



Yosef Dalbah

Omnidirektionale 3D-Umfeldmodel- lierung für Park- und Rangierfunktionen mit Time-of-Flight- Kameras



Audi-Dissertationsreihe, Band 108





Omnidirektionale 3D-Umfeldmodellierung für Park- und Rangierfunktionen mit Time-of-Flight-Kameras

Von der Carl-Friedrich-Gauß-Fakultät der Technischen Universität
Carolo-Wilhelmina zu Braunschweig zur Erlangung des Grades eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte

Dissertation

von
Dipl.-Inf. Yosef Dalbah
aus Essen

Eingereicht am: 06.01.2015
Disputation am: 24.04.2015
1. Referent: Prof. Dr.-Ing. Friedrich M. Wahl
2. Referent: Prof. Dr.-Ing. Bodo Rosenhahn



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

1. Aufl. - Göttingen: Cuvillier, 2015

Zugl.: (TU) Braunschweig, Univ., Diss., 2015

© CUVILLIER VERLAG, Göttingen 2015

Nonnenstieg 8, 37075 Göttingen

Telefon: 0551-54724-0

Telefax: 0551-54724-21

www.cuvillier.de

Alle Rechte vorbehalten. Ohne ausdrückliche Genehmigung des Verlages ist es nicht gestattet, das Buch oder Teile daraus auf fotomechanischem Weg (Fotokopie, Mikrokopie) zu vervielfältigen.

1. Auflage, 2015

Gedruckt auf umweltfreundlichem, säurefreiem Papier aus nachhaltiger Forstwirtschaft

ISBN 978-3-7369-9012-8

eISBN 978-3-7369-8012-9



Danksagung

Die vorliegende Arbeit entstand während meiner Zeit als Doktorand in der Vorentwicklung für Fahrerassistenzsysteme bei der Audi Electronics Venture GmbH und dem Institut für Robotik und Prozessinformatik an der Technischen Universität Braunschweig. An erster Stelle danke ich meinem Doktorvater Prof. Dr.-Ing. Friedrich M. Wahl für die Betreuung dieser Arbeit. Seine wissenschaftlichen Anregungen und wertvollen Denkanstöße haben sehr zum Gelingen dieser Arbeit beigetragen. Auch Herrn Prof. Dr.-Ing. Bodo Rosenhahn danke ich für die Übernahme des Zweitgutachtens.

Auf Seiten der Audi Electronics Venture GmbH danke ich meinen Abteilungsleitern Andreas Reich und Dr. Miklós Kiss für die notwendigen Freiräume zur Erstellung der Dissertation. Mein Dank gilt auch meinen Arbeitskollegen der Abteilung Vorentwicklung für Fahrerassistenzsysteme und insbesondere Dr.-Ing. Martin Roehder und Dr. rer. nat. Patrick Heinemann für die Betreuung der Arbeit und für viele intensive und unterstützende fachliche Gespräche.

Auch die angenehme und erfolgreiche Zusammenarbeit im Projekt mit Matthias Ewert, Martin Hempel, Nicolas Dingeldey und Michael Schlittenbauer hat zum Erfolg dieser Arbeit beigetragen. Insbesondere der stetige fachliche Austausch mit Herrn Dingeldey hat einen großen Anteil am Gelingen dieser Arbeit. Danke, Nico!

Ebenfalls bedanken möchte ich mich bei meinen ehemaligen Studenten, die ich bei Abschlussarbeiten oder Praktika im Unternehmen betreuen durfte. Bedanken möchte ich mich bei Mert Assoy, Dirk Koltermann, Stephan Rohr, Zifan Li, Alexander Bauer und Nick Schneider.

Zudem bedanke ich mich bei meiner Familie für die kompromisslose Unterstützung trotz des weiten Weges nach Ingolstadt.





