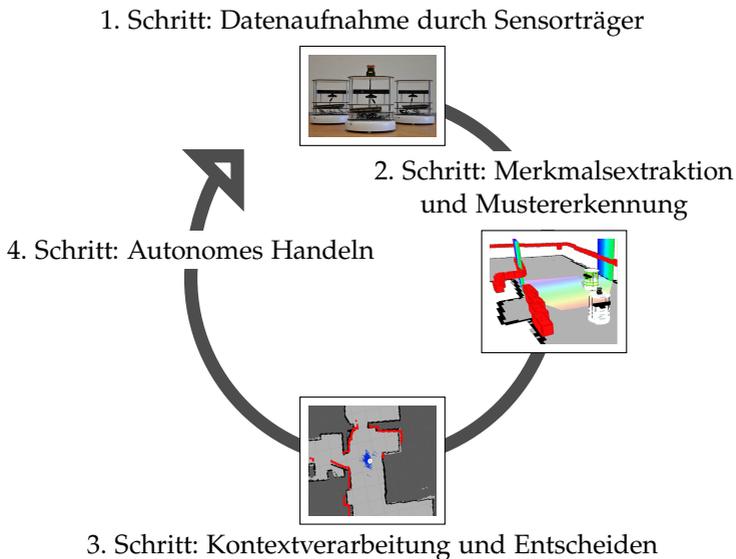


# 1 Einleitung

Die Wahrnehmung der eigenen Umgebung durch Maschinen wird in Zeiten autonomer Fahrzeuge, Haushaltsroboter, Drohnen, aber auch industrieller Zusammenarbeit mit Maschinen immer wichtiger. Was der Mensch durch seine Sinne wahrnehmen und mit Hilfe seines Gehirns verstehen kann, muss der Rechner durch Sensoren und Rechenkraft analysieren und daraus benötigte Informationen extrahieren. Die Fusion von Daten aus verschiedenen Sensoren zur Schätzung einer unbekanntem Anzahl an Objekten im Umfeld eines Systems ist daher, insbesondere in der Robotik und der Fahrerassistenz, ein vielbeachtetes Thema aufgrund der großen Anzahl an Einsatzmöglichkeiten und der hohen Bedeutung für eine sichere Mensch-Maschine-Interaktion [7, 27, 31, 61, 137].

Zusätzlich interessiert – beispielsweise für eine geeignete Wegplanung – nicht nur die Anzahl, sondern auch der Zustand – beispielsweise die Position – der Objekte, egal, ob es sich dabei um Maschinen, Personen oder andere Objekte handelt. Dabei lässt sich die Aufgabenstellung eines autonomen Systems, das selbständig sein Umfeld wahrnehmen und verstehen muss, um eigens geplante Handlungen durchführen zu können, in einen Kreislauf aus vier elementaren Schritten untergliedern, die in Abbildung 1.1 gezeigt sind und sich an dem Omnibus-Fusionsmodell [9, 122] anlehnen. Während in einem ersten Schritt die sensorische Datenaufnahme stattfindet, müssen in einem nächsten Schritt Merkmale aus den Rohdaten gewonnen und verarbeitet werden, um einem System zu ermöglichen, Teile der Umgebung wahrzunehmen. Zusätzlich werden multi-modal, aus mehreren, oft heterogenen Quellen gewonnene Daten im Sinne der Datenintegration [67] vorbereitet, um gemeinsam betrachtet zu werden. Außerdem werden bereits Sensordaten – insbesondere innerhalb eines Sensorträgers – fusioniert, um zum einen die Abstraktion zu erhöhen und somit zum anderen die Menge an auszutauschenden Daten zu reduzieren. In einem dritten Schritt können nun Informationen aus Vorwissen verschiedener Quellen und aktuelle

Sensordaten auch über verteilte Sensoren hinweg fusioniert werden, was eine Verfolgung beispielsweise des eigenen Zustands eines Sensorträgers oder auch umgebender Objekte ermöglicht und in einer Entscheidung über das aktuelle Umfeld resultiert. Nur wenn das System eine ausreichend genaue Beschreibung der Umgebung besitzt, ist es anschließend in der Lage, sinnvoll autonom zu handeln, ohne dabei sein Umfeld zu gefährden. Der Schwerpunkt der vorliegenden Arbeit liegt dabei auf der Wahrnehmung und der Modellierung der Umwelt in den Schritten zwei und drei.



**Abbildung 1.1** Kreislauf in einem autonomen System.

## 1.1 Problemstellung

Trotz intensiver Forschung auf dem Gebiet der Wahrnehmung und Fusion sind heutige Systeme noch weit von den Fähigkeiten des menschlichen Gehirns entfernt [61]. Begründet ist dies oft durch eine unangemessene Modellierung von Sensoren, vor allem jedoch durch die

Nichtberücksichtigung des Vorwissens, über das Menschen verfügen, da viele Informationen, die das Gehirn beispielsweise implizit beim Verfolgen von Objekten nutzt, in Zustandsschätzungen nicht ausreichend berücksichtigt werden. Zudem besitzen viele im Umfeld wahrgenommene Abläufe eine sehr hohe Komplexität [67]. Eine große Rolle spielt in diesem Zusammenhang auch die Frage, wie Daten aus verschiedenen Quellen miteinander fusioniert werden können, um diese komplexen Zusammenhänge zu verstehen, die teils nur durch verschiedene Arten von Information ein vollständiges Bild ergeben und eine Diversität der Sensoren erfordern. Insbesondere auf dem Gebiet der Bayes-optimalen Zustandsschätzung zeigt Mahlers Arbeit über eine Statistik endlicher Mengen [79] neue Wege auf, um Bayes-optimale Zustandsschätzer zu entwerfen, die probabilistisches Wissen über ihre Umwelt und über die für die Wahrnehmung eingesetzten Sensoren einbeziehen können. Dieser Ansatz weckte in den letzten Jahren das Interesse vieler Forschergruppen und liefert in der Theorie vielversprechende Ansätze [42, 101, 146, 149]. Dennoch sind gerade in diesem Bereich weitere Untersuchungen nötig, um die Leistungsfähigkeit solcher Ansätze auch in komplexen, realen Szenarien zu überprüfen und somit ihre Eignung als umfassendes Werkzeug für die Fusion nicht perfekter oder unsicherheitsbehafteter Daten beurteilen zu können [61].

Im Gegensatz zu traditionellen Ansätzen, die durch Erweiterungen mehrere Ein-Objekt-Filter zu Mehr-Objekt-Filtern kombinieren, bietet die Statistik endlicher Mengen einen Ansatz, die Bayes-Formel direkt auf das Multi-Objekt-Problem anzuwenden. Dies erfolgt unter Verwendung von speziellen Sensor- und Bewegungsmodellen, die den Anforderungen von Multi-Objekt-Szenarien, wie dem Hinzukommen/Verschwinden von Objekten in eine/aus einer Szene, entsprechen. Doch sind aufgrund der hohen Dimensionalität einer Lösung, die alle möglichen Kombinationen von Objekten berücksichtigen müsste, Näherungen nötig, um ein Multi-Objekt-Bayes-Filter sinnvoll einsetzen zu können. Näherungen, wie das Probability Hypothesis Density-Filter (PHD) [77, 106] und seine kardinalisierte Version (CPHD) [148] oder das Cardinality Balanced Multi-Target Multi-Bernoulli-Filter (CBMeMBer) [149], machen praktische Implementierungen mit Partikel-Darstellungen oder Gauß-Darstellungen möglich. Daher können diese Näherungen auch in Analogie zur Zustandsschät-

zung im Ein-Objekt-Fall als eine Art „Multi-Objekt-Kalman-Filter“ verstanden werden [81], die jedoch restriktive Annahmen bezüglich der Eigenschaften von Sensoren – beispielsweise deren Sichtfeldern – verlangen und somit wiederum einen Einsatz in der Praxis erschweren. Hingegen wird durch diesen Ansatz auf eine möglicherweise fehlerhafte Datenassoziation zwischen Messungen und Objekten verzichtet. In der vorliegenden Arbeit soll die verbleibende Lücke zwischen Annahmen für die theoretische Optimalität der auf endlichen Mengen basierenden Filter und einem praktischen Einsatz mittels Berücksichtigung der aktuellen Zustände der zu verfolgenden Objekte geschlossen werden. Am Beispiel der Verfolgung von Personen in Innenräumen sollen Modelle für nicht perfekte Sensoren und Detektionsverfahren entwickelt und evaluiert werden, die insbesondere den aktuellen Zustand des zu verfolgenden Objekts durch Zusatzinformationen wie die Klassifikation der Orientierung von Personen oder den Sensor-Objekt-Abstand berücksichtigen. Diese zusätzliche Information kann in die Fusion eingebracht werden und Probleme in den Modellannahmen der Multi-Objekt-Filter auflösen. Als Demonstrator dienen mehrere Multi-Sensor-Plattformen.

## 1.2 Eigener Beitrag

Nachfolgend werden die wesentlichen Beiträge der Arbeit kurz erläutert:

- **Systemaufbau:** Im Rahmen der Arbeit wurde ein System mit verteilten Sensorträgern konzipiert, das die Anwendungsmöglichkeiten der betrachteten Filter endlicher Mengen demonstriert. In Kombination mit einem Motion-Capture-System wurde eine hochgenaue Evaluation der Zustandsschätzungen ermöglicht und ein entsprechender Datensatz erstellt.
- **Vergleich verschiedener Näherungen des Multi-Objekt-Bayes-Filters an realen Daten:** Das System zur Evaluation ermöglicht des Weiteren erstmalig einen detaillierten Vergleich von Filtern basierend auf der Statistik endlicher Mengen an realen Daten mit Referenzdaten, der für die Verfolgung von Personen durchgeführt wurde.

- **Erweiterung um zustandsabhängige Detektionswahrscheinlichkeit:** Zudem sind während der Arbeit Sensoren umfassend charakterisiert worden und neue Sensormodelle entwickelt worden, die auf der Identifikation von Differenzen zwischen theoretischen Filterannahmen und realen Anwendungen beruhen. Diese ermöglichen eine detaillierte Modellierung des Sichtfelds und somit eine zustandsabhängige Beurteilung der Detektionswahrscheinlichkeit. Diese Modellierung wurde sowohl für Partikel-Implementierungen als auch für Gauß-Implementierungen in die entsprechenden Filter integriert und evaluiert.
- **Nutzung einer Orientierungsklassifikation:** Es wurde ein neuartiges Verfahren für Farb- und Tiefensensoren zur kontinuierlichen Orientierungsschätzung von Personen mit Hilfe von Gradienten in Einzelbildern entwickelt. Dieses berücksichtigt insbesondere die räumliche Beziehung von Pixeln durch die Beachtung der Richtung des Gradienten in Tiefenbildern und führt zu deutlichen Verbesserungen in der Schätzung. Zusammen mit einer vorgeschlagenen Fusionsstrategie wurde der Nutzen der Fusion dieser Orientierungsinformationen mit den konventionellen Messmodellen zur Verbesserung der Zustandsschätzung in Multi-Objekt-Szenarien untersucht.

## 1.3 Gliederung

In Kapitel 2 folgen Grundlagen zur Zustandsschätzung im Ein-Objekt-Fall sowie zu verschiedenen Detektions- und Klassifikationsverfahren. Darauf aufbauend stellt Kapitel 3 die Theorie der Statistik endlicher Mengen kurz vor und gibt einen Einblick in die gebräuchlichsten Approximationen des Multi-Objekt-Bayes-Filters zusammen mit einer Bewertungsmetrik. In Kapitel 4 werden die Erweiterungen bezüglich einer adaptiven Detektionswahrscheinlichkeit zusammen mit der Anwendung dieser auf Multi-Objekt-Filter vorgestellt. Kapitel 5 behandelt das vorgeschlagene Verfahren zur Schätzung von Orientierungen und die Fusion für eine richtungssensitive Verfolgung. Die Ergebnisse der vorgestellten

Verfahren werden in Kapitel 6 diskutiert und abschließend in Kapitel 7 zusammengefasst.

## 2 Grundlagen

Im folgenden Kapitel sollen die grundlegenden Methoden, die in späteren Kapiteln erweitert werden oder wichtige Bauteile für später vorgestellte Verfahren darstellen, eingeführt werden. Während der Schwerpunkt in diesem Kapitel auf der Beschreibung von Verfahren liegt, die sowohl in Ein-Objekt- als auch in Mehr-Objekt-Problemen eine grundlegende Rolle spielen und sich nicht auf Zustandsschätzung beschränken, behandelt Kapitel 3 insbesondere die theoretischen Grundlagen für die Filterung von Mehr-Objekt-Problemen.

### 2.1 Bewegungsmodelle

Ein sehr wichtiger Teil der Zustandsschätzung ist die Prädiktion der Zustände ausgehend vom geschätzten Zustand des letzten Zeitpunkts und somit im Zeitraum zwischen der zuletzt eingetroffenen und der neuen Messung. Da der Zeitraum zwischen dem Eintreffen zweier Messungen unbeobachtet ist, müssen Annahmen getroffen werden, ob und wie sich Zustände, wie beispielsweise Position oder Geschwindigkeit, in der Zwischenzeit verändern. Häufig verwendete Modelle treffen hierbei entweder keine Annahmen über eine Bewegung, die Annahme einer zufälligen Bewegung (siehe Abschnitt 2.1.1), die Annahme einer konstanten Geschwindigkeit (siehe Abschnitt 2.1.2) oder die Annahme einer konstanten Beschleunigung. Da die beiden letztgenannten Modelle in realen dynamischen Systemen in ihrer exakten Form kaum Relevanz besitzen, weil jede natürliche Bewegung Schwankungen unterlegen ist, werden sie in der Modellierung um mittelwertfreie Rauschanteile erweitert und führen zu einem Modell entsprechend der Brown'schen Molekularbewegung, einem Modell nahezu konstanter Geschwindigkeit und einem Modell nahezu konstanter Beschleunigung. Entsprechend der Intensität der gewählten Rauschanteile sind solche Modelle somit in begrenztem

Umfang auch in der Lage, die Bewegung manövrierender Objekte zu beschreiben.

Mathematisch können diese Modelle als gewöhnliche Differentialgleichungen beschrieben werden, die aufgrund ihrer rauschbedingten statistischen Eigenschaften als stochastische Differentialgleichungen (SDE – *Stochastic Differential Equations*) bekannt sind. Neben den entsprechenden Annahmen über die Bewegung hängt die Lösung der Differentialgleichung auch von der Koordinatendarstellung der Zustände – in, für gewöhnlich, kartesischer oder polarer Form – ab. Hierzu soll ein Überblick über die gebräuchlichsten Bewegungsmodelle im zweidimensionalen Raum sowie ihre Lösung und die Fortpflanzung der eingebrachten Unsicherheiten über die Zeit gegeben werden. Modelle höherer Ordnung, die beispielsweise Beschleunigungen berücksichtigen, spielen hauptsächlich bei trägeren Objekten, wie Automobilen oder Flugzeugen, eine Rolle. Sie besitzen jedoch bei der Verfolgung von Fußgängern aufgrund sehr kleiner Trägheit nur eine untergeordnete Bedeutung, weswegen für derartige Modelle auf beispielsweise [6, Seite 274 ff.] verwiesen wird.

### 2.1.1 Zufällige Bewegung

Beschreibt

$$\mathbf{x}(t) = [x(t) \quad y(t)]^T \quad (2.1)$$

den Zustand bzw. die Position eines Systems in einer zweidimensionalen  $x, y$ -Ebene, so lässt sich die zufällige Bewegung durch die Gleichung

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \mathbf{v}_{t+\Delta t} \quad (2.2)$$

beschreiben. Hierbei verändert sich der Zustand  $\mathbf{x}_t$ , also die Position  $\mathbf{x}$  zum Zeitpunkt  $t$ , bis zum nächsten Zeitschritt  $t + \Delta t$  entsprechend dem Rauschterm  $\mathbf{v}_{t+\Delta t}$ . Das Rauschen wird im Sinne einer Brown'schen Bewegung oder auch eines Wiener-Prozesses unabhängig für  $x$  und  $y$  als normalverteilt angesehen [134, Seite 29]. Die Rauschintensität  $q$  bestimmt hierbei das Prozessrauschen. Entsprechend der zeitlichen Dauer  $\Delta t$  zwischen zwei Zuständen kann hierbei die Varianz angepasst werden [6].

## 2.1.2 Konstante Geschwindigkeit

Beschreibt

$$\mathbf{x}(t) = [x(t) \quad y(t) \quad v_x(t) \quad v_y(t)]^T \quad (2.3)$$

den Zustand eines Systems mit Geschwindigkeit  $v_x$  in  $x$ -Richtung und Geschwindigkeit  $v_y$  in  $y$ -Richtung, so lässt sich ein Modell nahezu konstanter Geschwindigkeit mit der Differentialgleichung

$$\frac{\partial \mathbf{x}(t)}{\partial t} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{v}(t) \quad \text{mit} \quad \mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{und} \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad (2.4)$$

aufstellen. Durch diese einfache lineare Form lässt sich die Gleichung über die Laplace-Transformation [114] lösen und ergibt folgendes Ergebnis [130, Seite 124] für einen Zeitschritt der Dauer  $\Delta t$  von  $t$  nach  $t + \Delta t$  bzw. für entsprechende diskrete Zeitpunkte von  $k$  nach  $k + 1$ :

$$\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{v}_{k+1} \quad (2.5)$$

$$\text{mit} \quad \mathbf{F} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

$$\text{und} \quad E\{\mathbf{v}_{k+1}\mathbf{v}_{k+1}^T\} = \begin{bmatrix} \frac{1}{3}\Delta t^3 & 0 & \frac{1}{2}\Delta t^2 & 0 \\ 0 & \frac{1}{3}\Delta t^3 & 0 & \frac{1}{2}\Delta t^2 \\ \frac{1}{2}\Delta t^2 & 0 & \Delta t & 0 \\ 0 & \frac{1}{2}\Delta t^2 & 0 & \Delta t \end{bmatrix} \cdot q. \quad (2.7)$$

Für kleine Rauschintensitäten  $q$  spricht wird von einem nahezu konstanten Geschwindigkeitsmodell (*nearly constant velocity model (NCV)*) gesprochen. Das entsprechende Modell für eine Position in kartesischen Koordinaten mit Geschwindigkeitsvektor in Polarkoordinaten aus Orientierung und Geschwindigkeitsbetrag findet sich im Anhang A.

## 2.2 Messmodelle

Aufgabe des Messmodells ist es, die aus der Prädiktion durch Bewegungsmodelle entstehenden Zustände mit Messungen abzugleichen. Dazu liefert das Messmodell die Vorschrift, wie eine Messung aus einem Objekt unter Berücksichtigung der Messunsicherheit hervorgeht. Eine Messung  $\mathbf{z}_k$  zu einem diskreten Zeitpunkt  $k$  kann somit aus dem sensorspezifischen Modell  $\mathbf{h}_k(\mathbf{x})$  und einem Anteil des Sensorrauschens  $\mathbf{w}$  gebildet werden:

$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{w}_k. \quad (2.8)$$

Der probabilistische Zusammenhang zwischen Messung und Objekt entsteht durch die sogenannte Likelihood-Funktion, die die Wahrscheinlichkeitsdichte der Messung bedingt auf den Objektzustand angibt und sich aus dem Modell eines additiven Rauschterms aus (2.8) mittels der Wahrscheinlichkeitsdichte des Rauschens  $f_{\mathbf{w}_k}$  errechnen lässt [154]:

$$g(\mathbf{z}_k | \mathbf{x}_k) = f_{\mathbf{w}_k}(\mathbf{z}_k - \mathbf{h}_k(\mathbf{x}_k)). \quad (2.9)$$

Bei einer vom Ursprung abweichenden Sensorposition muss gegebenenfalls die Position des Sensors in Weltkoordinaten bei den ebenfalls in Weltkoordinaten präzidierten Objektkoordinaten berücksichtigt werden, so dass das Messmodell  $\mathbf{h}_k$  von zum Sensor relativen Koordinaten abhängt.

Eine explizite Betrachtung der Unsicherheit der Sensorposition wird hierbei nicht durchgeführt, ist aber, wie in [65] gezeigt, für die betrachteten Filter möglich.

## 2.3 Ein-Objekt-Bayes-Filter

Ziel eines Zustandsschätzers ist es, Zustände, die nicht direkt messbar sind, abzuschätzen. Zustandsbeobachter, wie beispielsweise der Luenerberger-Beobachter [75], ermöglichen dies für den deterministischen Fall. Spielen im System auch Messunsicherheiten eine Rolle, führt die Berücksichtigung dieser zufälligen Messunsicherheiten und auch Unsicherheiten im Systemmodell bei der Zustandsschätzung zu einer klaren

Verbesserung. Die Schätzung kann durch das Bayes-Filter realisiert werden, das den Satz von Bayes zur Grundlage hat und eine Bayes-optimale Lösung des Schätzproblems verfolgt. Kalman [56] lieferte einen Satz von Gleichungen, der, unter der Annahme normalverteilter Rauschprozesse mit bekannten Varianzen, die Optimallösung für die Zustandsschätzung in linearen Systemen darstellt und hierfür eine geschlossene Lösung ermöglicht. Das Kalman-Filter bildet darüber hinaus die Grundlage für eine Vielzahl weiterer Schätzalgorithmen. Um ebenso Nichtlinearitäten in Mess- und Systemmodellen auszugleichen, sind Erweiterungen entstanden, wie das erweiterte Kalman-Filter (EKF) oder das Unscented-Kalman-Filter (UKF). Im folgenden Abschnitt sollen auf Grundlage des Bayes-Filters das Kalman-Filter sowie dessen Erweiterungen auf nicht-lineare Systeme vorgestellt werden. Zusätzlich wird auch auf das auf der sequentiellen Monte-Carlo-Methode (SMC) basierende Partikelfilter eingegangen, das ohne die strengen Annahmen von Kalman eine Zustandsschätzung ermöglicht, jedoch, wie auch das Unscented-Kalman-Filter, nicht zwangsweise zu einer optimalen Lösung führt, da es auf einer zufälligen Approximation der Dichte beruht.

### 2.3.1 Modellierung des Schätzproblems

Allgemein wird das Schätzproblem entsprechend der Modelle aus Abschnitt 2.1 und 2.2 in Abhängigkeit des vergangenen bzw. aktuellen Zustands und der jeweiligen Rauschterme beschrieben [83]

$$\mathbf{x}_k = \mathbf{a}_{k-1}(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}) \quad (2.10)$$

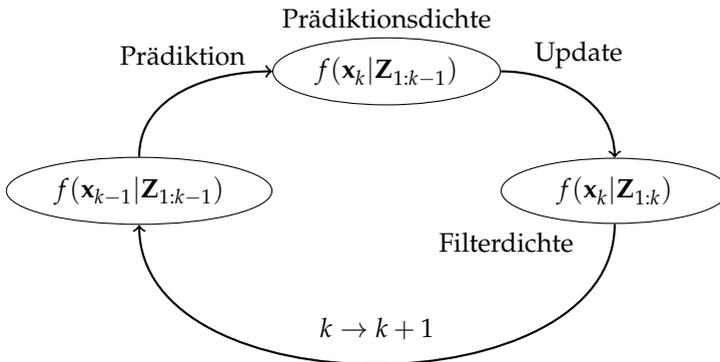
$$\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}_k, \mathbf{w}_k). \quad (2.11)$$

Dabei stellt die Funktion  $\mathbf{a}_{k-1}$  das die Dynamik des Zustands beinhaltende Systemmodell dar, während die Funktion  $\mathbf{h}_k$  das Messmodell repräsentiert und den aktuellen Zustand auf die Messgröße abbildet. Die Schreibweise ist hier analog zu Abschnitt 2.1 so zu verstehen, dass  $\mathbf{x}_{k-1}$  den Zustand  $\mathbf{x}(t_{k-1})$  am diskreten Zeitpunkt  $t_{k-1}$  bezeichnet.

### 2.3.2 Bayes-Filter

Das im Gegensatz zum Kalman-Filter nicht nur für die ersten beiden Momente anwendbare und somit allgemeinere Bayes-Filter basiert auf der

Beschreibung des Schätzproblems über Wahrscheinlichkeitsdichten [63]. Das Wissen über das System basiert hierbei auf den vorherigen Beobachtungen  $\mathbf{Z}_{1:k} = \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$  bis zum Zeitpunkt  $t_k$  und lässt sich als Dichte  $f(\mathbf{x}_k | \mathbf{Z}_{1:k})$ <sup>1</sup> beschreiben. Zusammen mit einer vor der ersten Messung angenommenen Anfangsdichte  $f(\mathbf{x}_0)$  arbeitet sowohl das Bayes-Filter als auch seine Approximationen des Kalman- oder des Partikel-Filters rekursiv, wobei stets unter Annahme des Bewegungsmodells eine Prädiktionsdichte (A-priori-Dichte)  $f(\mathbf{x}_k | \mathbf{Z}_{1:k-1})$  erzeugt wird. Im zweiten Schritt wird diese unter Berücksichtigung der über das Messmodell eingeflossenen neuen Messung  $\mathbf{z}_k$  korrigiert (*Update*-Schritt) und als Filter-Dichte (A-posteriori-Dichte)  $f(\mathbf{x}_k | \mathbf{Z}_{1:k})$  bezeichnet. Die Rekursion beginnt nun, indem im nächsten Schritt die vorherige Filterdichte als Anfangsdichte für eine Prädiktion im nächsten Zeitschritt verwendet wird. Der Ablauf ist in Abbildung 2.1 gezeigt.



**Abbildung 2.1** Bayes-Rekursion.

### Definition 2.1 (Bedingte Dichte)

Nach [21] ist eine bedingte Dichte zweier stetiger Zufallsvariablen  $x$  und  $y$  mit gemeinsamer Dichte  $f_{xy}(\xi, \zeta)$  für  $x$  unter der Bedingung

<sup>1</sup> Bei der Darstellung der Dichten wird entsprechend [113, Seite 111] falls möglich auf die Berücksichtigung der Abhängigkeit einer Zufallsvariable von dem Elementarereignis  $\xi$  verzichtet.

$y = \zeta$  durch

$$f_{x|y}(\xi|y = \zeta) = \frac{f_{xy}(\xi, \zeta)}{f_y(\zeta)} \quad (2.12)$$

definiert.

### Satz 2.1 (Bayes-Theorem)

Aus der Definition 2.1 folgt unmittelbar nach [113]

$$f_{xy}(x, y) = f_{x|y}(x|y) \cdot f_y(y) = f_{y|x}(y|x) \cdot f_x(x). \quad (2.13)$$

Hierbei stellen die beiden rechten Gleichungen von (2.13) den Satz von Bayes dar und stellen den Zusammenhang zwischen  $f_{x|y}(x|y)$  und  $f_{y|x}(y|x)$  her.

Aus Satz 2.1 lässt sich nun mithilfe von Marginalisierung und dem Gesetz der totalen Wahrscheinlichkeit die bedingte Dichte  $f_{x|y}(x|y)$  umschreiben zu

$$f_{x|y}(x|y) = \frac{f_{y|x}(y|x)f_x(x)}{f_y(y)} = \frac{f_{y|x}(y|x)f_x(x)}{\int_{-\infty}^{\infty} f_{y|x}(y|x)f_x(x)dx}. \quad (2.14)$$

Gleichung (2.14) ermöglicht das Vertauschen der Bedingtheit und lässt sich nutzen, um die prädierte Dichte in die Filterdichte zu übertragen. Unter Annahme der Markov-Eigenschaft erster Ordnung [154, Seite 23], die die Unabhängigkeit von der Zustands- und Messhistorie vor dem vergangenen Zeitpunkt fordert, lässt sich der Zusammenhang von prädiertem Zustand und vorherigem Zustand durch Marginalisierung über die *Chapman-Kolmogorov*-Gleichung [104] herstellen

$$f(\mathbf{x}_k|\mathbf{Z}_{1:k-1}) = \int f(\mathbf{x}_k|\mathbf{x}_{k-1})f(\mathbf{x}_{k-1}|\mathbf{Z}_{1:k-1}) d\mathbf{x}_{k-1}. \quad (2.15)$$

Hierbei ist  $f(\mathbf{x}_k|\mathbf{x}_{k-1})$  die durch das gewählte Bewegungsmodell (siehe Abschnitt 2.1) bestimmte Transitionsdichte und beschreibt die Wahrscheinlichkeit des nächsten Zustands.

Für die Bestimmung der *Update*-Gleichung wird die Formel nach Bayes (2.14) genutzt und von der gewünschten Dichte  $f(\mathbf{x}_k|\mathbf{Z}_{1:k})$  ausgegangen.

Diese lässt sich unter Aufspaltung aller Messungen in aktuelle und vergangene dann folgendermaßen nach [79, 177] darstellen

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k}) = \frac{f(\mathbf{Z}_{1:k} | \mathbf{x}_k) f(\mathbf{x}_k)}{f(\mathbf{Z}_{1:k})} = \frac{f(\mathbf{z}_k, \mathbf{Z}_{1:k-1} | \mathbf{x}_k) f(\mathbf{x}_k)}{f(\mathbf{z}_k, \mathbf{Z}_{1:k-1})}. \quad (2.16)$$

Die Anwendung der bedingten Wahrscheinlichkeit führt jeweils für Zähler und Nenner zu

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k}) = \frac{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1}, \mathbf{x}_k) f(\mathbf{Z}_{1:k-1} | \mathbf{x}_k) f(\mathbf{x}_k)}{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1}) f(\mathbf{Z}_{1:k-1})}. \quad (2.17)$$

Nach Umformung von  $f(\mathbf{Z}_{1:k-1} | \mathbf{x}_k)$  durch den Satz von Bayes erhält man

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k}) = \frac{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1}, \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{Z}_{1:k-1}) f(\mathbf{Z}_{1:k-1}) f(\mathbf{x}_k)}{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1}) f(\mathbf{Z}_{1:k-1}) f(\mathbf{x}_k)} \quad (2.18)$$

$$= \frac{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1}, \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{Z}_{1:k-1})}{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1})}. \quad (2.19)$$

Bei Einbeziehung der aktuellen Messung  $\mathbf{z}_k$  wird, wie bei der Prädiktion, die Markov-Eigenschaft angenommen, sodass diese Messung nur vom aktuellen Zustand  $\mathbf{x}_k$  abhängt. Entsprechend vereinfacht sich der Ausdruck zu

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k}) = \frac{f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{Z}_{1:k-1})}{f(\mathbf{z}_k | \mathbf{Z}_{1:k-1})}. \quad (2.20)$$

Ebenfalls durch Marginalisierung lässt sich ähnlich zu Gleichung (2.15) die Dichte  $f(\mathbf{z}_k | \mathbf{Z}_{1:k-1})$  als Aussage über die Wahrscheinlichkeit des Auftretens der aktuellen Messung unter Berücksichtigung aller möglichen Zustände interpretieren und stellt somit eine Normalisierung für die Wahrscheinlichkeit der aktuellen Messung zum aktuellen Zustand dar:

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k}) = \frac{f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{Z}_{1:k-1})}{\int f(\mathbf{z}_k | \mathbf{x}_k) f(\mathbf{x}_k | \mathbf{Z}_{1:k-1}) d\mathbf{x}_k}. \quad (2.21)$$

Dieses Ergebnis für die Korrektur der Prädiktion in Gleichung (2.15) nutzt neben der prädizierten Dichte die Likelihood-Funktion  $f(\mathbf{z}_k | \mathbf{x}_k)$ , die nach (2.9) eine dem Messmodell und dem dort hinterlegten Rauschmodell entsprechende Auftrittswahrscheinlichkeit der aktuellen Messung

bei gegebenem Zustand angibt. Die Likelihood-Funktion wird von hieran als  $g(\mathbf{z}_k | \mathbf{x}_k)$  bezeichnet.

Durch die Gleichungen (2.15) und (2.21) liegen somit die für die Rekursion benötigten Berechnungsvorschriften vor. Um aus den Dichten auf Schätzungen des Zustands zu schließen, werden Verfahren wie der *Maximum-a-posteriori*-Schätzer (MAP) zur Bestimmung des Maximums der Dichte

$$\hat{\mathbf{x}}_k^{\text{MAP}} = \underset{\mathbf{x}_k}{\operatorname{argmax}} f(\mathbf{x}_k | \mathbf{Z}_{1:k}), \quad (2.22)$$

oder der *A-posteriori*-Erwartungswert (*Expected-a-posteriori*, EAP) zur Bildung des Erwartungswerts

$$\hat{\mathbf{x}}_k^{\text{EAP}} = \int \mathbf{x}_k f(\mathbf{x}_k | \mathbf{Z}_{1:k}) d\mathbf{x}_k \quad (2.23)$$

verwendet [63, 79].

### 2.3.3 Kalman-Filter

Ausgehend von den zwei Grundannahmen [85] des Kalman-Filters: ein lineares System und mittelwertfreies, weißes, Gauß'sches Rauschen, beschreibt das Kalman-Filter (KF) [56] eine geschlossene Lösung der Bayes-Filter mit minimaler quadratischer Abweichung zwischen zu schätzendem Zustand und Schätzung und kann unter diesen Umständen als Optimalfilter bezeichnet werden [85]. Namensgeber Kalman [56] präsentierte somit 1960 einen zeitdiskreten und rekursiven Satz von Gleichungen, der das Bayes-Filter für die Praxis anwendbar machte und noch heute eine große Bedeutung für die digitale Signalverarbeitung besitzt.

Durch die Annahme eines linearen Systems lässt sich das Problem entsprechend Abschnitt 2.3.1 linear darstellen über

$$\mathbf{x}_k = \mathbf{A}_{k-1} \mathbf{x}_{k-1} + \mathbf{v}_{k-1} \quad (2.24)$$

$$\mathbf{z}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{w}_k \quad (2.25)$$

mit der aus dem Bewegungsmodell folgenden Transitionsmatrix  $\mathbf{A}$ , dem Messmodell  $\mathbf{H}$  sowie den mittelwertfreien, Gauß'schen, weißen Rauschtermen  $\mathbf{v}_k$  und  $\mathbf{w}_k$ . Weiterhin sind die Rauschterme zueinander unkor-

reliert und besitzen jeweils die als bekannt angenommenen Kovarianzmatrizen  $\mathbf{R}_k$  und  $\mathbf{Q}_k$ . Die Kovarianzmatrix des Zustands wird als  $\mathbf{P}$  bezeichnet.

**Definition 2.2 (Mehrdimensionale Normalverteilung [113])**

Eine mehrdimensionale Zufallsvariable  $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$  mit der Wahrscheinlichkeitsdichte

$$f_{\mathbf{x}}(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}) \quad (2.26)$$

$$= \frac{1}{(2\pi)^{\frac{d}{2}} |\boldsymbol{\Sigma}_{\mathbf{x}}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\right) \quad (2.27)$$

heißt multivariat normalverteilt. Eine mehrdimensionale Normalverteilung wird durch den Mittelwertvektor  $\boldsymbol{\mu}_{\mathbf{x}}$  und die Kovarianzmatrix  $\boldsymbol{\Sigma}_{\mathbf{x}}$  vollständig charakterisiert. Die Punkte gleicher Wahrscheinlichkeitsdichte beschreiben Ellipsoide in  $d$  Dimensionen.

Beide Terme (2.24) und (2.25) entsprechen nach Definition 2.2 einer Normalverteilung und sind somit vollständig beschrieben. Die in Abschnitt 2.3 beschriebenen Rekursionsgleichungen der Prädiktion und der Korrektur des Bayes-Filters werden somit ebenso durch die ersten beiden Momente beschrieben

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k-1}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \quad (2.28)$$

$$f(\mathbf{x}_k | \mathbf{Z}_{1:k}) = \mathcal{N}(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \quad (2.29)$$

wobei sich die unbekanntenen Parameter aus Kalmans Gleichungen ergeben:

**Prädiktions-Schritt**

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_{k-1} \hat{\mathbf{x}}_{k-1|k-1} \quad (2.30)$$

$$\mathbf{P}_{k|k-1} = \mathbf{Q}_{k-1} + \mathbf{A}_{k-1} \mathbf{P}_{k-1|k-1} \mathbf{A}_{k-1}^T, \quad (2.31)$$

### Filter-Schritt

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}) \quad (2.32)$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \quad (2.33)$$

$$\mathbf{S}_k = \mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}_k^T + \mathbf{R}_k \quad (2.34)$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T\mathbf{S}_k^{-1}. \quad (2.35)$$

Die Matrix  $\mathbf{K}_k$  steht für die Kalman-Verstärkung (Kalman-Gain-Matrix) und die Matrix  $\mathbf{S}_k$  bildet die Kovarianzmatrix des Innovationsresiduums  $\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k|k-1}$  ab (vgl. [118], Seite 8). Durch die Kalman-Verstärkung wird die Innovation, also der Unterschied des erwarteten Zustands zu der Messung, abhängig von der Güte der Messung bei der Schätzung des neuen Zustands miteinbezogen. Die durch die Prädiktion steigende Unsicherheit eines Zustands wird durch die Information der Messung wieder verringert. Folglich wird, unter den genannten Voraussetzungen, über die laufende Schätzung der Zustände  $\hat{\mathbf{x}}$  und deren Unsicherheiten  $\mathbf{P}$  die prädizierte und gemessene Information optimal im Sinne eines kleinsten quadratischen Fehlers zwischen Schätzung und tatsächlichem Zustand kombiniert.

### Erweitertes Kalman-Filter

Die ersten Anwendungsfälle fand das Kalman-Filter für nichtlineare Fälle aus der Raumfahrt in Form des erweiterten Kalman-Filters (EKF) [55]. So schreibt McElhoe [86] im Jahr 1966 von einem bemannten Flug Richtung Mars oder auch Venus unter Ausnutzung der Taylor-Approximation der Gleichungen nach Kalman. Heutzutage ist das erweiterte Kalman-Filter weit in der Navigation und Robotik verbreitet [137], denn für den Fall nichtlinearer Bewegungs- beziehungsweise Messmodelle

$$\mathbf{x}_k = \mathbf{a}(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \quad (2.36)$$

$$\mathbf{z}_k = \mathbf{h}(\mathbf{x}_k) + \mathbf{v}_k \quad (2.37)$$

sind die linearen Kalman-Gleichungen (2.30)-(2.35) nicht anwendbar. Führt man nun allerdings eine Taylor-Approximation, für den Fall multi-

variater Zustände mittels der Jacobi-Matrix, durch und bricht diese nach dem linearen Term ab, erhält man eine linearisierte Transitionsmatrix

$$\hat{\mathbf{A}}_{k-1} = \left. \frac{\partial f(\mathbf{x}_{k-1})}{\partial \mathbf{x}_{k-1}} \right|_{\mathbf{x}_{k-1} = \hat{\mathbf{x}}_{k-1|k-1}} \quad (2.38)$$

und eine linearisierte Messmatrix

$$\hat{\mathbf{H}}_k = \left. \frac{\partial h(\mathbf{x}_k)}{\partial \mathbf{x}_k} \right|_{\mathbf{x}_k = \hat{\mathbf{x}}_{k|k-1}} \quad (2.39)$$

Für das erweiterte Kalman-Filter können nun die nichtlinearen Anteile aus Gleichungen (2.36) und (2.37) für Prädiktion und Update des Zustands in Gleichungen (2.30) und (2.32) verwendet werden. In den übrigen Gleichungen der Kalman-Rekursion werden die linearisierten Näherungen  $\hat{\mathbf{A}}_{k-1}$  und  $\hat{\mathbf{H}}_k$  verwendet. Durch die Taylor-Approximation erster Ordnung verliert das Filter seine Optimalität. Abbildung 2.2(a) zeigt die Approximation im Vergleich zu Unscented-Kalman-Filter (UKF) und Partikel-Filter.

## Unscented-Kalman-Filter

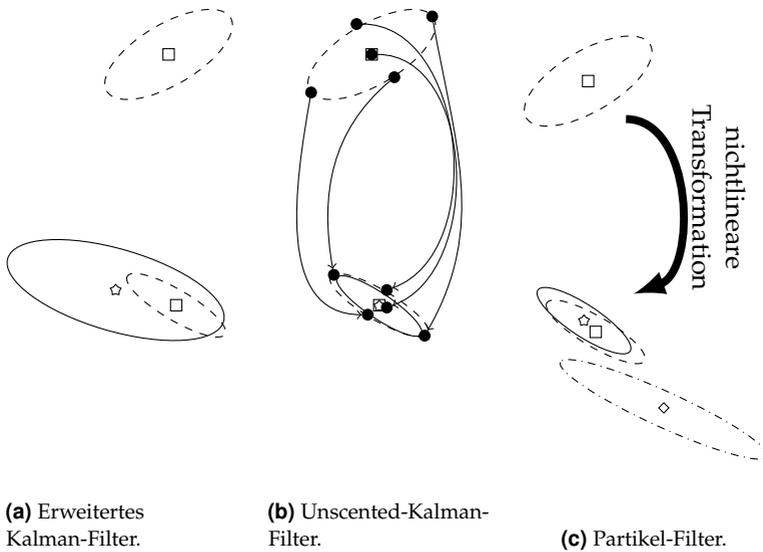
Trotz der weiten Verbreitung bescheinigen Julier und Uhlmann [55] nach über 35 Jahren der Verwendung dem erweiterten Kalman-Filter eine schwierige Handhabbarkeit und Probleme der vorgenommenen Linearisierung bei größeren Nichtlinearitäten. Gleichzeitig stellen sie die Unscented-Transformation für die Zustandsschätzung in nichtlinearen Systemen vor, die eingesetzt in den ursprünglichen Gleichungssatz nach Kalman das Unscented-Kalman-Filter ergibt, das in modernen Systemen gegenüber dem erweiterten Kalman-Filter bevorzugt wird [66, 79, 146]. Das Unscented-Kalman-Filter beruht ebenfalls auf der Beschreibung eines Systemzustands durch die ersten beiden Momente Mittelwert und Varianz, da diese oft ausreichen und höhere Momente einer praktischen Nutzung im Wege stehen könnten [55]. Des Weiteren werden die Probleme der Berechnung der für die Linearisierung nötigen Jacobi-Matrix durch Verwendung der Unscented-Transformation vermieden.

Grundidee bei der Unscented-Transformation ist die Auswahl spezieller Kombinationen von Sigma-Punkten, die es ermöglichen, eine

zu transformierende Verteilung durch wenige Punkte anzunähern, ohne die nichtlineare Transformationsfunktion anzunähern [55], wie es bei dem erweiterten Kalman-Filter der Fall ist. Diese Sigma-Punkte müssen so gewählt werden, dass sie in Mittelwert und Varianz mit der ursprünglichen Verteilung übereinstimmen. Neben weiteren Sigma-Punkt-Konstellationen wie in [3, 53, 110], die zum Teil auch die Berücksichtigung höherer Momente ermöglichen, wird von Julier [53] die Verwendung der Scaled-Unscented-Transformation für die Filterung empfohlen, die bei einer Zustandsdimension von  $N$  mit  $2N + 1$  symmetrischen Sigmapunkten auskommt. Hierzu werden Mittelpunkt und  $2N$  Punkte durch die Wurzel der Kovarianzmatrix entsprechend [151], beispielsweise über die Cholesky-Zerlegung, gewählt und gewichtet. Anschließend werden diese  $2N + 1$  Sigma-Punkte, wie in Abbildung 2.2(b), entsprechend den nichtlinearen Gleichungen für Transition (2.36) beziehungsweise Messmodell (2.37) transformiert. Aus den transformierten Punkten können anschließend transformierte Mittelwerte und die Kovarianzmatrizen berechnet werden, die für die Kalman-Filtergleichungen nötig sind. Obwohl das Unscented-Kalman-Filter zur Verbesserung der Schätzung bei Nichtlinearitäten führt, handelt es sich auch hier nicht um ein Optimalfilter. Außerdem führt es zu einem höheren Rechenaufwand als das erweiterte Kalman-Filter [137].

### 2.3.4 Partikelfilter

Partikelfilter versuchen die Zustandsschätzung ohne eine geschlossene Lösung wie die Kalman-Gleichungen durchzuführen. Sie beruhen auf sequentiellen Monte-Carlo-Verfahren, die wie auch das Kalman-Filter seit Mitte des letzten Jahrhunderts [46] bekannt sind, allerdings wegen ihres erhöhten Rechenaufwands erst durch das Aufkommen schnellerer Prozessoren in den letzten Jahren für die nichtlineare Zustandsschätzung verstärkt eingesetzt werden [118]. Während bei der Unscented-Transformation aus Abschnitt 2.3.3 wenige ausgewählte Punkte (Sigma-Punkte) verwendet werden, um bestimmte Momente zu repräsentieren und anschließend nichtlinearen Transformationen zu unterziehen, werden bei der Verwendung von Partikeln entsprechend der Monte-Carlo-Methode sehr viele Stichproben aus einer Wahrscheinlichkeitsdichte gezogen. Während dies natürlich einen erhöhten Rechenaufwand bedeutet,



**Abbildung 2.2** Gegenüberstellung verschiedener Ein-Objekt-Filter nach [151]. Die wahren Mittelwerte werden dabei durch Quadrate gekennzeichnet mit gestrichelten Kovarianzellipsen vor (oben) und nach (unten) einer nichtlinearen Transformation. Schätzungen von Mittelwerten sind durch Sterne und durchgezogene Kovarianzellipsen gekennzeichnet. Für das Partikel-Filter sind zwei Schätzungen gezeigt: Eine mit 10 Partikeln ( $\cdot - \cdot$ ) und eine weitere mit 100 Partikeln (durchgezogen) um den jeweiligen Mittelwert.

ermöglicht es jedoch gleichzeitig eine Repräsentation kompletter Dichten und beschränkt sich nicht auf einzelne Momente. Dennoch handelt es sich bei einer endlichen Anzahl an Partikeln ebenfalls um eine suboptimale Lösung, die jedoch keinen Beschränkungen der Nichtlinearität wie bei KF, EKF oder UKF unterliegt.

## Sequential Importance Sampling

Eine Monte-Carlo-Methode, die Basis für die Partikelfilterung ist, wird *Sequential Importance Sampling* (SIS) genannt [118]. Hierbei wird die Fil-

terdichte zum Zeitpunkt  $k$  durch

$$f(\mathbf{x}_k | \mathbf{z}_k) \approx \sum_{i=1}^N \omega_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i) \quad \text{mit} \quad \sum_{i=1}^N \omega_k^i = 1 \quad (2.40)$$

mit einer festen Anzahl von  $N$  Partikeln approximiert. Diese besitzen dabei jeweils das Gewicht  $\omega_k^i$  und repräsentieren einen Dirac-Impuls am Ort  $\mathbf{x}_k^i$ . Geht dabei  $N \rightarrow \infty$ , lässt sich zeigen, dass (2.40) gegen die wahre Filterdichte strebt [118]. Analog zu den Schritten für das Bayes-Filter in Abschnitt 2.3.2 und (2.20) lässt sich auch für das Partikel-filter eine Prädiktions-Gleichung bei gleichbleibendem Gewicht und eine Update-Gleichung angeben, die das Gewicht der Partikel aus dem letzten Zeitschritt  $k - 1$  aufgrund einer neu eingetroffenen Messung anpasst [118]:

$$\mathbf{x}_k^i \sim q(\mathbf{x}_k^i | \mathbf{x}_{k-1}, \mathbf{z}_k) \quad (2.41)$$

$$\omega_k^i \propto \frac{g(\mathbf{z}_k | \mathbf{x}_k^i) f(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i)}. \quad (2.42)$$

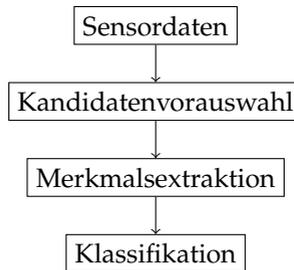
Hierbei stellt  $q(\bullet)$  die sogenannte Vorschlagsdichte (*Proposal Distribution*) dar, aus der Partikel gezogen werden. Das heißt die Verteilung der gezogenen Partikel entspricht im Grenzfall der gewählten Vorschlagsdichte. Diese hängt im optimalen Fall wie in (2.41) vom vorherigen Zustand und der aktuellen Messung ab, wird aber häufig suboptimal als Transitionsdichte entsprechend dem Bewegungsmodell gewählt, da das Ziehen von neuen Partikeln aus dieser Dichte in der Anwendung ohne großen Aufwand möglich sein muss [118]. Gleichung (2.42) ist die Gewichtungsfunktion, die mittels der Likelihood einer Messung die einzelnen Partikel nach deren Plausibilität bewertet.

## Resampling

Der SIS-Algorithmus führt nach einer gewissen Zeit zu einer Degenerierung der Partikel. Degenerierung bezeichnet den Zustand, in dem viele der Partikel ein nur noch sehr geringes Gewicht tragen und nur einige wenige Partikel einen wirklichen Beitrag zur Zustandsschätzung leisten. Deswegen wird das sogenannte *Resampling* angewendet, um



unter anderem Position, Größe, Entfernung, Winkel und Orientierung der Objekte erfassen. Abhängig von dem Informationsgehalt und der Güte der auswertenden Signalverarbeitung unterscheiden sich hierbei Informationsquellen in der Wahrscheinlichkeit einer Detektion und der Wahrscheinlichkeit von Falschalarmen. In den folgenden Abschnitten sollen grundlegende Detektionsverfahren diskutiert werden, die sich für den Einsatz von Laserscannern in einer Ebene und Tiefenkameras eignen und dem Schema nach Abbildung 2.4 folgen. Zusätzlich gehen die hier vorgestellten Verfahren davon aus, dass keine Annahme über die Bewegung der Sensoren getroffen werden kann. Somit beruhen alle Verfahren auf Auswertungen der Daten zu nur einem Zeitschritt, was eine uneingeschränkte Nutzung auch bei bewegten Sensoren sicherstellt.



**Abbildung 2.4** Ablauf eines Detektionsprozesses.

### 2.4.1 Bildbasierte Verfahren

Verfahren, die Grau- oder Farbwertbilder auswerten, um Objekte in bestimmten Bildbereichen zu detektieren, beruhen auf der Auswertung bestimmter Merkmale für die entsprechenden Bereiche. Das für die Analyse von potentiellen Personen entwickelte HOG-Merkmal soll daher in den folgenden Abschnitten vorgestellt werden.

#### HOG-Merkmal

*Histograms of Oriented Gradients* (HOG)-Merkmale haben seit ihrer Vorstellung durch Dalal und Triggs [22] eine sehr weite Verbreitung bei der



(a) Farbbild einer gehenden Person.



(b) Kantenwinkel des Grauwertbildes pro Zelle.

**Abbildung 2.5** Histogram of Oriented Gradients (HOG) Merkmal visualisiert über Kanten.

Detektion von Personen erlangt [37], da sie deutliche Vorteile gegenüber der Verwendung anderer Merkmale wie beispielsweise PCA-SIFT [59] oder Haar-Wavelets [89] für die Personendetektion aufweisen [22]. Trotz ihrer hauptsächlich Verwendung auf Grauwertbildern eignen sie sich ebenfalls für den Einsatz auf Tiefenbildern [132]. Das Merkmal an sich verwendet dabei Histogramme von Gradienten. Hierzu wird von Dalal und Triggs [22] die Verwendung eines  $128 \times 64$  Pixel großen Fensters vorgeschlagen, das anschließend in  $8 \times 16$  Zellen der Größe  $8 \times 8$  aufgeteilt wird. Zusätzlich wird eine Normalisierung über jeden der 105 (in der Anordnung  $7 \times 15$ ) Blöcke ausgeführt, die sich jeweils aus vier Zellen zusammensetzen und mit einer jeweiligen Überlappung von einer Zelle, beziehungsweise 8 Pixeln, in der horizontalen und vertikalen angeordnet sind. Abbildung 2.5(b) zeigt die Einteilung in Zellen von 2.5(a). Anschließend wird für jede Zelle ein Histogramm über die Gradientenwinkel erzeugt. Es hat sich gezeigt [22], dass bei der Auswertung von Intensitätsbildern eine betragsmäßige und somit ungerichtete Einteilung der Winkel von  $0^\circ$  bis  $180^\circ$  in neun Klassen für die Personendetektion ausreichend ist, da aus den Intensitäten, im Vergleich zu Tiefendaten [132], in Alltagsszenarien keine Information über Vorder- beziehungsweise Hin-

tergrund zu gewinnen ist. Somit ergibt sich für ein HOG-Merkmal eines  $128 \times 64$ -Bildausschnitts ein  $9 \times 4 \times 105 = 3780$  Elemente umfassender Merkmalsvektor, dessen Dimension sich aus Anzahl an Orientierungen pro Zelle, Zellen pro Block und Anzahl der Blöcke ergibt. Eine anschauliche Darstellung des Merkmals zusammen mit dem zugrundeliegenden Farbbild ist in Abbildung 2.5 zu sehen. Die Klassifikation eines Merkmalsvektors erfolgt mit einem linearen SVM-Klassifikator entsprechend Abschnitt 2.5.

**Verwendung gerichteter Gradienten** Während das soeben beschriebene Standard-HOG-Merkmal zwar eine Einteilung der Gradientenwinkel in neun Klassen von  $0^\circ$  bis  $180^\circ$  durchführt, gibt es auch Anwendungsfälle, in denen die Richtung im Bereich von  $0^\circ$  bis  $360^\circ$  wichtige Informationen liefert. Da eine Erweiterung des Histogramms um weitere neun Klassen eine Verdoppelung des Merkmalsvektors und dem damit verbundenen Rechenaufwand bedeuten würde, wird in [28] eine Abwandlung des HOG-Merkmals vorgeschlagen, die die  $p$  Richtungen der errechneten Gradienten  $\alpha_{\text{grad}}$  entsprechend

$$\alpha_{\text{gerichtet}} = \left\lfloor \frac{p \cdot \alpha_{\text{grad}}}{2\pi} + \frac{1}{2} \right\rfloor \bmod p \quad (2.43)$$

für gerichtete und

$$\alpha_{\text{ungerichtet}} = \left\lfloor \frac{p \cdot \alpha_{\text{grad}}}{\pi} + \frac{1}{2} \right\rfloor \bmod p \quad (2.44)$$

für ungerichtete Klassen bei Verwendung von Abrundungs- und Modulo-Operator einteilt. Zusätzlich wird die Dimensionalität des Merkmals durch eine Hauptkomponentenanalyse derart reduziert, dass das weiterhin aus vier Zellen bestehende, jedoch von  $9 \times 4$  auf  $18 \times 4$  Dimensionen gewachsene Merkmal nach Reduktion 36 Dimensionen hat und somit einen zum ursprünglichen Merkmal vergleichbaren Verarbeitungsaufwand besitzt. Während dieses erweiterte HOG-Merkmal in [28] nur auf unterschiedliche Verläufe von RGB-Bildern angewendet wurde, zeigt sich in dieser Arbeit, dass es sich auch sehr gut für Tiefenbilder einsetzen lässt.



**Abbildung 2.6** Ausschnitt eines Laserscans mit drei Personen neben einer Wand.

## 2.4.2 Laserbasierte Verfahren

Im Gegensatz zu Farbbildern, die sehr viele visuelle Informationen über Personen und andere Objekte liefern, bieten *Light detection and ranging* (LIDAR)-Sensoren, wie sie in der Robotik eingesetzt werden [137], häufig nur eine Abtastung der Umgebung aus Winkel und Entfernung in einer Ebene. Diese eignet sich zwar sehr gut für die Lokalisierung beispielsweise in Gebäuden, vergleiche Abschnitt 2.6, jedoch beschränkt sich die Information über beispielsweise das Vorhandensein von Personen, je nach Entfernung, auf nur wenige Punkte eines Beines, die näherungsweise auf einem Kreisbogen liegen, wie in Abbildung 2.6 gezeigt. Dennoch bieten Laserscanner für gewöhnlich sehr große Sichtfelder zwischen  $180^\circ$  und  $360^\circ$  und liegen somit weit über dem Erfassungsbereich gewöhnlicher Kameras bei hoher Entfernungsgenauigkeit.

Um in den wenigen Informationen dennoch Rückschlüsse über Personen zu schließen, wurde anfangs nach lokalen Minima im Scan gesucht, die zu Detektionen verbunden wurden [126]. Weitere Verbesserungen wurden erzielt, indem die Laserdaten vorab segmentiert wurden [4, 158] Hierbei wird der Reihe nach durch alle Punkte eines Scans iteriert, wobei ein neues Segment entsteht, sobald ein Punkt weiter als ein Schwellwert  $d_{\text{jump}}$ , der sogenannten *jumping distance* [4], von dem vorherigen

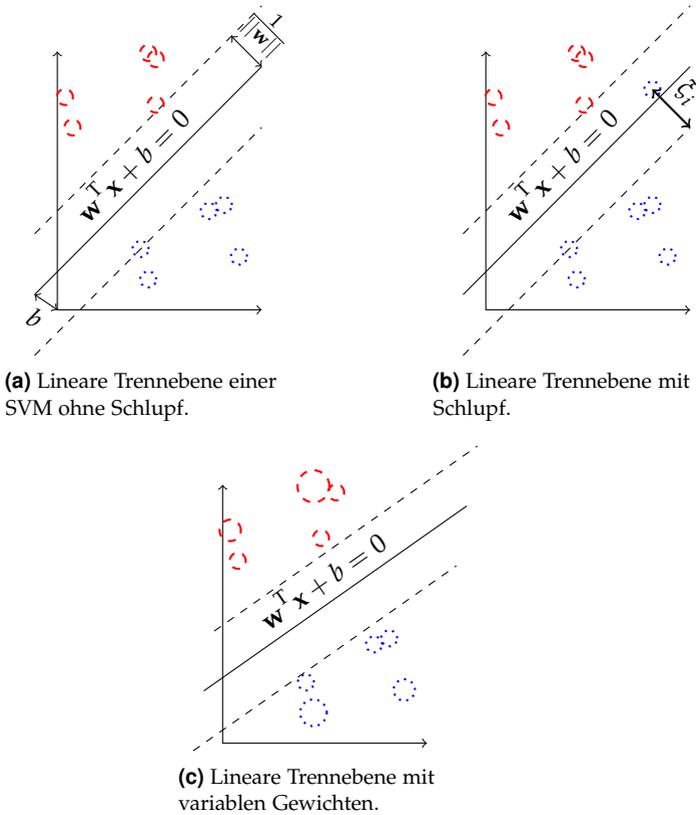
Punkt entfernt ist. Diese liegt für gewöhnlich zwischen 20 cm und 30 cm und lässt sich als Mindestabstand zur klaren Unterscheidung zweier Personen interpretieren. Die Segmente, deren Formen einem Kreisbogen entsprechen, werden anschließend als Detektionen eines Beines gewertet [158]. Eine deutliche robustere Klassifikation wurde von Arras et al. [4] durch die Verwendung einer Sammlung von Merkmalen wie Zirkularität, Linearität, Radius, Standardabweichung und Ähnlichen erreicht. Dieser Ansatz zeichnet sich gerade durch die angelernte Auswahl, Gewichtung und Klassifikation der Merkmale mittels eines Ada-Boost-Verfahrens [33] aus und ist dadurch den vorher genannten Verfahren überlegen [4].

## 2.5 Klassifikation

Teil des Detektionsprozesses ist es, nach der Bestimmung der Merkmale für das ganze Bild oder für bestimmte Regionen jede dieser Regionen daraufhin zu überprüfen, ob das gewonnene Merkmal ein gesuchtes Objekt darstellt oder nicht. Aufgrund der Komplexität der Umwelt und der daraus gewonnenen meist hochdimensionalen Merkmale, reichen einfache regelbasierte Entscheidung oft nicht aus, so dass maschinelle und auf der Wahrscheinlichkeitstheorie basierende Lernverfahren eingesetzt werden, um Klassifikationsfehler zu minimieren. Hier soll ein kurzer Überblick über das hauptsächlich in Kapitel 5 zum Einsatz kommende und für bildbasierte Personenklassifikation übliche [22, 25, 99, 105, 132] Klassifikationsverfahren der *Support Vector Machine* gegeben werden, während für eine generelle Übersicht über Klassifikationsprinzipien auf [12] verwiesen wird.

### 2.5.1 Support Vector Machine

Die *Support Vector Machine* (SVM) nach Vapnik [141] formt das Problem der Zuordnung von Merkmalen zu Klassen in ein konvexes Optimierungsproblem um und findet eine optimale Trennebene [12] unter Ausnutzung der Trainingsbeispiele, was im Folgenden für linear trennbare Daten vorgestellt werden soll.



**Abbildung 2.7** SVM-Trennebenen.

## Linear trennbare Daten

Geht man bei der Suche nach einer optimalen Trennebene von einer linearen Trennbarkeit der Daten aus, so lässt sich eine Entscheidungsfunktion

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2.45)$$

definieren. Dabei beschreibt  $\mathbf{w}$  den Normalenvektor der Trennebene und  $b$  einen Bias. Will man nun bei  $M$  gegebenen Trainingsbeispielen  $\mathbf{x}_i$  mit Klassenzugehörigkeit  $y_i \in \{-1, +1\}$  die optimale Trennebene finden, so soll der kleinste Abstand zwischen Ebene und Trainingsbeispielen unterschiedlicher Klassen maximal werden. Der Abstand ergibt sich durch die beiden Ebenen, die gerade noch alle Trainingsbeispiele trennen

$$\mathbf{w}^T \mathbf{x}_i + b = 1 \quad \text{und} \quad (2.46)$$

$$\mathbf{w}^T \mathbf{x}_i + b = -1 \quad (2.47)$$

zu  $\frac{1}{\|\mathbf{w}\|}$  (vgl. Abbildung 2.7(a)). Als quadratisches Optimierungsproblem lässt sich das Problem umschreiben zu [12]

$$\arg \min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.48)$$

mit der Nebenbedingung für alle  $i$  Trainingsbeispiele

$$y_i (\mathbf{w}^T \mathbf{x}_i - b) \geq 1. \quad (2.49)$$

Die durch Lösen des Optimierungsproblems mittels Lagrange-Multiplikatoren [12] gefundene Ebene stützt sich nur auf die namensgebenden *Support-Vektoren*, für die Gleichung (2.46) und (2.47) gelten und die auf den Rändern der Klassen, entsprechend Abbildung 2.7(a), liegen.

**Trennebene mit Ausreißern** In rauschbehafteten Fällen ist eine lineare scharfe Trennung oft aufgrund von Ausreißern nicht mehr möglich (vgl. Abbildung 2.7(b)). Um dennoch eine Trennebene zu finden, muss das Optimierungsproblem unter Hinzunahme von Schlupfvariablen  $\xi$  neu gestellt werden [142] und ergibt sich zu

$$\arg \min_{\mathbf{w}, \xi, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M \xi_i \quad (2.50)$$

mit dem Gewichtungssparameter  $C$  für den Einfluss von Ausreißern unter der Nebenbedingung

$$y_i(\mathbf{w}^T \mathbf{x}_i - b) \geq 1 - \xi_i, \quad \xi_i \geq 0. \quad (2.51)$$

**Gewichtete Trainingsbeispiele** In vielen Anwendungen kann eine zusätzliche Gewichtung einzelner Trainingsbeispiele das Klassifikationsergebnis verbessern (vgl. Abbildung 2.7(c)). Hierzu kann (2.50) entsprechend [159] um ein für jedes Trainingsbeispiel angepasstes Gewicht  $W_i$  erweitert werden

$$\operatorname{argmin}_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^M W_i \xi_i. \quad (2.52)$$

**Kernelfunktionen** Sollten Trainingsdaten sich nicht linear trennen lassen, so können durch Auswahl einer geeigneten Kernelfunktion  $\phi$  die Trainingsdaten der Dimension  $m$  in einen höherdimensionalen Raum der Dimension  $l$  abgebildet werden

$$\phi : \mathbb{R}^m \rightarrow \mathbb{R}^l, \quad \mathbf{x} \rightarrow \phi(\mathbf{x}). \quad (2.53)$$

Im höherdimensionalen Raum lässt sich eine lineare Entscheidungsfunktion finden mit

$$D_{\text{kernel}}(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b, \quad (2.54)$$

wobei häufig hierfür polynomiale oder Gauß'sche Kernel verwendet werden.

## 2.5.2 Platt Scaling

Wie auch bei SVMs liefern die Ergebnisse von Klassifikatoren oft keine probabilistische Aussage, sondern eigene Maße, wie beispielsweise für SVMs Abstände zu einer Trennebene oder wie bei Ada-Boost-Klassifikatoren Abstände zu Schwellwerten. Um eine weitergehende Beurteilung der Ergebnisse in einem probabilistischem Kontext durchzuführen, ist es daher oftmals sinnvoll, diese meist skalaren Ergebnisse

in eine Wahrscheinlichkeit der Klassenzugehörigkeit umzuformen [12]. In [109] wird dieses Problem über die Verwendung einer angelegten Sigmoidfunktion für SVMs gelöst, die den Ausgangswert des Klassifikators entsprechend skaliert. Dieses Vorgehen kann ebenfalls auf weitere Klassifikatoren wie *Decision Trees* [12] oder *Boosting* [33] angewendet werden [96]. Geht man wie im vorherigen Abschnitt 2.5.1 von dem Wert der Entscheidungsfunktion  $D(\mathbf{x})$  aus, so wird nach Platt [109] bei einem Zweiklassenproblem die Wahrscheinlichkeit des Merkmalsvektors  $\mathbf{x}$ , zur Klasse  $y = 1$  zu gehören, durch die Sigmoidfunktion

$$P(y = 1 | D(\mathbf{x})) = \frac{1}{1 + \exp(AD(\mathbf{x}) + B)} \quad (2.55)$$

bestimmt. Hierbei sind  $A$  und  $B$  die zu schätzenden Parameter der Funktion. Anhand eines Trainingsdatensatzes  $D(\mathbf{x}_i, t_i)$  mit

$$t_i = \frac{y_i + 1}{2} \quad (2.56)$$

lassen sich die Parameter  $A$  und  $B$  anhand der *Maximum-Likelihood-Methode* [63] durch Minimierung von

$$\operatorname{argmin}_{A,B} \left( - \sum_i t_i \log(p_i) + (1 - t_i) \log(1 - p_i) \right) \quad (2.57)$$

$$\text{mit } p_i = \frac{1}{1 + \exp(AD(\mathbf{x}_i) + B)} \quad (2.58)$$

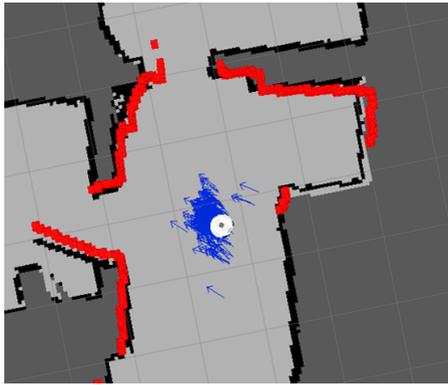
finden [109].

## 2.6 Lokalisierung

Existiert bereits eine Karte der statischen Umgebung, die beispielsweise unter Einsatz eines *Simultaneous Localization And Mapping* (SLAM)-Verfahrens entstanden ist [44, 90], sind Verfahren nötig, die die Lokalisierung einer Sensorplattform auf dieser Karte ermöglichen. Für die Lokalisierung in Gebäuden ist die Verwendung einer auf einem Gitter (*Grid*) basierenden Karte sinnvoll [157], da wenige markante Landmarken existieren, die für eine landmarkenbasierte Karte nötig wären [90].

Somit ist bei der Verwendung eines sogenannten *Occupancy-Grids*, also eines Belegungsgritters auf einer Karte, wie in Abbildung 2.8, für jeden Gitterpunkt festgelegt, ob er erreichbar, belegt oder unerreichbar ist.

Das Lokalisierungsproblem lässt sich ebenfalls als Zustandsschätzproblem beschreiben, mit Abgleich von Messdaten und Karte, vorzugsweise über Laserscanner, unbekannter Position und bekannten Steuerungsbeehlen. Soll auch eine globale Lokalisierung ohne gegebene Initialposition möglich sein, so werden Filter benötigt, die multimodale Verteilungen abbilden können. Daher bieten sich für diesen Fall Monte-Carlo-Methoden aus Abschnitt 2.3.4 an. Thrun et al. [138] schlagen eine robuste Methode vor, die für die Verwendung mit Laserscannern eine sehr weite Verbreitung gefunden hat [137] und in Abbildung 2.8 gezeigt ist. Darüber hinaus existieren auch Methoden für eine dreidimensionale Lokalisierung [24], die in dem hier verwendeten Szenario jedoch nicht notwendig sind.

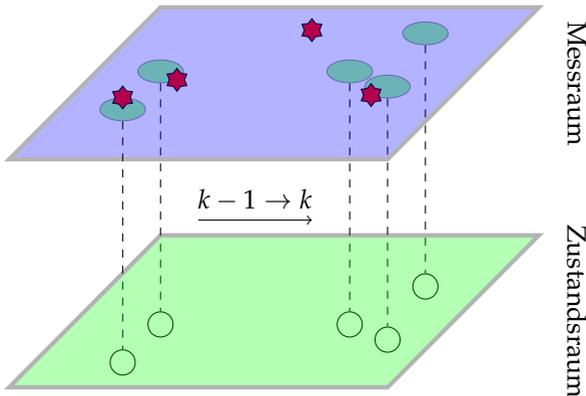


**Abbildung 2.8** Monte-Carlo-Lokalisierung eines Sensorträgers (weiß) auf gegebener Gebäude-Karte in Draufsicht. Die Laserdaten sind in Rot, die einzelnen Partikel in Blau, erreichbare Bereiche in Hellgrau, unerreichbare Bereiche in Dunkelgrau und Hindernisse in Schwarz gekennzeichnet.

## 3 Multi-Objekt-Filterung

Während im vorangegangenen Kapitel mit der Einführung von Bewegungsmodellen (Abschnitt 2.1), Sensormodellen (Abschnitt 2.2) und der Bayes-Rekursion (Abschnitt 2.3) Grundlagen für die Zustandsschätzung eines Objektes präsentiert wurden, werden in diesem Kapitel Verfahren und Modelle vorgestellt, die eine Verallgemeinerung auf den Fall von mehreren Objekten leisten. Im Multi-Objekt-Fall wird das Problem der Schätzung von Zuständen eines Objekts wie Position oder Geschwindigkeit dahingehend erschwert, dass nun weiterhin Zustände möglichst exakt geschätzt werden sollen, aber die Kardinalität, also die Anzahl der Objekte im Zustandsraum, beliebig zwischen keinem vorhandenen Objekt und einer großen Zahl an Objekten variieren kann und unbekannt ist. Ebenso muss diese Kardinalität über die Zeit nicht konstant bleiben, sondern kann sich von einem Zeitschritt  $k - 1$  bis zu einem nächsten Zeitschritt  $k$  verändern, so dass neue Objekte in einer beobachteten Szene erscheinen oder diese verlassen können (siehe Abbildung 3.1). Allgemein sind gerade im militärischen Bereich auch Fälle von Verschmelzen oder dem Aufteilen von Objekten denkbar, die jedoch in Anbetracht des fehlenden Bezugs zur Anwendung der Personenverfolgung in dieser Arbeit vernachlässigt werden sollen. Hierfür wird auf Literatur verwiesen [7, 79]. Zusätzlich muss in der Multi-Objekt-Filterung berücksichtigt werden, dass eine unbekannte Anzahl an Objekten auch eine variable Anzahl an Messungen erzeugt. So sind unterbliebene Detektionen (Fehldetektionen/Schlupf) und ebenso Falschalarne, bei denen beispielsweise ein unerwünschter Hintergrund als Objekt erkannt wurde, aufgrund von Fehlern im Detektionsprozess möglich. In Abbildung 3.1 wird diese Problemstellung verdeutlicht. Hierbei sind Objekte im Zustandsraum durch Kreise dargestellt mit entsprechender elliptischer Unsicherheit der Messung im Messraum und tatsächlichen Messungen in Sternform. Es ist zu erkennen, dass eine eindeutige Zuordnung von Messung zu Objekt, die sogenannte Datenassoziation, gerade in Fällen ausbleibender Messungen

oder beim Auftreten von Falschalarmen eine große Herausforderung darstellt [135] und somit sehr fehleranfällig sein kann.



**Abbildung 3.1** Veranschaulichung des Multi-Objekt-Falls von Zeitschritt  $k-1$  (links) zu Zeitschritt  $k$  (rechts) mit einem neugeborenen Objekt. Hierbei sind Objekte als Kreise mit elliptischen Unsicherheiten im Messraum dargestellt. Messungen sind als Sterne gezeigt.

### 3.1 Endliche Mengen

Während klassische Methoden versuchen, das Datenassoziationsproblem zu lösen, um Bayes-Konzepte aus dem Ein-Objekt-Fall anzuwenden [94, 116], führt die Suche nach einem optimalen Multi-Objekt-Bayes-Filter zu Konzepten, die auf der Auffassung des Multi-Objekt-Problems als Punktprozess [23] basieren und eine Alternative zu klassischen Verfahren bieten [103]. Hierdurch motiviert, schlug Mahler [76] die Darstellung von multiplen Objekte und Messungen als endliche Mengen vor. Diese Darstellung hat den Vorteil, dass keine Reihenfolge der Objekte zu einem Zeitpunkt besteht, die beispielsweise bei der Darstellung über einen großen Multi-Objekt-Vektor vorhanden wäre. Des Weiteren existiert auch eine Darstellung der leeren Menge, falls kein Objekt vorhanden ist. Als einzige Anforderung an die Menge verbleibt, dass die variable Objektanzahl endlich ist. Im Folgenden beschreibt der mengenwertige Zustand  $\mathcal{X}_k$

zum Zeitpunkte  $k$  die Menge aus  $N(k)$  Objekten mit Zustandsvektoren  $\mathbf{x}_{k,n}$ , die den Zustandsvektoren aus dem Ein-Objekt-Fall entsprechen,

$$\mathcal{X}_k = \{\mathbf{x}_{k,1}, \mathbf{x}_{k,2}, \dots, \mathbf{x}_{k,N(k)}\}, \quad (3.1)$$

sowie die mengenwertige Messung  $\mathcal{X}_k$  eines Zeitschritts  $k$  die Menge an  $M(k)$  Messungen mit den Einzel-Messvektoren  $\mathbf{z}_{k,m}$

$$\mathcal{Z}_k = \{\mathbf{z}_{k,1}, \mathbf{z}_{k,2}, \dots, \mathbf{z}_{k,M(k)}\}. \quad (3.2)$$

Die Definitionen der wichtigsten Klassen von endlichen Mengen sind im Anhang B zu finden.

## Statistik endlicher Mengen

Aus der Repräsentation der zu verfolgenden Objekte über endliche Mengen entstand, basierend auf der Lehre der Punktprozesse, die Statistik endlicher Mengen (*Finite Set Statistics* (FiSSt)). Ebenfalls durch Mahler wurde hierbei versucht, das mathematisch etablierte Konzept der Punktprozesse auf die Anforderungen im Bereich der physikalisch motivierten Multi-Objekt-Verfolgung herunterzubrechen und hierfür anwendbar zu machen [81]. Dieser Ansatz ermöglichte es in den folgenden Jahren auch Filter herzuleiten, die der Vision eines Multi-Objekt-Kalman-Filters [76] entsprechen. Dieses Multi-Objekt-Kalman-Filter soll somit eine möglichst optimale und gleichzeitige effiziente Lösung des Schätzproblems für den Multi-Objekt-Fall darstellen. Geeignete Ansätze und darauf abzielende Näherungen sollen nach einer kurzen Einführung in die Statistik endlicher Mengen in den nachfolgenden Abschnitten diskutiert werden.

Neben der Definition der endlichen Mengen in Abschnitt 3.1 beinhaltet FiSSt noch weitere Definitionen, die die Herleitung verschiedener Filter für die Anwendung ermöglichen. Hierzu soll nur ein kurzer Überblick gegeben werden, der die nötigsten Schreibweisen, Rechenvorschriften und Ansätze beinhaltet. Eine umfassende Beschreibung ist in [79] zu finden.

**Multi-Objekt-Dichte** Wurde im Ein-Objekt-Fall die Dichte eines Zustands, wie in (2.21), noch in Abhängigkeit eines Zustandsvektors  $\mathbf{x}$  bedingt auf einen Vektor aus Messungen  $\mathbf{Z}$  angegeben, so haben Dichten

im Multi-Objekt-Fall als Argument eine endliche Menge an Objekten  $\mathcal{X}$  und sind bedingt auf eine ebenfalls endliche Menge an Messungen  $\mathcal{Z}$ . Somit ergibt sich die Multi-Objekt-Dichte eines bereits mit aktuellen Messungen korrigierten Bayes-Filters entsprechend Abschnitt 2.3.2 zum Zeitschritt  $k$  zu

$$f_{k|k}(\mathcal{X}_k | \mathcal{Z}_k), \quad (3.3)$$

ohne hierbei schon näher auf die Gewinnung einer solchen Dichte einzugehen (Abschnitt 3.3).

**Mengen-Integral** Um mit Multi-Objekt-Dichten der Form (3.3) rechnen zu können, ist für viele Anwendungen, wie die Bildung von Momenten oder die Berechnung einer Wahrscheinlichkeiten für Teilbereiche, die Existenz eines Integrals über Mengen nötig. Mahler [79] adaptiert hierfür aus der statistischen Theorie der Gase und Flüssigkeiten das Mengen-Integral für die Verwendung in Multi-Objekt-Fällen. Wird für eine allgemeine Multi-Objekt-Dichte  $f(\mathcal{Y})$  einer beliebigen Menge  $\mathcal{Y}$  somit das entsprechende Integral über eine Region  $\mathcal{S}$  des Zustandsraums integriert, lässt sich das Integral entsprechend folgender Definition berechnen

$$\begin{aligned} \int_{\mathcal{S}} f(\mathcal{Y}) \delta\mathcal{Y} &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_{\underbrace{\mathcal{S} \times \dots \times \mathcal{S}}_n} f(\{\mathbf{y}_1, \dots, \mathbf{y}_n\}) \, d\mathbf{y}_1 \cdots d\mathbf{y}_n \quad (3.4) \\ &= f(\emptyset) + \int_{\mathcal{S}} f(\{\mathbf{y}_1\}) \, d\mathbf{y}_1 \\ &\quad + \frac{1}{2} \int_{\mathcal{S} \times \mathcal{S}} f(\{\mathbf{y}_1, \mathbf{y}_2\}) \, d\mathbf{y}_1 d\mathbf{y}_2 + \dots, \end{aligned} \quad (3.5)$$

wobei die Definition

$$\int_{\underbrace{\mathcal{S} \times \dots \times \mathcal{S}}_n} f(\{\mathbf{y}_1, \dots, \mathbf{y}_n\}) \, d\mathbf{y}_1 \cdots d\mathbf{y}_n \quad (3.6)$$

$$= n! \int_{\underbrace{\mathcal{S} \times \dots \times \mathcal{S}}_n} f(\mathbf{y}_1, \dots, \mathbf{y}_n) \, d\mathbf{y}_1 \cdots d\mathbf{y}_n \quad (3.7)$$

alle kombinatorischen Möglichkeiten einer Menge mit  $n$  Elementen ohne vorgegebene Reihenfolge berücksichtigt. In (3.4) deutet der Operator  $\delta$  auf die Verwendung des Mengen-Integrals im Gegensatz zu dem im Ein-Objekt-Fall üblichen Integral über den Zustandsraum mit dem Operator  $d$  hin.

Anschaulich lässt sich das Mengenintegral (3.4) als Summe der Wahrscheinlichkeiten aller Zustandsvektoren  $\mathbf{y}$  der Menge  $\mathcal{Y}$  interpretieren, wobei jeder Fall von  $n$  Objekten – und somit kein Objekt, ein Objekt, zwei Objekte usw. – seinen Beitrag zu der Wahrscheinlichkeit dieser Menge in der Region  $S$  beiträgt.

### Wahrscheinlichkeitserzeugende Funktionen und Funktionale

Während Integraltransformationen wie die Fourier- oder Laplace-Transformation zum Handwerkszeug jedes Ingenieurs gehören [79], sind auch solche Transformationen in der Statistik ein probates Mittel, um aufwändige Rechnungen zu vereinfachen. Ebenso lassen sich Konzepte wie charakteristische Funktionen oder momenterzeugende Funktionen auch auf die Analyse von endlichen Mengen und Mengendichten erweitern. So lässt sich beispielhaft für die Poisson-Verteilung einer Zufallsvariable  $x$  um Mittelwert  $\lambda$  mit nicht negativem reellen Parameter  $s$  die wahrscheinlichkeitserzeugende Funktion

$$G_x(s) = \exp(\lambda(s - 1)) \quad (3.8)$$

angeben. Wie zu erwarten, führt eine Ableitung nach  $s$  und Auswertung für  $s = 1$  zum ersten faktoriellen Moment, dem Mittelwert  $\lambda$  der Verteilung, wobei entsprechend eine  $r$ -te Ableitung zum  $r$ -ten faktoriellen Moment führt. Des Weiteren lässt sich auch die Wahrscheinlichkeit des Ereignisses  $x = r$  durch  $r$ -maliges Ableiten nach  $s$  ausgewertet an der Stelle  $s = 0$  mit der Normierung  $r!$  errechnen. Die Transformation hin zu wahrscheinlichkeitserzeugenden Funktionen beschreibt somit eine elegante Möglichkeit, Momente und Wahrscheinlichkeiten zu erzeugen.

Erweitert auf das Beispiel eines Poisson-Prozesses, bei dem der Parameter  $s$  nun keinen Skalar, sondern eine Funktion, beispielsweise eine räumliche Verteilung der Poisson-verteilten Anzahl von Objekten auf dem Zustandsraum, beschreibt, lässt sich zusammen mit der funktiona-

len Variable  $h$  das folgende wahrscheinlichkeitserzeugende Funktional finden

$$G_x(h) = \exp(\lambda(f(h) - 1)) \quad (3.9)$$

mit dem linearen Funktional

$$f(h) = \int h(x)s(x) dx. \quad (3.10)$$

Auch in diesem Fall lassen sich nun durch entsprechende Funktionalableitungen Momente und Wahrscheinlichkeiten bestimmter Zustände berechnen. Zusätzlich besitzen die erzeugenden Funktionen und Funktionale die Eigenschaft, dass die Summe zweier unabhängiger Zufallsvariablen/Zufallsprozesse als Produkt ihrer erzeugenden Funktionen/Funktionale beschrieben werden kann. Für endliche Mengen definiert sich ein wahrscheinlichkeitserzeugendes Funktional einer Mengendichte  $f(\mathcal{Y})$  im Falle vektorwertiger Zustände zu [79]

$$G_f[h] = \int h^{\mathcal{Y}} \cdot f(\mathcal{Y}) \delta\mathcal{Y}, \quad (3.11)$$

wobei  $h^{\mathcal{Y}} = 1$  für die leere Menge  $\mathcal{Y} = \emptyset$  und sonst  $h^{\mathcal{Y}} = \prod_{\mathbf{y} \in \mathcal{Y}} h(\mathbf{y})$  ist.

Für entsprechende Mengen-Ableitungen sei auf die Regeln in [79, Seiten 386 ff.] verwiesen.

## 3.2 Multi-Objekt-Modelle

In Abschnitt 2.1 wurde für den Ein-Objekt-Fall diskutiert, wie die Bewegung eines einzelnen Objektes modelliert werden kann. Für den Multi-Objekt-Fall muss nun zusätzlich berücksichtigt werden, dass mehrere Objekte vorhanden sein können, neu eine Szene betreten und auch verlassen können. Hier sollen entsprechend Mahler [79] und Vo [144] Modelle vorgestellt werden, die bei der Filtersynthese mit endlichen Mengen zum Einsatz kommen.

**Bewegungsmodell** Geht man davon aus, dass ein Objekt im letzten Zeitschritt  $t_{k-1}$  mit dem Zustand  $\mathbf{x}_{k-1}$  Element der Menge  $\mathcal{X}_{k-1}$  war,

lässt sich das weitere Existieren des Objektes im Zeitpunkt  $t_k$  mit der Wahrscheinlichkeit  $p_{S,k}(\mathbf{x}_{k-1})$  beschreiben. Somit handelt es sich um eine Bernoulli-verteilte Überlebenswahrscheinlichkeit (*probability of survival*  $p_S$ ) mit der Wahrscheinlichkeit, dass ein Objekt verschwindet von  $1 - p_S(\mathbf{x})$ , womit die Überlebenswahrscheinlichkeit auch abhängig von Position bzw. Zustand des Objektes im Intervall  $[0; 1]$  variieren kann. Neben der Wahrscheinlichkeit der Existenz eines Objektes wird zusätzlich sein räumlicher Zustand durch die dem Ein-Objekt-Fall analoge Transitionsdichte  $f_{k|k-1}(\mathbf{x}_k|\mathbf{x}_{k-1})$  beschrieben. Zusammengefasst lässt sich jedes zum letzten Zeitpunkt bestehende Objekt durch die Bernoulli-Zufallsmenge (siehe Anhang B.3)  $\mathcal{S}_{k|k-1}(\mathbf{x}_{k-1})$  beschreiben und enthält mit Wahrscheinlichkeit  $p_{S,k}$  ein entsprechend der Transitionsdichte verteiltes Objekt.

Eine sogenannte Geburt, das Erscheinen eines neuen Objekts, lässt sich ebenfalls durch eine Zufallsmenge  $\mathcal{B}$  beschreiben und wird als unabhängig von vorherigen Objekten angenommen.

Daraus folgt das Modell für die endliche Objekt-Menge im aktuellen Zeitschritt aus der Vereinigung aller weiterbestehenden und neuentstandenen Objekte (vergleiche Abbildung 3.1)

$$\mathcal{X}_k = \left[ \bigcup_{\mathbf{x}_{k-1} \in \mathcal{X}_{k-1}} \mathcal{S}_{k|k-1}(\mathbf{x}_{k-1}) \right] \cup \mathcal{B}_k. \quad (3.12)$$

**Sensormodell** Da Messungen im Multi-Objekt-Fall aufgrund nicht detektierter Objekte ausfallen können und auch fälschlich detektierte Objekte zu Falschalarmen (*clutter*) führen, ist ebenfalls eine Anpassung des Sensormodells nötig.

Ähnlich der Überlebenswahrscheinlichkeit lässt sich das Detektionsverhalten eines Systems durch eine Bernoulli-verteilte Detektionswahrscheinlichkeit  $p_{D,k}(\mathbf{x}_k)$  beschreiben. Hierbei wird davon ausgegangen, dass ein Objekt mit Zustand  $\mathbf{x}_k$  genau eine Messung  $\mathbf{z}_k$  mit möglicherweise zustandsabhängiger Detektionswahrscheinlichkeit  $p_{D,k}(\mathbf{x}_k)$  generiert und die Messung die auf einen Objektzustand  $\mathbf{x}_k$  bedingte Likelihood  $g_k(\mathbf{z}_k|\mathbf{x}_k)$  besitzt. Diese Likelihood kann dann wiederum einem Ein-Objekt-Sensor-Modell aus Abschnitt 2.2 entsprechen. Somit ergeben sich

alle Messungen detektierter Objekte zu einem Zeitschritt zu der Menge  $\mathcal{D}_k(\mathbf{x}_k)$ .

Neben der Wahrscheinlichkeit, ein Objekt zu detektieren, werden Falschalarne unabhängig davon oft als in der Anzahl Poisson-verteilte endliche Menge  $\mathcal{K}_k$  modelliert. Die Messungen zum Zeitpunkt  $k$  ergeben sich somit zu der Vereinigung aus detektierten Objekten und Falschalarmen zu

$$\mathcal{Z}_k = \left[ \bigcup_{\mathbf{x}_k \in \mathcal{X}_k} \mathcal{D}_k(\mathbf{x}_k) \right] \cup \mathcal{K}_k. \quad (3.13)$$

### 3.3 Multi-Objekt-Bayes-Filter

Aufgrund der Repräsentation und der Modelle aus den Abschnitten 3.1 und 3.2 kann wie auch für den Ein-Objekt-Fall nun durch die Verwendung der totalen Wahrscheinlichkeit und der Statistik endlicher Mengen aus dem vorherigen Zeitschritt eine A-priori-Dichte prädiert werden [79]

$$f_{k|k-1}(\mathcal{X}|\mathcal{Z}_{1:k-1}) = \int f_{k|k-1}(\mathcal{X}|\mathcal{X}') \cdot f_{k-1|k-1}(\mathcal{X}'|\mathcal{Z}_{1:k-1}) \delta\mathcal{X}'. \quad (3.14)$$

Durch die Verwendung des Satzes von Bayes ergibt sich die A-posteriori-Dichte, korrigiert durch die Messungen zum Zeitpunkt  $k$ , zu [79]

$$f_{k|k}(\mathcal{X}|\mathcal{Z}_k) = \frac{f_k(\mathcal{Z}_k|\mathcal{X}) \cdot f_{k|k-1}(\mathcal{X}|\mathcal{Z}_{1:k-1})}{f_k(\mathcal{Z}_k|\mathcal{Z}_{1:k-1})} \quad (3.15)$$

mit dem Normalisierungsterm der Wahrscheinlichkeit neuer Messungen gegeben der alten

$$f_k(\mathcal{Z}_k|\mathcal{Z}_{1:k-1}) = \int f_k(\mathcal{Z}_k|\mathcal{X}) \cdot f_{k|k-1}(\mathcal{X}|\mathcal{Z}_{1:k-1}) \delta\mathcal{X}. \quad (3.16)$$

Nach [79] lassen sich diese elementaren Gleichungen ebenfalls in eine Darstellung als wahrscheinlichkeitserzeugende Funktionale transformie-

ren, so dass ihre Entsprechungen für A-priori-Dichte

$$G_{k|k-1}[h] = \int \left( \int h^{\mathcal{X}} \cdot f_{k|k-1}(\mathcal{X}|\mathcal{X}') \delta\mathcal{X} \right) f_{k-1|k-1}(\mathcal{X}'|\mathcal{Z}_{1:k-1}) \delta\mathcal{X}' \quad (3.17)$$

und A-posteriori-Dichte

$$G_{k|k}[h] = \frac{\frac{\delta F}{\delta \mathcal{Z}_k}[0, h]}{\frac{\delta F}{\delta \mathcal{Z}_k[0,1]}} \quad (3.18)$$

lauten. Hierbei beschreibt als Grundelement  $F[g, h]$ , aus dem sich durch funktionale Ableitung das Ergebnis der Bayes-Formel ergibt, nun das von zufälligen Messungen und Objekten abhängige Funktional (Beweis in [79])

$$F[g, h] = \int h^{\mathcal{X}} \cdot \left( \int g^{\mathcal{Z}} \cdot f_k(\mathcal{Z}|\mathcal{X}) \delta\mathcal{Z} \right) \cdot f_{k|k-1}(\mathcal{X}) \delta\mathcal{X}. \quad (3.19)$$

### 3.4 Probability Hypothesis Density-Filter

Während traditionelle Multi-Objekt-Verfolgungsfiler eher dem Ansatz folgen, die Frage zu beantworten, wo sich, unter der Annahme einer bestimmten Anzahl, Objekte befinden, lässt sich das *Probability Hypothesis Density*(PHD)-Filter auch entsprechend Erdinc et al. [26] als Antwort auf die Frage „Ist an einem bestimmten Punkt ein Objekt?“ interpretieren. Das PHD-Filter nach Mahler [77] stellt somit eine skalierte Dichte, oder auch Intensität, auf dem Zustandsraum dar, die die räumliche Aufenthaltswahrscheinlichkeit von Objekten beschreibt, wobei das Integral über diese Intensität eine Schätzung der Objektanzahl angibt. Die Komplexität des Filters ist linear in der Objektanzahl  $n$  und der Anzahl der Messungen  $m$  mit  $\mathcal{O}(nm)$  [80].

Die Herleitung der Dichte und des resultierenden Filters beruht auf Approximation des Multi-Objekt-Bayes-Filters aus Abschnitt 3.3. Dies kann zum einen durch die Verwendung der durch Punktprozesse motivierten Statistik endlicher Mengen (Abschnitt 3.1) geschehen, aber auch alternativ durch die Aufteilung des Zustandsraums in diskrete Zellen,

wobei hier die Aufenthaltswahrscheinlichkeit pro Zelle berechnet wird [26]. Der Grenzwert von unendlich kleinen Zellen entspricht dabei exakt der Definition des PHD-Filters. Um genauer auf die in beiden Herleitungen impliziten Annahmen einzugehen, soll hier kurz der Ansatz unter Verwendung von FiSSt nach Mahler [77] verdeutlicht werden. Analog zu Abschnitt 3.1 kann die PHD  $D(\mathbf{x})$  als erstes Moment der Multi-Objekt-Bayes-Dichte aufgefasst werden und ergibt sich zur Intensitätsfunktion [79]

$$D(\mathbf{x}) = \int \delta_{\mathcal{X}}(\mathbf{x}) \cdot f(\mathcal{X}) \delta \mathcal{X} \quad (3.20)$$

bestehend aus der Multi-Objekt-Dichte  $f(\mathcal{X})$ , einem Mengenintegral gekennzeichnet durch den Operator  $\delta \mathcal{X}$  und der für eine Beschreibung im Zustandsraum nötigen Transformation  $\delta_{\mathcal{X}}(\mathbf{x})$  von Menge zu Vektor [23] mit  $\delta_{\mathcal{X}}(\mathbf{x}) = 0$  falls  $\mathcal{X} = \emptyset$  und sonst

$$\delta_{\mathcal{X}}(\mathbf{x}) = \sum_{\mathbf{w} \in \mathcal{X}} \delta_{\mathbf{w}}(\mathbf{x}) \quad (3.21)$$

mit  $\delta_{\mathbf{w}}(\mathbf{x})$  als Dirac-Dichte um einen Vektor  $\mathbf{w}$ . Ebenso lässt sich die PHD aus dem entsprechenden wahrscheinlichkeitserzeugenden Funktional der Multi-Objekt-Dichte  $G[h]$  durch Differentiation gewinnen

$$D(\mathbf{x}) = \frac{\delta G[h]}{\delta \mathbf{x}} [1]. \quad (3.22)$$

Hieraus wird ersichtlich, dass die Annahmen des Filters somit in der Beschaffenheit des Funktionals  $G[h]$  stecken und somit sowohl Intensitätsfunktion als auch Prädiktions- bzw. Korrekturgleichungen beeinflussen.

**Prädiktion** Unter der Annahme, dass einzelne Objekte sich unabhängig voneinander bewegen, neue Objekte entsprechend Abschnitt 3.2 unabhängig und Poisson-verteilt entstehen und sich die endliche Zufallsmenge des Mehrobjekt-Zustands durch eine Poisson-Zufallsmenge annähern lässt, kann das Funktional  $G_{k|k-1}$  zum prädizierten Zeitpunkt  $k$  als Produkt des Funktionals der neugeborenen Objekte und der durch das Bewegungsmodell prädizierten Objekte dargestellt werden [79] und

sich daraus die Prädiktionsgleichung der Intensitätsfunktion  $D_{k|k-1}(\mathbf{x}_k)$ <sup>1</sup> mit

$$\begin{aligned} D_{k|k-1}(\mathbf{x}_k) &= \frac{\delta G_{k|k-1}[h]}{\delta \mathbf{x}} [1] \\ &= b_k(\mathbf{x}_k) \\ &\quad + \int p_{S,k}(\mathbf{x}_{k-1}) f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}) D_{k-1|k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1} \end{aligned} \quad (3.23)$$

ableiten. Hierbei stellt  $b_k(\mathbf{x}_k)$  die Geburtendichte,  $p_{S,k}(\mathbf{x}_{k-1})$  die zustandsabhängige Überlebenswahrscheinlichkeit und  $f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1})$  die Transitionsdichte dar.

**Korrektur** Im Korrekturschritt wird zusätzlich angenommen, dass Objekte entsprechend dem Multi-Objekt-Messmodell unabhängig voneinander Messungen erzeugen und die endliche Menge an Falschalarme ebenfalls Poisson-verteilt ist. Nach [77] erhält man somit

$$\begin{aligned} D_{k|k}(\mathbf{x}_k) &= (1 - p_{D,k}(\mathbf{x}_k)) D_{k|k-1}(\mathbf{x}_k) \\ &\quad + \sum_{\mathbf{z}_k \in \mathcal{Z}_k} \frac{p_{D,k}(\mathbf{x}_k) g_k(\mathbf{z}_k | \mathbf{x}_k) D_{k|k-1}(\mathbf{x}_k)}{\kappa_k(\mathbf{z}_k) + \int p_{D,k}(\mathbf{x}_k) g_k(\mathbf{z}_k | \mathbf{x}_k) D_{k|k-1}(\mathbf{x}_k) d\mathbf{x}_k}. \end{aligned} \quad (3.24)$$

Hierbei stellt  $p_{D,k}(\mathbf{x}_k)$  die zustandsabhängige Detektionswahrscheinlichkeit,  $g_k(\mathbf{z}_k | \mathbf{x}_k)$  die Likelihood einer Messung  $\mathbf{z}_k$  und  $\kappa_k(\mathbf{z}_k)$  die Intensitätsfunktion der Falschalarme dar.

Eine Schätzung  $\hat{N}_k$  der wahren Anzahl  $N_k$  der Objekte zum Zeitpunkt  $k$  lässt sich aus der Intensitätsfunktion  $D_{k|k}(\mathbf{x}_k)$  durch Integration über den beobachteten Bereich  $\mathbb{X}$  des Zustandsraums gewinnen:

$$\hat{N}_k = \int_{\mathbb{X}} D_{k|k}(\mathbf{x}_k) d\mathbf{x}_k. \quad (3.25)$$

**Implementierung** Zwar erhält man durch die Approximation durch das erste Moment und die Herleitung der Filtergleichungen bereits entsprechende Filtergleichungen, jedoch sind numerische Ansätze wie das

<sup>1</sup> Die Schreibweisen  $D_{k|k-1}(\mathbf{x}_k)$  und  $D_{k|k}(\mathbf{x}_k)$  werden von hier an als Abkürzung für die auf die vergangenen Messungen bedingten Intensitätsfunktionen  $D_{k|k-1}(\mathbf{x}_k | \mathcal{Z}_{1:k-1})$  bzw.  $D_{k|k}(\mathbf{x}_k | \mathcal{Z}_{1:k})$  verwendet.

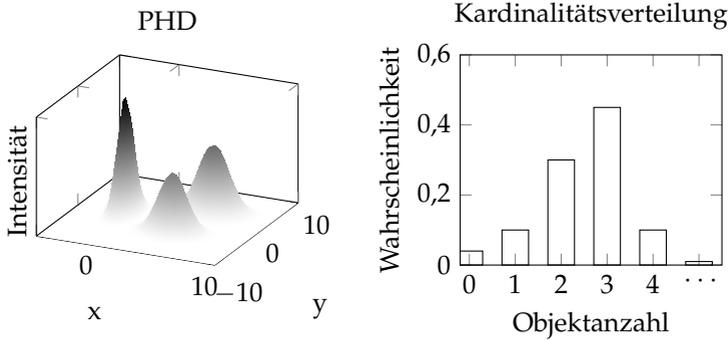
Partikel-PHD-Filter (SMC-PHD) [79, 119, 156] nötig, da keine geschlossene Lösung existiert. Durch weitere Annahmen einer Gauß-Darstellung der Intensitätsfunktion ist eine geschlossene Lösung, wie in [146], möglich. Die geschlossene Lösung wird als Gaussian-Mixture-PHD-Filter (GM-PHD) bezeichnet und angewendet [45, 50, 70]. Durch Verwendung der Unscented-Transformation (siehe Abschnitt 2.3.3) wird hierbei auch die Verwendung nichtlinearer Modelle ermöglicht (vergleiche Abschnitt 4.5.1).

Während das ursprüngliche Bewegungsmodell einen Geburtenprozess vorsieht, der allein auf den Modellannahmen von  $\mathcal{B}$  beruht, wurde in [119] ein Ansatz zur Effizienzsteigerung der PHD-Filter vorgestellt, der Geburten in Abhängigkeit der eintreffenden Messungen erzeugt und hier ebenfalls verwendet wird.

### **3.5 Cardinalized Probability Hypothesis Density-Filter**

Durch die Annahme einer Poisson-verteiltern endlichen Menge für die Beschreibung der PHD wird ersichtlich, dass die Kardinalität der PHD und somit die Schätzung der Anzahl der Objekte in dem betrachteten Zustandsraum einer Poisson-Verteilung unterliegt. Dies führt zu einem beträchtlichen Informationsverlust im Vergleich zu dem Multi-Objekt-Bayes-Filter, da die Verteilung der Kardinalität auf einen einzigen Parameter reduziert wird. Ebenso bedeutet dies, dass die Unsicherheit der Kardinalität starr an die Objektanzahl gebunden ist und somit für viele Objekte entsprechend ansteigt. Um diesen Informationsverlust auszugleichen, wurde das PHD-Filter durch Mahler [78] um eine zusätzliche Verteilung der Kardinalität erweitert, die auch höhere Momente als das erste Moment zulässt (vergleiche Abbildung 3.2). Hierbei basiert die zur PHD führende Approximation nicht mehr auf einer Poisson-verteiltern endlichen Menge, sondern auf einer endlichen Menge unabhängig und identisch verteilter Elemente ohne zwingend Poisson-verteilte Kardinalität. Dies gilt nicht nur für die zu schätzende PHD, sondern auch für die Mengen an Falschalarmen und an neugeborenen Objekten. Weiterhin bleibt die Annahme unabhängiger Objektbewegungen und ebenfalls

unabhängiger Messungen bestehen [144]. Die Komplexität des Filters ist linear in Anzahl der Objekte  $n$  und kubisch in der Anzahl der Messungen  $m$  mit  $\mathcal{O}(nm^3)$  [80].



**Abbildung 3.2** Darstellung der PHD-Intensitätsfunktion mit Kardinalitätsverteilung des CPHD-Filters.

**Prädiktion** Zusätzlich zur Gleichung der prädizierten PHD  $D_{k|k-1}(\mathbf{x}_k)$  wird beim CPHD-Filter auch eine Gleichung für die diskrete Wahrscheinlichkeit von  $n$  Objekten in Form von  $\rho_{k|k-1}(n)$  verwendet. Somit ergeben sich die Formeln der Prädiktion nach [148] zu<sup>2</sup>

$$\rho_{k|k-1}(n) = \sum_{j=0}^n \rho_{B,k}(n-j) \Pi_{k|k-1}[D_{k-1}(\mathbf{x}_{k-1}), \rho_{k-1}](j) \quad (3.26)$$

$$D_{k|k-1}(\mathbf{x}_k) = b_k(\mathbf{x}_k) + \int p_{S,k}(\mathbf{x}_{k-1}) f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}) D_{k-1}(\mathbf{x}_{k-1}) d\mathbf{x}_{k-1}, \quad (3.27)$$

<sup>2</sup> Die Schreibweisen  $\rho_k(n)$  oder  $D_{k-1}(\mathbf{x})$  sind beispielhaft als Abkürzungen von  $\rho_{k|k}(n)$  bzw.  $D_{k-1|k-1}(\mathbf{x})$  zu verstehen.

wobei

$$\begin{aligned} & \Pi_{k|k-1}[D(\mathbf{x}), \rho](j) \\ &= \sum_{l=j}^{\infty} C_j^l \frac{\langle p_{S,k}(\mathbf{x}), D(\mathbf{x}) \rangle^j \langle 1 - p_{S,k}(\mathbf{x}), D(\mathbf{x}) \rangle^{l-j}}{\langle 1, D(\mathbf{x}) \rangle^l} \rho(l). \end{aligned} \quad (3.28)$$

Hierbei wird  $\langle a(\mathbf{x}), b(\mathbf{x}) \rangle = \int a(\mathbf{x})b(\mathbf{x})d\mathbf{x}$  als Innenprodukt von zwei reellwertigen Funktionen  $a(\mathbf{x})$  und  $b(\mathbf{x})$  verwendet sowie  $C_j^l$  als Binomialkoeffizient  $C_j^l = \frac{l!}{j!(l-j)!}$ . Als zusätzliche Komponenten sind hier im Vergleich zur gleichen PHD-Filter-Intensitätsprädiktion (3.24) für die Kardinalitätsprädiktion Annahmen über die Kardinalitätsverteilung der neuentstehenden Objekte  $\rho_{B,k}(\bullet)$  nötig.