Inhaltsverzeichnis

Danksagung	I
Kurzfassung	V
Abstract	VII
1 Einleitung	1
1.1 Motivation	1
1.2 Heutige Fahrerassistenzsysteme für das Parken und Rangieren	2
1.3 Zukünftige Fahrerassistenzsysteme für das Parken und Rangieren	5
1.4 Anforderungen an die Umfeldwahrnehmung	7
1.5 Sensormessprinzipien zur Umfeldwahrnehmung	9
1.6 Vergleich der Sensoren	15
1.7 Zielsetzung	18
2 Stand der Technik und Methode	19
2.1 PMD-Kamera	19
2.2 Datenverarbeitung von Time-of-Flight-Kameras	36
2.3 Diskussion des Standes der Technik	48
2.4 Ableitung der Methode zur 3D-Umfeldmodellierung	51
3 Systemauslegung	55
3.1 Weitwinkel-PMD	55
3.2 Integration in Versuchsträger	58
3.3 Beleuchtung	60
3.4 Fehlereffekte der Tiefenmessung	70
3.5 Kalibrierung	75
3.6 Fusion der Sensordaten	83



4	Vorverarbeitung	87
4.1	Dynamische Sensoransteuerung	87
4.2	Bildglättung	89
4.3	Konfidenzberechnung	90
5	3D-Umfeldmodellierung	101
5.1	Datenstruktur	101
5.2	Integration von PMD-Messdaten	102
5.3	3D- <i>Occupancy-Grid</i>	107
5.4	Visualisierung	109
5.5	Informationsextraktion	110
6	3D-Bildverarbeitung	113
6.1	Flächenextraktion	113
6.2	Lokalisierung	121
6.3	Detektion dynamischer Objekte	141
7	Evaluierung	153
7.1	Versuchsaufbau	153
7.2	Referenzsystem	153
7.3	Evaluationskriterien	159
7.4	Ergebnisse	161
7.5	Diskussion	171
7.6	Eignung für zukünftige Fahrerassistenzfunktionen	173
8	Zusammenfassung	177
8.1	Fazit	177
8.2	Ausblick	178



Kurzfassung

Zukünftige Fahrerassistenzsysteme erhöhen die Anforderungen an die Umfeldwahrnehmung von Automobilen. Insbesondere automatisierte Fahrfunktionen sind auf hochgenaue und umfassende Umfeldinformationen angewiesen, um sichere Entscheidungen treffen zu können. Die sensorische Abtastung des Umfelds muss jederzeit möglich sein und das gesamte Fahrzeugumfeld abdecken. Um die Umgebung eines Fahrzeugs insbesondere im niedrigen Geschwindigkeitsbereich des Parkens und Rangierens detailliert abzumessen, sind die heute eingesetzten Sensoren allerdings nicht genügend präzise und bieten keine lückenlose Abdeckung.

Moderne Time-of-Flight-Kameras bieten großes Potenzial für eine hochgenaue und dichte sensorische Erfassung des nahen Fahrzeugumfelds. Time-of-Flight Kameras basieren auf der Messung der Lichtlaufzeit von modulierten Infrarotlichtpulsen und liefern dichte Tiefenbilder in hohen Frameraten.

In dieser Arbeit wird ein Konzept für eine 360°-Wahrnehmung im Nahbereich auf Basis von Weitwinkel-Time-of-Flight-Kameras erarbeitet. Durch eine intelligente Datenverarbeitung der Kameradaten kann die Fahrzeugumgebung in Echtzeit dreidimensional rekonstruiert werden. Schwerpunkte der Arbeit liegen zum einen auf der Systemauslegung der Time-of-Flight-Kameras, welche einen bisher unerreichten Sichtbereich von 180° pro Kamera ermöglicht und somit bei der Verwendung von vier Kameras zu einer vollständigen sensorischen Abdeckung des Fahrzeugumfelds führt. Zum anderen werden neuartige Verfahren zur Verarbeitung der Kameradaten präsentiert, welche von den Rohdaten der Kamera bis hin zur dreidimensionalen Rekonstruktion des Umfelds führen. Unter anderem werden neuartige Methoden für die Kalibrierung der Tiefenbilder und die Erstellung eines Konfidenzwertes für jedes Pixel vorgestellt. Indem 2D-Bildverarbeitung und 3D-Datenverarbeitung miteinander kombiniert werden, lassen sich Methoden zur Eigenlokalisierung und zur Erkennung von dynamischen Objekten entwickeln. Schließlich werden die verarbeiteten



Kurzfassung

3D-Punktclouds der Kameras in einem gemeinsamen Umfeldmodell akkumuliert und daraus ein 3D-Drahtgittermodell generiert. Dieses Modell kann anschließend genutzt werden, um Informationen über Hindernisse in der Umgebung zu gewinnen oder um dem Fahrer eine virtuelle Ansicht der Fahrzeugumgebung zu präsentieren.