**Korrektur** Während die zusätzliche Rekursion der Kardinalitätsverteilung  $\rho(n)$  in der Prädiktion keine direkten Auswirkungen auf die PHD  $D(\mathbf{x})$  hat, erfolgen die Verknüpfung und somit auch die positiven Effekte auf eine durch das CPHD-Filter verbesserte Schätzung der Zustände selbst durch die Kopplung im Korrekturschritt [144]. Diese ergeben sich nach [148] zu<sup>3</sup>

$$\rho_k(n) = \frac{Y_k^0[D_{k|k-1}, \mathcal{Z}_k](n) \rho_{k|k-1}(n)}{\langle Y_k^0[D_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1} \rangle} \quad (3.29)$$

$$D_k(\mathbf{x}_k) = [1 - p_{D,k}(\mathbf{x}_k)] \frac{\langle Y_k^1[D_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1} \rangle}{\langle Y_k^0[D_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1} \rangle} D_{k|k-1}(\mathbf{x}_k) \quad (3.30)$$

$$\begin{aligned} & + \sum_{\mathbf{z}_k \in \mathcal{Z}_k} \psi_{k,\mathbf{z}_k}(\mathbf{x}_k) \frac{\langle Y_k^1[D_{k|k-1}, \mathcal{Z}_k \setminus \{\mathbf{z}_k\}], \rho_{k|k-1} \rangle}{\langle Y_k^0[D_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1} \rangle} D_{k|k-1}(\mathbf{x}_k) \\ & \quad (3.31) \end{aligned}$$

<sup>3</sup> Abhängigkeiten von dem Zustand  $\mathbf{x}$  bzw. der Objektanzahl  $n$  werden aus Gründen der Übersichtlichkeit in unzweifelhaften Fällen weggelassen.

mit den Abkürzungen

$$Y_k^n[D, Z](n) = \sum_{j=0}^{\min(|Z|, n)} (|Z| - j)! \rho_{K,k}(|Z| - j) P_{j+u}^n \times \frac{\langle 1 - p_{D,k}, D \rangle^{n-(j+u)}}{\langle 1, D \rangle^n} e_j(\Xi_k(D, Z)) \quad (3.32)$$

$$e_j(Z) = \begin{cases} \sum_{S \subseteq Z, |Z|=j} \left( \prod_{\zeta \in S} \zeta \right), & j > 0 \\ 1, & j = 0 \end{cases} \quad (3.33)$$

$$\psi_{k, \mathbf{z}_k}(\mathbf{x}_k) = \frac{\langle 1, \kappa_k \rangle}{\kappa_k(\mathbf{z}_k)} g_k(\mathbf{z}_k | \mathbf{x}_k) p_{D,k}(\mathbf{x}_k) \quad (3.34)$$

$$\Xi_k(D, Z) = \{ \langle D, \psi_{k, \mathbf{z}} \rangle : \mathbf{z} \in Z \} . \quad (3.35)$$

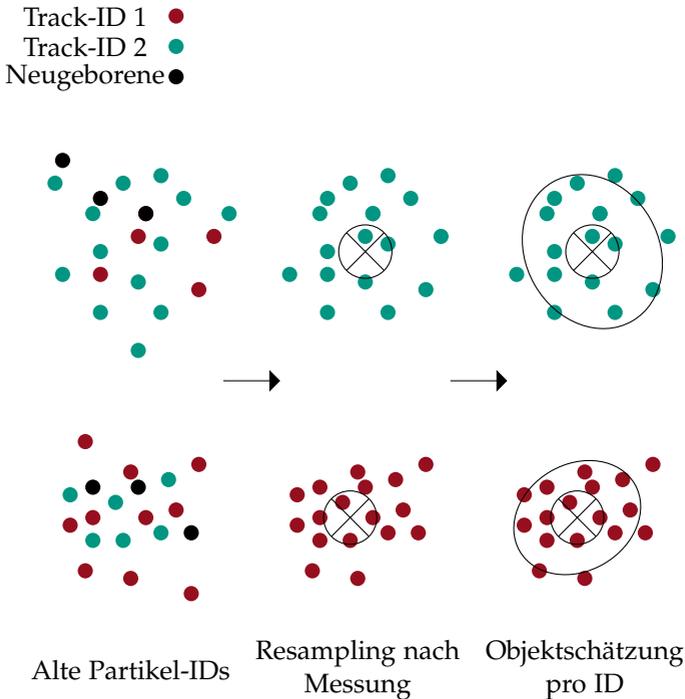
Hierbei stellt  $P_j^n = \frac{n!}{(n-j)!}$  den Permutationskoeffizienten und  $e_j(Z)$  die elementarsymmetrische Funktion dar. Analog zu der Verteilung neu entstehender Objekte ist ebenso eine Kardinalitätsverteilung von Falschalarmen  $\rho_{K,k}(\bullet)$  anzunehmen.

**Implementierung** Aus den Gleichungen (3.29)-(3.35) wird ersichtlich, dass gerade die Korrekturgleichungen des CPHD-Filters deutlich aufwendiger als die des PHD-Filters (3.25) sind. Dennoch lassen sich auch hier neben numerischen Methoden (SMC-CPHD-Filter) auch geschlossene Lösungen durch Verwendung von Gauß-Mischungen finden, die eine kubische Komplexität [148] aufweisen und eine deutliche Verbesserung gegenüber dem PHD-Filter oder auch traditionellen Filtern wie dem JPDA-Filter mit höherer Komplexität erzielen [148]. Zusätzlich existieren auch weitere Ansätze des CPHD-Filters, die für Szenarien mit einem Sensor eine lineare Komplexität für das CPHD-Filter aufweisen [80], jedoch zusätzliche Vereinfachungen in den Modellannahmen, beispielsweise bei Behandlung der Falschalarme, treffen. Das GM-CPHD-Filter lässt sich analog dem GM-PHD-Filter ebenfalls durch die Nutzung der EKF- bzw. UK-Erweiterung der Kalman-Gleichungen (UK-CPHD-Filter) im Falle von Nichtlinearitäten der Modelle einsetzen. Durch die globale Kardinalitätsverteilung kann es bei der CPHD-Filterung jedoch auch zur

sogenannten „*spooky action at a distance*“ [32] kommen. Dies bedeutet, dass Situationen entstehen können, in denen Gewichte von nicht detektierten Objekten, deren Existenz unsicherer wird, auf detektierte Objekte übergehen können, die räumlich weit von dem Ursprungsobjekt entfernt sind. Somit wird eine künstliche Korrelation hergestellt, die allerdings durch eine korrigierte Gewichtung weitgehend behoben werden kann, indem Objekte nicht mit zu hohen Gewichten versehen werden [101]. Die globale Kardinalitätsschätzung wird hierbei nicht verändert.

**Trajektorien** Die durch die Zeit propagierte PHD oder Intensitätsfunktion ist auch zusammen mit einer Kardinalitätsstatistik prinzipiell nur in der Lage, Informationen über Aufenthaltsorte von Objekten zu bestimmten Zeitpunkten zu liefern. Dies beinhaltet nicht automatisch eine Trajektorie, die die Positionen eines Objekts über die Zeit verbindet. Doch gerade die Information über den zeitlichen Verlauf der Position eines bestimmten Objekts oder einer Person ist für viele Anwendungen von großer Bedeutung, weshalb eine über die Zeit gleichbleibende Identifikationsnummer (ID) einzelner Objekte erforderlich ist. Abgesehen von einer Implementierung der PHD-Filter-Familie über Gauß-Mixturen, die sich einfach mit solchen IDs versehen lassen [102, 146], sind für die Verwendung von Partikeln der SMC-(C)PHD-Filter weitere Überlegungen notwendig. In [156] wird eine auf [119] basierende Möglichkeit für kontinuierliche Trajektorien bei SMC-PHD-Filtern vorgeschlagen, die hier auch für die CPHD-Filterung angewandt wurde. Der Grundgedanke von Wood et al. [156] ist, die Trajektorien entgegen dem ursprünglichen Ansatz nach Mahler [79] ohne Clustering-Methoden, gefolgt von einer Zuordnung der Schätzung zu den Trajektorien, zu gewinnen. Ein einfache und gleichzeitig effiziente Methode stellt die in [156] vorgestellte Benennung jedes verwendeten Partikels nach dem zugehörigen Objekt dar. Ein Clustering-Schritt wird hierbei nicht mehr benötigt, da die Korrektur entsprechend [119] zusammen mit einem Resampling-Schritt pro Messung ausgeführt wird, wodurch die Möglichkeit entsteht, IDs an das vorhergehende Cluster eines Resampling-Schritts zu vergeben. Hierzu wird die optimale Zuordnung zwischen Messungen und IDs mittels der Ungarischen oder auch Munkres-Methode [93] durchgeführt, wobei die Kostenfunktion für die Zuordnung eines Partikel-Clusters zu einem Ob-

jekt darauf basiert, wie hoch der Anteil der Partikel mit einer bestimmten ID in dem zu benennenden Cluster ist. Zusammen mit der Annahme, dass ein Objekt jeweils nur eine Messung verursacht, wird somit sichergestellt, dass ein Objekt A eher zu Cluster A als zu einem zweiten Cluster B eines anderen Objekts B zugeordnet wird. Dieses Vorgehen ist in Abbildung 3.3 gezeigt. Ebenso ist es somit im Gegensatz zum ursprünglichen SMC-PHD-Filter auch möglich, Positionsschätzungen eines zu einem Zeitschritt undetektierten Objekts zu ermöglichen [119], was in [156] gezeigt wurde. Labels von nicht detektierten Partikeln werden hierbei nicht verändert.



**Abbildung 3.3** Labeling von Partikeln durch Resampling pro Messung (Kreis mit Kreuz).

### 3.6 *Cardinality Balanced Multi-Target Multi-Bernoulli-Filter*

Ein von der Komplexität weniger aufwendiges Filter als das CPHD-Filter ist das *Cardinality Balanced Multi-Target Multi-Bernoulli* (CBMeMber)-Filter. Das CBMeMber-Filter [149] ist eine Erweiterung des MeMber-Filter [79], da es durch das in [149] vorgeschlagene *Cardinality-Balancing* eine fehlerhafte Schätzung der Kardinalität für Detektionswahrscheinlichkeiten verschieden von eins korrigiert. Grundgedanke bei dieser Gruppe an ebenfalls auf endlichen Mengen basierenden Filtern ist, die Approximation der Multi-Objekt-Dichte nicht über eine Intensitätsfunktion wie bei PHD und CPHD, sondern über die Verwendung von Multi-Bernoulli-verteilten endlichen Mengen, die auch für eine Annäherung der Rekursion des Multi-Objekt-Bayes-Filters genutzt werden können. Hierbei wird bei der Herleitung [79, 149] ebenfalls von unabhängiger Entstehung und Bewegung der beteiligten Objekte ausgegangen, allerdings folgt die Menge neu entstehender Objekte nicht etwa einer Poisson-Verteilung wie bei dem PHD-Filter, sondern ist Multi-Bernoulli-verteilt, bleibt aber unabhängig von den bestehenden Objekten. Ebenso wird angenommen, dass sich die A-posteriori-Dichte durch eine Multi-Bernoulli-Menge approximieren lässt. Falschalarme werden analog zum PHD-Filter als Poisson-verteilt modelliert. Eine weitere Annahme beinhaltet eine relativ hohe Detektionswahrscheinlichkeit und eher wenige Falschalarme [149]. Die Komplexität ist vergleichbar mit der des PHD-Filters [144].

**Prädiktion** Lässt sich die Multi-Objekt-Dichte zum Zeitpunkt  $k - 1$  durch die Multi-Bernoulli-Menge mit  $M_{k-1}$  Elementen

$$\pi_{k-1} = \{(r_{k-1}^{(i)}, p_{k-1}^{(i)}(\mathbf{x}_{k-1}))\}_{i=1}^{M_{k-1}} \quad (3.36)$$

nach [144] beschreiben, lässt sich die prädizierte Dichte

$$\pi_{k|k-1} = \{(r_{P,k|k-1}^{(i)}, p_{P,k|k-1}^{(i)}(\mathbf{x}_k))\}_{i=1}^{M_{k-1}} \cup \{(r_{B,k}^{(i)}, p_{B,k}^{(i)}(\mathbf{x}_k))\}_{i=1}^{M_{B,k}} \quad (3.37)$$

ebenfalls als Multi-Bernoulli-Menge angeben. Sie ist dabei die Vereinigung der prädizierten Menge und der  $M_{B,k}$ -elementigen Menge der

neu hinzukommenden Objekte. Jedes Element der prädizierten Multi-Bernoulli-Verteilung setzt sich dabei aus dem Wahrscheinlichkeitsparameter der Existenz  $r_{P,k|k-1}^{(i)}$  und einer Verteilung des Objekts im Zustandsraum  $p_{P,k|k-1}^{(i)}$  zusammen. Dies gilt analog auch für die neuentstandenen Objekte mit Index B. Die Prädiktion dieser beiden Eigenschaften erfolgt hierbei entsprechend

$$r_{P,k|k-1}^{(i)} = r_{k-1}^{(i)} \langle p_{k-1}^{(i)}(\mathbf{x}), p_{S,k}(\mathbf{x}) \rangle \quad (3.38)$$

$$p_{P,k|k-1}^{(i)}(\mathbf{x}_k) = \frac{\langle f_{k|k-1}(\mathbf{x}_k | \mathbf{x}_{k-1}), p_{k-1}^{(i)}(\mathbf{x}_{k-1}) p_{S,k}(\mathbf{x}_{k-1}) \rangle}{\langle p_{k-1}^{(i)}(\mathbf{x}), p_{S,k}(\mathbf{x}) \rangle}, \quad (3.39)$$

wobei  $f_{k|k-1}(\cdot|\cdot)$  und  $p_{S,k}(\cdot)$  analog den (C)PHD-Filter-Gleichungen aus den Abschnitten 3.4 und 3.5 gewählt sind. Zusammengefasst kann die prädizierte Menge auch wie folgt dargestellt werden:

$$\pi_{k|k-1} = \{(r_{k|k-1}^{(i)}, p_{k|k-1}^{(i)}(\mathbf{x}))\}_{i=1}^{M_{k|k-1}}. \quad (3.40)$$

**Korrektur** Werden nun neue Messungen berücksichtigt, lässt sich die neue Bernoulli-Menge zusammensetzen aus Elementen, deren Existenzwahrscheinlichkeit und räumliche Verteilung bereits durch Messungen korrigiert wurden und Elementen, die mit der Gegenwahrscheinlichkeit der Detektionswahrscheinlichkeit nicht detektiert wurden. Hierbei werden Elemente nach Korrektur (*Update*) durch den Index U gekennzeichnet und Elemente, die nicht durch Messungen repräsentiert werden, sind als „Alte Objekte“ (*Legacy*) beschrieben und durch den Index L bezeichnet.

$$\pi_k = \{(r_{L,k}^{(i)}, p_{L,k}^{(i)}(\mathbf{x}_k))\}_{i=1}^{M_{k|k-1}} \cup \{(r_{U,k}(\mathbf{z}_k), p_{U,k}(\mathbf{x}_k; \mathbf{z}_k))\}_{\mathbf{z}_k \in \mathcal{Z}_k}. \quad (3.41)$$

Entsprechend den oben getroffenen Annahmen lassen sich folgende Bestandteile von (3.41) analog zu den vorherigen Abschnitten mit Falschalarmintensität  $\kappa(\mathbf{z})$ , Detektionswahrscheinlichkeit  $p_D(\mathbf{x})$  und Likelihood

$g(\mathbf{z}|\mathbf{x})$  einer Einzelmessung herleiten [144]:

$$r_{L,k}^{(i)} = r_{k|k-1}^{(i)} \frac{1 - \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle}{1 - r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle} \quad (3.42)$$

$$p_{L,k}^{(i)}(\mathbf{x}_k) = p_{k|k-1}^{(i)}(\mathbf{x}_k) \frac{1 - p_{D,k}(\mathbf{x}_k)}{1 - \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle} \quad (3.43)$$

$$r_{U,k}(\mathbf{z}_k) = \frac{\sum_{i=1}^{M_{k|k-1}} \frac{r_{k|k-1}^{(i)} (1 - r_{k|k-1}^{(i)}) \langle p_{k|k-1}^{(i)}, \psi_{k,\mathbf{z}_k} \rangle}{(1 - r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle)^2}}{\kappa_k(\mathbf{z}_k) + \sum_{i=1}^{M_{k|k-1}} \frac{r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, \psi_{k,\mathbf{z}_k} \rangle}{1 - r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle}} \quad (3.44)$$

$$p_{U,k}(\mathbf{x}_k; \mathbf{z}_k) = \frac{\sum_{i=1}^{M_{k|k-1}} \frac{r_{k|k-1}^{(i)} p_{k|k-1}^{(i)}(\mathbf{x}_k) \psi_{k,\mathbf{z}_k}(\mathbf{x}_k)}{1 - r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle}}{\sum_{i=1}^{M_{k|k-1}} \frac{r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, \psi_{k,\mathbf{z}_k} \rangle}{1 - r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle}} \quad (3.45)$$

$$\psi_{k,\mathbf{z}_k}(\mathbf{x}_k) = g_k(\mathbf{z}_k|\mathbf{x}_k) p_{D,k}(\mathbf{x}_k). \quad (3.46)$$

Prädiktion und Korrektur der Bernoulli-Mengen lassen sich ebenso als die rekursive Beurteilung von  $i$  Objekthypothesen auffassen, die zu einem Anteil nur prädiziert und nicht detektiert und zum anderen Anteil prädiziert und durch neue Messungen korrigiert werden. Hierbei werden entsprechend die Existenzwahrscheinlichkeiten  $r$  angepasst. Für die in jeder Rekursion ansteigende Anzahl der A-posteriori-Objekthypothesen  $M_k$  aus erneut detektierten Objekten, *Legacy*-Hypothesen und neuen Messungen gilt

$$M_k = M_{k-1} + M_{B,k} + |\mathcal{Z}_k|. \quad (3.47)$$

Durch Berücksichtigung der Existenzwahrscheinlichkeit jeder Objekthypothese kann die erwartete Objektanzahl  $\hat{N}_k$  zum Zeitpunkt  $t_k$  zu

$$\hat{N}_k = \sum_{i=1}^{M_k} r_k^{(i)} \quad (3.48)$$

berechnet werden. Da diese Anzahl entgegen der Anzahl der Hypothesen jedoch nicht zwingend steigt, muss bei der Implementierung auf eine

Reduktion der Hypothesen aufgrund geringer Existenzwahrscheinlichkeit geachtet werden. Die Varianz der geschätzten Objektanzahl kann entsprechend der Varianz einer Multi-Bernoulli-Menge aus der Summe der Einzelvarianzen von Bernoulli-verteilten Objekten, dem Produkt der Existenzwahrscheinlichkeit und der Gegenwahrscheinlichkeit, bestimmt werden.

**Implementierung** Auch für das CBMeMber existieren geschlossene Lösungen bei Nutzung von Gauß-Mixturen und numerische Partikelvarianten [149], wobei die Partikelvariante nach Vo et al. [149] im Vergleich zu PHD und CPHD bessere Ergebnisse als beim Vergleich der Gauß-Implementierungen erreicht. Die Varianz der Kardinalität liegt im Allgemeinen zwischen der des PHD-Filters und der des CPHD-Filters [144].

**Trajektorien** Da die Beschreibung der CBMeMber-Approximation bereits als Rekursion von Objekthypothesen interpretiert werden kann, bietet es sich an, jedes Parameterpaar aus Aufenthaltswahrscheinlichkeit und Existenzwahrscheinlichkeit um eine ID oder auch Label  $l^{(i)}$  für  $i = 1, \dots, M$  zu erweitern. Während sich diese Label in der Prädiktion von Zeitschritt  $k - 1$  zu  $k$  nicht verändern und nur um neue Labels  $l_{B,k}^{(i)}$  für neu erschienene Objekte ergänzt werden, erhalten die Komponenten der *Legacy*- und *Update*-Objekte entsprechend (3.41) folgende Labels

$$l_{L,k}^{(i)} = l_{k|k-1}^{(i)} \quad (3.49)$$

$$l_{U,k}(\mathbf{z}_k) = l_{k|k-1}^{(n)} \quad \text{mit } n = \arg \max_i \frac{r_{k|k-1}^{(i)} (1 - r_{k|k-1}^{(i)}) \langle p_{k|k-1}^{(i)}, \psi_{k,\mathbf{z}_k} \rangle}{(1 - r_{k|k-1}^{(i)} \langle p_{k|k-1}^{(i)}, p_{D,k} \rangle)^2}. \quad (3.50)$$

## 3.7 Bewertung

Neben den unterschiedlichen Annahmen und Konzepten von verschiedenen Ein- und Multi-Objekt-Filtern ist für eine erfolgreiche Nutzung eine Bewertung der Qualität der Schätzergebnisse essentiell. Hierbei sind vor

allem zwei Aufgaben zu bewältigen. Zum einen müssen entsprechende Daten zur Verfügung stehen, die den wahren Verlauf der Objektzustände widerspiegeln. Während dies für den Vergleich mit simulierten Daten problemlos möglich ist [48, 65, 79, 119, 144], wird eine solche Bewertung bei der Beurteilung realer Daten schwieriger, da Referenzdaten aus einer weiteren Quelle nötig werden, die für das Schätzproblem als eine ausreichend gute Abbildung eines wahren Verlaufs der Zustände angesehen werden können. Für beide Ansätze der Evaluation ist ein Maß nötig, um die Qualität der Schätzung aufgrund der Referenzdaten zu beurteilen. Gegenüber dem Ein-Objekt-Fall muss bei dem Multi-Objekt-Fall zusätzlich zum Vergleich eines Zustands eines Objektes auch eine Bewertung der Schätzung der Kardinalität erfolgen. Der Vergleich der Zustände wird dadurch komplexer, so dass vorab eine Zuordnung zwischen geschätzten Objekten und Referenzobjekten erfolgen muss. Während eine Reihe von Maßen für die Beurteilung von anwendungsorientierten Multi-Objekt-Szenarien beispielsweise aus der Analyse von Videos [10, 95] existiert, ist bei der Analyse von Filtern, die auf der Statistik endlicher Mengen basiert, eine mathematisch konsistente Metrik sinnvoll, um einen Vergleich zu liefern, der auch unabhängig von den eingesetzten Sensoren im Falle von Multi-Sensor-Szenarien einsetzbar ist und sich gerade für die Beurteilung von Zustandsschätzung und Kardinalität eignet. Zusätzlich steht die Evaluation der Zustandsschätzung im Vordergrund, da eine gute Schätzung der Zustände Voraussetzung für das Nachvollziehen kompletter Trajektorien ist.

Eines der ersten Maße für Mengen, das entsprechend auch die Axiome einer Metrik (Symmetrie, positive Definitheit und Einhaltung der Dreiecksungleichung) erfüllt, ist die Hausdorff-Metrik [47], die jedoch kaum sensitiv gegenüber unterschiedlichen Kardinalitäten der zu vergleichenden Mengen ist [51]. Auf dieser Metrik aufbauend wurde daher die Wasserstein-Metrik als Grundlage des Vergleichs zweier Mengen in Multi-Objekt-Szenarien gewählt [51, 120, 124], die auch als *Earth-Movers-Distance* bekannt ist und anschaulich versucht, die minimalen Kosten zu ermitteln, die nötig sind, um eine Verteilung oder eine Menge in eine andere umzuwandeln. Hiervon ausgehend hat sich die *Optimal Subpattern* (OSPA)-Metrik für die Bewertung von Multi-Objekt-Filtern durchgesetzt [48, 120, 124], die im folgenden Abschnitt kurz erläutert werden soll.

## OSPA

Bei der OSPA-Metrik handelt es sich auch um ein intuitiv sehr gut nachvollziehbares Maß. Ein Vergleich zweier Mengen  $\mathcal{X}$  mit  $|\mathcal{X}| = m$  Elementen und  $\mathcal{Y}$  mit  $|\mathcal{Y}| = n$  wird dadurch erreicht, dass jeweils die Distanz der sich am meisten ähnelnden Objekte errechnet wird. Da dies nur möglich ist, falls eine übereinstimmende Anzahl von Objekten in beiden Mengen vorhanden ist, werden diese Distanzen für sogenannte *Optimal Subpattern* berechnet. Gegeben einer beliebigen Metrik für die Distanz  $d^{(c)}(\mathbf{x}, \mathbf{y})$  zweier Objekte und ihrer Zustandsvektoren  $\mathbf{x}$  und  $\mathbf{y}$ , wie zum Beispiel die euklidische oder die Mahalanobis-Distanz, die auf ein maximales Ergebnis  $c$  beschränkt ist, lässt sich die Distanz zweier Mengen der OSPA-Distanz  $\bar{d}_p^{(c)}(\mathcal{X}, \mathcal{Y})$  entsprechend [124] für  $m \leq n$  wie folgt definieren

$$\bar{d}_p^{(c)}(\mathcal{X}, \mathcal{Y}) := \left( \frac{1}{n} \left( \min_{\pi \in \Pi_n} \sum_{i=1}^m d^{(c)}(\mathbf{x}_i, \mathbf{y}_{\pi(i)})^p + c^p (n - m) \right) \right)^{\frac{1}{p}}, \quad (3.51)$$

und ansonsten

$$\bar{d}_p^{(c)}(\mathcal{X}, \mathcal{Y}) := \bar{d}_p^{(c)}(\mathcal{Y}, \mathcal{X}). \quad (3.52)$$

Für Fälle von  $n = m = 0$  ist die Metrik zu null definiert. Der Index  $\Pi_k$  beschreibt alle möglichen Permutationen einer Menge mit  $k$  Elementen. Für die optimale Zuordnung bei minimaler Distanz wird die ungarische Methode [93] verwendet. Bei der Metrik werden zwei Parameter  $p$  und  $c$  verwendet. Deren physikalische Interpretation ist für  $p$  die Ordnung der Metrik und beeinflusst, wie stark Ausreißer das Ergebnis beeinflussen und für  $c$ , als sogenannten *Cut-off-Parameter*, die Schwelle, ab der ein Objekt als ein Fehler in der Kardinalität gezählt wird und keinem Objekt mehr zugeordnet wird. Wie an der Summe in (3.51) gesehen werden kann, spaltet die Metrik sich in zwei Teile auf: Ein Anteil beschreibt die reine Distanz als Fehler in der Lokalisierung oder Zustandsschätzung eines Objektes und der andere Anteil berücksichtigt Kardinalitätsunterschiede. Somit lässt sich leicht eine getrennte Fehlerberechnung für Lokalisierung und Kardinalität angeben [124].

---

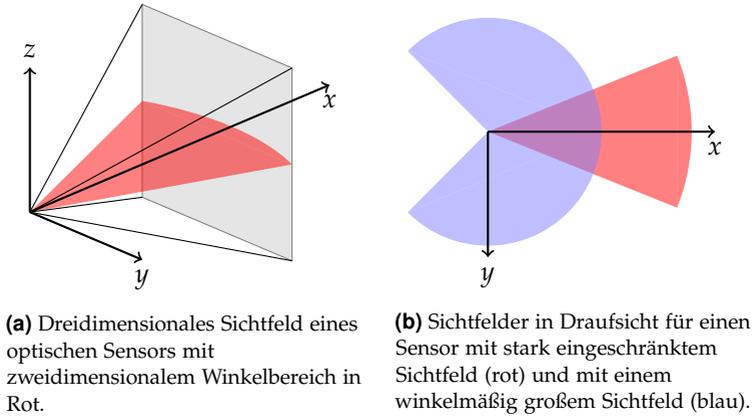
Um die OSPA-Metrik auch auf die Beurteilung von ganzen Trajektorien anzuwenden, existieren ebenso Ansätze nach [120] oder Erweiterungen bezüglich der Qualität einzelner Schätzungen [48].

## 4 Adaptive Detektionswahrscheinlichkeit

Das Problem, Personen in Sensordaten zu finden und zu verfolgen, hat eine sehr große Bedeutung insbesondere für die Robotik [126] oder auch Fahrerassistenzsysteme [100]. Im Gegensatz zu Szenarien ohne Falschalarne [155], die durch Hindernisse im Raum entstehen können, oder ohne die Möglichkeit einer aktiven Bereitschaft von Personen verfolgt zu werden, wie beispielsweise durch die Verwendung von Markern [129] oder zusätzlicher Ausrüstung [19], müssen Verfolgungsfiler eingesetzt werden, die mit verarbeiteten Sensorrohdaten arbeiten. Methoden, die wie in Abschnitt 2.4 Rohdaten verarbeiten und Detektionen relativ zu der Position des Sensors erzeugen, beruhen dabei auf einer langen Kette aus Signalverarbeitungsschritten, die alle einen Beitrag zu der Qualität der Detektion leisten. Unterschiedliche Randbedingungen, wie Beleuchtungsunterschiede, beliebige Umgebung, Verdeckungen, variable Erscheinung von Personen in Größe, Aussehen oder Kleidung und andere Faktoren führen zum einen zu aufwendigen Detektionsprozessen, die maschinelle Lernverfahren benötigen, zum anderen lässt sich die Detektion von Personen in beliebigen Umgebungen dennoch nicht zweifelsfrei lösen. Während andere Verfahren insbesondere in der Personenverfolgung auf Videosequenzen versuchen, eine Verfolgung durch eine Datenassoziation und anschließende Optimierung der Zuordnung offline [52] oder auch online [112] zu lösen, soll hier der allgemeinere und assoziationsfreie Ansatz der vorherigen Kapitel weiter verfolgt werden. Dabei wird das Problem über eine probabilistische Beschreibung des Detektionsverhaltens der verwendeten Sensoren und eine nachfolgende Signalverarbeitung beschrieben. Zudem tragen für Multi-Objekt-Filter nicht nur die vorhandenen Messungen Information, so dass ein Nicht-detektieren eines Objektes und somit auch ein entsprechendes Modell

für nicht erfolgte Messungen als „negative“ Information ebenfalls große Bedeutung besitzen [64].

Ein solches Modell der Detektionswahrscheinlichkeit kann anschließend in einem Verfolgungsalgorithmus genutzt werden, um Detektionen mehrerer Personen geeignet über verschiedene Sensoren hinweg und auch mit weiteren Informationsquellen wie Karten der Umgebung oder Annahmen über typische Bewegungsmuster (vgl. Abschnitt 2.1) zu fusionieren. Unabhängig davon, ob eine zentrale Fusion über die Sensoren hinweg oder eine verteilte Fusion verwendet wird, ist hier der Austausch von Informationen über die Detektionen der Sensoren entscheidend. Während für die Verfolgung oft eine konstante Detektionswahrscheinlichkeit [74, 83, 84, 126] im Umfeld eines Sensors angenommen wird, zeigt sich der Nutzen und die Notwendigkeit einer differenzierteren Betrachtung in [100] besonders bei sensorisch blinden Bereichen. Auch eine Diskrepanz zwischen korrekter Detektionswahrscheinlichkeit und im Filter angenommenen Parametern wurde als sehr wichtiges Kriterium für eine erfolgreiche Verfolgung von Objekten identifiziert [150]. Eine adaptive Anpassung an dynamische Objekte im Fall von Verdeckungen bei jedoch ansonsten einheitlicher Behandlung der Detektionswahrscheinlichkeit zeigt klare Vorteile in [117], vor allem bei der Verbesserung der Kardinalitätsschätzung. Die Überlappungen dieser Fälle in Szenarien mit verteilten Sensorträgern soll in den folgenden Abschnitten zusammen mit Vorschlägen einer geeigneten probabilistischen Modellierung vorgestellt werden. Das Ziel hierbei ist, durch die Fusion verteilter, heterogener Sensoren eine möglichst genaue Zustandsschätzung zu erreichen. Diese soll dabei durch entsprechende Modellierung der Sensoren in Multi-Objekt-Filtern erreicht werden. Die vorgeschlagenen Modellansätze werden hierzu zusammen mit ihrer Integration in die entsprechenden Filterstrukturen in den nächsten Abschnitten vorgestellt. Dabei orientieren sich die Ausführungen an einem Zeitschriftenbeitrag des Autors [161].



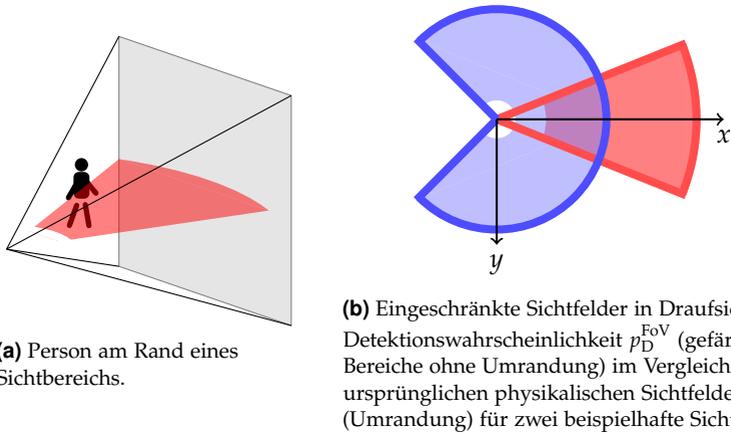
**Abbildung 4.1** Sichtfelder verschiedener Sensoren.

## 4.1 Sichtfelder

Bedingt durch den Aufbau von Sensoren ist es für gewöhnlich nur in einem begrenzten Bereich um einen Sensor herum möglich, Objekte zu erfassen. Gerade bei der optischen Erfassung der Umwelt ist durch die in Kameras eingesetzte Kombination aus Sensorchip und verwendeter Optik der Erfassungsbereich eingeschränkt, was vor allem für das horizontale und vertikale Sichtfeld, wie in Abbildung 4.1(a) gezeigt, eine Rolle spielt. In dem hier bearbeiteten Szenario ist für die Filterung vor allem der horizontale Erfassungswinkel interessant, da die Verfolgung von Objekten in der Ebene erfolgt und somit einer Draufsicht auf den dreidimensionalen Raum in Abbildung 4.1(b) entspricht, der hier für zwei unterschiedliche Sensoren gezeigt ist. Bei der Nutzung von Systemen, die auf der Laufzeitmessung von Laserimpulsen basieren (LIDAR vgl. Abschnitt 2.4.2), kann das Sichtfeld durch den Einsatz von rotierenden Spiegeln erheblich vergrößert werden, siehe Abbildung 4.1(b) in Blau. Dennoch ist eine komplette Rundumsicht bauartbedingt, wenn überhaupt, nur mit sehr hohem Aufwand und damit verbundenen Kosten für den Sensor möglich.

## Verarbeitungskette

Neben den rein physikalisch gegebenen Grenzen des Sichtbereichs eines Sensors muss für die Evaluation des maximalen Bereichs, in dem Detektionen erfolgen können, auch die Verarbeitungskette berücksichtigt werden. Somit wird das Sichtfeld auf den Bereich begrenzt, in dem die von einem Sensor erfassten Rohdaten eines zu detektierenden Objekts ausreichen, um eine Detektion durchzuführen. Hierbei können mehrere Faktoren eine Rolle spielen. Zum einen führt eine begrenzte Auflösung, bei Kameras beispielsweise durch die verwendete Anzahl an Pixeln oder bei Laserscannern durch die verfügbare Abtastung der Winkel, dazu, dass Objekte in weiter Entfernung einem Detektionsverfahren zu wenig Information liefern oder gar nicht mehr erfasst werden können. Zum anderen lassen sich, je nach Algorithmus, auch Objekte nur erfassen, sobald sie vollständig in dem physikalischen Sichtbereich stehen (vergleiche Abbildung 4.2), was beispielsweise zu einem für eine Detektion nötigen Mindestabstand zwischen Objekt zum Sensor oder zum Rand des Sichtfelds führen kann. Dies bringt weitere Einschränkungen mit sich, die über das Sichtfeld aufgrund von Sensor und Signalverarbeitung beschrieben werden können. Für die Modellierung wird hier angenommen, dass eine binäre Entscheidung getroffen werden kann, ob ein Objekt im Sichtfeld liegt oder nicht. Somit ergibt sich eine Detektionswahrscheinlichkeit nur aufgrund des Sichtfelds (*Field of View (FOV)*), definiert als sensorischer Sichtbereich  $p_D^{\text{FoV}}$ , zu eins, falls ein Objekt detektiert werden kann und zu null, falls es außerhalb des Bereichs liegt. Wie Abschnitt 4.5.1 jedoch zeigt, können Randeffekte nicht komplett vernachlässigt werden. Die Modellierung ist jedoch zulässig, da die hier vorgeschlagenen Filterverfahren implizit den Zustand eines Objektes aus mehreren Punkten schätzen und somit eine Mittelung erfolgt, die in Abschnitt 4.5.1 beschrieben wird. Das resultierende Sichtfeld ist im Vergleich zu Abbildung 4.1(b) in Abbildung 4.2(b) zu sehen.



(a) Person am Rand eines Sichtbereichs.

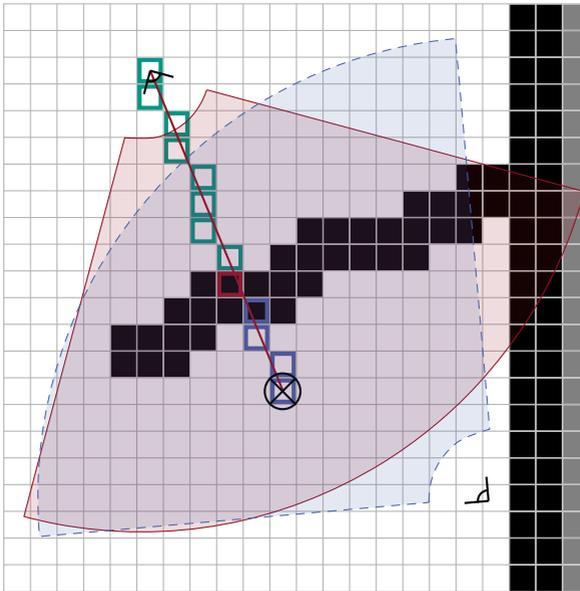
(b) Eingeschränkte Sichtfelder in Draufsicht mit Detektionswahrscheinlichkeit  $p_D^{\text{FoV}}$  (gefärbte Bereiche ohne Umrandung) im Vergleich zu ursprünglichen physikalischen Sichtfeldern (Umrandung) für zwei beispielhafte Sichtfelder.

**Abbildung 4.2** Durch die Signalverarbeitungskette eingeschränkte Sichtbereiche.

## 4.2 Verdeckung

Des Weiteren hängt die Möglichkeit, ein Objekt zu erfassen, auch stark von der Umgebung des Sensors ab. Während Verdeckungen in statischen Szenarien ohne eine mögliche Bewegung des Sensors durch Belegungs-gitter beschrieben werden können [29] oder Verfahren zur optimalen Abdeckung in verteilten Kamera-Netzwerken verwendet werden [57], ist gerade bei zum Teil überlappenden oder aneinandergrenzenden Sichtfeldern das Wissen des Fusionsfilters über das Sichtfeld eines jeden Sensors entscheidend. Im Falle mehrerer vorhandener Perspektiven einer Szene eignet sich das Wissen über Sichtfelder zur Wahl geeigneter Perspektiven [62, 72, 152], wodurch bereits Verdeckungen aufgelöst werden können. Doch soweit es die Fusionsstrategie unterstützt, können überlappende Sichtfelder auch gemeinsam genutzt werden. Dies ist für die Methoden der Statistik endlicher Mengen der Fall. Um gemeinsame Bereiche zu erkennen, muss jedoch festgestellt werden, welcher Sensor welche Bereiche in einem dynamischen Umfeld sowohl bei Eigenbewegung als auch bei Bewegung der zu verfolgenden Objekte erfassen kann und an welchen Stellen es zu Verdeckungen kommt, um Widersprüche zu vermeiden. Dazu werden in den folgenden Abschnitten Möglichkeiten zur Evaluierung des Sichtfelds im Falle von statischen und dynamischen

Verdeckungen vorgestellt. Dabei soll die genauere Betrachtung des Sichtfelds, die bisher hauptsächlich als vorgeschalteter Schritt für die Planung der Sensorpositionen [84, 87] benutzt wird, über die Modellierung einer Detektionswahrscheinlichkeit der Verdeckung (*Occlusion*)  $p_D^{\text{Occ}}$  für die Nutzung bei der Verfolgung selbst erfolgen.



**Abbildung 4.3** Statische Verdeckung im Falle verteilter Sensorträger.

### 4.2.1 Statische Verdeckungen

Wird nun entsprechend Abschnitt 4.1 von einem auf den physikalischen Eigenschaften des Sensors beruhenden Sichtfeld ausgegangen, kann es zu Verdeckungen kommen, wie in Abbildung 4.3 gezeigt. Zu sehen sind hier zwei Sensorträger auf einer Karte eines Gebäudes mit eingezeichneten maximalen Erfassungsbereichen. Zusätzlich ist durch den Kreis mit dem Kreuz ein zu erfassendes Objekt gekennzeichnet. Während der untere Sensor freie Sicht auf das Objekt besitzt, ist die Sicht des oberen Sensors durch ein in der Karte verzeichnetes Hindernis wie beispiels-

weise eine Wand beschränkt. Allerdings liegt das zu erfassende Objekt dennoch im theoretischen Sichtbereich beider Sensoren. Bei der Fusion der Detektionsergebnisse beider Sensorträger führt dieser Zustand zu Widersprüchen, da der obere Sensor das Objekt erfassen müsste, dies aber durch die Verdeckung nicht möglich ist. Eine adaptive Anpassung der Detektionswahrscheinlichkeit bei statischen Verdeckungen  $p_D^{\text{Occ,stat}}$  kann Abhilfe schaffen, da hierdurch das für die Fusion eingesetzte Filter Zusatzinformationen über derartige Verdeckungen nutzen kann.

Hierzu muss erfasst werden, an welchen Stellen sich Verdeckungen ergeben können. Im Falle von statischen Verdeckungen muss zur Bestimmung dieser für beliebige Positionen eines Sensorträgers ausgewertet werden, welche Objektpositionen aktuell zu erfassen sind. Für bewegliche Sensorträger setzt dies ein korrigiertes Sichtfeld nach jeder Eigenbewegung voraus. Hier eignen sich bei bekannter statischer Umgebung bzw. einer Karte und Informationen über die Lokalisierung, vgl. Abschnitt 2.6, Raytracing-Verfahren, um Verdeckungen festzustellen.

## Raytracing

Ziel des *Raytracing* ist es, dem Namen nach alle möglichen Strahlen vom Ursprung aus zu verfolgen und somit alle direkten Sichtverbindungen zwischen Sensor und Umgebung aufzufinden. Somit lassen sich durch die Auswertung des Verfahrens für prädiizierte Positionen von bereits verfolgten Objekten ebenfalls Objekte ausfindig machen, zu denen keine direkte Sichtverbindung besteht und bei denen somit eine Verdeckung vorliegt. Beispielhaft ist dies in Abbildung 4.3 gezeigt. Aus der Zeichnung wird ebenso klar, dass ein diskretes Raytracing vorgenommen wird, da für die maschinelle Berechnung eine Diskretisierung nötig wird. Verfahren für Raytracing sind hauptsächlich aus der Computergrafik zur Darstellung dreidimensionaler Objekte auf zweidimensionalen Bildschirmen bekannt [39]. Durch Einsatz des weit verbreiteten Bresenham-Algorithmus [15] kann dieses Problem effizient gelöst werden. Hierbei wird für die Diskretisierung genutzt, dass oft auch bereits Kartenmaterial, beispielsweise bei auf Gittern basierenden Karten (vgl. Abschnitt 2.6), in diskretisierter Form einer üblichen Auflösung von 5 cm pro Pixel vorliegt. Des Weiteren ist für die Analyse von Verdeckungen auch nicht immer ein Raytracing für das gesamte Sichtfeld nötig, sondern nur eine Betrachtung

tung der Gebiete, in denen sich Objekte aufgrund vorher prädizierter Schätzungen und somit auch Detektionen anderer Sensoren aufhalten können.

Die Detektionswahrscheinlichkeit aufgrund der statischen Verdeckung (*Occlusion*)  $p_D^{\text{Occ,stat}}$  wird nun analog zu Abschnitt 4.1 binär gewählt. Eine Berücksichtigung von Unsicherheiten der Objektposition, beispielsweise bei Darstellung von Objekten über Partikel-Darstellungen, ist ebenso möglich, was jedoch zu einem deutlich höheren Berechnungsaufwand führt.



**Abbildung 4.4** Farbbild einer Szene mit zwei Personen, von denen eine verdeckt wird.

## 4.2.2 Dynamische Verdeckungen

Auch wenn eine statische Umgebungskarte und die Position des Sensorträgers bekannt sind, kann es wie in Abbildung 4.4 zu dynamischen Verdeckungen kommen, die abhängig von dem Verhalten beweglicher Objekte, wie Personen, entstehen können. Diese können über die Berücksichtigung in der Detektionswahrscheinlichkeit analog zu statischen Verdeckungen ausgewertet werden [5]. Diese Detektionswahrscheinlichkeit ist wie auch in den vorherigen Abschnitten eine sensorabhängige Größe und muss berücksichtigen, wie sensitiv die verwendete Kombination aus physikalischem Sensor und Detektionsalgorithmus ist, da hier verschiedene Ansätze unterschiedliche Eigenschaften bezüglich einer

Detektion bei Teilverdeckungen besitzen können [22, 28]. Hierbei können über vorab bestimmte Ausmaße von Personen, die oft als kreisförmig mit einem Radius von 20 cm angenommen werden [74, 117], Regionen ermittelt werden, die von Personen verdeckt werden, die in vorherigen Zeitschritten bereits als Trajektorie existierten. Durch Nullsetzen der Detektionswahrscheinlichkeit  $p_D^{\text{Occ,dyn}}$  für diese aufgrund dynamischer Verdeckung „blinden“ Bereiche wird der entsprechende Sensor für die durch das Bewegungsmodell geschätzte Zeit der Verdeckung für die Positionsschätzung verdeckter Ziele nicht berücksichtigt. Für den Fall von variablen oder unbekanntem Ausmaßen lässt sich das Vorgehen auch durch eine gleichzeitige Schätzung der Ausdehnung der zu verfolgenden Objekte erweitern [42, 43], auf die in den weiteren Schritten allerdings nicht eingegangen wird, da sich im gewählten Szenario weder die Ausdehnungen ändern noch deutliche Abweichungen von den Modellannahmen der Ausdehnung auftraten. Die Modellierung von  $p_D^{\text{Occ,dyn}}$  beschränkt sich wiederum auf die Zustände sichtbar und verdeckt, so dass sich die gemeinsame Detektionswahrscheinlichkeit bei Verdeckung zu

$$p_D^{\text{Occ}} = p_D^{\text{Occ,stat}} \cdot p_D^{\text{Occ,dyn}} \quad (4.1)$$

ergibt.

### 4.3 Abstandsabhängige Detektionswahrscheinlichkeit

Während die Detektionswahrscheinlichkeit ohne Berücksichtigung von Verdeckung meist als konstant im Sichtfeld angenommen wird [74, 79, 117, 126], wird die tatsächlich oft auftretende Inhomogenität im Sichtfeld teilweise nur für die Optimierung von Merkmalen [41] und nicht für die Fusion genutzt. Während der Nutzen gerade für große Reichweiten im Automobilbereich bei Nutzung eines Sensorträgers bekannt ist [100] und auch bei einer Assoziation von Messungen zu Objekten nützlich sein kann [98], soll dieser Nutzen auf verteilte Sensorträger im Kontext der Statistik endlicher Mengen erweitert werden. Denn gerade bei der Verwendung von verteilten Sensorträgern in realen Anwendungen mit sich

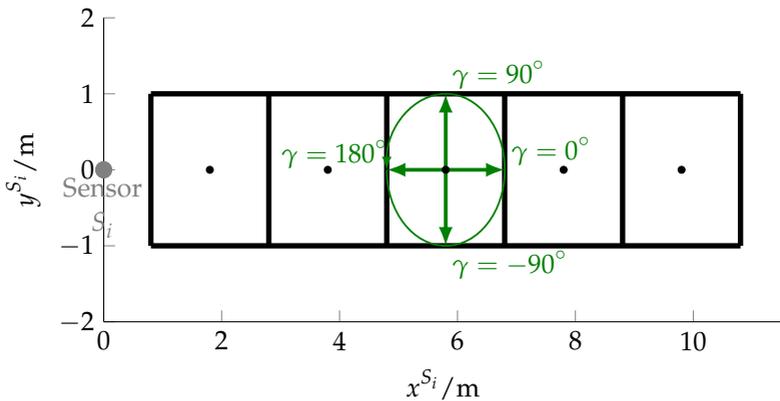
potentiell überlappenden Sichtfeldern spielt die richtige Modellierung der Sensoren eine wichtige Rolle für die Fusion von Detektionen verschiedener Sensoren und Sensorpositionen. Ein Grund sind, ähnlich den Verdeckungen auch, Widersprüche in den Detektionen unterschiedlicher Herkunft. Hält sich beispielsweise ein Objekt in einem Teil des Sichtfelds mit einer deutlich unterdurchschnittlichen Detektionsrate auf, tendiert ein Fusionsalgorithmus somit dazu, von einem verschwundenen Objekt auszugehen, wenn ein deutlicher Unterschied zwischen tatsächlich ankommender Rate an Detektionen und angenommener konstanter, beispielsweise über das gesamte Sichtfeld gemittelter, Detektionsrate besteht. Dies kann ebenso dazu führen, dass korrekte Detektionen eines Sensors, der diesen Bereich besser als angenommen abdeckt, als Falschalarme aufgefasst werden, da der erste Sensor entsprechend seinem Modell kein Objekt mehr auffindet.

Die Inhomogenitäten des Sichtfelds können mehrere Ursachen besitzen [41, 100], wobei der radiale Abstand zwischen Sensor und zu erfassendem Objekt, gerade bei begrenzten Sensorauflösungen, einen sehr starken Einfluss hat, der deswegen hier näher dargestellt werden soll. Darauf aufbauend kann ein Modell erstellt werden, das für die Fusion wichtige Rückschlüsse darüber liefert, wie gut die Sensor-/Objekterkennungsalgorithmus-Kombination in welchen Teilen des Sichtfelds funktioniert und somit welche Detektionsraten zu erwarten sind. Diese Detektionswahrscheinlichkeit wird abhängig von der Distanz zwischen Objekt und Sensor im Folgenden als  $p_D^{\text{Dist}}$  bezeichnet.