Die Evaluierung des Systems erfolgt auf der Basis eines Referenzsystems. Die Fahrzeugumgebung wird hochgenau vermessen und die Fahrzeugpose wird fortlaufend durch externe Sensorik bestimmt. Auf diese Weise wird das System durch die Referenzierung sowohl der unverarbeiteten Sensordaten als auch des entstehenden Umfeldmodells bewertet. Die Evaluierung zeigt, dass auf Basis der Weitwinkel-Time-of-Flight-Kameras eine Verbesserung der Umfeldwahrnehmung beim Parken und Rangieren erreicht werden kann.



Abstract

Future Advanced Driver Assistance Systems (ADAS) increase the necessity of the environment perception of automobiles. In particular, automated driving functionalities require highly accurate and comprehensive information about the environment to take safe decisions. At any given time scanning of the surrounding area must be enabled and needs to cover the whole area around the vehicle. Today's sensors do not cover the whole surrounding of the vehicle and are not precise enough to facilitate exact measuring of the vehicle's environment while parking and manoeuvring at slow speed.

Modern time-of-flight cameras offer great potential for precise and dense distance sensing of the close surrounding area of the vehicle. Based on the measurement of the propagation time of modulated near infrared light pulses, time-of-flight cameras deliver dense depth images with high frame-rates.

The present work presents a concept for a 360° perception of the near surroundings based on wide-angle time-of-flight cameras. The environment of the vehicle can be reconstructed three-dimensionally and in real time by intelligent processing of camera data. On the one hand, this work focuses on the system design of time-of-flight cameras that achieve an unprecedented visual field of 180° per camera, thereby enabling a full coverage of the surrounding area of the vehicle. On the other hand, the work presents new methods of processing camera data that range from raw data to three-dimensional reconstruction of the environment. New methods for the calibration of depth images and for the computation of per-pixel confidence values are introduced. The combination of two-dimensional image processing and three-dimensional data processing enables new methods of self-localization and detection of dynamic objects. The 3D point clouds are finally accumulated into a unified environment model and a 3D mesh is generated. The 3D environment model can be used to extract informa-



Abstract

tion about obstacles or to display a three-dimensional virtual view of the surroundings for the driver.

The evaluation of the system is based on a reference system. The vehicle environment is measured very accurately and the position of the vehicle is tracked permanently by external sensors. Thus, the system can be evaluated by referencing both the raw sensor data and the resulting environment model. The evaluation shows that wide-angle time-of-flight cameras lead to an improvement in vehicle perception necessary for parking and manoeuvring.



1 Einleitung

1.1 Motivation

Der Großteil heutiger Fahrerassistenzsysteme basiert auf der Wahrnehmung der Umgebung eines Fahrzeugs. Zur Unterstützung beim Parken kommen in erster Linie Ultraschallsensoren und Kamerasysteme zum Einsatz. Dabei vermessen ultraschallbasierte Einparkhilfen die Umgebung und geben dem Fahrer akustische oder visuelle Signale über die Abstände zu Hindernissen. Semiautomatische Einparkhilfen übernehmen gar die Querführung des Fahrzeugs, um in eine Parklücke zu rangieren. Der Fahrer muss lediglich das Gas- und Bremspedal betätigen. Kamerabasierte Systeme geben dem Fahrer eine Sicht auf die verdeckte Umgebung. Oft werden zur Unterstützung des Rangierens Hilfslinien in die Darstellung der Kamerabilder eingefügt oder auch mehrere Kamerabilder zu einer gemeinsamen Anzeige fusioniert.

Bei all diesen Systemen bleibt der Fahrer in der Verantwortung und muss jederzeit die Umgebung des Fahrzeugs überwachen. Dies liegt darin begründet, dass die verwendeten Sensoren keine vollständige Abdeckung des Fahrzeugumfelds bieten. Zum einen werden die Fahrzeugflanken nur unzureichend durch Sensoren abgedeckt und zum anderen wird nicht die gesamte Fahrzeughöhe sensorisch abgedeckt, sodass das Fahrzeug blind ist für herausragende Objekte. Es findet keine Vermessung der Höhe von Objekten statt.