### 4.3.1 Erfassung

Theoretische Ergebnisse haben für CPHD- und CBMemBer-Filter gezeigt [82, 145], dass eine automatische Erfassung von Parametern wie Detektionswahrscheinlichkeit und Falschalarmrate theoretisch möglich ist, sie gehen hierbei allerdings von idealen Bedingungen bezüglich Unabhängigkeiten der Messungen und rein zufällig verteilten Falschalarmen aus. Unter realen Bedingungen sind diese Annahmen jedoch höchstens näherungsweise erfüllt. Ebenso zeigt sich, dass eine den tatsächlichen Werten entsprechende Parametrisierung der verwendeten Modelle selbst unter idealen Bedingungen einer Schätzung deutlich überlegen ist. Während sich diese tatsächlichen Werte für die Simulationen beispielsweise aus [65,

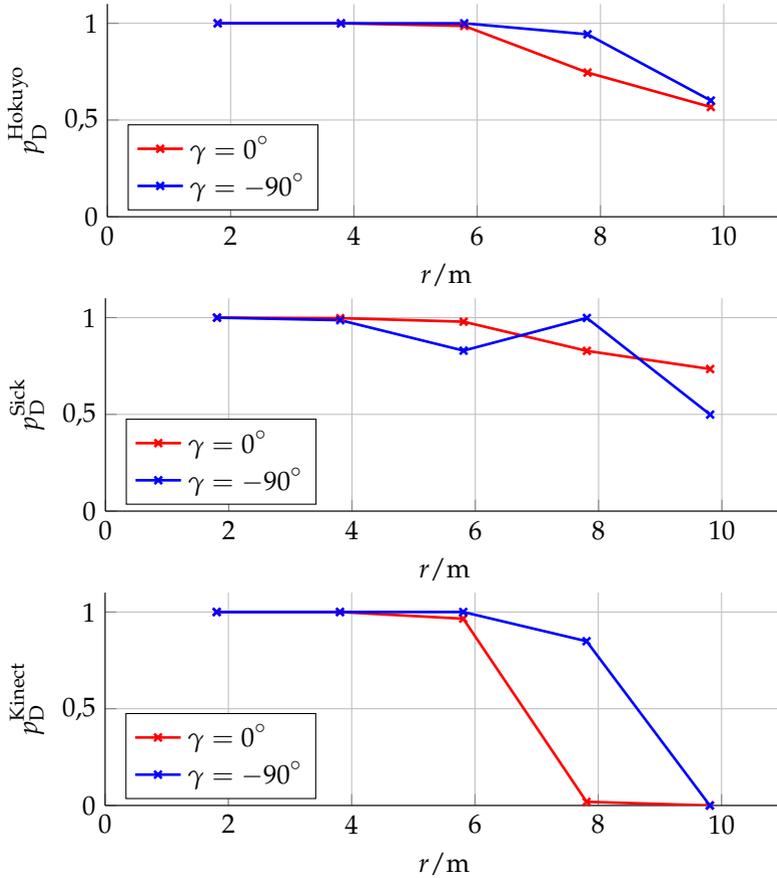
82, 145] leicht angeben lassen, können entsprechende Werte für reale Anwendungen mit Algorithmen nur aus Messungen hervorgehen, da eine analytische Lösung für beliebige Umgebungen und durch die Verwendung von auf Trainingsdaten beruhenden maschinellen Lernverfahren auch für einfachste Detektionsaufgaben beliebig komplex werden würde. Um für die in dem hier vorgestellten System verwendeten Algorithmen ebenfalls Aussagen über die Parametrierung der Detektionswahrscheinlichkeit in Abhängigkeit der relativen Position eines Objekts zum Sensor zu erhalten, wurden diese experimentell ausgewertet.



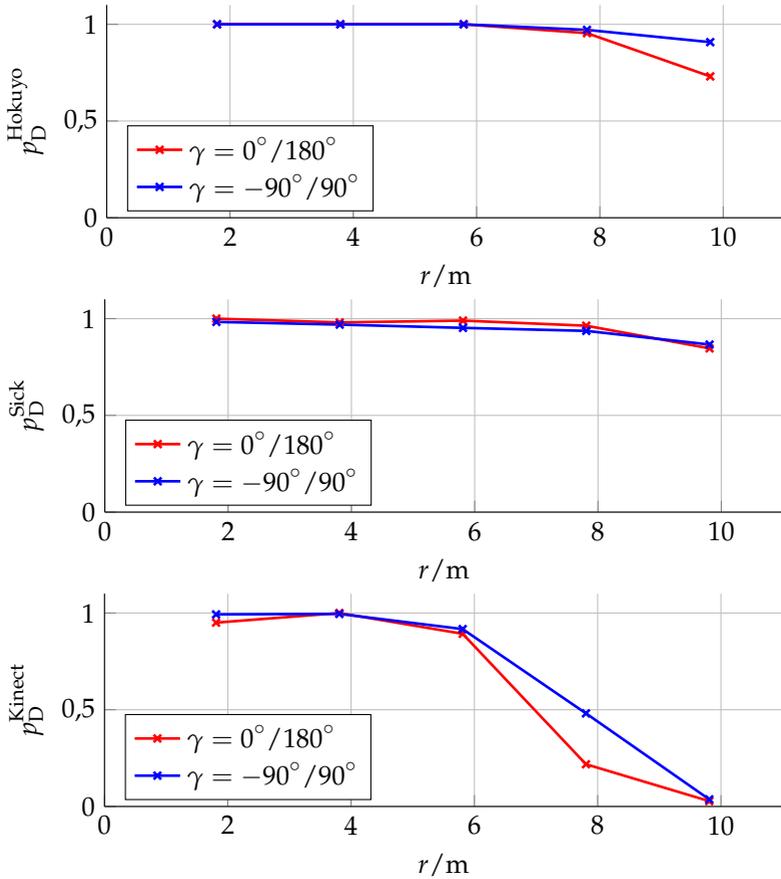
**Abbildung 4.5** Messbereiche zur Bestimmung der sensorabhängigen Detektionswahrscheinlichkeiten in zum Sensor relativen Koordinaten bei unterschiedlichen Orientierungen der Objekte zum Sensor.

In Anlehnung an [99] wurden entsprechende Charakteristika für das gewählte Szenario untersucht. Hierbei wurden sowohl Einfluss von Abstand zum Sensor als auch der Orientierung der zu erfassenden Personen auf das Detektionsergebnis beurteilt. Zur Berechnung der experimentellen Wahrscheinlichkeit  $p_D^{\text{Exp}}$ , mit der ein Objekt erfasst wurde, kann für jeden Sensor der Quotient aus korrekten Detektionen und tatsächlichen vorhandenen Objekten errechnet werden

$$p_D^{\text{Exp}} = \frac{\text{Anzahl korrekter Detektionen}}{\text{Anzahl der Messungen mit Person}}. \quad (4.2)$$



**Abbildung 4.6** Sensorabhängige Detektionswahrscheinlichkeiten in Abhängigkeit vom Abstand  $r$  und Winkel  $\gamma$  einer stehenden Person.



**Abbildung 4.7** Sensorabhängige Detektionswahrscheinlichkeiten in Abhängigkeit vom Abstand  $r$  und Winkel  $\gamma$  einer gehenden Person.

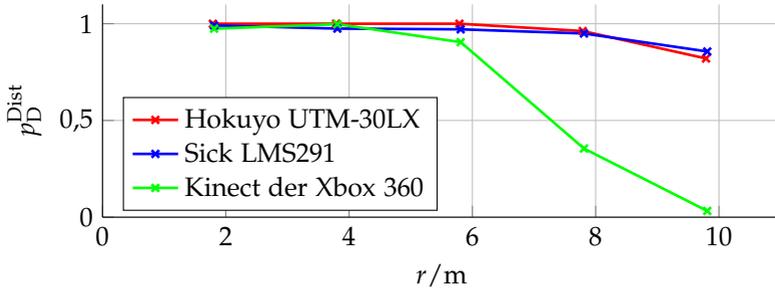
### 4.3.2 Auswertung

Diese Wahrscheinlichkeit wurde für jeden der fünf Messbereiche aus Abbildung 4.5 berechnet, die sich von etwa 1 m bis zu einer Entfernung von 11 m erstreckten, bei einer jeweiligen Länge von 2 m. Während der Messungen wurde auf ideale Bedingungen bezüglich Ausleuchtung oder Verdeckung geachtet, so dass bestmögliche Detektionswerte erreicht wurden. Die Detektionen wurden dabei jeweils von mehreren Personen für stehende Szenarien und Bewegung jeweils mit Orientierung  $\gamma$  von  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$  und  $-90^\circ$  erfasst, wobei keine nennenswerten Unterschiede in den um  $180^\circ$  verdrehten Orientierungen festgestellt wurden. Die restlichen Ergebnisse aus jeweils etwa 1000 Messungen sind für die drei verwendeten Sensoren – zwei Laserscanner der Typen Sick LMS291 und Hokuyo UTM-30LX, sowie den RGB-D-Sensor Microsoft Kinect – in Abbildung 4.6 für stehende Personen und in Abbildung 4.7 für gehende Personen abgebildet. Die Sensoren erzeugten dabei Detektionen entsprechend den in Abschnitt 2.4 behandelten Detektionsverfahren.

Bei allen Messungen ist zu erkennen, dass die bei kleiner Entfernung typischerweise nahe bei eins liegende Detektionswahrscheinlichkeit mit wachsender Entfernung der zu detektierenden Person zum Sensor absinkt. Die Ursachen sind hierfür hauptsächlich bei der begrenzten Auflösung zu suchen und somit einer steigenden Unsicherheit im Klassifikationsprozess zuzuschreiben, die häufiger zur Ablehnung von Objekten führt. Gerade bei dem Kinect-Sensor ist deutlich zu sehen, dass ab etwa 6 m ein starker Einbruch in der Wahrscheinlichkeit, dass Objekte detektiert werden, auftritt. Der eher sprunghafte Verlauf der stehenden Messungen lässt sich durch die geringe Varianz in den Posen erklären, was ebenso impliziert, dass selbst unter möglichst guten Umgebungsbedingungen Detektionsergebnisse weiteren Einflüssen unterliegen, wobei sich in den Messungen der Abstand als deutlichster Einfluss gezeigt hat. Um die Ergebnisse nicht nur von einigen wenigen Posen abhängig zu machen und auch auf das Szenario einer Verfolgung von sich bewegenden Fußgängern zu spezifizieren, wurden die Auswertungen der gehenden Personen für die in den folgenden Kapiteln durchgeführten Versuche verwendet. Dabei wurde der Einfluss der Orientierung für die Bestimmung der Detektionswahrscheinlichkeit herausgemittelt. Orientierungen werden gesondert in Kapitel 5 aufgegriffen. Die resultierenden

Kennlinien von  $p_D^{\text{Dist}}$  sind in Abbildung 4.8 angegeben.

Obwohl die absoluten Werte einer solchen Untersuchung im Anwendungsfall immer von der Wahl eines geeigneten Schwellwertes, dem Wert ab dem ein Klassifikationsergebnis als korrekte Detektion gewertet wird, abhängen und dieser auch für das verwendete Verfahren die Falschalarmrate mitbestimmt [4, 91], ist zu erkennen, dass Abhängigkeiten bezüglich des Abstandes bestehen. Dies zeigt sich unabhängig von der Wahl des Sensors oder des Detektionsverfahrens. Dieses Wissen kann im Vergleich zu einer undifferenzierten Annahme eines konstanten Detektionsverhaltens über das Sichtfeld, wie in Kapitel 6 gezeigt, nutzbringend als Zusatzinformationen für die Multi-Objekt-Filterung verwendet werden.



**Abbildung 4.8** Resultierende sensorabhängige Detektionswahrscheinlichkeiten.

## 4.4 Resultierende Detektionswahrscheinlichkeit

Für die Fusionsalgorithmen aus Kapitel 3 ist die Bestimmung einer gemeinsamen Detektionswahrscheinlichkeit nötig, die alle vorher beschriebenen Einflüsse berücksichtigt. Somit muss das Sensormodell, das hier für den physikalischen Sensor und das Detektionsverfahren gemeinsam gilt, für jede Position, beziehungsweise jeden Zustand, im Raum eine Detektionswahrscheinlichkeit  $p_D(\mathbf{x})$  kennen. Entsprechend den vorherigen Annahmen wird diese als Produkt aller betrachteten Einflussfaktoren definiert

$$p_D(\mathbf{x}) \approx p_D^{\text{FoV}}(\mathbf{x}) \cdot p_D^{\text{Dist}}(\mathbf{x}) \cdot p_D^{\text{Occ}}(\mathbf{x}). \quad (4.3)$$

Je nach Szenario und Umgebung bleibt eine Erweiterung des Modells möglich, die eventuelle konstante Korrekturfaktoren für beispielsweise schlechte Beleuchtung berücksichtigt. Obwohl es sich durch die in den Abschnitten 4.1 – 4.3 getroffenen Annahmen weiterhin um eine Näherung handelt, stellt dieses komplexere Modell eine deutliche Verbesserung gegenüber einem traditionell verwendeten konstanten Sichtfeld dar, wie sich in Abschnitt 6 zeigt.

## 4.5 Anwendung auf Statistik endlicher Mengen

Die sich überlappenden Effekte spielen gerade für verteilte multi-sensorielle Systeme eine große Rolle. Daher bietet sich bei diesen besonders die Verwendung von Filtern, die auf der Statistik endlicher Mengen entsprechend Kapitel 3 basieren, an, da hier versucht wird, unter Einbezug probabilistischer Sensormodelle eine optimale Lösung auf Basis getroffener Annahmen zu finden. Ebenso empfiehlt sich die Vermeidung einer expliziten Datenassoziation von Messung zu Objekten für die Anwendung bei Szenarien, in denen Verdeckungen vorkommen können, da diese Zuordnung in solchen Fällen fehleranfällig ist [18]. Das Hauptziel der Modellierung ist hierbei eine verbesserte Zustandsschätzung und somit das unmittelbare Ergebnis der Multi-Objekt-Filter. Hierzu können

für die Ermittlung von Trajektorien zusätzliche Informationen eingebracht werden, wie eine Re-Identifizierung von Personen [2, 91] oder ein auf höherer Ebene durchgeführtes Trackmanagement (vgl. Abschnitt 3.5), was jedoch nicht im Fokus dieser Arbeit liegen soll, da hierfür sehr anwendungsspezifische Annahmen zu treffen sind.

In diesem Abschnitt wird die Nutzung der in den vorherigen Abschnitten beschriebenen Modellierung einer zustandsabhängigen Detektionswahrscheinlichkeit für eine Nutzung in den Korrekturschritten der vorgestellten Filter aus Kapitel 3 erläutert, während die Auswertung in Kapitel 6 erfolgt.

Die harten Grenzen der Modellierung der Detektionswahrscheinlichkeit in den Abschnitten 4.1 und 4.2 werden bei der konkreten Anwendung der Filter durch die entsprechenden Umsetzungen relativiert. Somit sorgt bei einer Darstellung über Partikel, beispielsweise der PHD oder der Aufenthaltswahrscheinlichkeit des CBMeMBer-Filters, die Unsicherheit der Schätzung für die räumliche Verteilung der Partikel, was dazu führt, dass Partikel eines Objektes nicht zwingend alle die gleiche Detektionswahrscheinlichkeit erhalten. Somit werden einem zu verfolgenden Objekt, trotz teils harter Grenzen in der Modellierung, eine Vielzahl von Detektionswahrscheinlichkeiten zugeordnet, die sich bei Mittelung der zu einem Objekt zugeordneten Partikel zu einer differenzierten Beschreibung ergeben, gerade in Fällen von Objekten an Rändern der Sichtbereiche oder bei Verdeckungen. Gleiches lässt sich auch für die Sigmapunkte bei der Verwendung der Unscented-Transformation in Gauß-Mixtur-Implementierungen erreichen, jedoch entsprechend dem Vergleich von üblicherweise nur geringen Sigmapunkt-Anzahlen und wenigen Gauß-Mixtur-Komponenten bzw. Partikeln in abgeschwächter Form.

Im Folgenden soll nun für die verwendeten Filter dargestellt werden, an welchen Stellen der Filtergleichung und in welcher Form das erweiterte Modell eingebracht werden kann.

### 4.5.1 PHD

Während die hier verwendete Partikel-Variante des PHD-Filters nach [119], die eine Verbesserung der ursprünglichen Implementierung [147] darstellt, ohne Weiteres eine Integration einer nicht konstanten Detekti-

onswahrscheinlichkeit ermöglicht, ist bei der ursprünglichen Herleitung des Filters für eine Implementierung über Gauß-Mixturen [146] die Annahme von einer konstanten Detektionswahrscheinlichkeit im gesamten beobachteten Bereich nötig. Daher wird zuerst kurz auf die Verwendung der adaptiven Detektionswahrscheinlichkeit mit Partikel-basierter Implementierung eingegangen, gefolgt von den Erweiterungen für Gauß-Mixturen.

**Sequentielle Monte-Carlo-PHD** Für die Umsetzung der PHD-Korrektur aus Gleichung (3.25) mit sequentiellen Monte-Carlo (SMC)-Methoden und  $N$  Partikeln wird die PHD beschrieben durch

$$D_{k|k}(\mathbf{x}|\mathcal{Z}_{1:k}) \approx \sum_{n=1}^N \omega_{k|k}^n \delta_{\mathbf{x}_{k|k}^n}(\mathbf{x}). \quad (4.4)$$

Die korrigierten Gewichte  $\omega_{k|k}^n$  der durch Partikel repräsentierten Dirac-Impulse  $\delta_{\mathbf{x}_{k|k}^n}(\mathbf{x})$  im Zustandsraum ergeben sich entsprechend der Formel nach Ristic et al. [119] zu

$$\omega_{k|k}^n = (1 - p_D(\mathbf{x}_k^n))\omega_{k|k-1}^n + \sum_{\mathbf{z} \in \mathcal{Z}_k} \frac{p_D(\mathbf{x}_k^n)g(\mathbf{z}|\mathbf{x}_k^n)}{\kappa_k(\mathbf{z}) + \sum_n p_D(\mathbf{x}_k^n)g(\mathbf{z}|\mathbf{x}_k^n)\omega_{k|k-1}^n}, \quad (4.5)$$

mit der bereits zustandsabhängigen Detektionswahrscheinlichkeit entsprechend Gleichung (4.3) mit Likelihood und Falschalarmen entsprechend Kapitel 3. Hierbei ist  $\omega_{k|k-1}^n$  das prädizierte Gewicht des  $n$ -ten Partikels. Bei den Partikelimplementierungen wird entgegen Kalman-Korrekturschritten der zum Partikel gehörige prädizierte Zustand  $\mathbf{x}_{k|k-1}^n$  nicht korrigiert, da nur die Gewichtung des Partikels angepasst wird. Somit wird Gleichung (4.5) durch Einsetzen des Modells (4.3) für jedes Partikel gelöst. Wird nun für die Extraktion eines Objektes über alle Partikel eines Tracks (siehe Abschnitt 3.5) gemittelt, erfolgt ebenso eine Mittelung über verschiedene Detektionswahrscheinlichkeiten.

**Gaussian Mixture-PHD** Soll die PHD über Mixturen von Normalverteilungen dargestellt werden (GM-PHD), besteht die Hauptschwierigkeit

darin, sicherzustellen, dass diese Approximation auch nach Anwendung von Prädiktions- und Korrekturschritt der Bayes-Rekursion geeignet ist, also die prädizierten/korrigierten Gauß-Mixturen weiter als solche darstellbar sind. Neben den Anforderungen an lineare Bewegungs- und Messmodelle, die jedoch durch Verwendung von EKF- bzw. UKF-Approximationen umgangen werden können, erfordert dies entsprechend [146] auch eine konstante Detektionswahrscheinlichkeit

$$p_D(\mathbf{x}) = p_D = \text{const.} \quad (4.6)$$

Lässt sich die prädizierte PHD entsprechend [146] mit einer Anzahl  $J_{k|k-1}$  von durch  $\omega_{k|k-1}^{(i)}$  gewichteten Normalverteilungen beschreiben

$$D_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{J_{k|k-1}} \omega_{k|k-1}^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}), \quad (4.7)$$

kann die korrigierte PHD wie auch in Gleichung (3.25) und (4.5) aus detektierten und nicht detektierten Anteilen zusammengesetzt werden:

$$D_{k|k}(\mathbf{x}) = (1 - p_D)D_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z} \in \mathcal{Z}} D_{D,k}(\mathbf{x}; \mathbf{z}) \quad (4.8)$$

mit

$$D_{D,k}(\mathbf{x}; \mathbf{z}) = \sum_{j=1}^{J_{k|k-1}} \omega_k^{(j)}(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k}^{(j)}(\mathbf{z}), \mathbf{P}_{k|k}^{(j)}), \quad (4.9)$$

$$\omega_k^{(j)}(\mathbf{z}) = \frac{p_D \omega_{k|k-1}^{(j)}(\mathbf{z}) g_k^{(j)}(\mathbf{z})}{\kappa_k(\mathbf{z}) + p_D \sum_{l=1}^{J_{k|k-1}} \omega_{k|k-1}^{(l)} g_k^{(l)}(\mathbf{z})}. \quad (4.10)$$

Hierbei beschreibt  $D_D(\mathbf{x}; \mathbf{z})$  den detektierten Teil, und es ergeben sich  $\mathbf{m}_{k|k}^{(j)}(\mathbf{z})$  und  $\mathbf{P}_{k|k}^{(j)}$  entsprechend Kalman-, oder auch Unscented- bzw. erweiterten Kalman-Filterschritten mit der Likelihood einer Einzelmessung  $g_k^{(j)}(\mathbf{z})$ . Es lässt sich zeigen, dass für diesen Fall einer konstanten Detektionswahrscheinlichkeit die prädizierte Gaußmixture wieder zu einer Gauß-Mixture  $D_{k|k}(\mathbf{x})$  führt, was ansonsten nur für Modellierungen der

Detektionswahrscheinlichkeit aus Gauß-Mixturen gewährleistet werden kann [131], die mit einem hohen Aufwand verbunden sind.

Aufbauend auf [140] wurde in dieser Arbeit unabhängig von [50] eine zustandsabhängige Detektionswahrscheinlichkeit entwickelt, mit der jede einzelne Gauß-Komponente  $j$  eine von ihrem Mittelwert abhängige Detektionswahrscheinlichkeit  $p_D^{(j)}$  erhält. Diese bleibt somit pro Gauß-Komponente konstant und erhält somit die Möglichkeit, die PHD über eine Gauß-Mixtur darzustellen [50]. In [140] wird die Wahl der Detektionswahrscheinlichkeit stark durch die dort vorhandene Anwendung bestimmt, da eine allgemein geeignete Wahl als schwierig befunden wird. Ebenso führt der dort vorgestellte Ansatz zu einer Verdopplung der Gauß-Komponenten im Falle einer adaptiven Detektionswahrscheinlichkeit. Aber auch die Ermittlung der pro Gauß-Komponente konstanten Detektionswahrscheinlichkeit  $p_D^{(j)}$  durch Mitteln der Detektionswahrscheinlichkeit über die betrachtete Gauß-Komponente

$$p_D^{(j)} = \int p_D(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(j)}) d\mathbf{x} \quad (4.11)$$

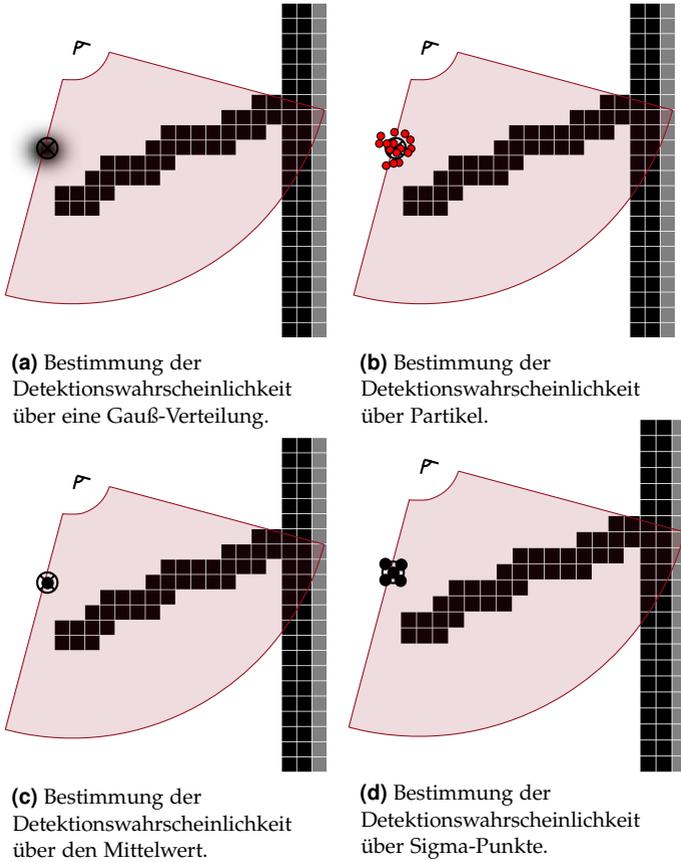
erweist sich hierbei als unpraktikabel, und die Annahme

$$p_D^{(j)} = p_D(\mathbf{m}_{k|k-1}^{(j)}) \quad (4.12)$$

zeigt, dass die Verwendung des Mittelwertes alleine eine zu grobe Näherung sein kann [140]. Nur unter der Bedingung nach [50]

$$\int p_D(\mathbf{x}) \mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}) d\mathbf{x} \approx \int p_D(\mathbf{m}) \mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{P}) d\mathbf{x}, \quad (4.13)$$

einer sich nur langsam um den Mittelpunkt der Gauß-Komponente veränderlichen Detektionswahrscheinlichkeit kann die Verwendung des Mittelwertes alleine als zulässig eingestuft werden. Die Methoden sind in Abbildung 4.9 dargestellt. Hierbei befindet sich ein Objekt mit einer gewissen Unsicherheit (Kreuz mit Kreis) am Rand eines Sichtfelds. Abbildung 4.9(a) zeigt exemplarisch die Überlagerung mit einer Gauß-Verteilung in Draufsicht zur Ermittlung der Detektionswahrscheinlichkeit entsprechend (4.11), während Abbildung 4.9(b) eine numerische Approximation durch Partikel (rote Punkte) vornimmt. Die Wahl des Mittelwertes



**Abbildung 4.9** Unterschiedliche Methoden zur Bestimmung der Detektionswahrscheinlichkeit der Gauß-Komponenten des GM-PHD-Filters.

(schwarzer Punkt) stellvertretend für das Objekt ist entsprechend Gleichung (4.12) in Abbildung 4.9(c) veranschaulicht.

Die Bestimmung der passenden Detektionswahrscheinlichkeit pro Gauß-Komponente lässt sich jedoch auch analog den Linearisierungsschritten für Sensor- und Bewegungsmodell aus Abschnitten 2.1 und 2.2 durch die Verwendung der Unscented-Transformation in Abschnitt 2.3.3 lösen. Die für das Unscented-Kalman-Update der ersten beiden Momente der Gauß-Mixturen bereits erforderlichen  $N_\sigma$  Sigmapunkte können somit einfach weiter verwendet werden, um eine Detektionswahrscheinlichkeit zu bestimmen. Diese hängt nicht nur vom Mittelwert ab, sondern berücksichtigt auch die Unsicherheit und führt beispielsweise bei Verdeckungen zu deutlich größeren Abstufungen und modelliert somit in gewisser Weise eine partielle Verdeckung. Für den Rand eines Sichtbereichs ist die Wahl der Sigmapunkte zur Bestimmung der Detektionswahrscheinlichkeit (schwarze Punkte) exemplarisch in Abbildung 4.9(d) gezeigt. Dabei ist zu berücksichtigen, dass die Unsicherheit im Allgemeinen nicht gleichbedeutend mit der Ausdehnung eines Objektes ist, wenn nicht auf Verfahren zur gleichzeitigen Ausdehnungsschätzung [42, 43] zurückgegriffen wird. Jedoch sind die Detektionsergebnisse sehr wohl abhängig von der Ausdehnung eines Objektes, so dass entstehende Ungenauigkeiten des nicht direkt messbaren Mittelpunktes des Objektes Teil der Messunsicherheit des Systems sind und sich somit auch in gewissem Maße auf die Unsicherheiten der Objekte übertragen. Zusätzlich kann die Wahl der Sigmapunkte in dieser Form keine Symmetrien ausnutzen [54], da diese zwar für normalverteilte Unsicherheiten gegeben sind, aber nicht unbedingt für die Detektionswahrscheinlichkeit. Die vorgeschlagene Näherung stellt somit einen Kompromiss zwischen (4.11) und (4.12) beziehungsweise (4.13) dar und ergibt sich entsprechend Abschnitt 2.3.3 aus den  $i \in \{1, \dots, N_\sigma\}$  Sigmapunkten  $\mathbf{m}_\sigma^{(j,i)}$  mit jeweiligem Gewicht  $\omega_\sigma^{(j,i)}$  der Sigmapunkte pro Gauß-Komponente mit Index  $j$  zu

$$p_D^{(j)} = \sum_{i=1}^{N_\sigma} \omega_\sigma^{(j,i)} p_D(\mathbf{m}_\sigma^{(j,i)}). \quad (4.14)$$

Wird nun die konstante Detektionswahrscheinlichkeit in den Gleichungen (4.8) und (4.10) derart ersetzt, dass jede Gauß-Komponente mit der variablen anstatt der konstanten Detektionswahrscheinlichkeit (4.14) be-

wertet wird, erhält man eine Variation des Gauß-Mixtur-PHD-Filters, die es ermöglicht, zustandsabhängige Sensor-Objekt-Beziehungen wie Abstand oder Verdeckungen zu berücksichtigen mit

$$D_{k|k}(\mathbf{x}) = \sum_{i=1}^{J_{k|k-1}} (1 - p_D^{(i)}) \omega_{k|k-1}^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}) + \sum_{\mathbf{z} \in \mathcal{Z}} D_{D,k}(\mathbf{x}; \mathbf{z}) \quad (4.15)$$

$$D_{D,k}(\mathbf{x}; \mathbf{z}) = \sum_{j=1}^{J_{k|k-1}} \omega_k^{(j)}(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k}^{(j)}(\mathbf{z}), \mathbf{P}_{k|k}^{(j)}) \quad (4.16)$$

und

$$\omega_k^{(j)}(\mathbf{z}) = \frac{p_D^{(j)} \omega_{k|k-1}^{(j)}(\mathbf{z}) g_k^{(j)}(\mathbf{z})}{\kappa_k(\mathbf{z}) + \sum_{l=1}^{J_{k|k-1}} p_D^{(l)} \omega_{k|k-1}^{(l)} g_k^{(l)}(\mathbf{z})} \quad (4.17)$$

bei gleichbleibender Gleichung (4.16). Hierbei ist anzumerken, wie auch für den Fall der Partikelvariante, dass die Anpassung der Detektionswahrscheinlichkeit auch für die Komponenten, die als nicht detektiert erachtet werden, wichtig ist, da auch eine ausbleibende und somit „negative“ Information für eine erfolgreiche Fusion hilfreich sein kann [64].

## 4.5.2 CPHD

**SMC-CPHD** Während beim CPHD-Filter ähnlich wie bei dem PHD-Filter im Falle der Partikel-Variante die entsprechenden Vorschriften der Implementierung der Originalbeiträge [79, 144] verwendet werden können, ist für die Verwendung von Gauß-Mixturen ebenfalls eine Näherung entsprechend dem vorherigen Abschnitt nötig.

**UK-CPHD** Entsprechend den vorherigen Abschnitten kann im Falle variabler Detektionswahrscheinlichkeiten ebenso die Korrektur der PHD und der Kardinalitätsverteilung unter Verwendung der Sigmapunkte der aus der Literatur bekannten Unscented-Kalman-Implementierung des

CPHD-Filters [148] erweitert werden. Lässt sich die prädierte PHD des CPHD-Filters weiterhin über Gleichung (4.7)

$$D_{k|k-1}(\mathbf{x}) = \sum_{i=1}^{J_{k|k-1}} \omega_{k|k-1}^{(i)} \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}) \quad (4.18)$$

darstellen, so kann für die korrekte Verwendung der variablen Detektionswahrscheinlichkeit nicht die aus der Literatur [144, 148] bekannte Korrekturformel verwendet werden, da von einer konstanten Detektionswahrscheinlichkeit ausgegangen wird, die in der Herleitung zu Vereinfachungen führt. Wird jedoch entsprechend die Detektionswahrscheinlichkeit jeder Gauß-Komponente einzeln bestimmt, muss entsprechender Vereinfachungsschritt der Herleitung in [78] und [144, Seite 198, Appendix E] ausgelassen werden. Der Korrekturschritt ergibt sich somit ausgehend von Gleichung (3.32) und der Darstellung über (4.18) unter Verwendung des Zusammenhangs

$$\langle \mathbf{1}, D_{k|k-1} \rangle = \langle \mathbf{1}, \omega_{k|k-1} \rangle = \sum_{i=1}^{J_{k|k-1}} \mathbf{1} \cdot \omega_{k|k-1}^{(i)} \quad (4.19)$$

zu

$$\begin{aligned} Y_k^u[D, \mathcal{Z}](n) &= \sum_{j=0}^{\min(|\mathcal{Z}|, n)} (|\mathcal{Z}| - j)! \rho_{K,k}(|\mathcal{Z}| - j) P_{j+u}^n \\ &\quad \times \frac{\langle \mathbf{1} - p_D, D \rangle^{n-(j+u)}}{\langle \mathbf{1}, D \rangle^n} e_j(\Xi_k(D, \mathcal{Z})) \end{aligned} \quad (4.20)$$

$$\begin{aligned} &= \sum_{j=0}^{\min(|\mathcal{Z}|, n)} (|\mathcal{Z}| - j)! \rho_{K,k}(|\mathcal{Z}| - j) P_{j+u}^n \\ &\quad \times \frac{\langle \mathbf{1} - p_D, \omega \rangle^{n-(j+u)}}{\langle \mathbf{1}, \omega \rangle^n} e_j(\Xi_k(\omega, \mathcal{Z})) \end{aligned} \quad (4.21)$$

$$= \Psi_k^u[\omega, \mathcal{Z}](n). \quad (4.22)$$

Mit Gleichung (4.19) und (4.22) lassen sich die Korrekturformeln der Gauß-Mixtur-CPHD-Implementierung entsprechend Abschnitt 3.5 angeben zu

$$\rho_k(n) = \frac{\Psi_k^0[\omega_{k|k-1}, \mathcal{Z}_k](n) \rho_{k|k-1}(n)}{\langle \Psi_k^0[\omega_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1}(n) \rangle}, \quad (4.23)$$

$$\begin{aligned} D_k(\mathbf{x}) &= \sum_{i=1}^{J_{k|k-1}} \frac{\langle \Psi_k^1[\omega_{k|k-1}, \mathcal{Z}_k](n) \rho_{k|k-1}(n) \rangle}{\langle \Psi_k^0[\omega_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1}(n) \rangle} \\ &\quad \cdot (1 - p_D^{(i)}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^{(i)}, \mathbf{P}_{k|k-1}^{(i)}) \\ &\quad + \sum_{\mathbf{z} \in \mathcal{Z}} \sum_{j=1}^{J_{k|k-1}} \omega_k^{(j)}(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_k^{(j)}(\mathbf{z}), \mathbf{P}_k^{(j)}) \end{aligned} \quad (4.24)$$

mit

$$\omega_{k|k-1} = [\omega_{k|k-1}^{(1)}, \dots, \omega_{k|k-1}^{(J_{k|k-1})}]^T \quad (4.25)$$

$$\omega_k^{(j)} = p_D^{(j)} \omega_{k|k-1}^{(j)} g_k^{(j)}(\mathbf{z}) \frac{\langle \Psi_k^1[\omega_{k|k-1}, \mathcal{Z}_k \setminus \{\mathbf{z}\}], \rho_{k|k-1} \rangle \langle \mathbf{1}, \kappa_k \rangle}{\langle \Psi_k^1[\omega_{k|k-1}, \mathcal{Z}_k], \rho_{k|k-1} \rangle \kappa_k(\mathbf{z})}, \quad (4.26)$$

wobei  $g_k^{(j)}(\mathbf{z})$  die Likelihood der Messung  $\mathbf{z}$  der  $j$ -ten Gauß-Komponente ist und  $\mathcal{Z} \setminus \{\mathbf{z}\}$  bedeutet, dass in diesem Ausdruck die Messung  $\mathbf{z}$  kein Element der Menge  $\mathcal{Z}$  ist. Die Gauß-Komponenten lassen sich weiterhin, wie auch die Mittelwerte und Varianzen der einzelnen Komponenten des GM-PHD-Filter-Updates, durch die Kalman-Filter-Gleichungen berechnen.

### 4.5.3 CBMeMber

Ähnlich den beiden vorangegangenen Filtern kann für die SMC-CBMeMber-Filter die Standardvariante für zustandsabhängige Detektionswahrscheinlichkeiten entsprechend [144] verwendet werden. Auf die Beschreibung einer zustandsabhängigen Gauß-Implementierung [144] wurde in dieser Arbeit verzichtet, da sich, wie Kapitel 6 zeigt, das CPHD-Filter für den

hier angenommenen Anwendungsfall als geeigneter herausstellt. Eine Anpassung der entsprechenden Gauß-Implementierung ist jedoch analog zu den vorangegangenen Filtern möglich.

## 5 Richtungssensitive Verfolgung

Während im vorherigen Kapitel mitunter Modellansätze vorgeschlagen wurden, die theoretische Näherungen und Messmodelle des optimalen Bayes-Filters mit realen Szenarien verknüpfen, soll diese Verknüpfung in diesem Kapitel fortgeführt und erweitert werden. Durch die Gewinnung zusätzlicher Information aus den Rohdaten der Sensoren soll insbesondere durch den Vorschlag eines Verfahrens zur Orientierungsschätzung aus Einzelbildern der Nutzen solchen Zusatzwissens für die Multi-Objekt-Verfolgung motiviert werden. Die vorgestellten Verfahren orientieren sich dabei an einem vom Autor veröffentlichten Zeitschriftenbeitrag [165].

In vielen Bereichen, wie zum Beispiel der Medizintechnik, der Sportwissenschaft oder der Animationstechnik, werden bereits Verfahren zum Posentracking von Menschen eingesetzt, um eine detaillierte Analyse menschlicher Bewegungsabläufe zu ermöglichen. Meist werden dafür Videoaufnahmen aus einer oder mehreren Perspektiven herangezogen [129]. Manche dieser Verfahren verwenden spezielle Marker, die am menschlichen Körper angebracht sind, oder benötigen sehr viele Kameras. Das markerlose Posentracking mit wenigen Kameras in unkontrollierten Umgebungen ist jedoch ein bisher ungelöstes Problem [88, 129]. Gerade solche Methoden wären für eine Vielzahl technischer Anwendungen in beispielsweise der Robotik, der Fahrerassistenz oder der Überwachung interessant und könnten neue Einsatzmöglichkeiten außerhalb des Labors erschließen, weshalb an dieser Stelle ein großer Forschungsbedarf besteht. Bei Anwendungen im Straßenverkehr stehen oftmals zusätzliche Umfellsensoren (z. B. Lidar, Radar) zur Verfügung. Die meisten Ansätze zur Fußgängererkennung beschäftigen sich bislang jedoch mit der Erkennung und Klassifikation von Fußgängern und erlauben nur eine eingeschränkte Inferenz über die Pose und Orientierung von Fuß-

gängern [37]. Falls eine Orientierungsschätzung beabsichtigt ist, wird diese teilweise von der Verfolgung losgelöst durchgeführt [38] oder ein vorhergehendes Tracking wird als Voraussetzung für eine Orientierungsschätzung benötigt [30]. Dabei lässt sich die Schätzung der Orientierung gut für eine Gefahrenanalyse, beispielsweise im Straßenverkehr, einsetzen [60], bei der auch spontane Änderungen der Bewegungsrichtung vorkommen können.

**Repräsentation menschlicher Bewegung** Um menschliche Bewegungen in Messdatensequenzen zu repräsentieren, werden globale und lokale Merkmale unterschieden [111, 153]. Anhand der Merkmale kann eine direkte Klassifikation der Bewegungen erfolgen.

**Globale Bewegungsmerkmale** repräsentieren Informationen über den gesamten Bereich einer Messdatenfolge, in der sich die Bewegung abspielt, sind allerdings störanfällig und blickwinkelabhängig. Die Interessensregion (*region of interest*, ROI) wird durch Personendetektion und -verfolgung ermittelt. Daraus werden *Templates* für verschiedene Bewegungen gebildet. Beim *Motion Energy Image* und bei *Space-Time Shapes* werden Körpersilhouetten aufeinanderfolgender Zeitschritte zusammengefügt und daraus Merkmale gebildet [13]. Ferner kann Bewegungsinformation durch den optischen Fluss in der ROI berücksichtigt werden. Sowohl in kontrollierten Umgebungen ohne Verdeckungen [111] als auch in Kombination mit Grammatikmodellen werden gute Ergebnisse erzielt – auch bei Aufnahmen geringer Auflösung oder aus großer Entfernung [153].

Zur Extraktion **lokaler Bewegungsmerkmale** werden nur ROIs einer Bildsequenz betrachtet. Aus der Literatur sind viele für die Bewegungsklassifikation geeignete lokale Merkmale bekannt: Neben den weit verbreiteten SIFT-, SURF- und HOG-Merkmalen (siehe Abschnitt 2.4.1) und ihren Erweiterungen auf den dreidimensionalen Fall seien hier vorwiegend die sog. *Space-Time Interest Points* (STIPs) [69] und die auf *Local Binary Patterns* (LBPs) basierenden *Local Trinary Patterns* [160] genannt. Lokale Merkmale sind robust gegenüber Verdeckungen und Blickwinkeländerungen [111]. Fehlende Informationen über die örtlich-räumliche Struktur der Bewegung lassen sich durch Verfolgung kompensieren.

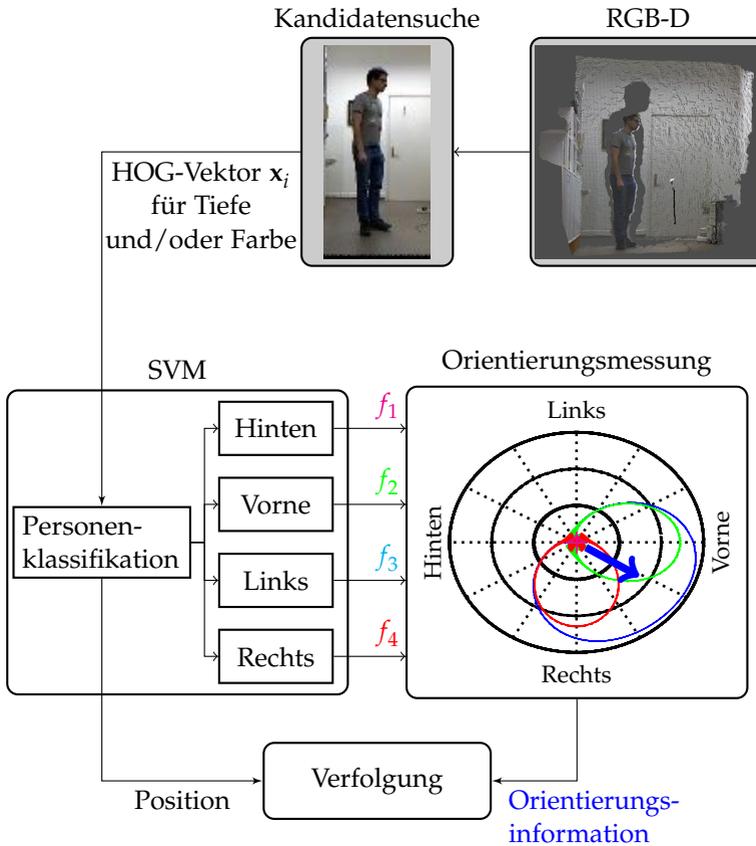
Bei der **direkten Klassifikation** wird die gesamte Bewegung durch einen Bewegungsdeskriptor dargestellt [111]. Bei Verwendung lokaler Merkmale müssen diese in den Bewegungsdeskriptor überführt werden. Die Klassifikation erfolgt meist mit diskriminativen Klassifikatoren, wie z. B. SVMs (siehe Abschnitt 2.5.1). Vorteilhaft an diesem Ansatz ist eine gewisse Invarianz gegenüber der Variabilität von Bewegungen.

**Herausforderungen in der Fußgänger-Verfolgung** Beim Verfolgen einer unbekanntem Anzahl an Objekten stellt die Zuordnung von zum Teil fehlerbehafteten Detektionen zu qualifizierten Objekten eine besondere Herausforderung dar. Anders als im Ein-Objekt-Fall muss im Multi-Objekt-Fall, wie in Kapitel 3 beschrieben, zusätzlich aus den Messdaten die Anzahl der vorhandenen Objekte geschätzt werden. Dies stellt eine erhöhte Anforderung an die Sensordatenverarbeitung dar – insbesondere wenn sich mehrere Objekte auf engem Raum bewegen. Des Weiteren sind die bei der Verfolgung von Personen üblicherweise verwendeten Bewegungsmodelle einer nahezu konstanten Geschwindigkeit lediglich eine grobe Näherung des tatsächlichen Bewegungsverhaltens. Richtungsänderungen müssen aus der Schätzung der Trajektorie über die Zeit erkannt werden, die jedoch auf einem nur näherungsweise zutreffenden Modell beruht und zusätzlich Messunsicherheiten unterliegt, weswegen hier Informationen über die Orientierung aus den Messdaten hinzugezogen werden sollen.

**Ansatz** In diesem Kapitel soll ein Ansatz vorgestellt werden, der Zusatzwissen aus einer Posenschätzung nutzt, um die Schätzung von Zuständen der Personen in einer Szene zu unterstützen. Auf der Grundlage lokaler Merkmale aus Einzelbildern kann bereits eine robuste Aussage über die Orientierung einer Person getroffen werden [25, 105], deren Auflösung mitunter der Hinzunahme von Tiefendaten hier verbessert wird. Zusätzlich wird die neu gewonnene Information in die Filterung nach dem Ansatz endlicher Mengen (Kapitel 3) eingebracht und auf nichtlineare Fälle anwendbar gemacht. Zur Umfeldwahrnehmung der Tiefendaten dient ein RGB-D-Sensor wie die Microsoft Kinect.

Im Folgenden werden aufbauend auf dem Verfahren zur Personendetektion (Abschnitt 5.1), die Kandidaten für nachfolgende Schritte sucht,

die vorgenommenen Erweiterungen der Orientierungsbestimmung (Abschnitt 5.2) erläutert. Darauf folgend wird das Modell für den Einbezug der zusätzlichen Messung in den Ansatz zur Multi-Objekt-Filterung vorgestellt (Abschnitt 5.3).



**Abbildung 5.1** Ablaufdiagramm des vorgestellten Verfahrens zur richtungsabhängigen Detektion und Verfolgung von Fußgängern mit vier Orientierungsklassen.

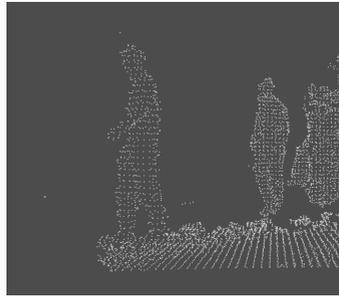
## 5.1 Kandidatensuche und Personendetektion

Um Personen nicht nur in einem Videobild, sondern auch im dreidimensionalen Raum verfolgen zu können, sind Sensoren nötig, die über die reinen Bildintensitäten hinaus noch Information über die Tiefe der beobachteten Szene liefern und somit Objekte im Raum lokalisieren können. Durch ihre günstige Verfügbarkeit bietet sich die Verwendung von RGB-D-Sensoren an. Die Tiefendaten bieten sich auch dazu an, die Personendetektion zu verbessern und den hierfür nötigen Rechenaufwand gegenüber einem klassischen, rechenintensiven *Sliding-Window*-Ansatz über alle möglichen Detektionsfenster [22] zu reduzieren [91, 132]. Der vorgestellte Ansatz basiert auf HOG-Merkmalen und erweitert den Ansatz nach [91], lässt sich jedoch auch auf andere Merkmale wie in [105, 107] anwenden. Das HOG-Merkmal hat sich im Bereich der Personendetektion als sehr gut geeignetes Merkmal herausgestellt und ist auf diesem Gebiet sehr weit verbreitet [22, 25, 28, 91, 99]. Eine Übersicht über das Verfahren ist in Abbildung 5.1 zu sehen.

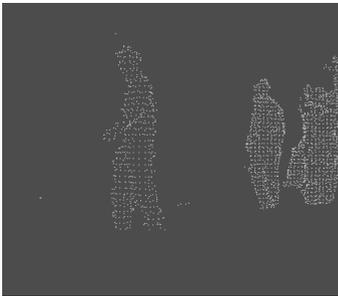
Zu Beginn des Detektionsschritts liegen die Tiefendaten des RGB-D-Sensors als 3D-Punktvolke mit  $x$ -,  $y$ - und  $z$ -Koordinate und RGB-Werten für jeden dieser Punkte vor. Oft wird hier auch von einer 2,5D-Aufzeichnung gesprochen, da die Aufnahmen nur aus der Perspektive des Sensors dreidimensional sind, jedoch durch Verdeckungen nicht die volle dreidimensionale Information der anderen Seite von Objekten oder weiteren Objekten im Hintergrund verfügbar ist, wie beispielsweise an den Schatten in Abbildung 5.2(a) zu sehen ist. Während sich die anschließende Verfolgung im hier gewählten Szenario auf die  $x,y$ -Ebene reduziert, spielt die Höhe  $z$  dennoch bei der Sensorrohdatenverarbeitung eine wichtige Rolle zur Skalierung von Kandidatenfenstern. Für die Vorverarbeitung der Punktvolke bietet es sich zur Reduktion der Rechenzeit an, die Punktmenge durch Unterabtastung zu reduzieren. Da die meisten Sensoren bei wachsenden Entfernungen eine abnehmende Auflösung aufweisen, eignet sich die sogenannte Voxel-Gitter-Filterung zudem, um eine Punktvolke mit äquidistanten Punkten zu erhalten (siehe Abbildung 5.2(b)). Um den Suchraum zu reduzieren und eine ROI festzulegen, werden nach [91] zuerst alle Punkte, die zum Boden gehören, entfernt. Um die entsprechende Ebenenbeschreibung des Bodens zu erhalten, muss einmalig diese Ebene definiert werden. Dies kann leicht unter der



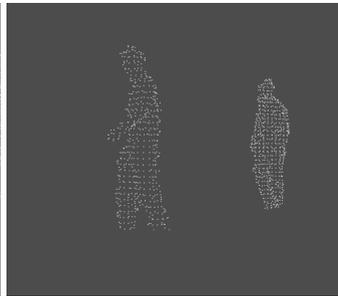
(a) RGB-D-Punktwolke einer Szene mit zwei Personen.



(b) Gefilterte Punktwolke mit äquidistantem Punktgitter.



(c) Gefilterte Punktwolke ohne Boden.



(d) Punktwolke aus den verbleibenden Kandidaten.

**Abbildung 5.2** Ablauf der Kandidatensuche.

Annahme einer konstanten und bekannten Ausrichtung des Sensors zum Boden durchgeführt werden. Das Ergebnis ist in Abbildung 5.2(c) zu sehen. Nun wird die verbliebene Punktwolke, deren Inhalt nun nicht mehr über den Boden miteinander verbunden ist, aufgrund der euklidischen Abstände zwischen den Punkten zu Gruppen zusammengefasst, die im Innenbereich beispielsweise zu detektierende Personen, Möbel oder Wände darstellen. Geht man nun davon aus, dass die zu detektierenden Personen eine gewisse Mindestgröße und eine maximale Größe aufweisen, lassen sich zum Beispiel Wände oder sehr kleine Punktgrup-

pen ausschließen. Auf der Grundlage der maximalen und minimalen Gruppengröße wird somit eine Vorauswahl getroffen (siehe Abbildung 5.2(d)). Für die verbleibenden – typischerweise wenigen – Punktgruppen wird nun jeweils ein Detektionsfenster von fester Größe der Breite 64 Pixel und Höhe 128 Pixel benötigt, für das der HOG-Merkmalsvektor berechnet werden kann. Zur Skalierung werden die Abmessungen der Punktgruppe in Höhe und Breite berücksichtigt, um für jede Person einen optimalen Rand des Detektionsfensters von etwa 16 Pixeln [22] zu erhalten. Eine Merkmalsberechnung ist somit weder für alle Bereiche eines Frames nötig, noch fällt die sonst nötige Berechnung verschiedener Skalen an, da diese durch die Nutzung von 3D-Daten bekannt ist. Diese Kandidatenfenster und die auf ihnen berechneten Merkmalsvektoren können nun durch einen vorab trainierten SVM-Klassifikator auf das Vorhandensein einer Person überprüft werden und erzeugen zusammen mit der Positionsinformation der entsprechenden Punktgruppe eine Detektion relativ zum Sensor, während Punktgruppen mit negativem Klassifikationsergebnis verworfen werden. Die resultierenden Detektionen des Beispiels in Abbildung 5.2 sind in Abbildung 5.3 durch grüne Rahmen gekennzeichnet.