Zukünftige Fahrerassistenzsysteme und insbesondere das automatisierte Fahren erhöhen die Anforderungen an die Umfeldwahrnehmung. Assistenzsysteme für das Parken und Rangieren werden in Zukunft stärker in das Fahrgeschehen eingreifen und dem Fahrer zunehmend Aufgaben abnehmen. Automatisierte Fahrsysteme übernehmen die Fahraufgabe sogar vollständig. Da die Verantwortung in derartigen Systemen nicht mehr bei einem Fahrer liegen kann, wird eine vollständige sensorische Absicherung der Umgebung benötigt.



1.2 Heutige Fahrerassistenzsysteme für das Parken und Rangieren

Das Ziel von heutigen Fahrerassistenzsystemen für das Parken und Rangieren besteht darin, den Fahrer in unübersichtlichen Situationen zu unterstützen. Sie kommen dann zum Einsatz, wenn der Fahrer keinen vollständigen Überblick über die Umgebung des Fahrzeugs hat. Im Folgenden werden verschiedene Ausprägungen von Fahrerassistenzsystemen für das Parken und Rangieren vorgestellt, die derzeit in Serienfahrzeugen verbaut werden.

1.2.1 Frühe Parkhilfen

Die ersten Assistenzsysteme zur Unterstützung beim Parken und Rangieren waren sogenannte Peilstäbe und Markierungen an Fahrzeugen. Diese werden als passive Systeme bezeichnet [1] und helfen dem Fahrer, die eigenen Fahrzeugabmessungen in Relation zu den Objekten in der Umgebung besser abzuschätzen. Beispielsweise bot die *S-Klasse* von *Mercedes Benz* ab 1991 einen Peilstab an, der beim Einlegen des Rückwärtsganges automatisch ausfährt. Auch den in den 1960er Jahren üblichen Heckflossen wird die Funktion zugeschrieben, dass sie dem Fahrer einen Überblick über die Proportionen des eigenen Fahrzeugs geben.

1.2.2 Messende Systeme

Messende Systeme zur Parkunterstützung vermessen die Umgebung mittels Sensoren und informieren den Fahrer über die Abstände zu Hindernissen. Die häufigsten Systeme zur Einparkunterstützung basieren auf Ultraschallsensoren [2]. Ultraschallsensoren werden in Kapitel 1.5.1 thematisiert. Abstände zu Hindernissen werden dem Fahrer in der Regel akustisch, zusätzlich oft auch optisch, vermittelt. Als Beispiel ist in Abbildung 1.1 die 8-Segment-Anzeige des *Audi A5* aus dem Baujahr 2007 dargestellt.



Abbildung 1.1: 8-Segment-Anzeige

1.2.3 Rückfahrkamera

Als Rückfahrkamera wird eine am Heck des Fahrzeugs angebrachte Kamera bezeichnet, die den hinteren Bereich des Fahrzeugs erfasst. Das Kamerabild wird dem Fahrer zur Unterstützung beim Parken und Rangieren angezeigt. Aktiviert wird die Anzeige in der Regel durch das Einlegen des Rückwärtsganges. Um dem Fahrer die Orientierung zu erleichtern, wird das Kamerabild gespiegelt wiedergegeben. Da die gesamte Breite des Fahrzeugs dargestellt werden soll, werden Weitwinkelobjektive mit einem großen Öffnungswinkel eingesetzt. Da diese stark verzerrte Aufnahmen liefern, wird dem Fahrer ein nachträglich entzerrtes Bild angezeigt. Im Display können zusätzlich Hilfslinien abgebildet werden, welche dem Fahrer eine Orientierung im Bild ermöglichen. Die Anzeige der Rückfahrkamera eines *Audi A6* aus dem Baujahr 2012 ist in Abbildung 1.2 dargestellt.