## 5.2 Orientierungsmessung

Für die Verarbeitung von Orientierungsinformationen wird der auf HOG-Merkmalen basierende Personendetektionsalgorithmus mit integrierter Richtungserkennung aus [25] erweitert. Im Gegensatz zu [34, 41, 105, 136] bietet dieser Ansatz eine kontinuierliche Schätzung der Orientierung unter Berücksichtigung von Intra-Klassenvarianzen der Trainingsdaten und führt aufgrund der verwendeten HOG-Merkmale zu Verbesserungen gegenüber der Verwendung von Haar-Merkmalen in [127]. Auch wird für das Verfahren keine vorher bekannte Trajektorie wie in [30] oder eine Startorientierung wie in [71] benötigt und bleibt dabei im Gegensatz zu [73] echtzeitfähig. In [16] werden mehrere kalibrierte Tiefensensoren verwendet, um die Orientierung anhand von Schulterpartien zu schätzen, wobei die Schätzung stark von den zu spezifizierenden Parametern einer vorgegeben Körpergeometrie abhängt und durch den Verzicht auf Lernverfahren kaum Invarianz bezüglich veränderter Posen zeigt. Eine

detailliertere Schätzung der Pose wird in [128] vorgeschlagen, ist allerdings nur auf den Winkelbereich der Orientierung von  $180^\circ$  ausgelegt, setzt eine hohe Auflösung der Tiefendaten voraus und benötigt einen hohen Trainings- und Rechenaufwand, wobei gerade bei gebräuchlichen RGB-D-Kameras nur in kleinen Bereichen des Sichtfelds hohe Auflösungen erreicht werden, die auch stark von den Beleuchtungsverhältnissen der Umgebung abhängen [132].

Im vorgestellten Ansatz wird die Herangehensweise aus [25] erweitert und für Tiefendaten angewendet. Dabei werden die Tiefendaten zusätzlich nach [91] genutzt, um Kandidatenbereiche auszuwählen und die zugehörigen Farb- und Tiefenbilder mittels mehrerer SVMs zu klassifizieren, um aus den Bildern eine Orientierung zu gewinnen. Die Klassifikationsergebnisse der vorgegebenen vier bzw. acht Hauptrichtungen werden daraufhin zur Verbesserung der Zustandsschätzung genutzt. Für die Schätzung der Orientierung wird der HOG-Merkmalvektor weiterverwendet, der durch die Kandidatensuche im vorherigen Abschnitt gefunden wurde, wobei bei höherem Rechenaufwand auch die Kombination mit anderen in der Literatur vorgeschlagenen Merkmalen für Orientierungsschätzung in Tiefendaten [107] denkbar wäre, deren Nutzen bisher allerdings nur auf künstlichen Daten vorgestellt wurde. Daher wurde sich hier bei der Wahl der evaluierten Merkmale auf Variationen des HOG-Merkmals beschränkt.

Zur besseren Unterscheidung des Orientierungsergebnisses vor und nach der in Abschnitt 5.3 folgenden Fusion mit Positionsdaten und dem zeitlichen Kontext wird die Orientierung nach der Fusion als Orientierungsschätzung bezeichnet. Die Orientierung, die sich entsprechend der nächsten Abschnitte aus Einzelbildern ergibt, soll als Orientierungsmessung bezeichnet werden, da sie für das Fusionsfilter eine Messung darstellt, vorab allerdings ebenfalls von Klassifikatoren geschätzt wird.

### 5.2.1 Experten-Klassifikatoren

In Anlehnung an [25] wird hier ebenfalls eine Mischung von „Experten“-Klassifikatoren gewählt, die auch für die Schätzung von Kopfposen erfolgreich eingesetzt wurde [125]. Allerdings unterscheidet sich der vorgestellte Ansatz deutlich von [25] durch die genaue Funktion und das Anlernen der Expertenmeinungen. In dieser Arbeit wird die Mischung

der Expertenmeinung nur für die Bestimmung der Orientierung und nicht zur Detektion eingesetzt und durch eine kontinuierliche Gewichtung der Trainingsdaten sehr gezielt auf einzelne Richtungen angelernt. Die  $N$  verwendeten Experten werden durch lineare SVM-Klassifikatoren realisiert, die wie in [25] auf  $N$  Hauptrichtungen, die mit  $r$  indiziert werden, angelernt werden. Die SVM-Klassifikatoren nach [25] zeigen dabei nach [136] eine mit einem *Random-Decision-Forest* Klassifikator vergleichbare Leistung. Für  $N = 4$  Hauptrichtungen werden die Orientierungen Hinten, Vorne, Links und Rechts verwendet. Ebenso kann der Klassifikator in Anlehnung an [34, 127] auf acht gleichmäßig verteilte Klassen erweitert werden, wobei die zusätzlichen Klassen zwischen den Hauptrichtungen liegen. Das Ergebnis jeder Klassifikation kann dabei basierend auf den Trainingsdaten nach Abschnitt 2.5.2 in eine Wahrscheinlichkeit  $f_r(\mathbf{x}_i)$  für die Klassenzugehörigkeit des Merkmalsvektors  $\mathbf{x}_i$  eines zu testenden Bildes transformiert werden. Diese wird als Gewicht für die Bestimmung der Wahrscheinlichkeitsdichte  $f_{\text{Ori}}$  der kontinuierlichen Orientierung  $\theta$  verwendet:

$$f_{\text{Ori}}(\theta|\mathbf{x}_i) = \frac{1}{N} \sum_{k=1}^N f_r(\mathbf{x}_i) q_r(\theta). \quad (5.1)$$

Hierbei wird, wie auch in [25], jeder Experte zusammengesetzt aus der durch seine Wahrscheinlichkeit bestimmten Gewichtung  $f_r$  und seinem Modell

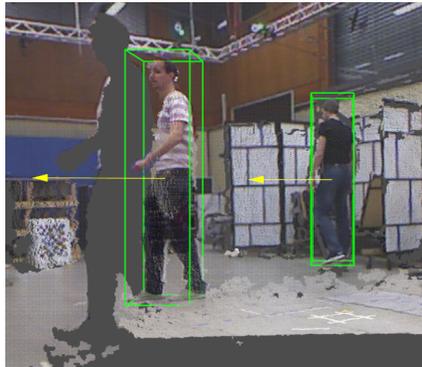
$$q_k(\theta) = \mathcal{N}(\theta | \mu_r, \sigma_{\text{Exp}}^2), \quad (5.2)$$

das durch eine Normalverteilung mit Mittelwert  $\mu_r$  der jeweiligen Richtung der gewählten Anzahl an Klassen und einer für alle Klassen konstanten Standardabweichung eines Expertenmodells  $\sigma_{\text{Exp}}$  beschrieben ist. Die beste Schätzung der gemessenen Orientierung kann hierbei aus der Dichte  $f_{\text{Ori}}(\theta|\mathbf{x})$  in Gleichung (5.1) gewonnen werden, indem beispielsweise das Maximum gewählt wird [25]. In dieser Arbeit hat sich jedoch die Integration der Dichte in einem Bereich als robuster erwiesen. Dabei hat sich als geeignete beste Orientierung der Mittelpunkt eines Winkelbereichs, der das höchste Gewicht vereinigt, herausgestellt. Der ausgewertete Winkelbereich besitzt gerade die Ausdehnung einer Klasse,

die bei acht Klassen einem Winkelbereich von  $WB = \pi/4$  entspricht. Der gemessene Wert der Orientierung  $\theta_{\max}$  ergibt sich somit zu

$$\theta_{\max}(\mathbf{x}_i) = \arg \max_{\theta} \int_{\theta-WB}^{\theta+WB} f_{\text{Ori}}(\theta' | \mathbf{x}_i) d\theta'. \quad (5.3)$$

Die beste Orientierung kann somit für jeden Merkmalsvektor  $\mathbf{x}_i$  zusammen mit den zugehörigen Positionsinformationen ermittelt werden. Eine beispielhafte Superposition von vier Experten mit bester Schätzung ist in Abbildung 5.1 zu sehen. Zudem zeigt sich, dass die durch Superposition entstehende Dichte der Orientierung auch eine Aussage über die Unsicherheit der Messung, die in Abschnitt 5.3.3 näher beschrieben wird, zulässt. Beste Orientierungen im dreidimensionalen Fall sind in Abbildung 5.3 gezeigt.

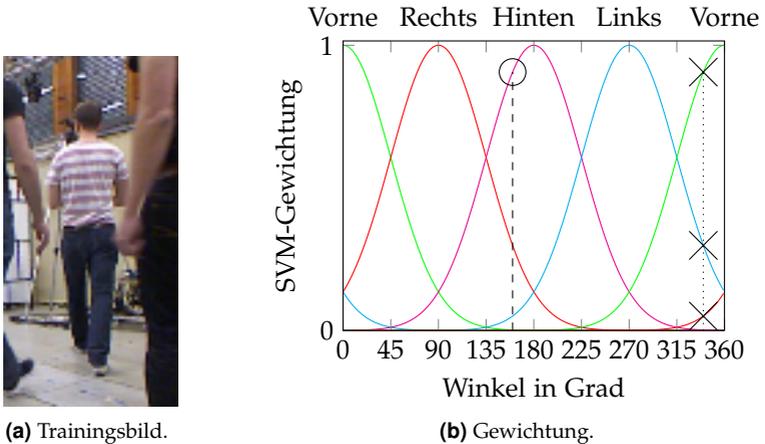


**Abbildung 5.3** Detektionen in RGB-D-Punktwolke.

## 5.2.2 Lernstrategie

Das Anlernen der  $N$  Klassifikatoren spielt eine entscheidende Rolle für die Bestimmung der Orientierung. Im Gegensatz zu [25] und der reinen Personendetektion aus Abschnitt 5.1 bestehen negative Trainingsbeispiele hier nicht aus Bildern ohne Fußgänger, sondern aus Bildern von Fußgängern anderer Richtungen, um eine bestmögliche Abgrenzung zu erreichen. Des Weiteren sollen zum einen beim Training Orientierungsunterschiede innerhalb einer Klasse von Richtungen berücksichtigt werden.

Zum anderen soll auch eine stärkere Abgrenzung von Richtungsklassen gegenüber der um  $180^\circ$  verschobenen, gegenüberliegenden Klasse im Vergleich zu – im Fall von vier Klassen – den beiden verbleibenden, um  $90^\circ$  verschobenen Richtungsklassen erreicht werden.



**Abbildung 5.4** Positive (Kreise) und negative (Kreuze) SVM-Gewichte für ein Trainingsbeispiel mit der Orientierung  $160^\circ$  der Klasse „Hinten“ bei Verwendung von 4 Klassen. Dabei ist die gepunktete Linie zur Auswahl der negativen Gewichte  $180^\circ$  entfernt von der wahren Orientierung des Trainingsbildes, die gestrichelt gekennzeichnet ist.

## Gewichtung

Hierzu wird ein mit kontinuierlichen Orientierungen versehener Trainingsdatensatz verwendet und das normalverteilte Modell der Experten  $q_r$  erneut aufgegriffen, das sowohl positive Trainingsbeispiele in einer gewichteten SVM bei größerer Abweichung von der idealen Haupttrichtung geringer gewichtet, als auch Negativbeispiele umso stärker gewichtet, je mehr sie der gegenüberliegenden Richtung entsprechen. Dazu kann das positive Gewicht  $w_{\text{pos}}$  eines Trainingsbeispiels mit der Orientierung  $\theta_T$  über den Parameter  $\sigma_{\text{SVM}}$  bei Beachtung der Periodizität des Winkels errechnet werden:

$$w_{k,\text{pos}}(\theta_T) = \mathcal{N}(\theta_T | \mu_r, \sigma_{\text{SVM}}^2). \quad (5.4)$$

Negative Gewichte errechnen sich zu

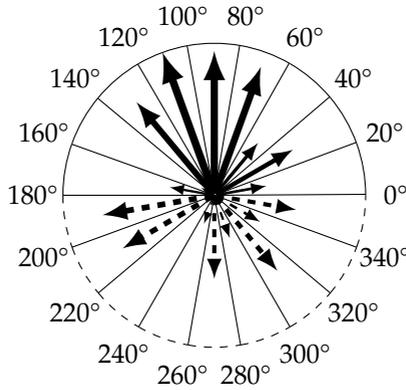
$$w_{k,\text{neg}}(\theta_T) = \mathcal{N}(\theta_T + 180^\circ | \mu_r, \sigma_{\text{SVM}}^2), \quad (5.5)$$

wobei immer ein positives Gewicht für das Trainingsbeispiel aus den zugehörigen  $N$  Trainingsklassen ausgewählt und das Trainingsbeispiel in den Klassifikatoren der restlichen Klassen mit den entsprechenden negativen Gewichten bewertet wird.

Abbildung 5.4 zeigt die Gewichtung für den Fall von vier Trainingsklassen. Dabei wird ein Beispielbild mit einer Orientierung von  $160^\circ$  in der zugehörigen Klasse „Hinten“ positiv mit einer Gewichtung entsprechend dem Kreis in den gewichteten SVM-Klassifikator eingebracht. Die Klassifikatoren der anderen Klassen nehmen hierbei die mit einem Kreuz an der um  $180^\circ$  gegensätzlichen Stelle gekennzeichneten Gewichte für ein Negativbeispiel an. Somit wird einerseits die harte Klassenunterscheidung zwischen zwei Klassen abgeschwächt, andererseits werden auch fehleranfällige Unterscheidungen wie Vorne/Hinten durch gezieltes Training vermieden.

### 5.2.3 Anwendung des HOG-Merkmals

Während das HOG-Merkmal typischerweise insbesondere im Kontext von Personen auf RGB-Bildern mit 9 Orientierungsklassen der Gradienten eingesetzt wird [22, 25, 99, 132], hat sich gezeigt, dass bei anderen Objektklassen eine gerichtete Klasseneinteilung von Vorteil ist [28]. Hierbei wird, wie in Abbildung 5.5 gezeigt, die Einteilung nicht nur im Bereich von  $180^\circ$ , sondern für  $360^\circ$  vorgenommen und liefert 18 Klassen, die die Richtung der Gradienten berücksichtigen. Wird entsprechend ein HOG-Merkmal, wie in [132], nicht mehr auf RGB-, sondern auf Tiefendaten angewendet, wird die Wichtigkeit der Richtung der Gradienten anhand der ersten beiden Zeilen von Abbildung 5.6 deutlich. Während die Gradientenrichtung für RGB-Bilder nicht eindeutig ein Merkmal beispielsweise für die Unterscheidung von Hinter- und Vordergrund sein muss, ist in den Tiefenbildern die Richtung eines Gradienten automatisch ein Hinweis auf die räumliche Beziehung der Pixel, die in [132] keine Berücksichtigung findet. In der gegebenen Darstellung 5.6 entsprechen dabei helle Pixel einem weit entfernten und dunkle einem näher gelegenen Punkt.



**Abbildung 5.5** Klassen der Gradientenorientierung des HOG-Merkmals ohne (durchgezogen) und mit (gestrichelt) Berücksichtigung der Richtung.

## Ausschneiden

Durch den vorhergehenden Schritt der Kandidatensuche (Abschnitt 5.1) ist es darüber hinaus noch möglich, den die Person beschreibenden Teil der Tiefendaten herauszufiltern. Dazu werden die während der Kandidatensuche geformten  $N$  Cluster auf ihre minimalen  $V_{\min}^i$  und maximalen Punkte  $V_{\max}^i$  hin untersucht. Diese beschreiben jeweils den Punkt des Clusters mit den kleinsten beziehungsweise größten  $x, y, z$ -Koordinaten im Raum und lassen somit Rückschlüsse über die Ausdehnung des Clusters zu. Somit werden nur die zum Cluster gehörenden Pixel mit Positionen  $x$  und  $y$  des originalen Tiefenbildes  $t_{\text{voll}}(x, y)$  in das Trainingsbeziehungsweise zur Laufzeit in den auszuwertenden  $i$ -ten Bildausschnitt  $t_{\text{cut}}(x, y)$  übernommen:

$$t_{\text{cut}}^i(x, y) = \begin{cases} t_{\text{voll}}(x, y) & \text{für } V_{\min}^i \leq t_{\text{voll}}(x, y) \leq V_{\max}^i \\ \text{const.} & \text{sonst} \end{cases} \quad (5.6)$$

Hierbei wird  $t_{\text{voll}}(x, y)$  sowohl als Tiefenbild als auch als Punkt mit Pixelkoordinate  $x, y$  und Tiefenkoordinate  $z = t_{\text{voll}}(x, y)$  verstanden. Davon abweichende Punkte und die zugehörigen Pixel werden gesondert behandelt, da sie keine relevante Tiefeninformation enthalten. Die daraus resultierenden Bilder sind in der dritten Zeile von Abbildung 5.6 zu

sehen, wobei nicht verwendete Pixel weiß gefärbt sind. Analog ist dies auch für Farbbilder, wie in der fünften Zeile gezeigt, möglich. Durch das Vorgehen werden die so für Training und Test errechneten Merkmale unabhängig vom Hintergrund.

## Normierung

Bereits in [132] zeigt sich, dass eine gesonderte Normierung für HOG-Merkmale auf Tiefendaten sinnvoll ist, da die Pixelwerte keine RGB-Farbtintensitäten, sondern Entfernungen codieren. Während in [132] eine Normierung auf den gesamten Erfassungsbereich vorgenommen wird, macht der zuvor beschriebene Auswahlsschritt eine Normierung pro Kandidatenbild möglich, so dass die vorgefundenen Tiefenwerte auf den kompletten Wertebereich ausgedehnt werden können

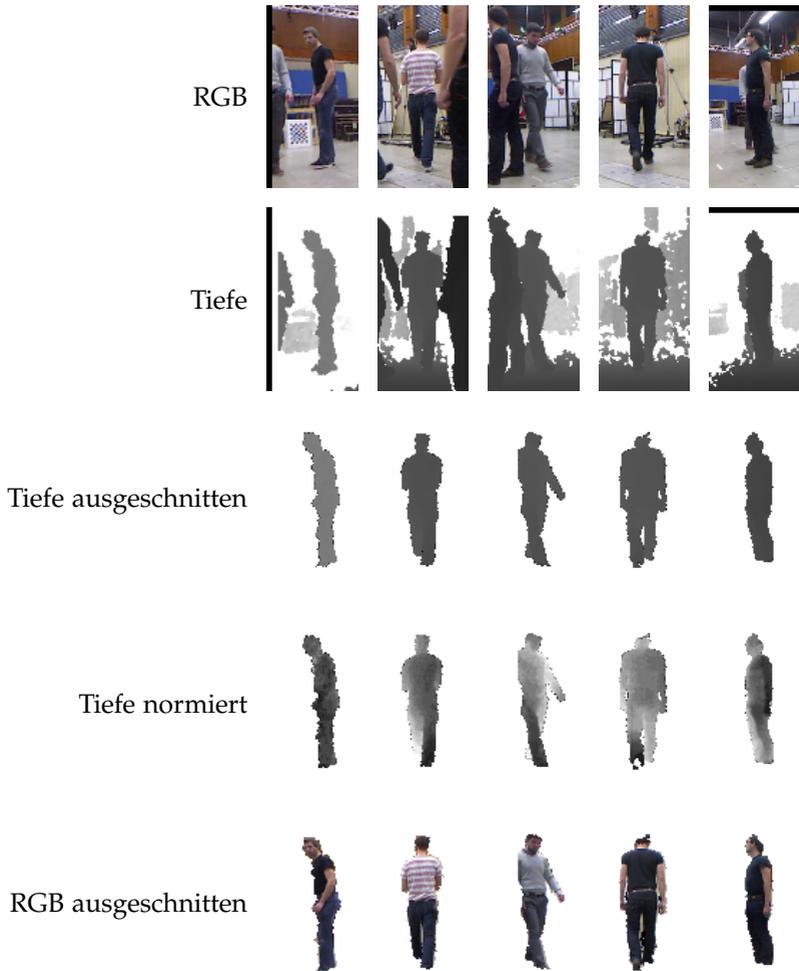
$$t_{\text{cut,norm}}^i(x, y) = \frac{t_{\text{cut}}^i(x, y) - \min_{x,y} t_{\text{cut}}^i(x, y)}{\max_{x,y} t_{\text{cut}}^i(x, y) - \min_{x,y} t_{\text{cut}}^i(x, y)}. \quad (5.7)$$

Das Ergebnis ist in der vierten Zeile von Abbildung 5.6 zu sehen. Die Normierung erreicht somit eine Invarianz gegenüber der Entfernung von Trainings- und Testbildern und verstärkt gleichzeitig die Sichtbarkeit der unterschiedlichen Entfernungen einzelner Körperteile. Im Fall der Tiefenbilder weisen die Daten ebenfalls auch kaum personenspezifische Merkmale wie Kleidung auf, so dass für die Bestimmung unerwünschte Unterschiede der Daten in Training und Test kaum Berücksichtigung finden.

Pixel, die im Auswahlsschritt nicht berücksichtigt wurden, werden für die Normierung ebenfalls nicht verwendet und am Ende der Vorverarbeitung auf null gesetzt.

## 5.3 Fusion für die Verfolgung

Ziel einer geeigneten Repräsentation der Orientierung ist eine im Filter zu erreichende Korrektur der auf einem Modell basierenden Prädiktion. Die Literatur zur Verfolgung von Fußgängern geht zumeist von einem Modell konstanter Geschwindigkeit aus [20, 91, 100, 132]. Auch wenn



**Abbildung 5.6** Auswahl verschiedener RGB- und Tiefenbilder mit und ohne Hintergrund bei entsprechender Skalierung. Dunkle Pixel stehen für kurze Entfernung zum Sensor, während helle Pixel weit weg sind.

zusätzliche Erweiterungen um soziales menschliches Verhalten [49, 117] genutzt werden, bleiben weiterhin unvorhersehbare Wechsel der Bewegungsrichtung, Orientierungswechsel im Stand oder die Orientierung bei einer Initialisierung unberücksichtigt, da diese Information bei der reinen Detektion der Position fehlt und nur über zeitliche Bewegung erschlossen werden kann. Gerade im Fall von unsicherheitsbehafteten Messungen oder unbewegten Objekten kann dieser Ansatz jedoch zu Fehlern führen und lässt sich auch durch gleichzeitige Verwendung multipler Bewegungsmodelle [36], bei beispielsweise unveränderter Position, aufgrund fehlender Information nicht auflösen.

Um die aus den vorherigen Abschnitten gewonnene Information der Orientierung nutzen können, muss diese in das Multi-Objekt-Filter eingebracht werden. Dazu muss das auf der alleinigen Nutzung der Position relativ zum Sensor basierende Messmodell (vgl. Abschnitt 2.2) erweitert werden, um auch eine zum Sensor relative Orientierung zu berücksichtigen. Da durch den Aufbau des vorgestellten Verfahrens zu jedem Personenkandidaten eine Dichteverteilung der Orientierung aus Gauß-Mixturen vorliegt, bieten sich für die Fusion einer prädizierten Orientierung der zu verfolgenden Personen und der Messung im Korrekturschritt der Filterung je nach Implementierung unterschiedliche Möglichkeiten an. Diese Beschreibung muss im Falle eines effizienten Kalman-Filter-basierten Ansatzes auf die Verwendung der ersten beiden Momente reduziert werden, die sich jedoch für verschiedene Kandidaten innerhalb eines Zeitschritts unterscheiden kann.

### 5.3.1 Messmodell

Abhängig von verwendeter Beschreibung des Zustandsraums kann eine prädizierte Messung mit Orientierungsinformation  $\hat{\mathbf{z}}_{\text{ori}}$  aus dem ursprünglichen Messmodell für die Positionsinformation aus Abschnitt 2.2 und einer prädizierten Orientierung  $\hat{\theta}$  aus dem Messmodell der Orientierung  $h_{\text{Ori}}$  zusammengesetzt werden. Das Bewegungsmodell wird dabei entsprechend Abschnitt 2.1 beibehalten. Hierbei berechnet sich  $\hat{\theta}$  bei kartesischer Darstellung der Geschwindigkeit in  $x$ - und  $y$ -Richtung mit vom Bewegungsmodell prädizierter Geschwindigkeit  $v_{x,k|k-1}$  und

$v_{y,k|k-1}$  im Zustandsraum aus

$$\hat{\theta} = h_{\text{Ori,kartesisch}}(v_{x,k|k-1}, v_{y,k|k-1}) = \arctan\left(\frac{v_{y,k|k-1}}{v_{x,k|k-1}}\right). \quad (5.8)$$

Bei polarer Darstellung ergibt sich die prädierte Orientierung direkt zu

$$\hat{\theta} = h_{\text{Ori,polar}}(\theta_{k|k-1}) = \theta_{k|k-1}. \quad (5.9)$$

Je nachdem, wie der eingesetzte Sensor im Raum orientiert ist, muss entsprechend die eigene Orientierung des Sensorträgers in einem für alle Sensoren gemeinsamen Koordinatensystem berücksichtigt werden.

### 5.3.2 Korrekturschritt und Repräsentation der Unsicherheit

Für die Korrektur ist neben der prädierten Orientierung aus dem vorherigen Abschnitt ebenso eine Beurteilung der Wahrscheinlichkeit, dass eine neue Messung bei gegebenen prädierten Objekten auftritt, nötig. Diese wird durch die in Abschnitt 2.3.2 eingeführte Likelihood dargestellt. Durch diese kann die prädierte Orientierung mit der gemessenen Orientierung verglichen werden, die entsprechend Gleichung (5.3) bestimmt wurde.

Während für die Repräsentation dieser Likelihood bei Positionskorrekturen für gewöhnlich eine konstante normalverteilte Unsicherheit verwendet wird [91, 117, 155], kann in diesem Fall, neben einer sich aus Vermessung des Verfahrens ergebenden konstanten Unsicherheit, auch eine aus dem Verfahren selbst stammende Beurteilung der Genauigkeit der Messung verwendet werden. Diese Beschreibung resultiert aus der Verwendung der  $N$  Expertenklassifikatoren aus Abschnitt 5.2.1, die in (5.1) auch als Likelihood interpretiert werden kann. Diese gibt an, wie wahrscheinlich bei den vorhandenen prädierten Orientierungen eine bestimmte Ausrichtung eines gegebenen Merkmalsvektors eines Testbildes ist. Diese Likelihood wird hier nun in mehreren Formen verwendet.

#### Kontinuierliche Likelihood

Wird eine Gewichtung mit dem ursprünglichen Expertenmodell vorgenommen, kann eine eintreffende Messung mit dem Merkmalsvektor  $\mathbf{x}_j$

der Messung bei einer prädizierten Orientierung  $\hat{\theta}_i$  entsprechend mit (5.1) gewichtet werden

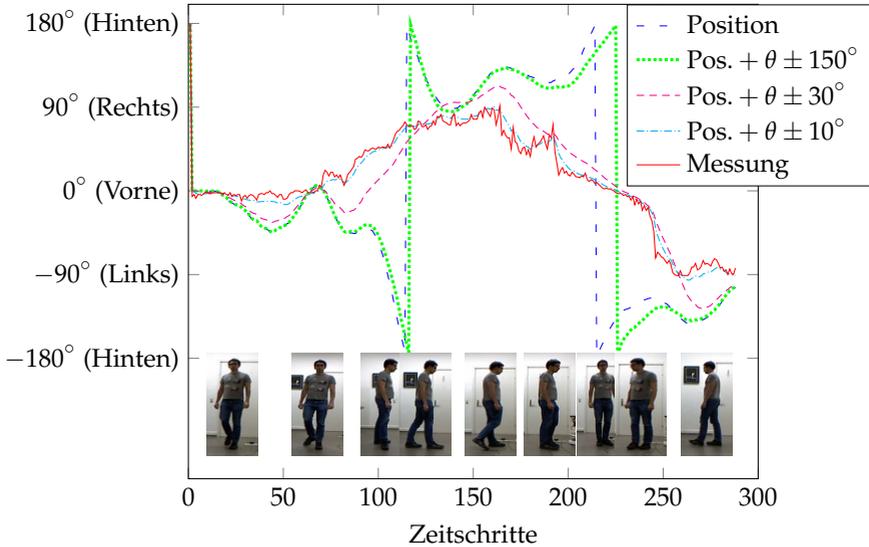
$$g_{\text{Ori}}(\mathbf{x}_j|\hat{\theta}_i) = \frac{1}{N} \sum_{k=1}^N f_r(\mathbf{x}_j) q_r(\hat{\theta}_i), \quad (5.10)$$

die sich speziell für die Verwendung mit Partikel-Filter-Varianten anbietet. Dies führt allerdings zu erhöhtem Aufwand, da im Vergleich zu den nachfolgenden Methoden pro Klasse dem Fusionsfilter alle nötigen Parameter zur Beschreibung der Verteilung  $q_r$  mit Gewicht zur Beschreibung der Dichte bekannt sein müssen. Bei Annahme einer von der Positionsmessung unabhängigen Orientierungsmessung ergibt sich nach [133] die resultierende Likelihood aus dem Produkt der Likelihood-Funktion für Position (2.9) und der Likelihood-Funktion der Orientierung (5.10).

### Adaptive normalverteilte Likelihood

Für die Nutzung der Orientierungsmessung in einem recheneffizienten Kalman-Ansatz ist eine Reduktion der Dichte der Orientierung auf die ersten beiden Momente sinnvoll. Im Zuge des Kalman-Update-Schritts kann das erste Moment entsprechend dem Messmodell aus Abschnitt 5.3.1 berechnet werden. Die daraus erhaltene Abweichung zwischen  $\hat{\theta}$  und der gemessenen Orientierung  $\theta_{\text{max}}$  aus Abschnitt 5.2 kann nun als weiteres Element dem Kalman-Innovationsvektor zugeführt werden und zusammen mit einer Unsicherheit, die sich aus den Ungenauigkeiten des Schätzalgorithmus ergibt, in den Kalman-Gleichungen genutzt werden. Dabei wird, wie im vorigen Abschnitt, von einer von der Positionsmessung unabhängigen Orientierungsmessung ausgegangen. Basierend auf den Unsicherheiten für Position und Orientierung kann somit eine Korrektur des prädizierten Zustands unter Berücksichtigung der zusätzlichen Information erfolgen. Die Unsicherheit  $\sigma_{\text{Ori, adapt}}$  berechnet sich dabei unter Berücksichtigung der Zirkularität aus der Varianz der Dichte (5.10) für einen Merkmalsvektor  $\mathbf{x}_j$  mit  $f_{\text{Ori}}(\theta|\mathbf{x}_j)$  entsprechend (5.1) zu

$$\sigma_{\text{Ori, adapt}}^2 = \int_0^{2\pi} (\theta' - \hat{\theta}_{\text{max}}(\mathbf{x}_j))^2 f_{\text{Ori}}(\theta|\mathbf{x}_j) d\theta'. \quad (5.11)$$



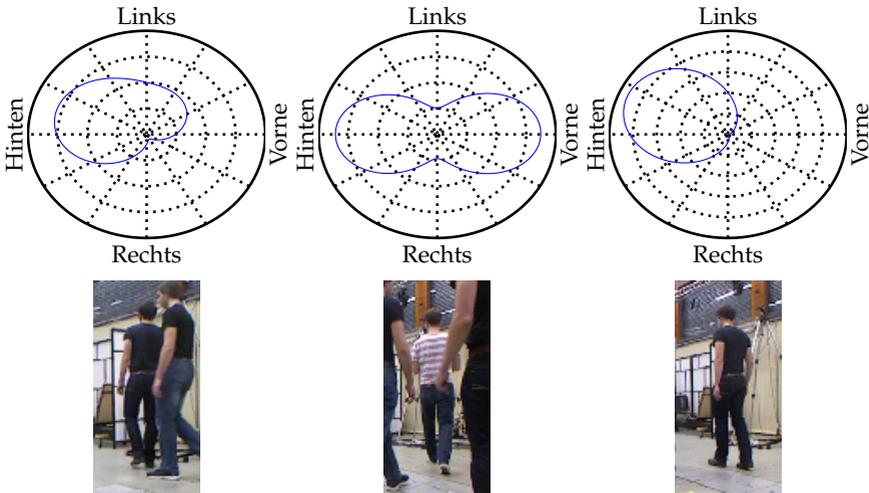
**Abbildung 5.7** Die Abbildung zeigt die Fusion von Orientierungs- und Positionsmessung zur Orientierungsschätzung bei variierender Unsicherheit für ein Ein-Personen-Szenario. Dazu ist sowohl die Schätzung der Orientierung alleine aufgrund von Positionsinformationen (blau) und alleine auf Orientierungsmessungen (rot) gezeigt, als auch Fusionsergebnisse für verschieden angenommene Unsicherheiten der Orientierungsmessung zwischen  $10^\circ$  und  $150^\circ$ . Die Farbbilder beschreiben dabei den tatsächlichen Verlauf in der Szene.

Bei der Annahme einer normalverteilten Unsicherheit der Orientierungsmessung wird die Messung durch die zwei Parameter (5.3) und (5.11) charakterisiert. Da die sich hieraus errechnende Standardabweichung mindestens so groß wie die Standardabweichung eines Experten mit  $\sigma_{\text{Exp}}$  entsprechend der Klassenbreite ist, wird die verwendete Standardabweichung um diesen Wert korrigiert

$$\sigma_{\text{Ori, adapt, korr}} = \sigma_{\text{Orient, adaptiv}} - \sigma_{\text{Exp}} \cdot \quad (5.12)$$

### Konstante normalverteilte Likelihood

Zusätzlich besteht die Möglichkeit, analog der Positionsmessunsicherheit, die Unsicherheit des Verfahrens über die Standardabweichung des Fehlers aus einem Testdatensatz zu beurteilen, die zu einer konstanten Unsicherheit  $\sigma_{\text{Ori, konst}}$  aller Messungen führt.



**Abbildung 5.8** Dichte der Orientierungsschätzung für beispielhafte Bilder.

### 5.3.3 Rolle der Unsicherheit

Am Beispiel aus Abbildung 5.7 soll verdeutlicht werden, welchen Einfluss die im Messmodell angenommene Unsicherheit für die Fusion spielt. Die Abbildung zeigt dabei ein Ein-Personen-Beispiel, in dem Positionsinformationen und Orientierungsinformationen fusioniert werden, während die angenommene Unsicherheit der Orientierung jeweils für die gesamte Szene variiert wird. Hierbei ist in Rot die Messung der Orientierung aus Einzelbildern gezeigt und in Blau die aus der zeitlich veränderten Position geschätzte Orientierung der in den Bildern gezeigten Person aufgetragen. Durch die drei unterschiedlichen Wahlen der für den Abschnitt als konstant angenommenen Orientierungsunsicherheit von  $10^\circ$ ,  $30^\circ$  und

150° ist gut zu erkennen, wie sich das Ergebnis der Fusion aus Position und Orientierung beeinflussen lässt. Somit stellt die Unsicherheit der Orientierung ein effektives Mittel dar, um die Schätzung eher auf dem Einfluss der Orientierungsmessung oder auf den Ergebnissen der Positionsmessung beruhen zu lassen. Dies ermöglicht es, die Schätzung der Orientierung wie im gezeigten Beispiel deutlich zu verbessern, hängt aber von der jeweiligen Güte der Messungen ab.

In Abbildung 5.8 sind zudem verschiedene Beispiele von Orientierungsmessungen gezeigt. Gut zu sehen sind die unterschiedlichen aufgetragenen Dichten der Orientierungsmessung. Diese geben somit Aufschluss über die Unsicherheit der Messung, die nicht konstant sein muss und den Einsatz einer adaptiven Fusion zwischen stärkeren Anteilen der Orientierungs- bzw. Positionsmessung motiviert. Eine detailliertere Evaluation einer adaptiven Unsicherheit erfolgt in Abschnitt 6.4.2.



# 6 Ergebnisse

Ziel dieses Kapitels ist es, die vorgestellten Verfahren zu evaluieren. Hierzu wird in Abschnitt 6.1 das entwickelte System mit seinen Komponenten vorgestellt, das die Basis der Vergleiche liefert. Das zur Gewinnung von Referenzdaten eingesetzte Motion-Capture-System wird in Abschnitt 6.2 vorgestellt. Diese Referenzdaten sind insbesondere zur Beurteilung der Zustandsschätzung sinnvoll. Anschließend werden die Verfahren aus den Kapiteln 3 und 4 in Abschnitt 6.3 evaluiert und diskutiert, gefolgt von der Evaluation der Kombination dieser mit den Methoden aus Kapitel 5 in Abschnitt 6.4.

## 6.1 Systemaufbau

Der Aufbau des nachfolgend vorgestellten Systems ist auf die Verfolgung von Personen in Innenräumen ausgelegt. Alle Algorithmen wurden so gewählt und entwickelt, dass keine Einschränkungen bei bewegten Sensoren entstehen. Somit wird nicht von konstanten Hintergründen ausgegangen, die eine einfache Segmentierung von Vorder- und Hintergrund ermöglichen würden. Aufgrund der limitierten Bandbreite in Drahtlosnetzwerken, die hier zur Kommunikation dienen, ist eine echtzeitfähige Übertragung beispielsweise einer oder gar mehrerer Punktwolken mit zusätzlichen RGB-Informationen nicht in Echtzeit möglich, so dass lediglich die auf den Sensorplattformen erzeugten Detektionen übertragen werden. Die Detektionsalgorithmen aus den vorherigen Kapiteln nutzen dabei lediglich einzelne Bilder bzw. Scans, um auch robust gegenüber Eigenbewegung zu sein. Im Folgenden wird zunächst auf die eingesetzten Sensorträger eingegangen. Der vorgestellte Aufbau des Systems eignet sich jedoch für beliebige Sensorkonfigurationen.

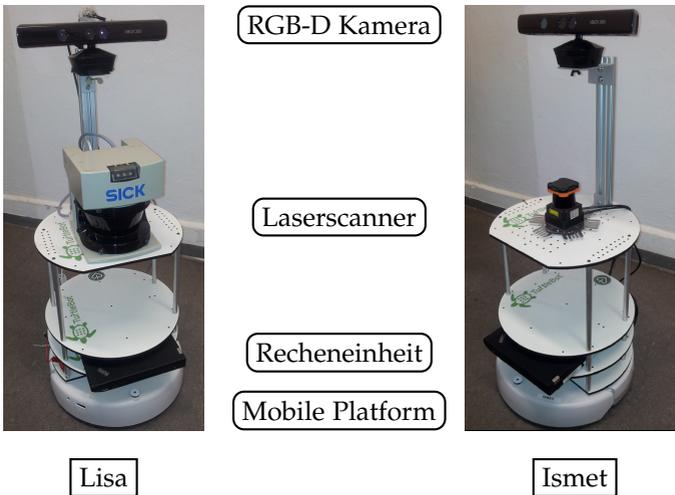


Abbildung 6.1 Sensorträger.

### 6.1.1 Sensorträger

Als Sensorträger wurden für den Gebrauch im Haushalt dimensionierte mobile Plattformen (*TurtleBots*) gewählt, die Ende 2010 vom Hersteller *Willow Garage* vorgestellt wurden. Die mobile Basiseinheit ist hierbei ein *iRobot Create*. Durch einen Aufbau ist es einer solchen Plattform möglich, verschiedene Sensoren aufzunehmen und mit einer eigenen Recheneinheit Daten zu verarbeiten und über (W)-LAN mit anderen Geräten zu kommunizieren. Die eingesetzte Konfiguration der Sensorträger ist in Abbildung 6.1 zu sehen. Hierbei ist Sensorträger *Ismet* aus einer mobilen Plattform, dem *TurtleBot*, einem *Microsoft Kinect*-Sensor und einem *Hokuyo UTM-30LX*-Laserscanner (Winkelauflösung  $0,25^\circ$ ) aufgebaut. Sensorträger *Lisa* besteht ebenfalls aus einem *TurtleBot* mit einem *Kinect*-Sensor und ist mit einem *Sick LMS291* Laserscanner (Winkelauflösung  $0,5^\circ$ ) ausgestattet, um auch unterschiedliche Eigenschaften von ähnlichen Sensoren berücksichtigen zu können. Die 1D-Laserscanner sind jeweils etwa auf Kniehöhe in etwa 50 cm angebracht. Die Kinect-Sensoren befinden sich in etwa 80 cm über dem Boden und weisen einen leichten Neigungswinkel nach oben auf, um auch größere Objekte in möglichst

kleiner Entfernung zu erfassen. Die von den Sensoren erfassten Rohdaten werden von der Recheneinheit auf jedem Sensorträger zu Detektionen verarbeitet.

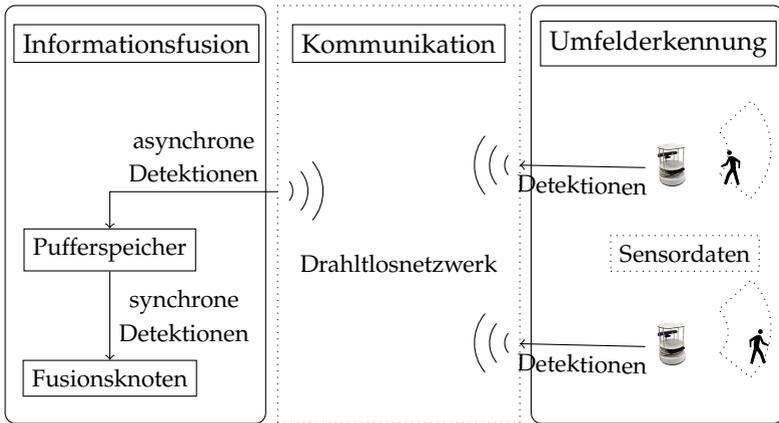


Abbildung 6.2 Systemaufbau.

## 6.1.2 Informationsaustausch

Die von den Sensorträgern erzeugten Detektionen werden entsprechend Abbildung 6.2 an einen Fusionsknoten gesendet. Das Gesamtsystem lässt sich dabei in den Rohdaten verarbeitenden Teil der Umfelderkennung, den Kommunikationskanal und den Fusionsteil untergliedern. Dabei sind die über das Funknetzwerk zu übertragenden Messungen weiterhin im Koordinatensystem der jeweiligen Sensorplattform gegeben, so dass dem Fusionsknoten auch die Position des Sensors als Messung bekannt sein muss. Die Messungen werden alle auf der Sensorplattform mit einem über alle Sensorplattformen synchronen Zeitstempel versehen, der am Fusionsknoten genutzt werden kann, um Verzögerungen in der Übertragung, zum Beispiel durch einen Pufferspeicher, auszugleichen. Somit kann für die Fusion von einer chronologisch geordneten Menge an Messungen ausgegangen werden. Für das hier vorgestellte Szenario ist ein vollständiger Austausch der Detektionen zwischen Sensorträgern

und Fusionsknoten möglich. Für den Fall sehr vieler Sensorträger ist die Verwendung einer verteilten Fusion entsprechend [8, 97] notwendig.

### 6.1.3 Eingesetzte Detektionsverfahren

**Laserbasierte Detektionen** Das eingesetzte Verfahren zur Detektion von Personen in Laserdaten entspricht dem Verfahren aus Abschnitt 2.4.2 nach [4]. Dabei wurde für jeden der eingesetzten Lasersensoren im Gegensatz zu [164] ein eigener Klassifikator auf Basis von jeweils etwa 2500 Scans nach [4] trainiert. Die Detektionen der Lasersensoren werden mit einer Rate von etwa 15 Hz erzeugt.

**RGB-D-Detektionen** Die Detektionen der RGB-D-Sensoren werden entsprechend dem in Abschnitt 5.1 vorgestellten, auf [91] basierenden Verfahren, erzeugt. Die reinen Detektoren für Personen wurden hierbei wie in [91] auf dem INRIA-Datensatz [22] trainiert. Das Training der Orientierungen ist Bestandteil der Evaluierung in Abschnitt 6.4. Die Detektionen werden mit einer Rate von etwa 12 Hz erzeugt. Eventuell entstehende Interferenzen, die durch die an Streifenprojektion [11] angelehnten strukturierten Muster im Nah-Infrarotbereich entstehen können, werden mit der Methode nach [17] unterdrückt.

### 6.1.4 Fusion

Im Fusionsknoten werden die eintreffenden Messungen chronologisch mit den in Kapiteln 3–5 vorgestellten Algorithmen abgearbeitet. Somit findet die Fusion zwischen den Sensorträgern rein auf Detektionsebene statt, während große Teile der Daten, wie Farb- und Tiefendaten, bereits während der Detektion innerhalb eines Sensorträgers fusioniert werden.

### 6.1.5 Implementierung

Zur Umsetzung des Projekts wurde auf verschiedene Bibliotheksfunktionen zurückgegriffen, auf die im Folgenden eingegangen wird. Die Rohdatenverarbeitung der Sensoren und die Kommunikation ist in der

Programmiersprache C++ implementiert. Zwecks schnellerer Entwicklung wurde auf *Matlab* für die Implementierung und Erweiterung der Filterung im Fusionsknoten zurückgegriffen.

Für die Kommunikation und die Steuerung der Sensorträger bzw. die Einbindung der Sensoren wurde das *Robot Operating System (ROS)* [115] verwendet. Dieses Open-Source-Projekt bietet neben vielen integrierten Treibern für Sensoren eine eigene Kommunikationsstruktur, die gerade für die Kommunikation zwischen verschiedenen Sensoren, aber auch zwischen örtlich verteilten Sensorträgern den Austausch von Rohdaten, Steuerbefehlen und Detektionen vereinfacht und parallele und verteilte Berechnungen ermöglicht.

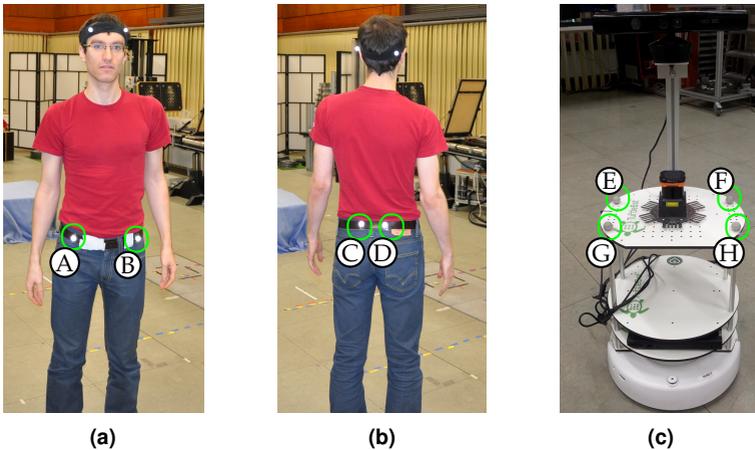
Für die in Kapitel 5 entwickelten Verfahren wurden die Funktionen des Open-Source-Projekts *Point Cloud Library (PCL)* [123] für effiziente Manipulationen und Filterungen von Punktwolken genutzt. Zur Verarbeitung von Farb- und Tiefenbildern wurden die *Open Computer Vision (OpenCV)* [14] Bibliothek sowie die *VLFeat* [143] Open-Source-Bibliothek verwendet. Die Kalibrierung von Kinect zu Laserscanner wurde mittels der Kalibrierungsfunktionen nach [58] durchgeführt.

Die Implementierung der Fusion macht Gebrauch eines am Institut für Industrielle Informationstechnik in *Matlab* entwickelten Filtertools [65], das verschiedene grundlegende Strukturen zur Repräsentationen verschiedener Zustandsräume, Sensormodelle und Dichten, sowie Filter zur Datenfusion bietet und in dieser Arbeit um die vorgestellten Methoden erweitert wurde, jedoch nur für wenig komplexe Fälle Echtzeitfähigkeit bietet.

## 6.2 Experimentelle Evaluierung

Die genaue Evaluierung eines Multi-Objekt-Szenarios spielt eine wichtige Rolle bei der Beurteilung der vorgestellten Verfahren. Manuelle Annotationsverfahren wie in [132] bieten oft nur unvollständige Referenzdaten, da nur die Pixelpositionen in einem Video markiert werden, ohne eine absolute Position anzugeben. Andere Datensätze lassen wiederum keine Rückschlüsse über Orientierungen zu [155]. Unter Zuhilfenahme aufwändiger Sensorik, wie 2D-Laserscannern, ist umfassende manuelle Annotation zwar möglich [35], diese unterliegt aber weiterhin aufgrund

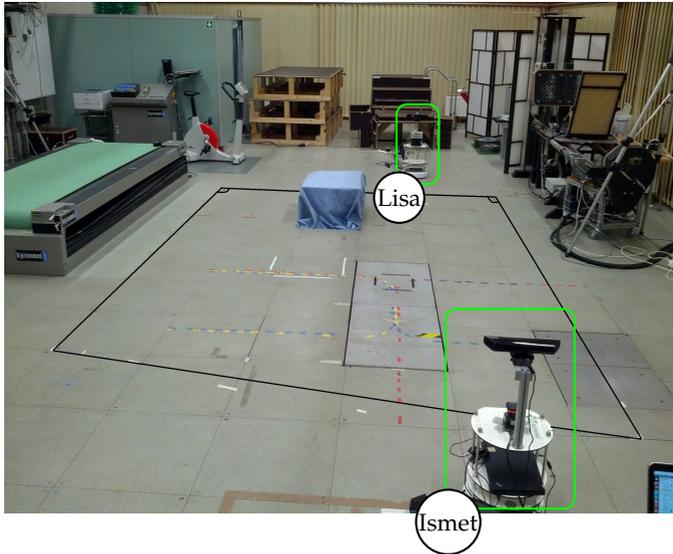
des manuellen Anteils Varianzen bezüglich der Genauigkeit. Aktive Systeme, wobei hier aktiv bedeutet, dass die Testpersonen durch das Tragen von zusätzlicher Ausrüstung aktiv bei der Referenzdatenaufzeichnung beteiligt sind, sind in der Lage, ähnlich zu [129] (keine Multi-Objekt-Szenarien) oder [92] (keine Orientierungsdaten) Referenzdaten mit einer deutlich höheren Genauigkeit aufzuzeichnen. Ohne ein solches Referenzsystem bleibt die Evaluierung im Multi-Objekt-Fall meist auf die Beurteilung der Kardinalität beschränkt, wenn anstatt von Simulationen reale Experimente bewertet werden sollen. Doch gerade verlässliche Vergleiche von Filterergebnissen für Realdaten sind entscheidend bei der Beurteilung der Filteransätze.



**Abbildung 6.3** Positionen der Marker bei Testperson mit vier Kopf- und vier Hüftmarkern und bei Sensorträger mit vier Markern.

### 6.2.1 Motion-Capture-Systeme

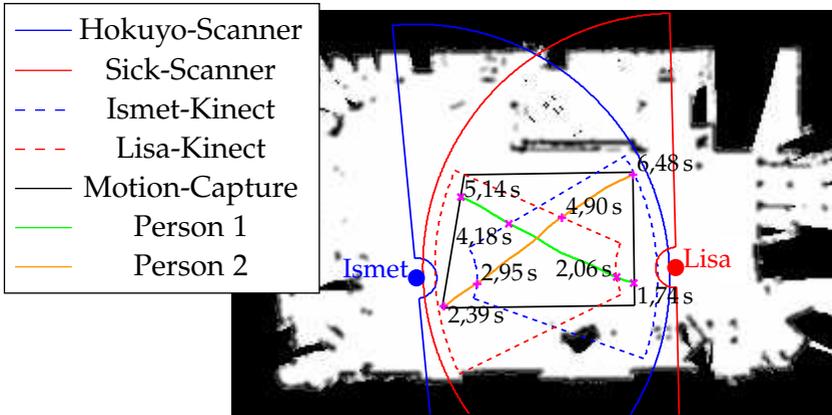
Ein Referenzsystem benötigt eine Genauigkeit, die dem zu überprüfenden Verfahren überlegen ist. Da funkbasierte Systeme diese Genauigkeit nicht in der gewünschten Genauigkeit liefern können [1], wurde auf ein optisches System zurückgegriffen, das mit Hilfe von Infrarotlicht reflektierenden Markern arbeitet und dabei Genauigkeiten im Millimeterbereich erreicht [129].