1.2.4 Umgebungskameras

Bei dem sogenannten *Surround-View-System* wird ein virtuelles Bild aus den Daten mehrerer Kameras berechnet. Dieses System wird bei Audi *Top View*, bei Nissan *Around View* und bei Volkswagen *Area View* genannt. Kameras mit Weitwinkelobjektiven erfassen das Umfeld des Fahrzeugs und werden zu einer virtuellen Darstellung des Fahrzeugs und der Umgebung aus der Vogelperspektive fusioniert. Für die Berechnung dieses Bildes



Abbildung 1.2: Rückfahrkamera eines *Audi A6*

ist eine genaue Kalibrierung der Kameras notwendig. Die Bildpunkte der einzelnen Kameras werden hierzu bei bekannten optischen Abbildungseigenschaften auf die Bodenebene projiziert und in einem gemeinsamen Bild zusammengefügt. Eine beispielhafte Anzeige des *Top-View*-Systems eines *Audi A8* aus dem Baujahr 2013 zeigt Abbildung 1.3. In diesem System wird das Kamerabild aus vier Weitwinkel-Kameras zusammengefügt, die sich an der Front, am Heck und an den beiden Außenspiegeln befinden.

1.2.5 Semiautomatische Einparksysteme

Semiautomatische Einparksysteme unterstützen den Fahrer durch aktive Lenkeingriffe während des Einparkvorgangs. Zunächst wird beim Vorbeifahren an einer Parklücke der Freiraum vermessen. Dies kann beispielsweise durch Radarsensoren geschehen [3]. Wenn die vermessene Parklücke groß genug ist, um das Fahrzeug hineinzumanövrieren, wird dies dem Fahrer angezeigt. Durch Einlegen des Rückwärtsganges wird das System aktiviert und das Fahrzeug in die Parklücke geführt. Die Umgebung wird während des Parkvorgangs durch Ultraschallsensoren erfasst. Eine Trajektorienplanung berechnet den optimalen Weg, um das Fahrzeug in die Parklücke hineinzumanövrieren. Um der berechneten Trajektorie zu folgen, werden Informationen über die Bewegung des Fahrzeugs berücksichtigt. Während des Einparkvorgangs erhält der Fahrer Anweisungen zum



Abbildung 1.3: *Top-View*-Darstellung eines *Audi A8*

Einlegen des Vorwärts- und Rückwärtsganges sowie zum Gas geben. Das Fahrzeug lenkt selbstständig. Es erfolgt eine automatische Querführung. Dennoch muss der Fahrer weiterhin das Umfeld überwachen und bleibt somit in der Verantwortung. Die eingesetzte Sensorik liefert keine zentimetergenaue Vermessung der Umgebung. Daher muss beim automatischen Rangieren ein Sicherheitsbereich eingehalten werden, sodass die Systeme nicht in enge Parklücken rangieren können, obwohl es durch einen geübten Fahrer möglich wäre. Semiautomatische Systeme befinden sich bei verschiedenen Herstellern als optionale Sonderausstattung im Angebot. Bei dem Modell *Audi Q3* wird das System beispielsweise als *Parkassistent* bezeichnet [3].

1.3 Zukünftige Fahrerassistenzsysteme für das Parken und Rangieren

Die zukünftigen Fahrerassistenzsysteme für das Parken und Rangieren lassen sich unterteilen in Systeme zur Assistenz des Fahrers beim selbstständigen Fahren und in Systeme zum autonomen Fahren.



1.3.1 Park- und Rangierassistentenz

Bei Systemen zur Park- und Rangierassistentenz wird der Fahrer während seiner Fahraufgabe unterstützt, und die Verantwortung für das Fahren liegt beim Fahrer. Das Fahrzeug kann hierbei allerdings Teilaufgaben übernehmen und so dem Fahrer die Fahraufgabe erleichtern. Beispielsweise bietet es sich an, dass die heutigen semiautomatischen Einparksysteme derart erweitert werden, dass das Fahrzeug die Längsführung übernimmt, während der Fahrer sich ausschließlich der Überwachung der Umgebung des Fahrzeugs widmet [4]. Eine weitere Ausprägung zukünftiger Assistenzsysteme ist die Verbesserung der Darstellung der Umgebungsinformationen. Das in Abschnitt 1.2.4 vorgestellte *Surround-View*-System zur Anzeige der Fahrzeugumgebung leidet unter fehlenden Höheninformationen, sodass Objekte, welche die Fahrbahn überragen, verzerrt dargestellt werden. Im Beispiel aus Abbildung 1.3 kann beobachtet werden, dass das benachbarte Fahrzeug stark verzerrt wird. Durch eine genauere Umgebungswahrnehmung im dreidimensionalen Bereich können Ansichten generiert werden, die realistischere Darstellungen ermöglichen. Eine Verbesserung der Umfeldwahrnehmung kann zudem zu genaueren Abstandswarnsystemen führen. Auch sind Systeme denkbar, die automatisch auf Hindernisse durch Notbremsungen reagieren. Notwendig für solche Systeme ist eine umfassendere Umfeldwahrnehmung, die den Nahbereich des Fahrzeugs in einer hohen Genauigkeit vermessen kann.