**Abbildung 6.4** Motion-Capture-Areal mit Sensorträgern.

## 6.2.2 Aufbau

In dieser Arbeit wurde ein *Vicon MX* Motion-Capture-System mit 12 Kameras verwendet. Damit können Marker, die zuvor auf den Objekten angebracht werden, präzise lokalisiert und verfolgt werden. Somit unterscheidet sich dieses Verfahren grundlegend von den in dieser Arbeit vorgestellten Verfahren. Für die Evaluierung wurde somit jede Testperson mit acht Markern versehen, wobei nur, entsprechend Abbildungen 6.3(a) und 6.3(b), die Marker A–D an der Hüfte ausgewertet wurden. Die Sensorträger wurden jeweils mit vier Markern E–H entsprechend Abbildung 6.3(c) versehen. Diese Anordnung ermöglicht, im Gegensatz zu [92], zu jedem Zeitpunkt eine Auswertung der Orientierung. Der gesamte Aufbau mit den beiden Sensorträgern und einem Erfassungsbereich von etwa  $3,5\text{ m} \times 5\text{ m}$  ist in Abbildung 6.4 zu sehen. Die Referenzdaten des Motion-Capture-Systems werden mit einer Rate von 200 Hz erfasst. Die Zeitstempel von Referenz- und Testsystem werden über ein gemeinsames Triggersignal synchronisiert.



**Abbildung 6.5** Laboraufbau mit eingezichnetem Motion-Capture-Bereich auf einer Occupancy-Grid-Karte. Ebenso ist die räumliche Anordnung der Sensorträger mit entsprechenden Sichtfeldern und ein beispielhafter Bewegungsverlauf zweier Testpersonen gezeigt.

## 6.3 Evaluierung der adaptiven Detektionswahrscheinlichkeit

### 6.3.1 Versuchsbeschreibung

Der in den Versuchen verwendete Versuchsaufbau nutzt die beiden in Abschnitt 6.1.1 vorgestellten Sensorträger mit jeweils einem Laser-scanner und einer RGB-D-Kamera, wobei alle Sensoren bei einer vorab gewählten Falschalarmrate unterschiedliche Detektionscharakteristiken entsprechend Kapitel 4, wie beispielsweise bezüglich des Sensor-Objekt-Abstands (Abbildung 4.8) oder des Sichtfelds, aufweisen. Die räumliche Anordnung im Motion-Capture-Labor ist in Abbildung 6.5 dargestellt und zeigt die Einschränkungen der Sichtfelder. Die Versuche stellen eine verbesserte und von einer größeren Datenbasis ausgehende Auswertung der vom Autor vorgestellten Evaluierungen in [161] dar.

Die Evaluation des in Kapitel 4 vorgestellten Ansatzes der adaptiven Detektionswahrscheinlichkeit und der Vergleich der in Kapitel 3 beschriebenen Filter wird sowohl auf der Grundlage teil-simulierter

Daten, aber auch auf Realdaten durchgeführt. In beiden Versuchsanordnungen werden bis zu zwei Personen in Szenen verfolgt, die während sich kreuzender Geraden sehr enge Begegnungen beinhalten und Verdeckungen von Sensoren verursachen, sowie verschiedene Sichtfelder der Sensoren durchschreiten. Die Beschränkung auf geradlinige Bewegung erfolgt hier, da insbesondere der Einfluss des Sensormodells und nicht die Bedeutung des Bewegungsmodells in diesem Schritt beurteilt werden soll. Daher wird ein nahezu konstantes Bewegungsmodell (siehe Abschnitt 2.1) verwendet, das mit der tatsächlichen Bewegung der Personen übereinstimmt. Beispielhaft sind die Bewegungspfade einer kreuzenden Szene in Abbildung 6.5 gezeigt. Weitere Evaluationen mit aufwändigeren Modellen wie *Social-Force*-Modellen [49] wären daher ebenso möglich, würden aber mit den Effekten der Sichtfeldmodellierung, die hier im Fokus der Evaluierung stehen, interferieren. Im Falle der SMC-basierten Filter werden pro Objekt etwa 1000 Partikel angesetzt, was sicherstellt, dass eine ausreichende Repräsentation der PHD vorliegt.

Für die realen und die teil-simulierten Ergebnisse wurde jeweils das adaptive Modell mit einer konstanten Detektionswahrscheinlichkeit im gesamten Überwachungsbereich und einer konstanten Detektionswahrscheinlichkeit bei Berücksichtigung der Sichtfelder verglichen. Die konstante Detektionswahrscheinlichkeit entspricht dem Durchschnitt der Detektionsrate ohne Berücksichtigung von Verdeckungen im Motion-Capture-Bereich. Beide Modelle wurden zwecks Verbesserung der Kardinalitätsschätzung jeweils mit dem szenario-spezifischen Wert von  $p_D^{Szen} = 0,9$  multipliziert, der die im Vergleich zu Abbildung 4.8 nicht ideale Umgebung berücksichtigt, was jedoch auch zu einer höheren Standardabweichung führt [161]. Das adaptive Modell berücksichtigt entgegen dem konstanten Modell Verdeckungen, Sichtfelder und Sensoreigenschaften entsprechend Kapitel 4. Die Überlebenswahrscheinlichkeit ist für alle Versuche  $p_S = 0,99$ . Die im Bewegungsmodell angenommenen Standardabweichungen für Geschwindigkeit und Orientierung betragen experimentell  $0,25 \text{ m/s}$  beziehungsweise  $11^\circ$ .

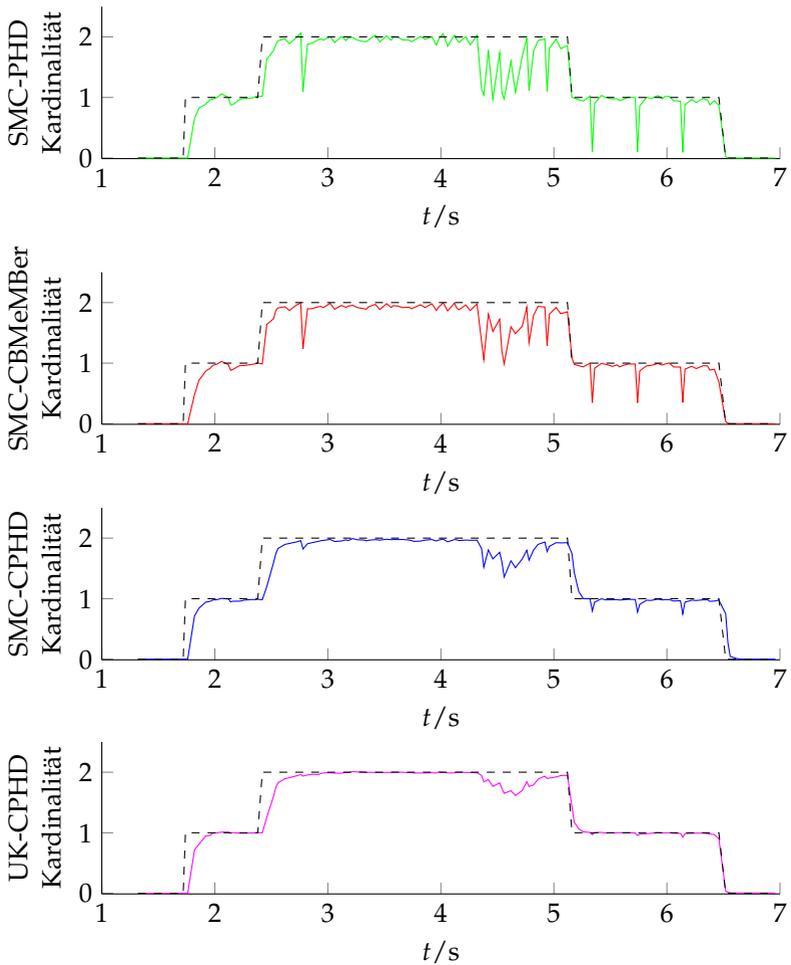
Als Gütekriterium der Filterergebnisse wird die OSPA-Metrik aus Abschnitt 3.7 verwendet. Der *Cut-off*-Parameter  $c$  wird hierbei zu  $0,3 \text{ m}$  gewählt und die Ordnung  $p$  zu 1. Die OSPA-Metrik wird aufgrund der Referenzdaten nur innerhalb des Motion-Capture-Bereichs berech-

net, so dass auch nur Messungen in diesem Bereich verwendet werden. Um durch die Wahl der Ordnung  $p$  eine gute Interpretation der OSPA-Metrik zu ermöglichen, wurde die Auswertung auf die  $x,y$ -Positionen beschränkt, sodass der OSPA-Fehler eine Aussage über den durchschnittlichen Schätzfehler der Position kombiniert mit Fehlern der Kardinalität (Anzahl der Personen) ausdrückt, ohne Geschwindigkeitsdifferenzen einzurechnen.

### 6.3.2 Simulation

Der erste Teil der Auswertung basiert auf teil-simulierten Daten. Hierbei werden die echten, durch das Motion-Capture-System aufgezeichneten Trajektorien der Testpersonen verwendet und anhand dieser Bewegungsverläufe die Messungen mit Falschalarmen sowie Fehldetektionen simuliert. Die Simulationen der Messungen, die die Sichtfeldeigenschaften wie in Kapitel 4 berücksichtigten, wurden entsprechend einer konstanten Annahme des Sichtfelds überprüft und schrittweise mit Berücksichtigung der erweiterten Modellierung ausgewertet. Für jede Simulation wurden Mittelungen über 100 Monte-Carlo (MC)-Simulationen durchgeführt. Der Vorteil der Simulation ist, dass Fehldetektionen und Falschalarme nicht wiederholt an gleichen Stellen auftreten und eine Bewertung der Verfahren weniger von einem Einzelfall abhängt und mehr auf den Modellannahmen basiert.

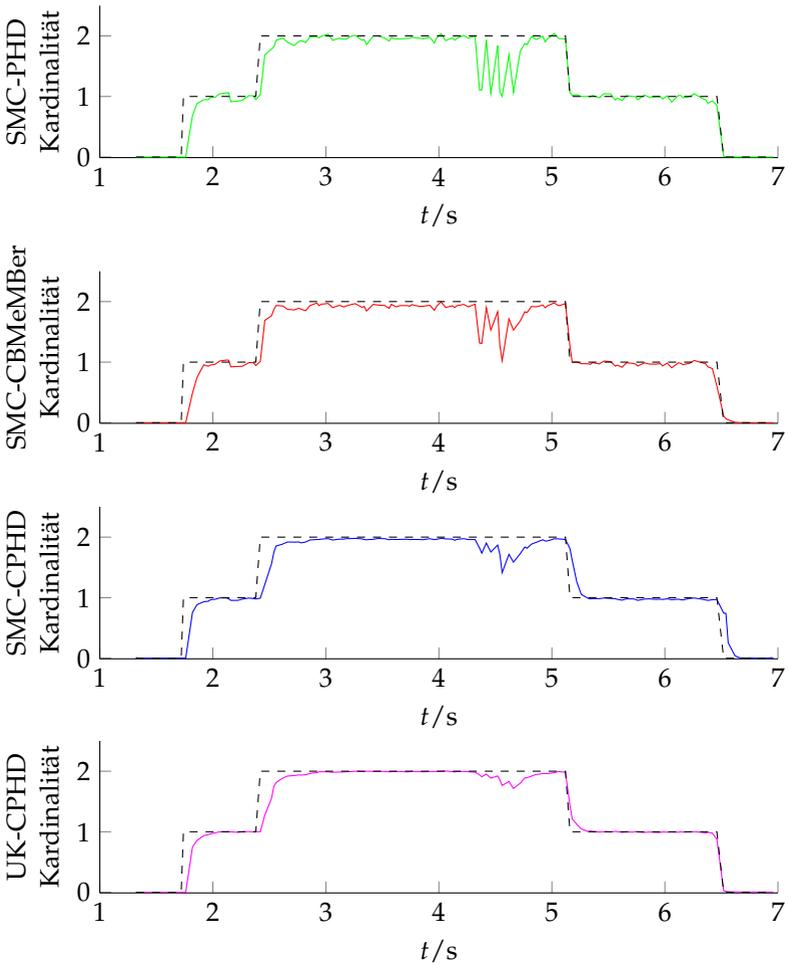
Im Falle der Annahme einer konstanten Detektionswahrscheinlichkeit in den Filtern, deren Kardinalitätsschätzung in Abbildung 6.6 gezeigt ist, ist ersichtlich, dass erwartungsgemäß alle Filter stark davon beeinflusst werden, dass das korrekte Sichtfeld und Verdeckungen nicht entsprechend berücksichtigt werden. Dies zeigt sich durch sich wiederholende Peak-förmige Abfälle der Kardinalitäten an Stellen, an denen das Filter ein Objekt im Sichtfeld vermutet, es sich aber außerhalb des Sichtfelds befindet. Ebenso zeigt sich in dem Beispiel etwa um den Zeitpunkt 4,5 s, während eine Person die zweite Person verdeckt, dass alle Filter Einbrüche in der Kardinalitätsschätzung aufweisen und die Kardinalität fehlerhaft geschätzt wird, obwohl nur die Sicht eines Sensorträgers verdeckt ist. Zusätzlich zeigt sich die unterschiedliche Robustheit der Filter gegenüber Fehldetektionen. Während das PHD-Filter auf jede Fehlmessung direkt mit einem Einbruch der Kardinalität reagiert, prägen sich



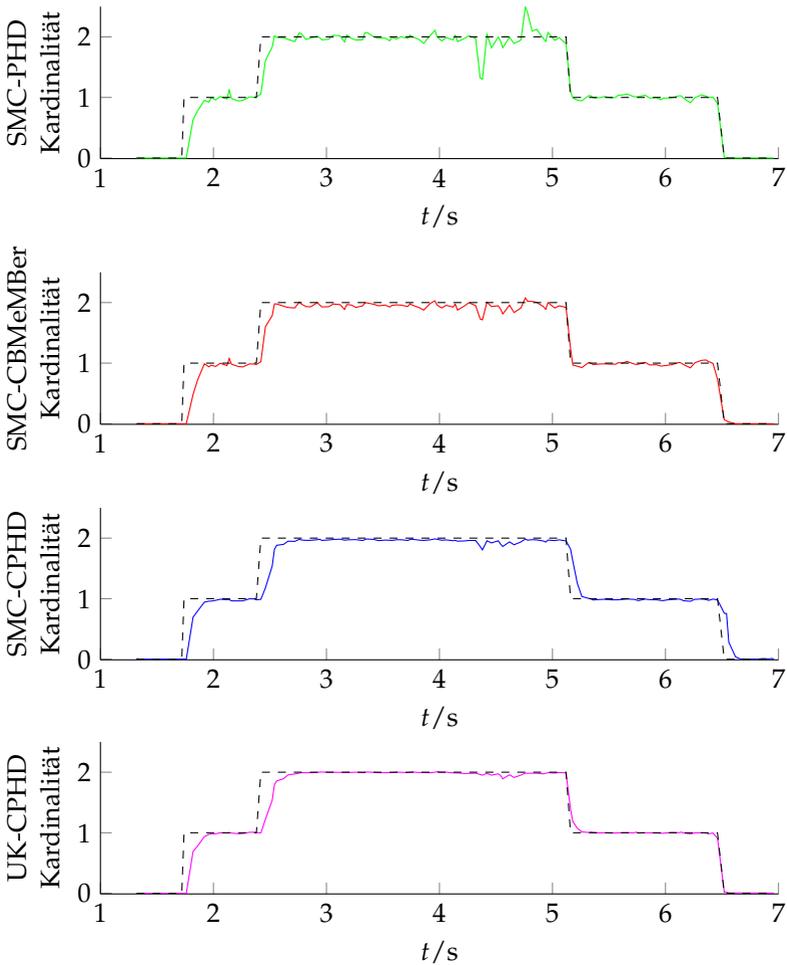
**Abbildung 6.6** Monte-Carlo-Durchschnitt für eine konstante Detektionswahrscheinlichkeit bei wahrer Kardinalität (---) und Schätzung (—) für 100 MC-Durchläufe der Simulation.

diese Einbrüche etwas schwächer beim CBMeMber-Filter und deutlich schwächer bei den kardinalisierten PHD-Filtern aus. Die aus der zusätzlichen Schätzung der Verteilung der Kardinalität stabilisierte Schätzung führt gleichzeitig jedoch auch zu einem trägeren Verhalten bei dem Hinzukommen neuer Objekte. Gerade bezüglich ausbleibender Messungen zeigt das UK-CPHD-Filter die am stärksten ausgeprägte Trägheit, da hier im Vergleich zu den SMC-basierten Filtern keine große Vielfalt an Partikeln besteht. Das Ergebnis bei Berücksichtigung der Sichtfelder (vergleiche Abbildung 6.5) ist in Abbildung 6.7 zu sehen. Hier kann man deutlich die fehlenden Einbrüche erkennen, wobei erst die zusätzliche Modellierung von Verdeckungen (Abbildung 6.8) dazu führt, dass die geschätzte Kardinalität der wahren Kardinalität (in Schwarz) weitestgehend entspricht. Die Berücksichtigung des Detektionsprozesses mit einer zustandsabhängigen Detektionswahrscheinlichkeit kann somit Wechsel zwischen Sichtfeldern und auch Verdeckungen ausgleichen, die bei einer konstanten Modellierung zu Fehlern führen.

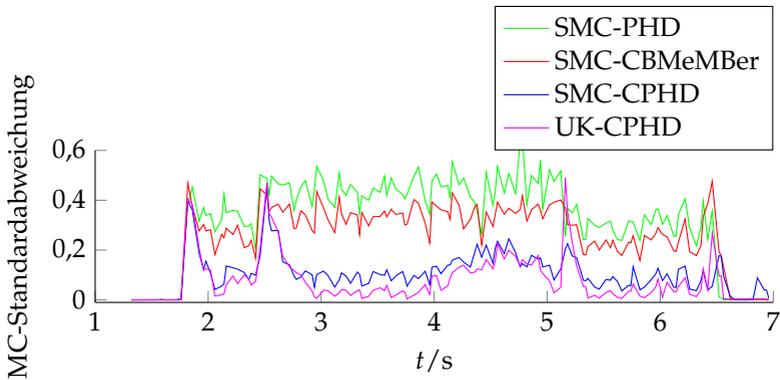
Die Auswertung zeigt gleichzeitig die unterschiedlichen Eigenschaften der hier verglichenen Filter und ihre Eignung für das gewählte Szenario. Wie Abbildung 6.9 zeigt, besitzt das PHD-Filter die höchste Standardabweichung errechnet aus 100 MC-Durchläufen bei der Berechnung der Kardinalität, was den Ergebnissen anderer Untersuchungen entspricht [149]. Die CPHD-Algorithmen besitzen dabei die geringsten Standardabweichungen. Abbildung 6.10 zeigt den zugehörigen OSPA-Fehler. Die mittleren OSPA-Fehler als kombinierte Auswertung von Kardinalitäts- und Lokalisierungsfehlern sind in Tabelle 6.1 aufgeführt. Hierbei wurden vier Laufszene von jeweils zwei Personen ähnlich dem soeben vorgestellten Beispiel untersucht und über deren Fehler gemittelt. Die besten Ergebnisse erreichen auch hier die CPHD-Approximationen mit einem durchschnittlichen Fehler unter 8 cm. Beim Vergleich der Filterimplementierungen fällt ebenso auf, dass insbesondere die kardinalisierten Filter weniger von einzelnen ausfallenden Messungen, die beispielsweise beim Verlassen einer Person eines Sichtfelds entstehen, beeinflusst werden (vergleiche Abbildung 6.6), während die Berücksichtigung von Verdeckungen, die meist mehrere Sensoren betrifft, für alle Filter deutliche Verbesserungen mit sich bringt.



**Abbildung 6.7** Monte-Carlo-Durchschnitt für eine konstante Detektionswahrscheinlichkeit mit Berücksichtigung von  $p_D^{\text{FoV}}$  bei wahrer Kardinalität (---) und Schätzung (—) für 100 MC-Durchläufe der Simulation.



**Abbildung 6.8** Monte-Carlo-Durchschnitt für eine adaptive Detektionswahrscheinlichkeit bei wahrer Kardinalität (---) und Schätzung (—) für 100 MC-Durchläufe der Simulation.



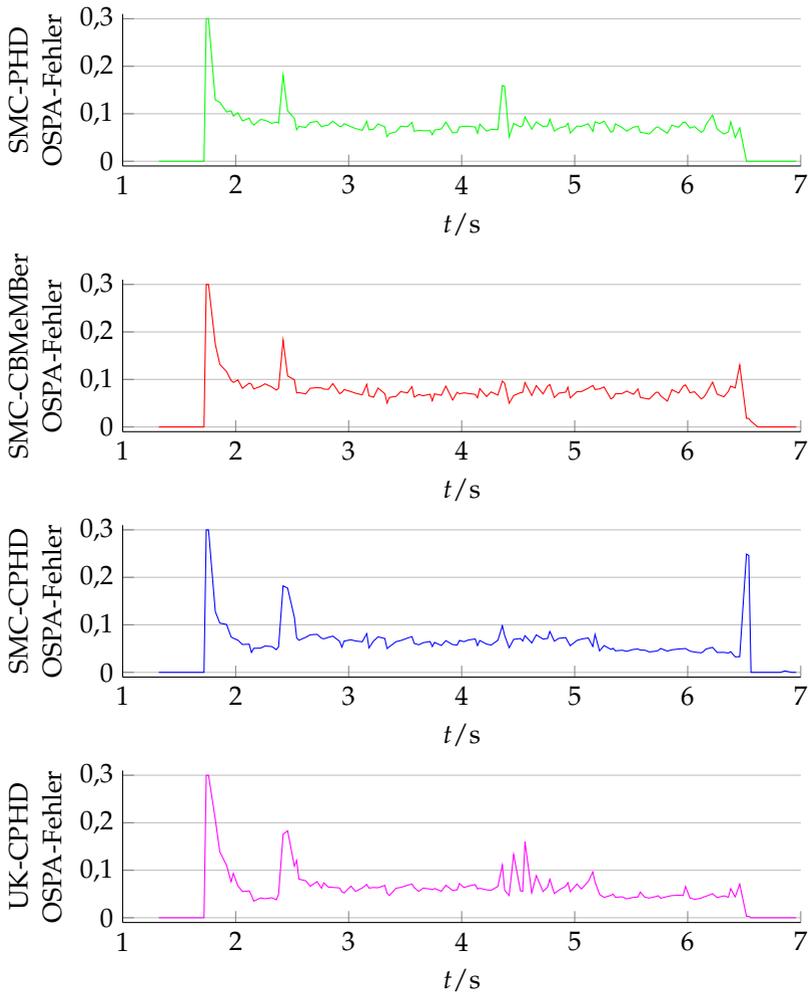
**Abbildung 6.9** Monte-Carlo-Standardabweichung der Kardinalitätsschätzung für eine adaptive Detektionswahrscheinlichkeit für 100 MC-Durchläufe der Simulation.

**Tabelle 6.1** Mittlere OSPA-Fehler in der Simulation.

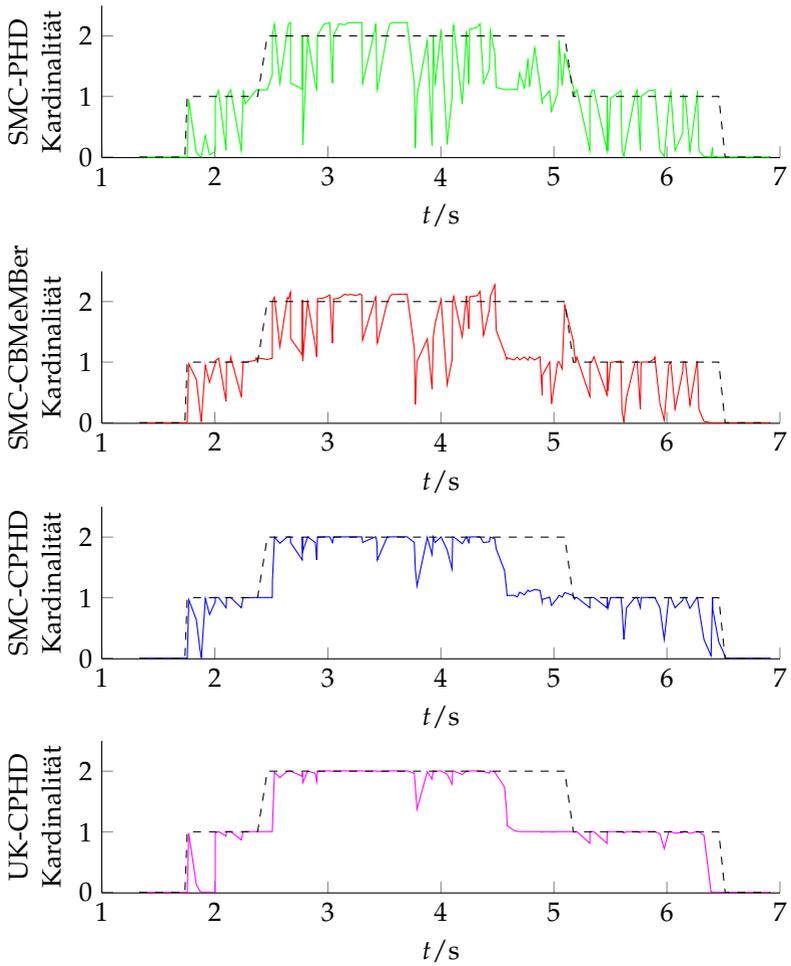
	SMC- PHD	SMC- CBMeMBer	SMC- CPHD	UK-CPHD
konst. $p_D$	10,42 cm	9,91 cm	8,53 cm	8,82 cm
konst. $p_D$ mit $p_D^{\text{FoV}}$	9,63 cm	9,17 cm	8,36 cm	8,41 cm
adaptive $p_D$	8,76 cm	8,69 cm	7,31 cm	7,80 cm

### 6.3.3 Realdaten

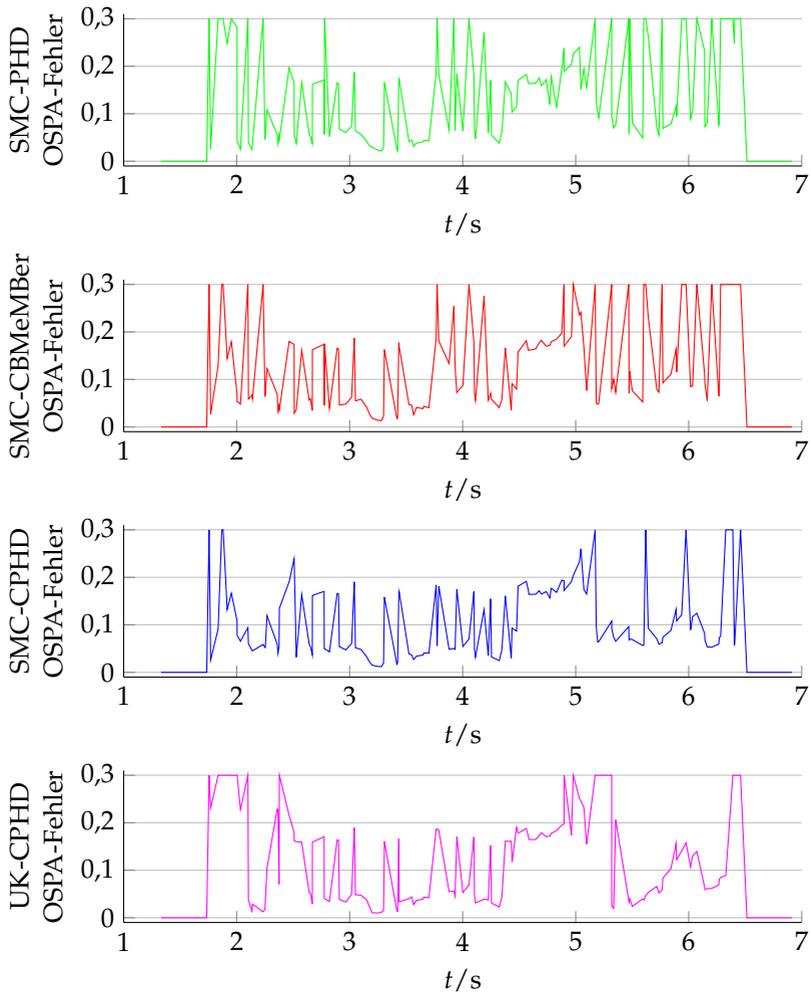
Um die Resultate der Simulationen zu verifizieren, wurden in einem zweiten Schritt die echten Sensordaten für die gleichen Trajektorien ausgewertet und mit den Referenzdaten verglichen. Da die Aufzeichnung der Sensordaten nicht wiederholt werden kann, weil eine Wiederholung einer exakt gleichen Trajektorie durch die Testperson nicht möglich ist, wird bei der Auswertung immer auf dieselben Sensordaten zurückgegriffen und nicht wie bei der Simulation auf für jeden Durchlauf neu simulierte Fehldetektionen oder Falschalarme. Auch hier werden vier verschiedene Kreuzungsszenen ausgewertet, wobei die erste zusätzlich anhand von Graphen analysiert wird.



**Abbildung 6.10** Monte-Carlo-OSPA-Fehler für eine adaptive Detektionswahrscheinlichkeit für 100 MC-Durchläufe der Simulation in cm.



**Abbildung 6.11** Monte-Carlo-Durchschnitt für eine konstante Detektionswahrscheinlichkeit bei wahrer Kardinalität (---) und Schätzung (—) für 100 MC-Durchläufe der Realdaten.



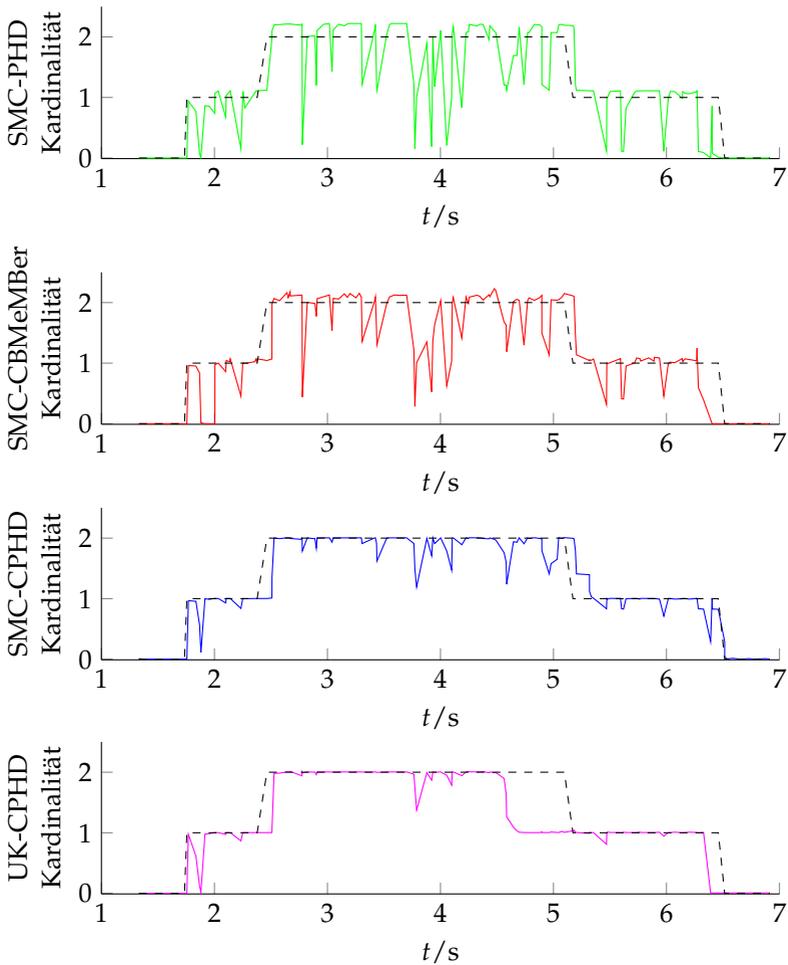
**Abbildung 6.12** Monte-Carlo-OSPA-Fehler für eine konstante Detektionswahrscheinlichkeit für 100 MC-Durchläufe der Realdaten in cm.

Abbildung 6.11 zeigt die Ergebnisse für das Modell der konstanten Detektionswahrscheinlichkeit. Wie auch für die Simulation lassen sich starke Einbrüche durch fehlerhafte Annahmen über die Sichtbarkeit von Objekten durch Vernachlässigung der Sichtfelder sehen und im Bereich von 4,5 s führt die Verdeckung zu fehlerhaften Schätzungen. Ebenso zeigen auch hier die Filter ohne verbesserte Kardinalitätsschätzung deutlich größere Fehler, die auch in der Auswertung der OSPA-Fehler in Abbildung 6.12 zu sehen sind.

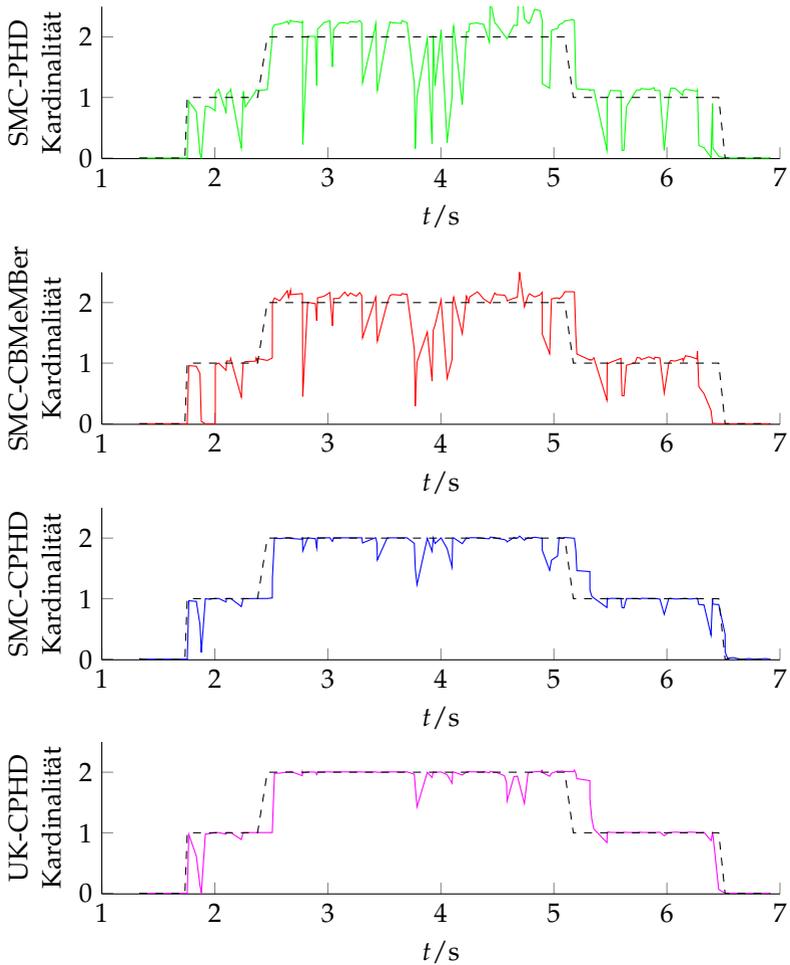
Werden die Sichtfelder entsprechend berücksichtigt (Abbildung 6.13) verschwinden die durch unterschiedliche Sichtbereiche entstehenden Einbrüche. Allerdings bleibt gerade für PHD- und CBMeMber-Filter die Kardinalitätsschätzung auch außerhalb des Bereichs der Verdeckung fehlerhaft. Die Ursache hierfür sind im Detektionsprozess auftretende Fehldetektionen. Während die Mittelung bei der Simulation von zufällig auftretenden Fehlern diese Einbrüche unterdrückt, bleiben sie bei den realen Messungen erhalten und lassen sich in der Filterung nur durch den Einsatz zusätzlicher Kardinalitätsstatistiken wie bei den CPHD-Filtern unterdrücken.

Auch die mit der Modellierung der Verdeckung komplettierte adaptive Detektionswahrscheinlichkeit (Abbildung 6.14) führt bei allen Verfahren zu einer Verbesserung, wobei sich, wie auch bei der Simulation, die CPHD-Filter insgesamt als die am besten geeigneten Methoden herausstellen. Die Verdeckung um die Zeit 4,5 s ist noch einmal vergrößert in Abbildung 6.16 dargestellt, wobei zu sehen ist, dass speziell das SMC-CPHD-Filter die beste Schätzung erreicht. Theoretisch lässt sich das insgesamt bessere Abschneiden der CPHD-Filter wie auch für die Simulation durch die Herleitungen der Filter [77, 148, 149] und die Erweiterungen durch die volle Kardinalitätsverteilung begründen, während das PHD-Filter nur den Mittelwert der Kardinalität und das CBMeMber eine parametrisierte Näherung der Kardinalität propagiert.

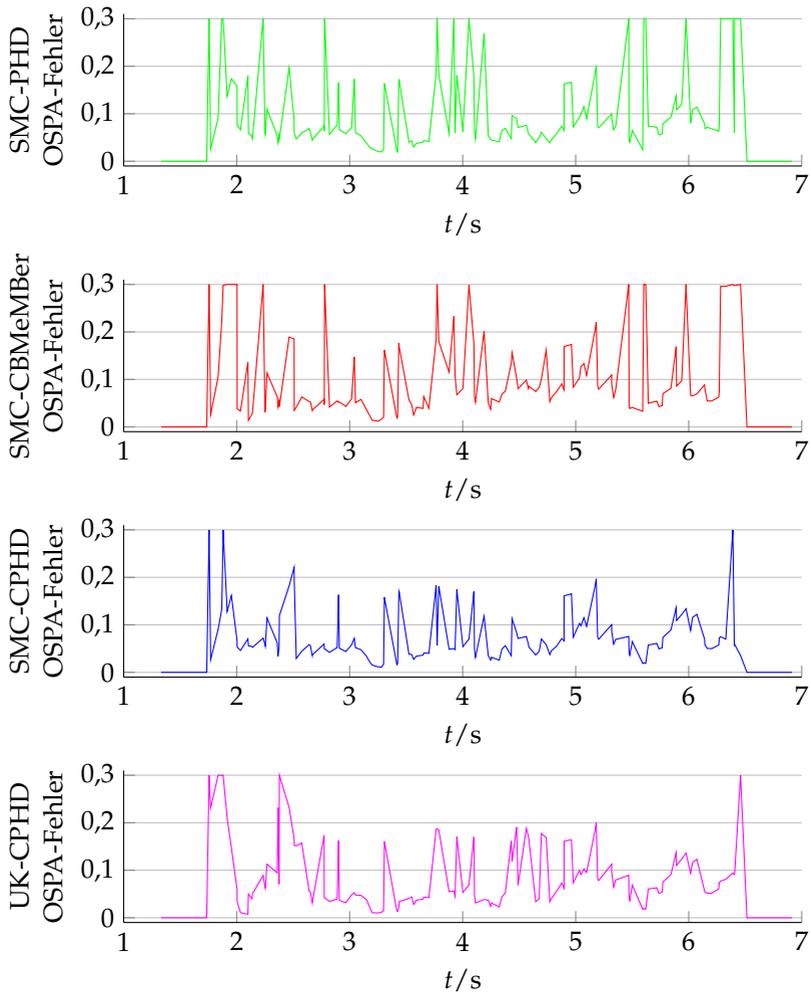
Obwohl sich entsprechend der Simulation alle Verfahren im Mittel dem wahren Verlauf der Kardinalität annähern sollten, leiden die Ergebnisse aller Schätzer, gerade bei der Untersuchung der Daten nur eines Durchlaufs, unter der Annahme aller Verfahren, dass Messungen unabhängig voneinander sind. Diese Annahme wird jedoch beispielsweise um die Zeitschritte bei 3,8 s verletzt. Hier sind die Objekte so nah



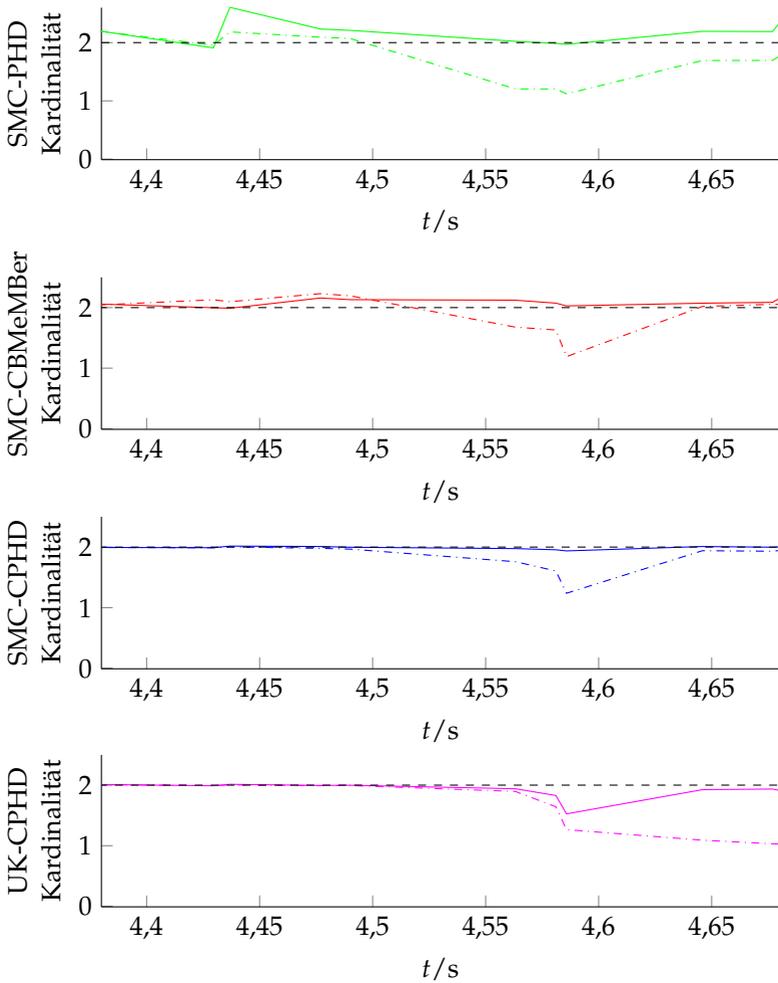
**Abbildung 6.13** Monte-Carlo-Durchschnitt für eine konstante Detektionswahrscheinlichkeit mit Berücksichtigung der Sichtfelder bei wahrer Kardinalität (---) und Schätzung (—) für 100 MC-Durchläufe der Realdaten.



**Abbildung 6.14** Monte-Carlo-Durchschnitt für eine adaptive Detektionswahrscheinlichkeit bei wahrer Kardinalität (---) und Schätzung (—) für 100 MC-Durchläufe der Realdaten.



**Abbildung 6.15** Monte-Carlo-OSPA-Fehler für eine adaptive Detektionswahrscheinlichkeit für 100 MC-Durchläufe der Realdaten in cm.



**Abbildung 6.16** Monte-Carlo-Vergleich der Kardinalität ohne (---) und mit (—) Berücksichtigung von  $p_D^{\text{Occ}}$  bei wahrem Verlauf (---) bei 100 MC-Durchläufen der Realdaten.

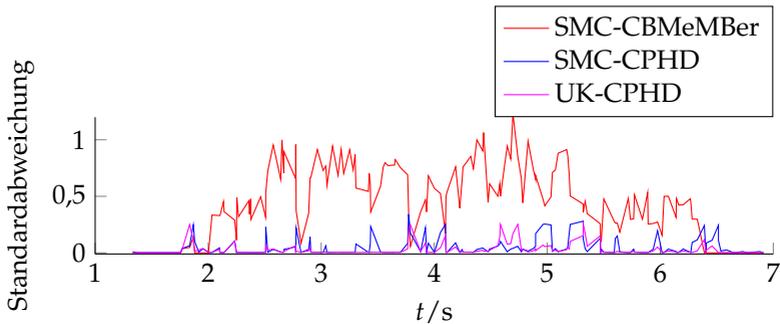
beieinander, dass die Detektionsverfahren, teilweise durch Fehler in der Segmentierung der Laserdaten, keine klare Trennung mehr vornehmen können und ein Objekt die Detektion des anderen beeinflusst, was sich in Fehlern der Kardinalitätsschätzung zeigt, obwohl alle Objekte sichtbar sind. Diese Fehler zeigen sich auch deutlich in der OSPA-Metrik (Abbildung 6.15), die sich ansonsten zu Fehlern im Bereich von 5 cm errechnet, wenn keine Fehler der Kardinalität vorliegen. Die mittleren OSPA-Fehler sind in Tabelle 6.2 gezeigt. Obwohl auch für die adaptive Detektionswahrscheinlichkeit noch Einbrüche in der Kardinalität zu sehen sind, soll an dieser Stelle darauf hingewiesen werden, dass eine Verfolgung eines Objekts dennoch möglich ist, da sich beispielsweise der OSPA-Fehler nur anhand der Objekte berechnet, die ein Gewicht über 0,5 besitzen. Für ein Trackmanagement wie beispielsweise in Abschnitt 3.6 verschwinden diese Objekte insbesondere bei den CPHD-Filtern jedoch nicht, so dass sie im nächsten Schritt bei erneuter Detektion mit gleicher ID berücksichtigt werden können.

Für die Standardabweichung der Kardinalität einer einzelnen Schätzung (Abbildung 6.17) wird die aus den Verfahren resultierende Standardabweichung aus einem Durchlauf verwendet. Da allerdings das PHD-Filter neben der geschätzten Kardinalität keine weitere Information über die Verteilung dieser propagiert, wird diese nicht in Abbildung 6.17 dargestellt. Deutlich ist zu sehen, dass die CPHD-Filter die kleinsten Unsicherheiten liefern. Zusätzlich ist ein wichtiger Aspekt bei der Analyse der Unsicherheit bzw. der Standardabweichung die Betrachtung der Schwankungen zwischen mehreren Durchläufen der gleichen Messung. Abbildung 6.18 zeigt die unterschiedlichen Standardabweichungen der Verfahren aus 100 Durchläufen. Während man sieht, dass SMC-PHD- und SMC-CBMeMber-Verfahren zu bestimmten Zeitpunkten deutliche Abweichungen aufweisen können, sind diese für die CPHD-Verfahren deutlich kleiner. Für das UK-CPHD-Verfahren sind sie exakt null, da hierbei kein zufälliges Verrauschen der Partikel vorgenommen wird, wie dies für die SMC-Filter geschieht, um die nötige Vielfalt der Partikel im Resampling-Schritt zu erhalten. Somit ist das Ergebnis beim UK-CPHD exakt wiederholbar, was für eine Anwendung Vorteile mit sich bringt.

Um explizit auf die Bedeutung der adaptiven Detektionswahrscheinlichkeit  $p_D^{\text{Dist}}$  einzugehen, wurde ein weiteres Szenario gewählt. Da das

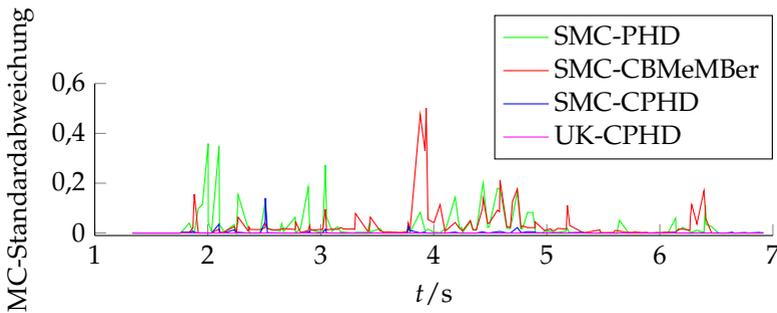
**Tabelle 6.2** Mittlere OSPA-Fehler in Realdaten.

	SMC- PHD	SMC- CBMeMber	SMC- CPHD	UK-CPHD
konst. $p_D$	15,08 cm	14,19 cm	12,31 cm	12,97 cm
konst. $p_D$ mit $p_D^{\text{FoV}}$	12,57 cm	12,50 cm	10,74 cm	11,49 cm
adaptive $p_D$	11,82 cm	12,12 cm	10,09 cm	10,56 cm

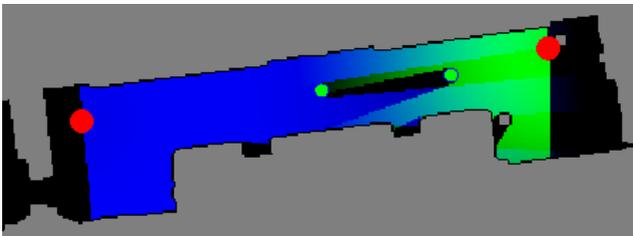
**Abbildung 6.17** Aus dem Verfahren errechnete Standardabweichung der Kardinalität für eine adaptive Detektionswahrscheinlichkeit der Realdaten.

Motion-Capture-System nur einen relativ kleinen Raum abdeckt, wird hier anhand der bereits in [164] veröffentlichten Ergebnisse ein Szenario ohne Auswertung durch das Motion-Capture-System auf der Basis von Laserdaten gewählt. Daher werden zum Vergleich Trajektorien aus den unterschiedlichen Schätzungen verwendet. Die beispielhaften Ergebnisse zeigen die vom PHD-Filter erzeugten Trajektorien, da dieses die höchste Sensitivität gegenüber einer Diskrepanz zwischen wahrer und angenommener Detektionswahrscheinlichkeit besitzt und die Grundlage der CPHD-Schätzung bildet. Der Aufbau entspricht Abbildung 6.19 mit den Sensorpositionen in Rot und den zu verfolgenden Fußgängern in Grün. Ebenso sind in Grün und Blau die jeweils von den Sensorpositionen ausgehend mit wachsendem Abstand zwischen Sensor und Objekt abfallenden Detektionsraten gezeigt, sowie die durch Verdeckungen verursachte Detektionsrate. Abbildung 6.20 zeigt in der obigen Abbildung

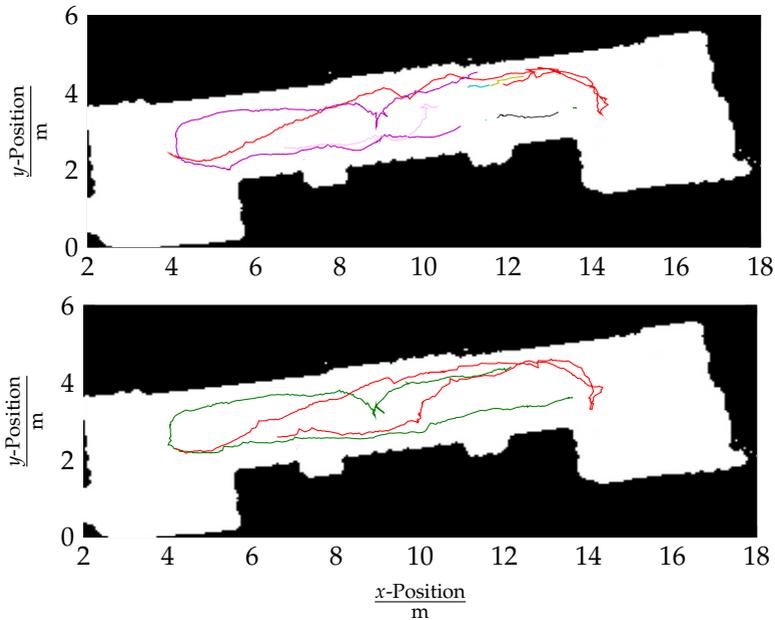
die entstehenden Trajektorien ohne Berücksichtigung der adaptiven Detektionswahrscheinlichkeit und in der unteren Abbildung die korrekten Trajektorien bei deren Berücksichtigung mit einer Farbe pro Objekt. Gerade im Bereich von etwa 12 m ist eine besondere Anfälligkeit für verlorene und reinitialisierte Objekte (unterbrochene Trajektorie/Farbwechsel) zu erkennen, der sich mit dem am schlechtesten von den Sensoren erfassten Bereich deckt.



**Abbildung 6.18** Monte-Carlo-Standardabweichung der Kardinalität für eine adaptive Detektionswahrscheinlichkeit für 100 MC-Durchläufe der Realdaten.



**Abbildung 6.19** Zustandsabhängige Detektionswahrscheinlichkeit  $p_D$ . Die Positionen der Lasersensoren sind in Rot auf der Umgebungskarte gezeigt. Objekte sind beispielhaft in Grün gezeigt. Die Detektionswahrscheinlichkeit der Sichtfelder der Sensoren ist in Grün und Blau aufgezeichnet.



**Abbildung 6.20** Geschätzte farblich kodierte Trajektorien von zwei Personen mittels eines SMC-PHD Filters nach dem Aufbau entsprechend Abbildung 6.19. Die obere Abbildung zeigt das Ergebnis einer konstanten Detektionswahrscheinlichkeit und die untere das Ergebnis der vorgestellten adaptiven Modellierung.

### 6.3.4 Wahl der Detektionswahrscheinlichkeit pro Gauß-Komponente

In diesem Abschnitt soll gesondert auf den Effekt der Verwendung der adaptiven Detektionswahrscheinlichkeit unter Berücksichtigung der in Abschnitt 4.5.1 vorgestellten Berechnung der Detektionswahrscheinlichkeit pro Gauß-Komponente eingegangen werden. Für das CPHD-Filter wurde dabei verglichen, wie sich die Anwendung der adaptiven Detektionswahrscheinlichkeit auf die Schätzungen auswirkt. Tabellen 6.3 und 6.4 zeigen für die simulierten und die realen Szenarien den Gewinn durch die Nutzung der Eigenschaften der Sigmapunkte bei der Berechnung der Detektionswahrscheinlichkeit. Während für die OSPA-Fehler der linken Spalte der Tabelle die Detektionswahrscheinlichkeit nur aufgrund des

Mittelwertes berechnet wurde, zeigt die mittlere Spalte Fehler für die Detektionswahrscheinlichkeit einer Gauß-Komponente, berechnet über die gewichteten Sigmapunkte analog zu Tabellen 6.1 und 6.2. Entsprechend wurden als Referenz auch nochmals die Ergebnisse bei Nutzung des SMC-CPHD-Filters in der rechten Spalte mit aufgenommen. Für die Auswertung mit konstanter Detektionswahrscheinlichkeit sind die Ergebnisse wie zu erwarten ähnlich und unterscheiden sich nur sehr geringfügig durch die zufälligen Fehldetektionen und Falschalarme der Simulation. Die Ergebnisse bestätigen die Verbesserungen gegenüber der von Ulmke et al. [140] und Hendeby und Karlsson [50] diskutierten, durch den Mittelwert bestimmten, Detektionswahrscheinlichkeit. Ebenso zeigt sich auch, dass die Unscented-Transformation eine Verwendung für sich schnell ändernde Detektionswahrscheinlichkeiten ermöglicht und Fehler erreicht, die insbesondere bei den realen Daten deutlich näher an den besseren Ergebnissen der SMC-basierten Methoden liegen. Ein Grund hierfür ist mitunter die größere Robustheit der UK-Filter, die sich durch die kleine Standardabweichung in Abbildung 6.17 gezeigt hat. Zudem müssen für die vorgeschlagene Berechnung der Detektionswahrscheinlichkeit keine zusätzlichen Sigmapunkte berechnet werden, da diese bereits für die entsprechenden Unscented-Kalman-Schritte benötigt werden. Dadurch lässt sich der Rechenaufwand des UK-CPHD-Filters gering halten; dieser ist für die vorgestellten Szenarien etwa halb so hoch im Vergleich zum SMC-CPHD-Filter.

**Tabelle 6.3** OSPA-Fehler für Wahl der Sigmapunkte bei 100 MC-Durchläufen für die Simulation.

	UK-CPHD mit Mittelwert	Sigmapunkt UK-CPHD	SMC-CPHD
konst. $p_D$	8,81 cm	8,82 cm	8,53 cm
konst. $p_D$ mit $p_D^{\text{FoV}}$	8,51 cm	8,41 cm	8,36 cm
adaptive $p_D$	7,99 cm	7,80 cm	7,31 cm

**Tabelle 6.4** OSPA-Fehler für Wahl der Sigmapunkte für Realdaten.

	UK-CPHD mit Mittelwert	Sigmapunkt UK-CPHD	SMC-CPHD
konst. $p_D$	12,97 cm	12,97 cm	12,31 cm
konst. $p_D$ mit $p_D^{\text{FoV}}$	12,03 cm	11,49 cm	10,74 cm
adaptive $p_D$	11,52 cm	10,56 cm	10,09 cm

## 6.4 Evaluierung der richtungssensitiven Verfolgung

In diesem Abschnitt werden zum einen die Ergebnisse des Verfahrens zur Schätzung von Orientierungen aus Einzelbildern aus Kapitel 5 vorgestellt und bewertet. Zum anderen folgt eine Bewertung des Nutzens der richtungssensitiven Verfolgung in einem Multi-Objekt-Szenario. Dabei wird die Schätzung der Orientierung aus Einzelbildern im Folgenden als Orientierungsmessung bezeichnet, um sie von der Schätzung der Orientierung durch einen Zustandsschätzer über den zeitlichen Verlauf hinweg und fusioniert mit Positionsinformationen zu unterscheiden. Die Evaluierung der Orientierungsmessung erfolgt der Übersichtlichkeit halber zuerst im Ein-Objekt-Fall und wurde bereits in ähnlicher Form vom Autor in [165] veröffentlicht. Anschließend folgt die Evaluation im Multi-Objekt-Fall zusammen mit der Fusion durch die vorgestellten Filter in Abschnitt 6.4.2.

Für die Evaluation ist für ein Verfahren, das eine kontinuierliche Orientierung als Ergebnis der Auswertung eines Bildes liefert, eine Datenbasis nötig, die eine Vielzahl von Bildern mit kontinuierlicher Orientierungsangabe liefert. Da viele Datensätze nur grob quantisierte Orientierungsrichtungen verwenden [25, 139], wurde hier auf die KITTI-Datenbank [35] zurückgegriffen, da sie aus einer Vielzahl an unterschiedlichen Personen in unterschiedlichen Orientierungen in realistischer Umgebung mit manueller Annotation der kontinuierlichen Orientierung besteht. Dies ermöglicht somit auch über Richtigklassifikationsraten hinaus eine Auswertung von Orientierungsfehlern. Zusätzlich wurde ebenfalls auf einen eigens erstellten Referenzdatensatz mit Farb- und Tiefendaten sowie

Annotationen der Orientierung mittels eines Motion-Capture-Systems zurückgegriffen, um eine gemeinsame Datenbasis für den Vergleich von auf Tiefen- und RGB-Daten basierenden Merkmalen mit exakten Annotationen zu schaffen.

Aus der KITTI-Datenbank wurden alle unverdeckten oder nur zu kleinen Teilen verdeckten und sich zu mindestens 90 % im Sensorsichtfeld befindenden Personen ab einer Mindestgröße von 90 Pixeln in der Höhe ausgeschnitten und auf die Fenstergröße des verwendeten Detektionsfensters von  $128 \times 64$  Pixeln skaliert. Die Auftrennung der resultierenden 1286 Bilder von Personen in Trainings- und Testdaten erfolgte zufällig je zur Hälfte, wobei alle Bilder durch eine zweifache Kreuzvalidierung einmal als Test- und ein weiteres Mal als Trainingsdaten verwendet wurden. In allen Untersuchungen wurde wie in [25] der Parameter  $\sigma_{\text{Exp}} = 45^\circ$  gewählt. Entsprechend wurde auch der SVM-Parameter  $\sigma_{\text{SVM}} = 45^\circ$  gewählt. Für das Vergleichsverfahren nach [25] hat sich gezeigt, dass die ausgewerteten Ergebnisse mit etwa den halben Standardabweichungen zu besseren Werten führen, die hier dementsprechend angepasst wurden.

### 6.4.1 Orientierungsmessung

Zur Beurteilung der Orientierungsmessung wurden die aus den Testdaten geschätzten Orientierungen durch das vorgestellte Verfahren mit vier (Tabelle 6.5) und mit acht Trainingsklassen (Tabelle 6.6) sowie durch das Referenzverfahren nach [25] mit vier Trainingsklassen (Tabelle 6.7) wieder in vier Hauptrichtungen klassifiziert und mit den wahren Orientierungen der Bilder, ebenfalls in vier Klassen eingeteilt, verglichen. Gerade die Zuordnung in rechte und linke Orientierung erreicht hier in allen Fällen sehr gute Ergebnisse, wobei es bei Zuordnungen zur vorderen Richtung zu deutlich schlechteren Klassifikationsergebnissen kommt, welche sich durch die im Vergleich zur oft charakteristischen Beinstellung von Links/Rechts wenig ausgeprägte Charakteristik von Orientierungen wie Vorne und Hinten begründen lässt. Dennoch ist eine Verbesserung durch die adaptiven negativen und positiven Gewichte im vorgestellten Verfahren für vier und acht Klassen zu sehen, die sich auch an der verbesserten durchschnittlichen Richtigklassifikationsrate der vorgestellten Verfahren von etwa 79 % im Gegensatz zu 65 % für das Vergleichsverfahren zeigt, wobei im ursprünglichen Ansatz des Ver-

gleichsverfahrens zusätzlich Konturinformationen verwendet wurden, die hier nicht berücksichtigt wurden, da sie in den untersuchten Daten zu keiner Verbesserung führten.

**Tabelle 6.5** Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 79,39% für das vorgestellte Verfahren mit vier Trainingsklassen.

		Schätzung			
		Hinten	Vorne	Links	Rechts
Wahr	Hinten	83 %	7 %	6 %	4 %
	Vorne	9 %	70 %	9 %	12 %
	Links	9 %	6 %	83 %	2 %
	Rechts	4 %	10 %	3 %	83 %

**Tabelle 6.6** Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 78,54% für das vorgestellte Verfahren mit acht Trainingsklassen.

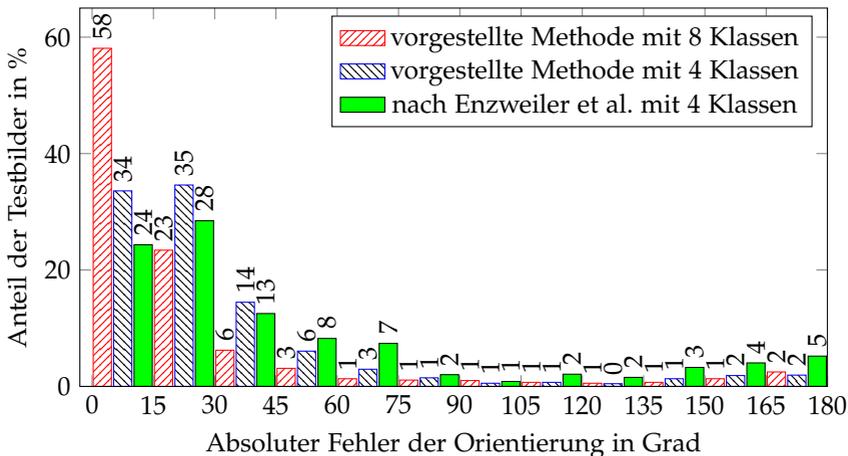
		Schätzung			
		Hinten	Vorne	Links	Rechts
Wahr	Hinten	83 %	6 %	7 %	4 %
	Vorne	7 %	69 %	9 %	15 %
	Links	8 %	8 %	82 %	2 %
	Rechts	5 %	11 %	1 %	83 %

In Abbildung 6.21 sind die Beträge der absoluten Fehler der kontinuierlichen Orientierungsschätzung als Histogramm angegeben. Es ist gut zu sehen, dass eine Fehlschätzung um  $90^\circ$  seltener vorkommt als eine Verwechslung mit der gegensätzlichen Seite. Der mittlere absolute Winkelfehler lässt sich zu  $24,4^\circ$  für das vorgestellte Verfahren mit acht Klassen, zu  $31,8^\circ$  für vier Klassen und zu  $48,7^\circ$  für das Referenzverfahren mit ebenfalls vier Klassen bestimmen. Ebenso ist gut zu erkennen, dass die Erhöhung der Klassenanzahl im Training von vier auf acht einen sehr positiven Effekt hat, der sich nicht in der Klassifikation in die vier Ergebnisklassen aus Tabellen 6.5 und 6.6, jedoch in einem verbesserten

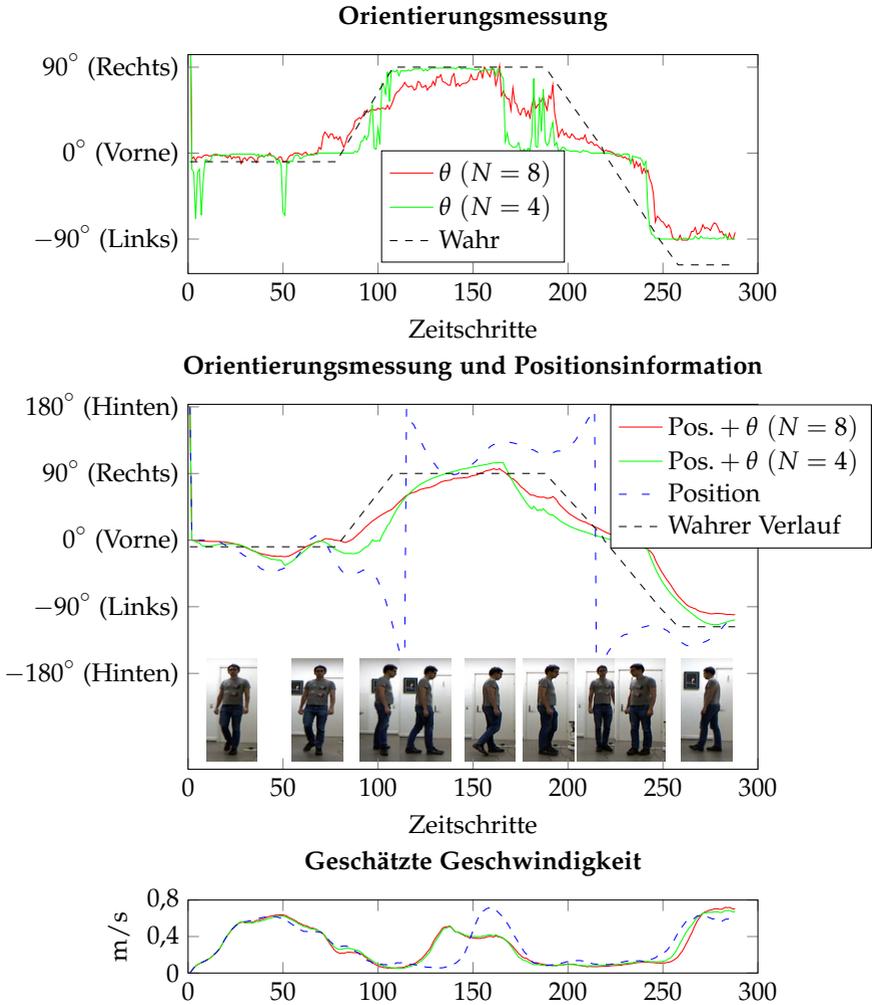
**Tabelle 6.7** Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 65,09% für das Vergleichsverfahren [25].

		Schätzung			
		Hinten	Vorne	Links	Rechts
Wahr	Hinten	73 %	11 %	10 %	6 %
	Vorne	28 %	33 %	13 %	26 %
	Links	14 %	4 %	79 %	3 %
	Rechts	7 %	5 %	5 %	83 %

absoluten mittleren Winkelfehler zeigt. Darüber hinaus liefert das Verfahren mit acht Klassen, entsprechend Abbildung 6.21, den größten Anteil von Bildern, nämlich 58 %, die einen Fehler in der Orientierung zwischen  $0^\circ$  und  $15^\circ$  besitzen, im Gegensatz zu der vorgestellten Variante mit vier Klassen, die hier nur bis zu 34 % erreicht. Das Histogramm zeigt jedoch auch, dass weiterhin auch für das beste Verfahren etwa ein Fünftel der Testbilder Fehler über  $30^\circ$  aufweisen.



**Abbildung 6.21** In Bereiche aufgeteilter absoluter Orientierungsfehler und der Anteil der sich hierin befindenden Testbilder für die verglichenen Verfahren.



**Abbildung 6.22** Im oberen Teil der Abbildung sind Orientierungsmessungen zu sehen. Der untere Teil zeigt die Fusion des Orientierungswinkels aus der Position mit und ohne Orientierungsmessung.

## Vergleich der Trainingsmethoden im Ein-Objekt-Fall

Zur Validierung der Ergebnisse der unterschiedlichen Klassifikatoren für die Schätzung der Orientierung aus Einzelbildern wurde für die Fusion von zeitlichen Positionsinformationen mit den Orientierungen auf das Beispiel aus Abschnitt 5.3.3 zurückgegriffen, einem selbst erzeugten Ein-Personen-Szenario mit häufigen Richtungswechseln aus [165]. Hierfür wurden manuell die wahren Orientierungen (schwarz) annotiert. Abbildung 6.22 zeigt im oberen Teil die Ergebnisse für die vorgestellte Orientierungsmessung mit vier (grün) bzw. acht (rot) Klassen. Bei Verwendung von acht Klassen fällt auf, dass die Verläufe sich weniger sprunghaft verändern, da die Schätzung mit vier Klassen dazu tendiert, weniger Abstufungen in der Orientierung zu treffen. Der untere Teil des Graphen zeigt die Unscented-Kalman-Filterung mit kartesischen Positionen und polarer Geschwindigkeit mit Rot für acht und mit Grün für vier Klassen fusioniert mit den aus Abschnitt 5.1 gewonnenen Positions- und sich somit über die Zeit ergebenden Geschwindigkeitsinformationen. Dabei wurde im Gegensatz zur in den nachfolgenden Abschnitten diskutierten adaptiven Varianz eine konstante Unsicherheitsannahme der Orientierung entsprechend den im vorigen Abschnitt ermittelten Messfehlern getroffen. Dies ermöglicht dabei eine Vergleichbarkeit der Klassifikatoren, ohne den Einfluss der Unsicherheit zu berücksichtigen. In Blau ist die Orientierungsschätzung ohne Hinzunahme von Orientierungsinformationen zu sehen. Farbbilder veranschaulichen ausgewählte Zeitschritte.

Während alle Filter aufgrund des verwendeten Modells nahezu konstanter Geschwindigkeit ein eher träges Verhalten aufzeigen, ist deutlich zu sehen, dass die Orientierungsinformation die Schätzung der Drehrichtung korrigiert, die bei der Filterung ohne Orientierungsinformation nur auf Bewegung basiert und somit an unbewegten Richtungswechseln zu Fehlern führen kann. Diese Fehler werden erst durch eine auf die Richtungsänderung folgende Bewegung korrigiert, wie gut um die Zeitschritte 100 und 220 herum zu sehen ist. Hingegen wird die, beispielsweise um Zeitschritt 180, fehlerhafte Orientierungsdetektion (grün/rot) durch die vorhergehende eindeutige Bewegungsrichtung ausgeglichen. Somit ergibt sich für die Testsequenz ein durchschnittlicher absoluter Winkelfehler in der Orientierung von  $20,0^\circ$  für acht Klassen,  $21,7^\circ$  für vier

Klassen und  $55,2^\circ$  ohne die Verwendung von Zusatzinformation. Der erhöhte Zusatzaufwand durch die Verwendung vieler Klassifikatoren lässt sich dadurch relativieren, dass die Auswertung der Klassifikatoren komplett parallelisierbar ist und nur einfache lineare Kernel eingesetzt werden. Zusätzlich nutzen alle Klassifikatoren dasselbe Merkmal, womit die aufwändige Merkmalsbestimmung pro Person und Zeitschritt nur einmal durchgeführt werden muss.

### Nutzung von Tiefendaten

Um auch den Einfluss der Tiefendaten und die in Abschnitt 5.2.3 vorgeschlagene Vorverarbeitung sowie die Verwendung von HOG-Merkmalen bei Berücksichtigung der Richtung der Gradienten beurteilen zu können, wurde ein eigens erstellter Datensatz aus etwa 9100 Trainingsbildern eines 3-Personen-Szenarios und 13600 Bildern aus einem 4-Personen-Szenario mit Motion-Capture-Daten verwendet (vergleiche Abbildung 5.6). Der aufgenommene Test-Datensatz besteht aus Bildern von vier Personen jeweils mit RGB- und Tiefenbild, die durch den Personendetektionsalgorithmus detektiert wurden und ist jeweils mit der Orientierung entsprechend dem Referenzsystem versehen. Zur Evaluation wurden fünf verschiedene Merkmalskonstellationen verwendet, jeweils mit gerichteter und ungerichteter HOG-Merkmals-Berechnung. Dabei wurde für alle Merkmale die vorgeschlagene Trainingsmethode mit acht Orientierungsklassen verwendet.

**Tabelle 6.8** Mittlerer absoluter Winkelfehler (MAE) der Orientierungsmessung nach Verfahren.

Trainings- und Testmerkmale	MAE ungerichtet	MAE gerichtet
RGB-HOG	$19,7^\circ$	$21,9^\circ$
Tiefen-HOG unausgeschnitten	$29,8^\circ$	$33,5^\circ$
Tiefen-HOG ausgeschnitten unnormiert	$26,4^\circ$	$25,7^\circ$
Tiefen-HOG	$21,7^\circ$	$14,9^\circ$
RGB-HOG ausgeschnitten	$18,3^\circ$	$17,7^\circ$

In Tabelle 6.8 sind die mittleren absoluten Fehler (*Mean Absolute Error*, MAE) angegeben. Während sich für die Verwendung des HOG-Merkmals auf RGB-Bildern ein geringer Unterschied zwischen gerichtetem und ungerichtetem HOG-Merkmal zeigt (19,7° zu 21,9°) mit Präferenz für das gebräuchlichere ungerichtete RGB-HOG-Merkmal, sieht man, dass die Verwendung der Tiefendaten ohne Vorverarbeitung (Tiefen-HOG unausgeschnitten) deutlich schlechtere Ergebnisse zeigt. Zwar bringt das Ausschneiden der entsprechenden Bereiche (Tiefen-HOG ausgeschnitten) eine Verbesserung, aber erst die vorgeschlagene Normierung (5.7) (Tiefen-HOG) führt zu Ergebnissen, die für den ungerichteten Fall vergleichbar mit den RGB-HOG-Merkmalen sind. Doch gerade die Verwendung der gerichteten Tiefen-HOG-Merkmale erreicht eine deutliche Verbesserung, da Wissen über näher und weiter entfernte Teile der Person berücksichtigt werden können, was einen wichtigen Einfluss auf die Orientierungsschätzung besitzt. Die Kombination von RGB- und Tiefen-HOG-Merkmal durch ein ausgeschnittenes RGB-HOG-Merkmal führt zu Verbesserungen gegenüber dem RGB-HOG-Merkmal ohne Ausschneiden. Jedoch liefert es schlechtere Ergebnisse als das Tiefen-HOG-Merkmal, wobei auch, wie im Fall des Tiefen-HOG-Merkmals, ohne Normierung der deutlichste Zugewinn durch das Ausschneiden für die gerichteten Varianten zu sehen ist.

**Tabelle 6.9** Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 78,84 % für das vorgestellte Verfahren mit ungerichteten HOG-RGB-Merkmalen.

		Schätzung			
		Hinten	Vorne	Links	Rechts
Wahr	Hinten	84,1 %	1,4 %	4,9 %	9,6 %
	Vorne	1,4 %	79,9 %	10,1 %	8,6 %
	Links	16,3 %	10,5 %	72,3 %	0,9 %
	Rechts	11,5 %	12,3 %	0,3 %	76 %

Werden die Konfusionsmatrizen für eine Klassifikation in vier Richtungen für RGB-HOG (ungerichtet, Tabelle 6.9), Tiefen-HOG (ungerichtet, Tabelle 6.10) und Tiefen-HOG (gerichtet, Tabelle 6.11) unter Berücksichtigung des Fehlerhistogramms in Abbildung 6.23 untersucht und mit Tabelle 6.6 bzw. Abbildung 6.21 verglichen, zeigt sich, dass die Ergebnisse

auf dem eigenen Datensatz für große Winkelfehler durchgehend besser als auf dem KITTI-Datensatz sind. Gründe hierfür liegen in den größeren Variationen innerhalb des KITTI-Datensatzes in Bezug auf Beleuchtung und Vielfalt an Personen. Ein weiterer Grund kann aber auch die automatisierte Auswahl der Detektionsfenster sein, da für die Auswertungen des KITTI-Datensatzes die manuell annotierten Referenzfenster als Basis gewählt wurden und die Auswahl der Detektionsfenster in dem eigenen Datensatz automatisch durch den in Abschnitt 5.1 beschriebenen Cluster-Algorithmus durchgeführt wird. Zusätzlich wird ein größerer Datensatz für das Training verwendet, der mehr Trainingsbeispiele für jede Richtung bietet.

**Tabelle 6.10** Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 78,47% für das vorgestellte Verfahren mit ungerichteten HOG-Tiefen-Merkmalen.

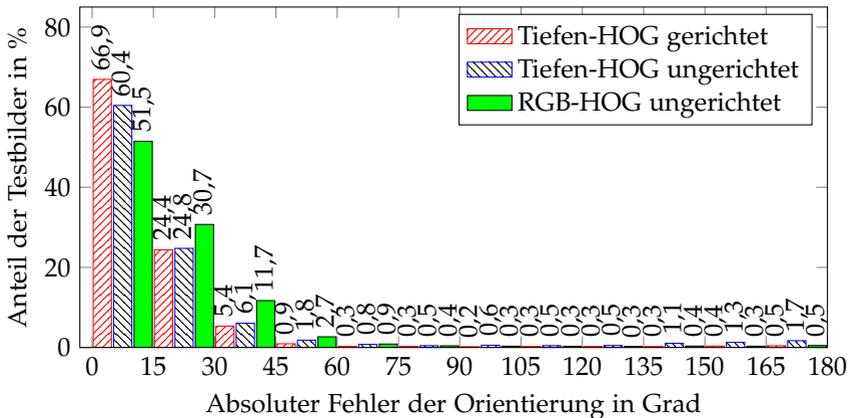
		Schätzung			
		Hinten	Vorne	Links	Rechts
Wahr	Hinten	80,8 %	6,1 %	5,7 %	7,3 %
	Vorne	8 %	78,1 %	6,3 %	7,6 %
	Links	12,3 %	10,8 %	76,4 %	0,5 %
	Rechts	13 %	9,1 %	0,4 %	77,5 %

**Tabelle 6.11** Konfusionsmatrix der Klassenzugehörigkeit bei einer Richtigklassifikationsrate von 82,40% für das vorgestellte Verfahren mit gerichteten HOG-Tiefen-Merkmalen.

		Schätzung			
		Hinten	Vorne	Links	Rechts
Wahr	Hinten	86,7 %	1,8 %	4,5 %	7 %
	Vorne	1,6 %	84,9 %	5,8 %	7,8 %
	Links	14 %	10,2 %	75,1 %	0,7 %
	Rechts	10,3 %	9,9 %	0,2 %	79,6 %

Vergleicht man die verschiedenen Merkmale in Abbildung 6.23, zeigt sich, dass Vertauschungen um 180° mit dem gerichteten tiefenbasierten Ansatz dreimal seltener vorkommen als bei Verwendung von ungerichteten Tiefen-Merkmalen und auch etwa 91 % der Testbilder einen absoluten

Fehler kleiner als  $30^\circ$  aufweisen, während dies bei den ungerichteten Tiefen-Merkmalen 86 % und bei den RGB-Merkmalen nur etwa 82 % der Testbilder sind. Somit erzielen die Tiefen-HOG-Merkmale die besten Ergebnisse und sind dabei aufgrund der Kandidatenwahl invariant gegenüber Hintergründen und durch Verwendung der Tiefendaten unabhängig von beispielsweise personenspezifischen Farbmerkmalen.



**Abbildung 6.23** In Bereiche aufgeteilter absoluter Orientierungsfehler und der Anteil der sich hierin befindenden Testbilder für die verglichenen Verfahren.

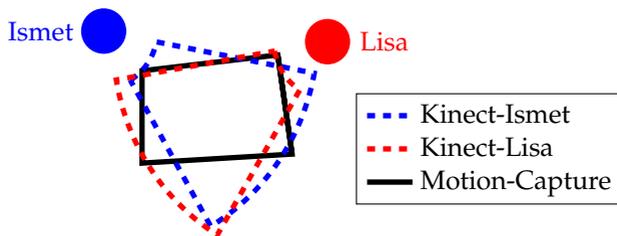
## 6.4.2 Fusion von Position und Orientierung im Multi-Objekt-Fall

Im Folgenden soll der Einfluss der Orientierungsmessung auf den Multi-Objekt-Fall anhand von Motion-Capture-Referenzdaten ausgewertet werden.

### Szenario

Zur Evaluierung werden verschiedene Szenarien mit bis zu vier Personen herangezogen, wobei auch die Szene der Testdaten für die Bewertung der Orientierungsmessungen herangezogen wird, nicht jedoch die Trainingsdaten. Die Szenen sollen eine zufällige Bewegung der Personen

abbilden, beinhalten zum einen Teil sowohl Drehungen mit gleichzeitiger Positionsänderung und Drehungen auf der Stelle und zum anderen Teil nur Szenen mit Bewegung ohne Stehenbleiben, um die unterschiedlichen Einflüsse von Bewegung auf die Orientierungsschätzung beurteilen zu können. Bei allen Szenen treten Verdeckungen sehr häufig auf. Zusätzlich werden neben den Orientierungsmessungen auch die Positionsmessungen eingesetzt und fusioniert. Die für die Auswertung eingesetzten beiden Kinect-Sensoren wurden entsprechend Abbildung 6.24 ausgerichtet und der Motion-Capture-Bereich auf den von den Sensoren abgedeckten Bereich eingeschränkt. Bei der Ausrichtung wurde darauf geachtet, einen möglichst großen Bereich durch die beiden Kinect-Sensoren abzudecken und gleichzeitig eine unterschiedliche Perspektive auf die Personen zu erreichen. Für dieses Szenario werden entsprechend den vorherigen Abschnitten sowohl Partikel- als auch Unscented-Kalman-basierte Ansätze bei unterschiedlichen Merkmalskonfigurationen und Verwendungen der Unsicherheit der Orientierungsmessung evaluiert. Das Detektionsverfahren entspricht in den Eigenschaften und der Abstrakte dem Verfahren der vorherigen Auswertungen, erweitert um die Orientierungsmessung.



**Abbildung 6.24** Laboraufbau mit eingezeichnetem Motion-Capture-Bereich und den Sichtfeldern der Kinect-Sensoren.

## OSPA-Kriterium

Um die unterschiedlichen Ansätze mittels eines Maßes zu bewerten, wurde neben der getrennten Auswertung von Positions- und Orientierungsfehler der Schätzung erneut das OSPA-Maß herangezogen, um neben der Qualität der Schätzung der Zustände auch Kardinalitätsfehler

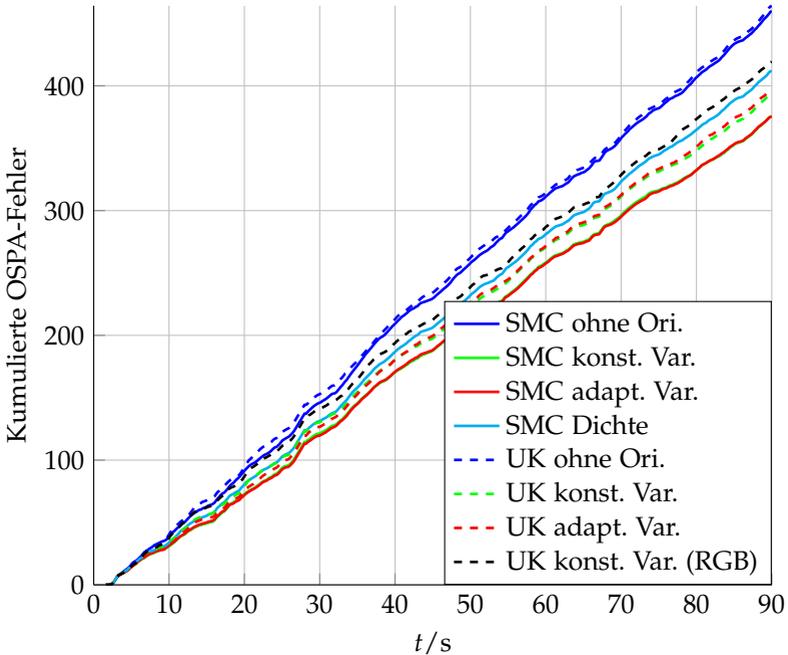
mit in die Beurteilung einzubeziehen. Im Vergleich zu Abschnitt 6.3 werden zusätzlich zu  $x$ - und  $y$ -Koordinate gleicher Einheit zusätzlich noch Orientierungsfehler mit eingebracht, die aufgrund ihrer Einheit verschieden gewichtet werden müssen. Um den Einfluss aller größer vergleichbar zu machen, wurden die einzelnen Fehler von Position und Orientierung für das Maß aus Abschnitt 3.7 so gewichtet, dass sich die gewichteten mittleren Fehler nominal ähneln. Somit wird die Position pro Koordinate  $x$  und  $y$  jeweils mit eins und die Orientierung mit 0,2 für die Berechnung des Fehlers gewichtet.

**Tabelle 6.12** OSPA-Positionsfehler und Winkelfehler in Abhängigkeit des gewählten Filters und der Behandlung der Orientierung.

Verwendung der Orientierung	UK-CPHD		SMC-CPHD	
	Position	Winkel	Position	Winkel
Ohne	7,7 cm	22,6°	7,8 cm	22,5°
konstante Unsicherheit	7,9 cm	9,9°	8,4 cm	10,1°
adaptive Unsicherheit	8,5 cm	10,2°	8,1 cm	11,5°
Dichte	–	–	7,7 cm	14,9°
RGB	7,9 cm	11,6°	8,5 cm	11,9°

### Nutzung der Orientierung

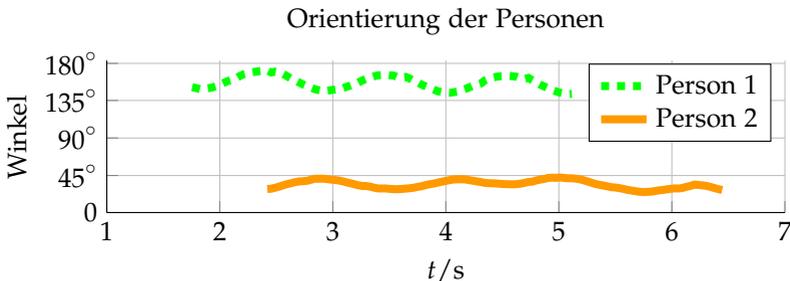
Abbildung 6.25 zeigt die OSPA-Fehler für das Vier-Personen-Szenario mit Orientierungswechsel mit und ohne Bewegung für das SMC-CPHD- und UK-CPHD-Filter. Dabei wird der kumulative OSPA-Fehler betrachtet, um zum einen eine bessere Anschaulichkeit zu erreichen und zum anderen darzustellen, dass die Fehler sich kaum kreuzen und somit die Verfahren im Vergleich zueinander insgesamt besser oder schlechter abschneiden. Betrachtet man die Fehler der in Blau gezeichneten Filter ohne Orientierungsschätzung, sieht man, dass diese für beide Implementierungen sehr ähnliche Ergebnisse liefern, was auch von den mittleren Winkelfehlern und mittleren Positionsfehlern in Tabelle 6.12 gestützt wird. In der Tabelle sieht man jedoch auch den nur auf Positionsmessungen basierenden relativ großen Winkelfehler der beiden Verfahren von im Mittel 22,6°.



**Abbildung 6.25** Vergleich der Verfahren für das 4-Personen-Szenario bei Drehungen mit und ohne Bewegung mit Unscented-Kalman- (gestrichelt) und Partikel-Implementierungen (durchgezogen). Dabei wird der Einsatz einer konstanten Unsicherheit, einer adaptiven Unsicherheit und der Einsatz der kompletten Dichte (nur SMC-CPHD) verglichen für gerichtete HOG-Merkmale auf Tiefen-Daten. Zusätzlich ist noch ein Filter konstanter Unsicherheit für die Verwendung von RGB-Bildern gezeigt. Als Bewertungskriterium wird ein kumulierter OSPA-Fehler über Position, Orientierung und Kardinalität verwendet.

Beim Vergleich der verschiedenen Verfahren mit Orientierungsmessung zeigt sich erwartungsgemäß, dass die Verwendung der ungerichteten RGB-HOG-Merkmale wie auch in den vorigen Auswertungen schlechter abschneidet als die der gerichteten Tiefen-HOG-Merkmale. Die konstanten Unsicherheiten wurden hierbei jeweils an den Standardabweichungen der Orientierungsmessungen orientiert, die sich zwischen  $15^\circ$  und  $22^\circ$  für Tiefen- bzw. Farbmerkmale bewegen. Die konstante Wahl der Unsicherheit zeigt sich dabei der adaptiven Unsicherheitsbehandlung ebenbürtig, wenn nicht sogar teilweise überlegen, was sich

auch durch den kleineren Winkelfehler von im Mittel etwa  $10^\circ$  für UK- und SMC-CPHD-Filter bei konstanter Unsicherheit gegenüber  $11,5^\circ$  bei Verwendung des SMC-CPHD-Filters mit adaptiver Unsicherheit zeigt, wobei berücksichtigt werden muss, dass für das adaptive Verfahren keine vorherige Abschätzung einer Unsicherheit nötig ist. Generell lässt sich jedoch ein besseres Abschneiden der SMC-basierten Filter analog zu Abschnitt 6.3 beobachten, die trotz vergleichbarer Winkel- und Positionsfehler einen kleineren OSPA-Fehler erreichen, was sich durch eine bessere Kardinalitätsschätzung als verbleibende Komponente des Fehlers begründen lässt. Dennoch liegt der Kardinalitätsfehler aufgrund der vielen Verdeckungen im Durchschnitt für alle Verfahren bei etwa einem Objekt, was durch Fehldetektionen der Personendetektion begründet ist und nicht allein durch die Modellierung von Verdeckungen aufgelöst werden kann. Betrachtet man hingegen die Schwankung einer Referenzmessung, die aufgrund der Positionierung der Marker an der Hüfte beim Geradeauslaufen – beispielhaft für zwei Personen mit Standardabweichung von  $8^\circ$  in Abbildung 6.26 zu sehen – entsteht, zeigt sich die Qualität der Schätzung bei Fusion der Orientierungsmessungen.



**Abbildung 6.26** Orientierungsreferenz von zwei geradeauslaufenden Personen aus Motion-Capture-Daten.

Sieht man die nur für das SMC-CPHD-Filter auswertbare Repräsentation der Unsicherheit über die Dichte an, so zeigt sich, dass diese in der Gesamtbewertung zwischen den Verfahren ohne Orientierung und den Verfahren mit konstanter bzw. adaptiver Orientierung angesiedelt ist. Die Begründung hierfür ist in der Repräsentation der Dichte aus den gaußförmigen „Expertenmeinungen“ zu finden. Eine Dichte, wie beispielsweise

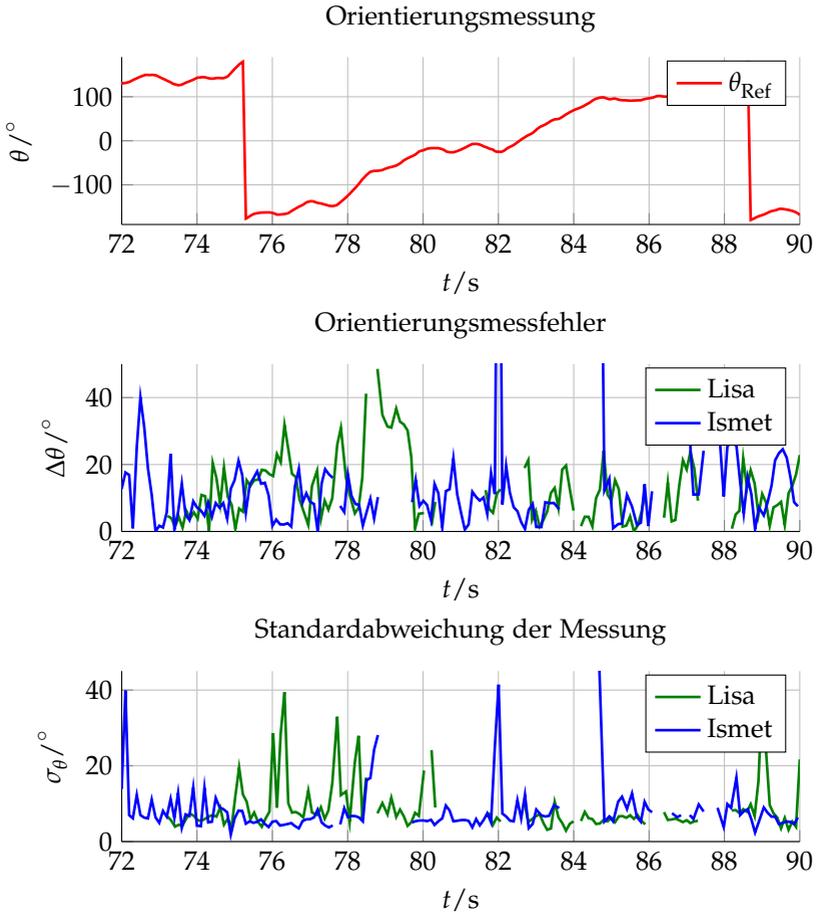
in Abbildung 5.8, weist auch für vermeintlich sichere Messungen, die beispielsweise nur auf dem Gewicht eines Klassifikators basieren, die Standardabweichung von  $\sigma_{\text{Exp}}$  auf. Diese entspricht der Klassenbreite und ist somit als Maß der Unsicherheit eine oft zu pessimistische Schätzung. Für die adaptiven (mit Korrektur) und konstanten Verfahren ist diese Unsicherheit jedoch meist deutlich kleiner, was einen stärkeren Einfluss der Orientierungsmessung auf das Schätzergebnis zur Folge hat. Auf der anderen Seite kommt es, wie sich auch in Tabelle 6.12 zeigt, für geringeren Einfluss der Orientierung auch zu nur verhältnismäßig kleinen Fehlern der Position, da Widersprüche zwischen Orientierungsmessung und einer aus der Position folgenden Orientierung zu Gunsten einer besseren Position entschieden werden.

Die Problematik eines adaptiven Ansatzes der Unsicherheit wird weiter in Abbildung 6.27 deutlich. Dort sind neben der Referenzorientierung  $\theta_{\text{Ref}}$  die von beiden Sensorträgern generierten Messfehler in der Orientierung  $\Delta\theta$  und die zugehörigen Unsicherheiten  $\sigma_\theta$  einer adaptiven Unsicherheit für eine Person in einem 18-sekündigen Ausschnitt der Szene dargestellt. Zwar ist zu erkennen, dass große Fehler in der Messung meistens auch an erhöhte Unsicherheiten geknüpft sind, jedoch fehlerhafte Messungen nicht immer zwingend für eine erhöhte Standardabweichung sorgen und umgekehrt. Die adaptive Unsicherheit liegt die meiste Zeit unter der konstanten Annahme von etwa  $15^\circ$  für das gesamte Szenario, so dass auch falsche Messungen noch stärker in die Schätzung einfließen, was sich auch in Tabelle 6.12 durch erhöhte Winkelfehler zeigt. Trotz eines Korrelationskoeffizienten zwischen Fehler und Standardabweichung von 0,48 lässt sich kein weiterer Gewinn in der vorgeschlagenen Fusion erzielen.

Bei dem Vergleich zu einer ebenfalls aus vier Personen bestehenden Szene, die allerdings nur aus sich bewegenden Personen besteht, die nicht stoppen, zeigt sich insbesondere der Vorteil der Fusion bei unbewegten Szenen beziehungsweise Drehungen ohne Positionsänderung. In der ausschließlich bewegten Szene verbessert sich die Orientierungsschätzung von  $15,6^\circ$  ohne Orientierungsmessung auf zwischen  $9,8^\circ$  (UK-CPHD bei konstanter Unsicherheit) und  $10,3^\circ$  (UK-CPHD bei adaptiver Unsicherheit). Während im Falle von Bewegung die Orientierungsschätzung aufgrund der Position bereits gute Ergebnisse erzielt, hilft die

Zusatzinformation gerade in Situationen ohne Bewegung, Fehler in der Zustandsschätzung zu korrigieren.

Somit lassen sich zusammenfassend durch die Fusion von Position und Orientierung aus Einzelmessungen über die Zeit deutliche Verbesserungen der Orientierungsschätzung erreichen. Während die Messungen einen mittleren absoluten Fehler von  $14,9^\circ$  besitzen und die Schätzung der Orientierung nur aufgrund der Position zu einem Fehler von  $15,6^\circ$  für Szenen mit durchgehender Bewegung und  $22,5^\circ$  bei zusätzlichen Drehungen im Stand führt, lässt sich durch die Fusion auch im Multi-Objekt-Fall ein Fehler von etwa  $10^\circ$  erreichen. Dabei besteht der zusätzliche Berechnungsaufwand lediglich in der Auswertung linearer SVM-Klassifikatoren und Erweiterung der Messungen um die Orientierungsinformation bzw. die Aktualisierung der Likelihood-Funktion.



**Abbildung 6.27** Ausschnitt aus Referenzorientierung, Fehlern der gemessenen Orientierung und adaptiver Standardabweichung für eine Person des Vier-Personen-Szenarios.



# 7 Zusammenfassung

Die Zustandsschätzung einer unbekanntem Anzahl an Objekten stellt trotz theoretisch Bayes-optimaler Multi-Objekt-Filter durch die große Anzahl an Modellannahmen dieser Filter eine große Herausforderung dar, da einfache Charakterisierungen der Sensoren bezüglich des Detektionsverhaltens in praktischen Anwendungen nur unzureichende Ergebnisse liefern. In dieser Arbeit wurde die Eignung von Filtermethoden auf Basis endlicher Mengen ohne explizite Datenassoziation für den praktischen Einsatz in Multi-Objekt-Multi-Sensor-Szenarien untersucht. Als Anwendungsszenario wurde auf die Verfolgung von Fußgängern in Innenräumen eingegangen, das mittels eines Aufbaus verteilter Sensorträger evaluiert wurde. Die Untersuchungen basieren dabei auf Filtern, wie dem PHD-Filter, die das optimale Bayes-Filter unter Verwendung verschiedener Modellannahmen annähern und somit eine Analogie des Kalman-Filters für die Ein-Objekt-Zustandsschätzung im Multi-Objekt-Fall darstellen und versuchen, die gleichzeitige Schätzung von Objektanzahl und Objektzuständen bestmöglich durchzuführen.

In der vorliegenden Arbeit wurden nötige Modellerweiterungen für die verteilte Sensorfusion identifiziert, die diese theoretischen Ansätze unter Einbeziehung des Wissens über die verwendeten Sensoren und über die Nutzung zusätzlicher Information aus den Sensordaten für die Praxis nutzbar machen. Die Erweiterungen lassen sich in eine adaptive Modellierung der Detektionswahrscheinlichkeit und eine richtungssensitive Verfolgung untergliedern, deren Eigenschaften inklusive der Erkenntnisse nun kurz zusammengefasst werden.

## 7.1 Adaptive Detektionswahrscheinlichkeit

Die Untersuchung und der Vergleich von drei Näherungen des optimalen Multi-Objekt-Bayes-Filters – des PHD-, CPHD- und CBMeMber-Filters –

zeigen, dass die Wahl der angenommenen Detektionswahrscheinlichkeit großen Einfluss auf die Schätzung der unbekanntes Objektanzahl hat. Dabei wird in der Literatur meist von einer konstanten Detektionswahrscheinlichkeit ausgegangen, was jedoch gerade in Szenarien mit teilweise überlappenden und begrenzten Sichtfeldern zu Widersprüchen und somit zu Fehlern führt. Zur Vermeidung dieser Fehler wurde eine adaptive Detektionswahrscheinlichkeit vorgeschlagen, die Sichtfelder, statische sowie dynamische Verdeckungen und auch den Abstand zwischen Objekt und Sensor miteinbezieht. Es zeigt sich, dass die vorgestellte Modellierung widersprüchliche Annahmen in der Fusion zwischen verteilten Sensoren auflösen kann und zu Verbesserungen der Kardinalitätsschätzung führt. Dieses Modell wurde sowohl für Monte-Carlo-basierte Partikelansätze der Filter als auch für Gauß-Mixtur-Ansätze mit Unscented-Kalman-Filtern in die rekursiven Filterstrukturen integriert. Für die Ermittlung der Detektionswahrscheinlichkeit für Gauß-Ansätze wurde dabei eine Vorgehensweise unter der Verwendung von Sigma-Punkten vorgeschlagen, die weitere Verbesserungen der Schätzfehler erzielt, indem nicht nur der Mittelwert für die Berechnung der Detektionswahrscheinlichkeit einer Gauß-Komponente herangezogen wird, sondern auch eine differenziertere Berechnung über die Unscented-Transformation vorgenommen wird.

Anhand von simulativen sowie realen Experimenten und unter Nutzung von Motion-Capture-Referenzdaten wurde für das Szenario der Fußgängerfolgung gezeigt, dass das CPHD-Filter zwar die höchste Komplexität besitzt, aber die größte Robustheit gegenüber Fehldektionen aufweist und somit den über die OSPA-Metrik beurteilten kleinsten Schätzfehler erzielt. Während die Partikel-Implementierungen tendenziell zu kleineren Fehlern führen, zeigen die Unscented-Kalman-Implementierungen im Einzelfall robustere und reproduzierbare Ergebnisse bei geringerer Rechenzeit. Die vorgeschlagenen Erweiterungen erzielen unabhängig von der gewählten Implementierung Verbesserungen insbesondere bei der Kardinalitätsschätzung, indem widersprüchliche Annahmen über die Sichtbarkeit von Objekten durch geeignete Modelle ausgeglichen werden.

Problematisch bleiben jedoch weiterhin gehäufte Fehldetektionen oder Messungen, die die Grundannahme der Näherung von einer Messung

pro Objekt und somit die Unabhängigkeit der Objekte verletzen. Für die hieraus resultierenden Fehler müssen, falls sie nicht durch eine auf Trajektorien-Ebene stattfindende Zuordnung oder durch verbesserte Detektionsalgorithmen aufgelöst werden können, Näherungen gefunden werden, die diese Objektinteraktionen berücksichtigen.

## 7.2 Richtungssensitive Verfolgung

Im zweiten Teil der Arbeit wurde zudem der Nutzen von Orientierungsinformationen für die Schätzung der Zustände untersucht. Da die Bewegung von Personen nicht vollständig durch die Position beschrieben wird und die häufige Modellierung der Bewegung durch eine nahezu konstante Geschwindigkeit nur eine grobe Näherung ist, wurde ein Verfahren zur Gewinnung von Orientierungen von Personen relativ zu einer Kamera aus Einzelbildern erweitert. Das auf der Klassifikation von HOG-Merkmalen mittels SVMs beruhende System wird dabei vor allem durch eine gezielte Gewichtung, die die kontinuierliche Orientierung der Trainingsdaten berücksichtigt, verbessert. Hierbei führt die Hinzunahme von Tiefendaten mittels RGB-D-Sensoren zu Verbesserungen für die Bestimmung der Orientierung. Dabei bewirkt vor allem die Verwendung gerichteter Gradienten bei der Berechnung der Merkmale eine deutliche Verbesserung der Schätzung. Die Fusion dieser Orientierungsmessungen mit den Positionsdaten über die Zeit wurde anschließend mittels der zuvor diskutierten Filter durchgeführt und erreicht deutliche Verbesserungen der Orientierungsschätzung in Ein- und Multi-Objekt-Szenarien, was an einem eigens angefertigten Datensatz mit Referenzdaten für Position und Orientierung verifiziert wurde. Eine sich aus dem Verfahren ergebende Beschreibung der Unsicherheit der Orientierungsmessung ermöglicht die parameterfreie Fusion der Orientierung mit der Position bei vergleichbaren Ergebnissen zu mit einer a priori bestimmten konstanten Beschreibung der Messunsicherheit.

## 7.3 Ausblick

Wie Fehler in der Zustandsschätzung – beispielsweise durch voneinander abhängige Objekte – zeigen, bedürfen spezielle Szenarien einer weitergehenden Verarbeitung. Dies könnte durch eine möglicherweise szenario-spezifische Verwaltung der Trajektorien oder die Verwendung weiterer Merkmale, die eine Wiedererkennung von Objekten nach dem Verlassen aller Sichtfelder ermöglicht, geschehen. Von theoretischer Seite her wird die Erforschung von Verfahren, die bei vertretbarer Komplexität auch Interaktionen von Messungen und Objekten modellieren, ein wichtiger Schritt in Richtung der praktischen Anwendung derartiger Multi-Objekt-Filter sein. Bezüglich der Modellierung lässt sich das vorgeschlagene Konzept einer adaptiven Detektionswahrscheinlichkeit auch auf andere Einflüsse wie Beleuchtung und weiteres Wissen über die Szene oder Objekte übertragen. Ebenso ist das automatische Anlernen von Sensor-modellen ein interessanter Anknüpfungspunkt weiterer Forschung.

Für die Messung der Orientierung wäre ein weiterführender Ansatz, Regression nicht nur für die Bestimmung der Expertengewichte zu verwenden, sondern auch eine direkte Ausgabe der Orientierung zu ermöglichen, ohne eine vorherige Diskretisierung in verschiedene Klassen durchzuführen. Dies könnte auch eine direkte Aussage über die Unsicherheit der Messung ermöglichen, um eine Fusion von Position und Orientierung weiter zu verbessern. Ebenso lässt sich dieser Ansatz auch auf Unsicherheiten in der Position übertragen, um eine optimale Balance des Einflusses der verschiedenen Modalitäten zu erhalten und somit auch positive Effekte für die Positionsschätzung zu erreichen. Während die meisten Versuche in dieser Arbeit in statischen Laborumgebungen stattgefunden haben, wäre ein nächster Schritt die Erweiterung der betrachteten Posen, Objekte und Szenarien sowie der zugehörigen Detektionsalgorithmen für Szenarien mit Eigenbewegung der Sensoren und komplexere Umgebungen.

# Anhang



# A Bewegungsmodell mit konstanter polarer Geschwindigkeit

Beschreibt

$$\mathbf{x}(t) = [x(t) \quad y(t) \quad \theta(t) \quad v(t)]^T \quad (\text{A.1})$$

den Zustand eines Systems mit Geschwindigkeitsbetrag  $v(t)$  und Orientierung des Objekts  $\theta(t)$ , so wird in [68] gezeigt, dass sich das Modell konstanter Geschwindigkeit bei polarem Geschwindigkeitsvektor entsprechend Abschnitt 2.1 nach Diskretisierung der Zeit mit Dauer  $\Delta t$  zwischen Zeitschritt  $t_1$  und  $t_2$  bzw.  $k-1$  und  $k$  und diskretem Zustandsvektor zu

$$\mathbf{x}_k = \begin{bmatrix} x_k \\ y_k \\ \theta_k \\ v_k \end{bmatrix} = \begin{bmatrix} x_{k-1} + v_{k-1} \cos(\theta_{k-1})\Delta t - \sigma_\theta^2 v_{k-1} \cos(\theta_{k-1}) \frac{\Delta t^2}{4} \\ y_{k-1} + v_{k-1} \sin(\theta_{k-1})\Delta t - \sigma_\theta^2 v_{k-1} \sin(\theta_{k-1}) \frac{\Delta t^2}{4} \\ \theta_{k-1} \\ v_{k-1} \end{bmatrix} + \mathbf{v}_{k-1} \quad (\text{A.2})$$

mit der Kovarianzmatrix

$$\mathbf{Q}_{k-1} = \begin{bmatrix} a & b & c & d \\ b & e & f & g \\ c & f & h & 0 \\ d & g & 0 & i \end{bmatrix} \quad (\text{A.3})$$

$$a = \left[ \sigma_{\theta}^2 v_{k-1}^2 \sin^2(\theta_{k-1}) + \sigma_v^2 \cos^2(\theta_{k-1}) \right] \frac{\Delta t^3}{3} \quad (\text{A.4})$$

$$b = \left[ \sigma_v^2 - \sigma_{\theta}^2 v_{k-1}^2 \right] \sin(\theta_{k-1}) \cos(\theta_{k-1}) \frac{\Delta t^3}{3} \quad (\text{A.5})$$

$$c = -\sigma_{\theta}^2 v_{k-1} \sin(\theta_{k-1}) \frac{\Delta t^2}{2} \quad (\text{A.6})$$

$$d = \sigma_v^2 \cos(\theta_{k-1}) \frac{\Delta t^2}{2} \quad (\text{A.7})$$

$$e = \left[ \sigma_{\theta}^2 v_{k-1}^2 \cos^2(\theta_{k-1}) + \sigma_v^2 \sin^2(\theta_{k-1}) \right] \frac{\Delta t^3}{3} \quad (\text{A.8})$$

$$f = \sigma_{\theta}^2 v_{k-1} \cos(\theta_{k-1}) \frac{\Delta t^2}{2} \quad (\text{A.9})$$

$$g = \sigma_v^2 \sin(\theta_{k-1}) \frac{\Delta t^2}{2} \quad (\text{A.10})$$

$$h = \sigma_{\theta}^2 \Delta t \quad (\text{A.11})$$

$$i = \sigma_v^2 \Delta t \quad (\text{A.12})$$

berechnen lässt. Die Zeit  $\Delta t$  beschreibt dabei die Differenz zwischen zwei eintreffenden Sensormessungen.

# B Klassen endlicher Zufallsmengen

Eine endliche Zufallsmenge  $\mathcal{X}$  ist sowohl in ihren Elementen als auch in ihrer Kardinalität zufällig. Somit lässt sie sich durch ihre Kardinalitätsverteilung  $\rho(n) := P(|\mathcal{X}| = n)$  und die symmetrischen gemeinsamen Verteilungen  $p_n(\mathbf{x}_1, \dots, \mathbf{x}_n)$ , mit  $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathcal{X}$  charakterisieren als Wahrscheinlichkeitsdichte  $\pi(\mathcal{X})$  [121]

$$\pi(\{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = n! \cdot \rho(n) \cdot p_n(\mathbf{x}_1, \dots, \mathbf{x}_n). \quad (\text{B.1})$$

Im Folgenden sind die wichtigsten Klassen von endlichen Zufallsmengen in Anlehnung an [177] kurz vorgestellt, während für eine ausführlichere Einführung auf [65, 144] verwiesen wird.

## B.1 Endliche IID-Cluster-Zufallsmenge

Eine IID-Cluster-Zufallsmenge  $\mathcal{X}$  (*independent, identically distributed cluster random finite set* (RFS)) nennt man eine Menge, deren Kardinalität die Bedingung entsprechend dem Erwartungswert  $\hat{N}$  der Intensitätsfunktion  $D(\mathbf{x})$

$$\hat{N} = \sum_{n=0}^{\infty} n\rho(n) = \int D(\mathbf{x})d\mathbf{x}, \quad (\text{B.2})$$

erfüllt. Jedes Element der endlichen Menge  $\mathcal{X}$  ist dabei unabhängig von den anderen und gleich verteilt mit der Wahrscheinlichkeitsdichte  $p(\mathbf{x}) = D(\mathbf{x})/\hat{N}$ .

## B.2 Endliche Poisson-Zufallsmenge

Man spricht von einer Poisson-verteilten Zufallsmenge, sobald die Kardinalität Poisson-verteilt mit Erwartungswert  $\hat{N}$

$$\hat{N} = \int D(\mathbf{x}) \, d\mathbf{x} \quad (\text{B.3})$$

ist. Die Kardinalität ergibt sich somit zu

$$\rho(n) = \frac{\hat{N}^n e^{-\hat{N}}}{n!}. \quad (\text{B.4})$$

Dies stellt einen Spezialfall der *IID-Cluster-Zufallsmenge* dar.

## B.3 Endliche Bernoulli-Zufallsmenge

Eine endliche Bernoulli-Zufallsmenge  $\mathcal{X}$  ist entweder leer mit Wahrscheinlichkeit  $1 - r$  oder besitzt mit der Gegenwahrscheinlichkeit  $r$  genau ein Element. Dieses Element ist gemäß einer Wahrscheinlichkeitsdichte  $p(\mathbf{x})$  verteilt, so dass gilt

$$\pi(\mathcal{X}) = \begin{cases} 1 - r, & \mathcal{X} = \emptyset \\ r \cdot p(\mathbf{x}), & \mathcal{X} = \{\mathbf{x}\} \end{cases}. \quad (\text{B.5})$$

## B.4 Endliche Multi-Bernoulli-Zufallsmenge

Eine endliche Multi-Bernoulli-Zufallsmenge  $\mathcal{X}$  ist die Vereinigung von  $M$  unabhängigen Bernoulli-verteilten Zufallsmengen  $\mathcal{X}^{(i)}$  mit  $i = 1, \dots, M$ :

$$\mathcal{X} = \bigcup_{i=1}^M \mathcal{X}^{(i)}. \quad (\text{B.6})$$

# Literaturverzeichnis

- [1] **Al Nuaimi, K. und Kamel, H.** *A survey of indoor positioning systems and algorithms*. In: *International Conference on Innovations in Information Technology (IIT)*. 2011, S. 185–190.
- [2] **An, L., Kafai, M., Yang, S. und Bhanu, B.** *Person Re-Identification with Reference Descriptor*. In: *IEEE Transactions on Circuits and Systems for Video Technology* PP.99 (2015), S. 1–12.
- [3] **Arasaratnam, I. und Haykin, Simon.** *Cubature Kalman Filters*. In: *IEEE Transactions on Automatic Control* 54.6 (2009), S. 1254–1269.
- [4] **Arras, K.O., Mozos, O.M. und Burgard, W.** *Using Boosted Features for the Detection of People in 2D Range Data*. In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2007, S. 3402–3407.
- [5] **Arras, K.O., Grzonka, S., Luber, M. und Burgard, W.** *Efficient people tracking in laser range data using a multi-hypothesis leg-tracker with adaptive occlusion probabilities*. In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2008, S. 1710–1715.
- [6] **Bar-Shalom, Y., Li, X.R. und Kirubarajan, T.** *Estimation with Applications to Tracking and Navigation: Theory Algorithms and Software*. New York: Wiley, 2004.
- [7] **Bar-Shalom, Y., Willett, P.K., Willett, P.K. und Tian, X.** *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing, 2011.
- [8] **Battistelli, G., Chisci, L., Fantacci, C., Farina, A. und Graziano, A.** *Consensus CPHD Filter for Distributed Multitarget Tracking*. In: *IEEE Journal of Selected Topics in Signal Processing* 7.3 (2013), S. 508–520.
- [9] **Bedworth, M. und O'Brien, J.** *The Omnibus model: a new model of data fusion?* In: *Aerospace and Electronic Systems Magazine, IEEE* 15.4 (2000), S. 30–36.

- [10] **Bernardin, Keni und Stiefelhagen, Rainer.** *Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics*. In: *Journal on Image Video Processing* 2008 (2008), S. 1–10.
- [11] **Beyerer, Jürgen, Puente León, Fernando und Frese, Christian.** *Automatische Sichtprüfung: Grundlagen, Methoden und Praxis der Bildgewinnung und Bildauswertung*. Berlin Heidelberg: Springer, 2012.
- [12] **Bishop, Christopher M.** *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer New York, 2006.
- [13] **Bo, Liefeng und Sminchisescu, C.** *Structured output-associative regression*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. 2009, S. 2403–2410.
- [14] **Bradski, G.** *The OpenCV Library*. In: *Dr. Dobb's Journal of Software Tools* (2000).
- [15] **Bresenham, J. E.** *Algorithm for Computer Control of a Digital Plotter*. In: *IBM Systems Journal* 4.1 (1965), S. 25–30.
- [16] **Brscic, D., Kanda, T., Ikeda, T. und Miyashita, T.** *Person Tracking in Large Public Spaces Using 3-D Range Sensors*. In: *IEEE Transactions on Human-Machine Systems* 43.6 (2013), S. 522–534.
- [17] **Butler, D. Alex, Izadi, Shahram, Hilliges, Otmar, Molyneaux, David, Hodges, Steve und Kim, David.** *Shake'N'Sense: Reducing Interference for Overlapping Structured Light Depth Cameras*. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Austin, Texas, USA: ACM, 2012, S. 1933–1936.
- [18] **Cielniak, Grzegorz, Duckett, Tom und Lilienthal, Achim J.** *Data association and occlusion handling for vision-based people tracking by mobile robots*. In: *Robotics and Autonomous Systems* 58.5 (2010), S. 435–443.
- [19] **Colombo, A, Fontanelli, D., Macii, D. und Palopoli, L.** *Flexible Indoor Localization and Tracking Based on a Wearable Platform and Sensor Data Fusion*. In: *IEEE Transactions on Instrumentation and Measurement* 63.4 (2014), S. 864–876.

- [20] **Correa, Javier, Liu, Jindong und Yang, Guang-Zhong.** *Real Time People Tracking in Crowded Environments with Range Measurements.* In: *Social Robotics.* Hrsg. von **Herrmann, Guido, Pearson, Martin J., Lenz, Alexander, Bremner, Paul, Spiers, Adam und Leonards, Ute.** Bd. 8239. Lecture Notes in Computer Science. Springer International Publishing, 2013, S. 471–480.
- [21] **Cramer, Erhard und Kamps, Udo.** *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik.* Berlin: Springer, 2007.
- [22] **Dalal, N. und Triggs, B.** *Histograms of oriented gradients for human detection.* In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).* Bd. 1. 2005, 886–893 vol. 1.
- [23] **Daley, Daryl J. und Vere-Jones, David.** *An introduction to the theory of point processes. vol. I. , Elementary theory and methods.* Probability and its applications. New York, Berlin, Paris: Springer, 2003.
- [24] **Endres, Felix, Hess, J., Engelhard, Nikolas, Sturm, J. und Burgard, Wolfram.** *Online-6D-SLAM für RGB-D-Sensoren.* In: *at-Automatisierungstechnik Methoden und Anwendungen der Steuerungs-, Regelungs- und Informationstechnik 60.5* (2012), S. 270–278.
- [25] **Enzweiler, M. und Gavrila, D.M.** *Integrated pedestrian classification and orientation estimation.* In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR).* 2010, S. 982–989.
- [26] **Erdinc, O., Willett, P. und Bar-Shalom, Y.** *The Bin-Occupancy Filter and Its Connection to the PHD Filters.* In: *IEEE Transactions on Signal Processing* 57.11 (2009), S. 4232–4246.
- [27] **Faouzi, Nour-Eddin El, Leung, Henry und Kurian, Ajeesh.** *Data fusion in intelligent transportation systems: Progress and challenges – A survey.* In: *Information Fusion* 12.1 (2011). Special Issue on Intelligent Transportation Systems, S. 4–10.
- [28] **Felzenszwalb, P.F., Girshick, R.B., McAllester, D. und Ramanan, D.** *Object Detection with Discriminatively Trained Part-Based Models.* In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.9 (2010), S. 1627–1645.

- [29] **Fleuret, F., Berclaz, J., Lengagne, R. und Fua, P.** *Multicamera People Tracking with a Probabilistic Occupancy Map*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30.2 (2008), S. 267–282.
- [30] **Flohr, F., Dumitru-Guzu, M., Kooij, J.F.P. und Gavrila, D.M.** *Joint probabilistic pedestrian head and body orientation estimation*. In: *IEEE Intelligent Vehicles Symposium (IV)*. 2014, S. 617–622.
- [31] **Fradi, Hajer und Dugelay, Jean-Luc.** *Towards crowd density-aware video surveillance applications*. In: *Information Fusion* 24 (2015), S. 3–15.
- [32] **Franken, D., Schmidt, M. und Ulmke, M.** *SSpooky Action at a Distance in the Cardinalized Probability Hypothesis Density Filter*. In: *IEEE Transactions on Aerospace and Electronic Systems* 45.4 (2009), S. 1657–1664.
- [33] **Freund, Yoav und Schapire, Robert E.** *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*. In: *Journal of Computer and System Sciences* 55.1 (1997), S. 119–139.
- [34] **Gandhi, T. und Trivedi, M.M.** *Image based estimation of pedestrian orientation for improving path prediction*. In: *IEEE Intelligent Vehicles Symposium (IV)*. 2008, S. 506–511.
- [35] **Geiger, A, Lenz, P, Stiller, C und Urtasun, R.** *Vision Meets Robotics: The KITTI Dataset*. In: *International Journal of Robotics Research* 32.11 (2013), S. 1231–1237.
- [36] **Georgescu, R. und Willett, P.** *The Multiple Model CPHD Tracker*. In: *IEEE Transactions on Signal Processing* 60.4 (2012), S. 1741–1751.
- [37] **Geronimo, D., Lopez, A.M., Sappa, A.D. und Graf, T.** *Survey of Pedestrian Detection for Advanced Driver Assistance Systems*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.7 (2010), S. 1239–1258.
- [38] **Glas, Dylan F., Miyashita, Takahiro, Ishiguro, Hiroshi und Hagita, Norihiro.** *Laser-Based Tracking of Human Position and Orientation Using Parametric Shape Modeling*. In: *Advanced Robotics* 23.4 (2009), S. 405–428.

- [39] **Glassner, Andrew S.** *An introduction to ray tracing*. Burlington: Morgan Kaufmann, 1989.
- [40] **Gordon, N.J., Salmond, D.J. und Smith, A.F.M.** *Novel approach to nonlinear/non-Gaussian Bayesian state estimation*. In: *IEEE Proceedings on Radar and Signal Processing* 140.2 (1993), S. 107–113.
- [41] **Goto, K., Kidono, K., Kimura, Y. und Naito, T.** *Pedestrian detection and direction estimation by cascade detector with multi-classifiers utilizing feature interaction descriptor*. In: *IEEE Intelligent Vehicles Symposium (IV)*. 2011, S. 224–229.
- [42] **Granstrom, K., Lundquist, C. und Orguner, O.** *Extended Target Tracking using a Gaussian-Mixture PHD Filter*. In: *IEEE Transactions on Aerospace and Electronic Systems* 48.4 (2012), S. 3268–3286.
- [43] **Granstrom, K. und Orguner, U.** *A phd Filter for Tracking Multiple Extended Targets Using Random Matrices*. In: *IEEE Transactions on Signal Processing* 60.11 (2012), S. 5657–5671.
- [44] **Grisetti, G., Stachniss, C. und Burgard, W.** *Improved Techniques for Grid Mapping With Rao-Blackwellized Particle Filters*. In: *IEEE Transactions on Robotics* 23.1 (2007), S. 34–46.
- [45] **Guldogan, M.B., Lindgren, D., Gustafsson, F., Habberstad, H. und Orguner, U.** *Multiple target tracking with Gaussian mixture PHD filter using passive acoustic Doppler-only measurements*. In: *15th International Conference on Information Fusion (FUSION)*. 2012, S. 2600–2607.
- [46] **Hammersley, J. M. und Morton, K. W.** *Poor Man's Monte Carlo*. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 16.1 (1954), S. 23–38.
- [47] **Hausdorff, Felix.** *Grundzüge der Mengenlehre*. Leipzig: Verlag von Veit & Comp., 1914.
- [48] **He, Xiaofan, Tharmarasa, R., Kirubarajan, T. und Thayaparan, T.** *A Track Quality Based Metric for Evaluating Performance of Multitarget Filters*. In: *IEEE Transactions on Aerospace and Electronic Systems* 49.1 (2013), S. 610–616.
- [49] **Helbing, D. und Molnar, P.** *Social force model for pedestrian dynamics*. In: *Physical Review* 51 (1995), S. 4282–4286.

- [50] **Hendeby, G. und Karlsson, R.** *Gaussian mixture PHD filtering with variable probability of detection*. In: *17th International Conference on Information Fusion (FUSION)*. 2014, S. 1–7.
- [51] **Hoffman, J.R. und Mahler, R.P.S.** *Multitarget miss distance via optimal assignment*. In: *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans* 34.3 (2004), S. 327–336.
- [52] **Hofmann, M., Wolf, D. und Rigoll, G.** *Hypergraphs for Joint Multi-view Reconstruction and Multi-object Tracking*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. 2013, S. 3650–3657.
- [53] **Julier, S.J.** *The scaled unscented transformation*. In: *Proceedings of the American Control Conference*. Bd. 6. 2002, 4555–4559 vol.6.
- [54] **Julier, S.J.** *The spherical simplex unscented transformation*. In: *Proceedings of the American Control Conference*. Bd. 3. 2003, 2430–2434 vol.3.
- [55] **Julier, S.J. und Uhlmann, J.K.** *Unscented filtering and nonlinear estimation*. In: *Proceedings of the IEEE* 92.3 (2004), S. 401–422.
- [56] **Kalman, R. E.** *A New Approach to Linear Filtering And Prediction Problems*. In: *ASME Journal of Basic Engineering* (1960).
- [57] **Kamal, A.T., Ding, C., Morye, A.A., Farrell, J.A. und Roy-Chowdhury, AmitK.** *An Overview of Distributed Tracking and Control in Camera Networks*. In: *Wide Area Surveillance*. Hrsg. von **Asari, Vijayan K.** Bd. 6. *Augmented Vision and Reality*. Springer Berlin Heidelberg, 2014, S. 207–234.
- [58] **Kassir, A. und Peynot, T.** *Reliable Automatic Camera-Laser Calibration*. In: *Australasian Conference on Robotics and Automation*. 2010.
- [59] **Ke, Yan und Sukthankar, R.** *PCA-SIFT: a more distinctive representation for local image descriptors*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. Bd. 2. 2004, S. 506–513.
- [60] **Keller, C.G. und Gavrila, D.M.** *Will the Pedestrian Cross? A Study on Pedestrian Path Prediction*. In: *IEEE Transactions on Intelligent Transportation Systems* 15.2 (2014), S. 494–506.

- [61] **Khaleghi, Bahador, Khamis, Alaa, Karray, Fakhreddine O. und Razavi, Saiedeh N.** *Multisensor data fusion: A review of the state-of-the-art*. In: *Information Fusion* 4 (2013), S. 28–44.
- [62] **Khan, S. und Shah, M.** *Consistent labeling of tracked objects in multiple cameras with overlapping fields of view*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25.10 (2003), S. 1355–1360.
- [63] **Kiencke, U., Schwarz, M. und Weickert, T.** *Signalverarbeitung: Zeit-Frequenz-Analyse und Schätzverfahren*. 1. Aufl. München: Oldenbourg Verlag, 2008.
- [64] **Koch, Wolfgang.** *On exploiting negative sensor evidence for target tracking and sensor data fusion*. In: *Information Fusion* 8.1 (2007). Special Issue on the Seventh International Conference on Information Fusion-Part {II}, S. 28–39.
- [65] **Kruse, Marco.** *Mehrobjekt-Zustandsschätzung mit verteilten Sensorträgern am Beispiel der Umfeldwahrnehmung im Straßenverkehr*. Diss. Karlsruhe, 2013.
- [66] **Kruse, Marco und Puente León, Fernando.** *Radargestützte Mehrobjekt-Verfolgung mit dem PHD-Filter*. In: *tm - Technisches Messen* 78.4 (2011), S. 190–195.
- [67] **Lahat, D., Adali, T. und Jutten, C.** *Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects*. In: *Proceedings of the IEEE* 103.9 (2015), S. 1449–1477.
- [68] **Laneuville, Dann.** *Polar versus Cartesian velocity models for maneuvering target tracking with IMM*. In: *IEEE Aerospace Conference*. 2013, S. 1–15.
- [69] **Laptev, Ivan.** *On Space-Time Interest Points*. In: *International Journal of Computer Vision* 64.2 (2005), S. 107–123.
- [70] **Li, Wenling, Jia, Yingmin, Du, Junping und Yu, Fashan.** *Gaussian mixture {PHD} filter for multi-sensor multi-target tracking with registration errors*. In: *Signal Processing* 93.1 (2013), S. 86–99.

- [71] **Liem, M.C. und Gavrila, D.M.** *Person appearance modeling and orientation estimation using Spherical Harmonics*. In: *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. 2013, S. 1–6.
- [72] **Lin, Daw-Tung und Huang, Kai-Yung.** *Collaborative Pedestrian Tracking and Data Fusion With Multiple Cameras*. In: *IEEE Transactions on Information Forensics and Security* 6.4 (2011), S. 1432–1444.
- [73] **Liu, Wu, Zhang, Yongdong, Tang, Sheng, Tang, Jinhui, Hong, Richang und Li, Jintao.** *Accurate Estimation of Human Body Orientation From RGB-D Sensors*. In: *IEEE Transactions on Cybernetics* 43.5 (2013), S. 1442–1452.
- [74] **Luber, M., Stork, J.A., Tipaldi, G.D. und Arras, K.O.** *People tracking with human motion predictions from social forces*. In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2010, S. 464–469.
- [75] **Luenberger, D.G.** *Observing the State of a Linear System*. In: *IEEE Transactions on Military Electronics* 8.2 (1964), S. 74–80.
- [76] *Random-set approach to data fusion*. Bd. 2234. 1994, S. 287–295.
- [77] **Mahler, R.** *Multitarget Bayes filtering via first-order multitarget moments*. In: *IEEE Transactions on Aerospace and Electronic Systems* 39.4 (2003), S. 1152–1178.
- [78] **Mahler, R.** *PHD filters of higher order in target number*. In: *IEEE Transactions on Aerospace and Electronic Systems* 43.4 (2007), S. 1523–1543.
- [79] **Mahler, R.** *Statistical Multisource-Multitarget Information Fusion*. London: Artech House, 2007.
- [80] **Mahler, R.** *Linear-complexity CPHD filters*. In: *13th International Conference on Information Fusion (FUSION)*. 2010, S. 1–8.
- [81] **Mahler, R.** *„Statistics 102“ for Multisource-Multitarget Detection and Tracking*. In: *IEEE Journal of Selected Topics in Signal Processing* 7.3 (2013), S. 376–389.
- [82] **Mahler, R., Vo, Ba-Tuong und Vo, Ba-Ngu.** *CPHD Filtering With Unknown Clutter Rate and Detection Profile*. In: *IEEE Transactions on Signal Processing* 59.8 (2011), S. 3497–3513.

- [83] **Mählich, M.** *Filtersynthese zur simultanen Minimierung von Existenz-, Assoziations- und Zustandsunsicherheiten in der Fahrzeugumfelderfassung mit heterogenen Sensordaten.* Dissertation. Universität Ulm, 2009.
- [84] **Mavrinac, Aaron und Chen, Xiang.** *Modeling Coverage in Camera Networks: A Survey.* In: *International Journal of Computer Vision* 101.1 (2013), S. 205–226.
- [85] **Maybeck, Peter S.** *Stochastic models, estimation, and control.* New York: Academic Press, 1979.
- [86] **McElhoe, Bruce A.** *An Assessment of the Navigation and Course Corrections for a Manned Flyby of Mars or Venus.* In: *IEEE Transactions on Aerospace and Electronic Systems* AES-2.4 (1966), S. 613–623.
- [87] **Mittal, Anurag und Davis, Larry S.** *A General Method for Sensor Planning in Multi-Sensor Systems: Extension to Random Occlusion.* In: *International Journal of Computer Vision* 76.1 (2008), S. 31–52.
- [88] **Moeslund, Thomas B., Hilton, Adrian, Krüger, Volker und Sigal, Leonid,** Hrsg. *Visual Analysis of Humans.* London: Springer, 2011.
- [89] **Mohan, A., Papageorgiou, C. und Poggio, T.** *Example-based object detection in images by components.* In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.4 (2001), S. 349–361.
- [90] **Montemerlo, M., Thrun, S., Koller, D. und Wegbreit, B.** *Fast-SLAM 2.0: An Improved Particle Filtering Algorithm for Simultaneous Localization and Mapping that Provably Converges.* In: *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI).* IJCAI. 2003.
- [91] **Munaro, M., Basso, F. und Menegatti, E.** *Tracking people within groups with RGB-D data.* In: *IEEE International Conference on Intelligent Robots and Systems (IROS).* 2012, S. 2101–2107.
- [92] **Munaro, Matteo und Menegatti, Emanuele.** *Fast RGB-D people tracking for service robots.* In: *Autonomous Robots* 37.3 (2014), S. 227–242.

- [93] **Munkres, James.** *Algorithms for the Assignment and Transportation Problems.* In: *Journal of the Society of Industrial and Applied Mathematics* 5 (1957), S. 32–38.
- [94] **Musicki, D. und Evans, R.** *Joint integrated probabilistic data association: JIPDA.* In: *IEEE Transactions on Aerospace and Electronic Systems* 40.3 (2004), S. 1093–1099.
- [95] **Nawaz, T., Poiesi, F. und Cavallaro, A.** *Measures of Effective Video Tracking.* In: *IEEE Transactions on Image Processing* 23.1 (2014), S. 376–388.
- [96] **Niculescu, Alexandru und Caruana, Rich.** *Obtaining Calibrated Probabilities from Boosting.* In: *Proc. 21st Conference on Uncertainty in Artificial Intelligence.* AUAI Press, 2005.
- [97] **Olfati-Saber, R.** *Kalman-Consensus Filter : Optimality, stability, and performance.* In: *Proceedings of the 48th IEEE Conference on Decision and Control.* 2009, S. 7036–7042.
- [98] **O'Rourke, S.M. und Swindlehurst, A.L.** *Limited field-of-view multimodal sensor adaptation for data association.* In: *IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM).* 2012, S. 241–244.
- [99] **Otto, Carola.** *Fusion of data from heterogeneous sensors with distributed fields of view and situation evaluation for advanced driver assistance systems.* Diss. Karlsruhe: Karlsruhe Institute of Technology, 2013.
- [100] **Otto, C., Gerber, W., Puente León, F. und Wirnitzer, J.** *A Joint Integrated Probabilistic Data Association Filter for pedestrian tracking across blind regions using monocular camera and radar.* In: *IEEE Intelligent Vehicles Symposium (IV).* 2012, S. 636–641.
- [101] **OuYang, Cheng, Ji, Hong-Bing und Tian, Ye.** *Improved Gaussian Mixture CPHD Tracker for Multitarget Tracking.* In: *IEEE Transactions on Aerospace and Electronic Systems* 49.2 (2013), S. 1177–1191.
- [102] **Panta, K., Clark, D.E. und Vo, Ba-Ngu.** *Data Association and Track Management for the Gaussian Mixture Probability Hypothesis Density Filter.* In: *IEEE Transactions on Aerospace and Electronic Systems* 45.3 (2009), S. 1003–1016.

- [103] **Panta, Kusha, Vo, Ba-Ngu, Singh, Sumeetpal und Doucet, Arnaud.** *Probability hypothesis density filter versus multiple hypothesis tracking.* In: *Defense and Security.* International Society for Optics und Photonics. 2004, S. 284–295.
- [104] **Papoulis, Athanasios.** *Probability, Random Variables and Stochastic Processes.* Bd. 3. New York: McGraw-Hill, 1991.
- [105] **Pérez Grassi, Ana, Frolov, Vadim und Puente León, Fernando.** *Information fusion to detect and classify pedestrians using invariant features.* In: *Information Fusion 12.4* (2011). Special Issue on Information Fusion for Cognitive Automobiles, S. 284–292.
- [106] **Petetin, Y., Morelande, M. und Desbouvries, F.** *Marginalized particle PHD filters for multiple object Bayesian filtering.* In: *IEEE Transactions on Aerospace and Electronic Systems* 50.2 (2014), S. 1182–1196.
- [107] **Pierard, Sebastien, Leroy, Damien, Hansen, Jean-Frederic und Van Droogenbroeck, Marc.** *Estimation of Human Orientation in Images Captured with a Range Camera.* In: *Advanced Concepts for Intelligent Vision Systems.* Hrsg. von **Blanc-Talon, Jacques, Kleihorst, Richard, Philips, Wilfried, Popescu, Dan und Scheunders, Paul.** Bd. 6915. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, S. 519–530.
- [108] **Pitt, Michael K. und Shephard, Neil.** *Filtering via Simulation: Auxiliary Particle Filters.* In: *Journal of the American Statistical Association* 94.446 (1999), S. 590–599.
- [109] **Platt, John C.** *Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods.* In: *Advances in large margin classifiers.* Cambridge: MIT Press, 1999, S. 61–74.
- [110] **Ponomareva, Ksenia und Date, Paresh.** *Higher order sigma point filter: A new heuristic for nonlinear time series filtering.* In: *Applied Mathematics and Computation* 221 (2013), S. 662–671.
- [111] **Poppe, Ronald.** *A survey on vision-based human action recognition.* In: *Image and Vision Computing* 28.6 (2010), S. 976–990.

- [112] **Possegger, H., Mauthner, T., Roth, P.M. und Bischof, H.** *Occlusion Geodesics for Online Multi-object Tracking*. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CV-PR)*. 2014, S. 1306–1313.
- [113] **Puente León, Fernando und Kiencke, Uwe.** *Messtechnik*. 9. Aufl. Berlin: Springer-Verlag, 2012.
- [114] **Puente León, Fernando, Kiencke, Uwe und Jäkel, Holger.** *Signale und Systeme*. 5. Aufl. München: Oldenbourg, 2011.
- [115] **Quigley, Morgan, Conley, Ken, Gerkey, Brian, Faust, Josh, Foote, Tully, Leibs, Jeremy, Wheeler, Rob und Ng, Andrew Y.** *ROS: an open-source Robot Operating System*. In: *ICRA workshop on open source software 3.2* (2009).
- [116] **Reid, D.B.** *An algorithm for tracking multiple targets*. In: *IEEE Transactions on Automatic Control* 24.6 (1979), S. 843–854.
- [117] **Reuter, S. und Dietmayer, K.** *Pedestrian tracking using Random Finite Sets*. In: *14th International Conference on Information Fusion (FUSION)*. 2011, S. 1–8.
- [118] **Ristic, B., Arulampalam, S. und Gordon, N.** *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston, London: Artech House, 2004.
- [119] **Ristic, Branko, Clark, Daniel und Vo, Ba-Ngu.** *Improved SMC implementation of the PHD filter*. In: *Proc. 13th Int. Conf. Information Fusion*. Edinburgh, 2010.
- [120] **Ristic, B., Vo, Ba-Ngu, Clark, D. und Vo, Ba-Tuong.** *A Metric for Performance Evaluation of Multi-Target Tracking Algorithms*. In: *IEEE Transactions on Signal Processing* 59.7 (2011), S. 3452–3457.
- [121] **Ristic, B., Vo, Ba-Tuong, Vo, Ba-Ngu und Farina, A.** *A Tutorial on Bernoulli Filters: Theory, Implementation and Applications*. In: *IEEE Transactions on Signal Processing* 61.13 (2013), S. 3406–3430.
- [122] **Ruser, Heinrich und Puente León, Fernando.** *Informationsfusion - Eine Übersicht*. In: *Technisches Messen* 74.3 (2007), S. 93–102.
- [123] **Rusu, R.B. und Cousins, S.** *3D is here: Point Cloud Library (PCL)*. In: *IEEE International Conference on Robotics and Automation (ICRA)*. 2011, S. 1–4.

- [124] **Schuhmacher, D., Vo, Ba-Tuong und Vo, Ba-Ngu.** *A Consistent Metric for Performance Evaluation of Multi-Object Filters.* In: *IEEE Transactions on Signal Processing* 56.8 (2008), S. 3447–3457.
- [125] **Schulz, Andreas, Damer, Naser, Fischer, Mika und Stiefelhaagen, Rainer.** *Combined Head Localization and Head Pose Estimation for Video-Based Advanced Driver Assistance Systems.* In: *Pattern Recognition.* Hrsg. von **Mester, Rudolf und Felsberg, Michael.** Bd. 6835. *Lecture Notes in Computer Science.* Berlin Heidelberg: Springer, 2011, S. 51–60.
- [126] **Schulz, Dirk, Burgard, Wolfram, Fox, Dieter und Cremers, Armin B.** *People Tracking with a Mobile Robot Using Sample-Based Joint Probabilistic Data Association Filters.* In: *International Journal of Robotics (IJRR)* 22.2 (2003), S. 99–116.
- [127] **Shimizu, H. und Poggio, T.** *Direction estimation of pedestrian from multiple still images.* In: *IEEE Intelligent Vehicles Symposium (IV).* 2004, S. 596–600.
- [128] **Shotton, Jamie, Sharp, Toby, Kipman, Alex, Fitzgibbon, Andrew, Finocchio, Mark, Blake, Andrew, Cook, Mat und Moore, Richard.** *Real-time Human Pose Recognition in Parts from Single Depth Images.* In: *ACM Communications* 56.1 (2013), S. 116–124.
- [129] **Sigal, Leonid, Balan, Alexandru O. und Black, Michael J.** *HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion.* In: *International Journal of Computer Vision* 87.1-2 (2010), S. 4–27.
- [130] **Simon, D.** *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* New York: Wiley, 2006.
- [131] **Song, Taek Lyul, Musicki, D. und Kim, Yong.** *Multi-target tracking with target state dependent detection.* In: *15th International Conference on Information Fusion (FUSION).* 2012, S. 324–329.
- [132] **Spinello, L. und Arras, K.O.** *People detection in RGB-D data.* In: *IEEE International Conference on Intelligent Robots and Systems (IROS).* 2011, S. 3838–3843.
- [133] **Springer, Melvin Dale.** *The Algebra of Random Variables.* New York: Wiley, 1979.

- [134] **Steele, J.M.** *Stochastic Calculus and Financial Applications*. Applications of mathematics : stochastic modelling and applied probability. Springer, 2001.
- [135] **Stiller, Christoph, Puente León, Fernando und Kruse, Marco.** *Information fusion for automotive applications - An overview*. In: *Information Fusion* 12.4 (2011). Special Issue on Information Fusion for Cognitive Automobiles, S. 244–252.
- [136] **Tao, Junli und Klette, R.** *Integrated Pedestrian and Direction Classification Using a Random Decision Forest*. In: *IEEE International Conference on Computer Vision Workshops (ICCVW)*. 2013, S. 230–237.
- [137] **Thrun, Sebastian, Burgard, Wolfram und Fox, Dieter.** *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. Cambridge: The MIT Press, 2005.
- [138] **Thrun, Sebastian, Fox, Dieter, Burgard, Wolfram und Dellaert, Frank.** *Robust Monte Carlo localization for mobile robots*. In: *Artificial Intelligence* 128.1-2 (2001), S. 99–141.
- [139] **Tosato, D., Spera, M., Cristani, M. und Murino, V.** *Characterizing Humans on Riemannian Manifolds*. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013), S. 1972–1984.
- [140] **Ulmke, M., Erdinc, O. und Willett, P.** *Gaussian mixture cardinalized PHD filter for ground moving target tracking*. In: *10th International Conference on Information Fusion (FUSION)*. 2007, S. 1–8.
- [141] **Vapnik, Vladimir.** *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics)*. Secaucus, NJ, USA: Springer-Verlag, 1982.
- [142] **Vapnik, Vladimir N.** *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag, 1995.
- [143] **Vedaldi, Andrea und Fulkerson, Brian.** *VLFeat: An open and portable library of computer vision algorithms*. In: *Proceedings of the international Conference on Multimedia*. ACM. 2010, S. 1469–1472.
- [144] **Vo, B. T.** *Random Finite Sets in Multi-Object Filtering*. Dissertation. University of Western Australia, 2008.

- [145] **Vo, Ba Tuong, Vo, Ba Ngu, Hoseinnezhad, R. und Mahler, R.P.S.** *Multi-Bernoulli filtering with unknown clutter intensity and sensor field-of-view*. In: *Conference on Information Sciences and Systems (CISS)*. 2011, S. 1–6.
- [146] **Vo, Ba-Ngu und Ma, Wing-Kin.** *The Gaussian Mixture Probability Hypothesis Density Filter*. In: *IEEE Transactions on Signal Processing* 54.11 (2006), S. 4091–4104.
- [147] **Vo, Ba-Ngu, Singh, Sumeetpal und Doucet, A.** *Sequential Monte Carlo implementation of the PHD filter for multi-target tracking*. In: *14th International Conference on Information Fusion (FUSION)*. Bd. 2. 2003, S. 792–799.
- [148] **Vo, Ba-Tuong, Vo, Ba-Ngu und Cantoni, Antonio.** *Analytic Implementations of the Cardinalized Probability Hypothesis Density Filter*. In: *IEEE Transactions on Signal Processing* 55.7 (2007), S. 3553–3567.
- [149] **Vo, Ba-Tuong, Vo, Ba-Ngu und Cantoni, Antonio.** *The Cardinality Balanced Multi-Target Multi-Bernoulli Filter and Its Implementations*. In: *IEEE Transactions on Signal Processing* 57.2 (2009), S. 409–423.
- [150] **Vo, Ba-Tuong, Vo, Ba-Ngu, Hoseinnezhad, R. und Mahler, R.P.S.** *Robust Multi-Bernoulli Filtering*. In: *IEEE Journal of Selected Topics in Signal Processing* 7.3 (2013), S. 399–409.
- [151] **Wan, E.A. und Van der Merwe, R.** *The unscented Kalman filter for nonlinear estimation*. In: *IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*. 2000, S. 153–158.
- [152] **Wang, Xiaogang.** *Intelligent multi-camera video surveillance: A review*. In: *Pattern Recognition Letters* 34.1 (2013). Extracting Semantics from Multi-Spectrum Video, S. 3–19.
- [153] **Weinland, Daniel, Ronfard, Remi und Boyer, Edmond.** *A survey of vision-based methods for action representation, segmentation and recognition*. In: *Computer Vision and Image Understanding* 115.2 (2011), S. 224–241.
- [154] **Wendel, J.** *Integrierte Navigationssysteme: Sensordatenfusion, GPS und Inertiale Navigation*. 2. überarbeitete Auflage. München: Oldenbourg Verlag, 2011.

- [155] **Wenzl, K., Ruser, H. und Kargel, C.** *Performance Evaluation of a Decentralized Multitarget-Tracking Algorithm Using a LIDAR Sensor Network With Stationary Beams.* In: *IEEE Transactions on Instrumentation and Measurement* 62.5 (2013), S. 1174–1182.
- [156] **Wood, Trevor, Clark, Daniel und Ristic, Branko.** *Efficient resampling and basic track continuity for the SMC-PHD filter.* In: *Proc. Cognitive Systems with Interactive Sensors.* Crawley, UK, 2010, S. 1–10.
- [157] **Wurm, Kai M., Stachniss, Cyrill und Grisetti, Giorgio.** *Bridging the Gap Between Feature- and Grid-based SLAM.* In: *Robot. Auton. Syst.* 58.2 (2010), S. 140–148.
- [158] **Xavier, J., Pacheco, M., Castro, D., Ruano, A. und Nunes, U.** *Fast Line, Arc/Circle and Leg Detection from Laser Scan Data in a Player Driver.* In: *IEEE International Conference on Robotics and Automation (ICRA).* 2005, S. 3930–3935.
- [159] **Yang, Xulei, Song, Qing und Cao, Aize.** *Weighted support vector machine for data classification.* In: *IEEE International Joint Conference on Neural Networks.* Bd. 2. 2005, S. 859–864.
- [160] **Yeffet, L. und Wolf, L.** *Local Trinary Patterns for human action recognition.* In: *IEEE 12th International Conference on Computer Vision.* 2009, S. 492–497.

## Eigene Veröffentlichungen

- [161] **Pallauf, J., Wagner, J. und Puente León, Fernando.** *Evaluation of State-Dependent Pedestrian Tracking Based on Finite Sets.* In: *IEEE Transactions on Instrumentation and Measurement* 64.5 (2015), S. 1284–1292.
- [162] **Pallauf, Johannes und Puente León, Fernando.** *Verteilte Verfolgung von Personen mit dem PHD-Filter unter Einsatz von Laserscannern.* In: *XXVII. Messtechnisches Symposium des Arbeitskreises der Hochschullehrer für Messtechnik e.V.* Hrsg. von **Knapp, Wolfgang und Gebhardt, Michael.** Shaker Verlag, Aachen, 2013, S. 63–74.

- [163] **Pallauf, Johannes und Puente León, Fernando.** *Richtungsabhängige Personendetektion und -verfolgung.* In: *Forum Bildverarbeitung 2014.* Hrsg. von **Puente León, Fernando und Heizmann, Michael.** Karlsruhe: KIT Scientific Publishing, 2014, S. 273–283.
- [164] **Pallauf, Johannes und Puente León, Fernando.** *State-dependent and distributed pedestrian tracking using the (C)PHD filter.* In: *IEEE International Instrumentation and Measurement Technology Conference.* Montevideo, Uruguay, 2014.
- [165] **Pallauf, Johannes und Puente León, Fernando.** *Richtungssensitive Fußgänger-Verfolgung durch Klassifikation von Bildmerkmalen.* In: *tm - Technisches Messen* 82.5 (2015), S. 242–250.
- [166] **Pallauf, Johannes, Gomes, Pedro Tiago, Brás, Susana, Cunha, João Paulo Silva und Coimbra, Miguel Tavares.** *Associating ECG features with firefighter's activities.* In: *IEEE Engineering in Medicine and Biology Society Conference.* Boston, MA: IEEE, 2011, S. 6009–6012.
- [167] **Vater, Sebastian, Pallauf, Johannes und Puente León, Fernando.** *Referenzdatenbestimmung für die 3D-Kopfposenschätzung unter Verwendung eines Motion-Capture-Systems.* In: *XXIX. Messtechnisches Symposium 2015.* Hrsg. von **Fröhlich, Thomas und Manske, Eberhard.** Berlin: De Gruyter Oldenbourg, 2015, S. 115–122.
- [168] **Ye, Can, Pallauf, Johannes, Kumar, B. V. K. Vijaya und Coimbra, Miguel Tavares.** *Customizing the training dataset to an individual for improved heartbeat recognition performance in long-term ECG signals.* In: *IEEE Engineering in Medicine and Biology Society Conference.* Boston, MA: IEEE, 2011, S. 3322–3325.

## Betreute studentische Arbeiten

- [169] **Barsch, Marwin.** *Zustandsabhängige Modellierung der Sensoreigenschaften für verteilte Sensorfusion.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2013.

- [170] **Barth, Dominik.** *Evaluierung des unverdeckten Sichtfeldes eines mobilen Roboters für die verteilte Sensorfusion.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2013.
- [171] **Bultmann, Simon.** *Nutzen einer Orientierungsschätzung für die Multi-Personen-Verfolgung.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2015.
- [172] **Finkenbein, Felix.** *Poseninvariante Erkennung von Fußgängern mittels nicht linearer Merkmalsintegration.* Masterarbeit. Karlsruher Institut für Technologie (KIT), 2015.
- [173] **Köpf, Florian.** *Fusion von Tiefen- und RGB-Daten für die Personendetektion.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2013.
- [174] **Ries, Lennart.** *Verfolgung ausgedehnter Objekte mit dem PHD-Filter.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2014.
- [175] **Steyer, Sascha.** *Lidar-basierte Objektverfolgung unter Verwendung eines Roboter-Betriebssystems ROS.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2012.
- [176] **Thomé, Sebastian.** *Tiefendaten-basierte Personendetektion.* Masterarbeit. Karlsruher Institut für Technologie (KIT), 2013.
- [177] **Wagner, Jörg.** *Evaluierung des Cardinality Balanced Multi-Target Multi-Bernoulli Filters für die Umfelderkennung in Innenräumen.* Masterarbeit. Karlsruher Institut für Technologie (KIT), 2014.
- [178] **Weiler, Martin.** *Orientierungsschätzung in RGB-D-Daten für die Personendetektion.* Bachelorarbeit. Karlsruher Institut für Technologie (KIT), 2014.





# **Forschungsberichte aus der Industriellen Informationstechnik (ISSN 2190-6629)**

**Institut für Industrielle Informationstechnik  
Karlsruher Institut für Technologie (KIT)**

---

Hrsg.: Prof. Dr.-Ing. Fernando Puente León, Prof. Dr.-Ing. habil. Klaus Dostert

Die Bände sind unter [www.ksp.kit.edu](http://www.ksp.kit.edu) als PDF frei verfügbar oder als Druckausgabe bestellbar.

- Band 1      Pérez Grassi, Ana  
**Variable illumination and invariant features for detecting and classifying varnish defects.** (2010)  
ISBN 978-3-86644-537-6
- Band 2      Christ, Konrad  
**Kalibrierung von Magnet-Injektoren für Benzin-Direkteinspritzsysteme mittels Körperschall.** (2011)  
ISBN 978-3-86644-718-9
- Band 3      Sandmair, Andreas  
**Konzepte zur Trennung von Sprachsignalen in unterbestimmten Szenarien.** (2011)  
ISBN 978-3-86644-744-8
- Band 4      Bauer, Michael  
**Vergleich von Mehrträger-Übertragungsverfahren und Entwurfskriterien für neuartige Powerline-Kommunikationssysteme zur Realisierung von Smart Grids.** (2012)  
ISBN 978-3-86644-779-0
- Band 5      Kruse, Marco  
**Mehrobjekt-Zustandsschätzung mit verteilten Sensorträgern am Beispiel der Umfeldwahrnehmung im Straßenverkehr** (2013)  
ISBN 978-3-86644-982-4
- Band 6      Dudeck, Sven  
**Kamerabasierte In-situ-Überwachung gepulster Laserschweißprozesse** (2013)  
ISBN 978-3-7315-0019-3
- Band 7      Liu, Wenqing  
**Emulation of Narrowband Powerline Data Transmission Channels and Evaluation of PLC Systems** (2013)  
ISBN 978-3-7315-0071-1

Hrsg.: Prof. Dr.-Ing. Fernando Puente León, Prof. Dr.-Ing. habil. Klaus Dostert

- Band 8      Otto, Carola  
**Fusion of Data from Heterogeneous Sensors with Distributed Fields of View and Situation Evaluation for Advanced Driver Assistance Systems.** (2013)  
ISBN 978-3-7315-0073-5
- Band 9      Wang, Limeng  
**Image Analysis and Evaluation of Cylinder Bore Surfaces in Micrographs.** (2014)  
ISBN 978-3-7315-0239-5
- Band 10     Michelsburg, Matthias  
**Materialklassifikation in optischen Inspektionssystemen mithilfe hyperspektraler Daten.** (2014)  
ISBN 978-3-7315-0273-9
- Band 11     Pallauf, Johannes  
**Objektsensitive Verfolgung und Klassifikation von Fußgängern mit verteilten Multi-Sensor-Trägern.** (2016)  
ISBN 978-3-7315-0529-7



ISSN 2190-6629  
ISBN 978-3-7315-0529-7

ISBN 978-3-7315-0529-7



9 783731 505297 >