1.3.2 Automatisches Parken und Rangieren

Bei automatischen Parksystemen fährt das Fahrzeug selbstständig in die Parklücke. Beispiele hierfür sind die von der *Audi AG* vorgestellten prototypischen Systeme *Audi Parkpilot* und *Audi Garagenpilot* [5]. Bei diesen Systemen muss sich der Fahrer nicht mehr im Fahrzeug befinden und überwacht die Fahrzeugumgebung von außerhalb. Er kann das System jederzeit bei Gefahr unterbrechen. Durch die visuelle Überwachung kann die Verantwortung weiterhin beim Fahrer bleiben. Auch *Volkswagen* [6], *BMW* [7],[8] und *Renault* [9] forschen an ähnlichen Systemen zum automatischen Einparken.

Für vollständig autonome Parksysteme muss eine vollständige Absicherung der Fahrzeugumgebung gewährleistet werden. Um dies zu ermög-



lichen, ist eine vollständige sensorische Erfassung des Fahrzeugumfelds notwendig. Erst dann kann die Verantwortung für den Einparkvorgang vom Fahrzeug übernommen werden. Auf diese Weise könnte der Fahrer auch in unübersichtlichen Rangiersituationen jederzeit die Kontrolle an das Fahrzeug übergeben.

1.4 Anforderungen an die Umfeldwahrnehmung

Im vorhergehenden Abschnitt 1.3 wurden Perspektiven zukünftiger Fahrerassistenzsysteme für das Parken und Rangieren vorgestellt. Aus diesen Systemen ergeben sich neue Anforderungen an die Umfeldwahrnehmung, die nachfolgend aufgelistet werden.

360°-Abdeckung

Insbesondere autonome Fahrfunktionen benötigen eine vollständige Absicherung der Fahrzeugumgebung. Das Fahrzeug muss das Umfeld über den gesamten Fahrzeugumfang kennen. Hierzu wird eine lückenlose sensorische Abdeckung benötigt. Insbesondere muss eine sensorische Abdeckung auch in vertikaler Richtung gegeben sein.

Reichweite

Für Park- und Rangierfunktionen werden im Vergleich zu anderen Fahrerassistenzsystemen nur geringe Reichweiten für die sensorische Erfassung benötigt. Beim Parken bewegt sich das Fahrzeug in einer niedrigen Geschwindigkeit. Dennoch muss eine Reaktion auf Objekte möglich sein. Daher wird eine sensorische Erfassung mit einer Reichweite von etwa 3 m benötigt. Die Reichweite der Sensorik kann im Seitenbereich geringer sein.

Robustheit gegenüber Störeffekten

Die Sensorik eines Fahrzeugs muss auch unter verschiedenen Witterungsbedingungen ausreichend genau messen. Aufgrund der notwendigen Abdeckung in alle Richtungen muss die Sensorik auch bei Dunkelheit funktionie-



1 Einleitung

ren, da im Regelfall nur der Bereich vor dem Fahrzeug durch Scheinwerfer beleuchtet wird.

Absolute Messgenauigkeit

Für zentimetergenaues Rangieren wird eine entsprechende Genauigkeit der Abstandssensorik benötigt. Die notwendige Genauigkeit hängt von der Distanz eines Hindernisses zum Fahrzeug ab. In unmittelbarer Nähe zum Fahrzeug soll die Genauigkeit wenige Zentimeter betragen.

Hochgenaue 3D-Umfeldmodellierung

Die Daten der Sensorik sollen eine hochgenaue 3D-Umfeldmodellierung ermöglichen. Eine dreidimensionale Modellierung wird insbesondere für die nächste Generation der in Kapitel 1.2.4 präsentierten *Surround-View-Systeme* benötigt. Auch für die Messung von Über- und Unterfahrbarkeit von Objekten wird eine hochgenaue 3D-Modellierung benötigt.

Verbesserung der Lokalisierung durch Sensordaten

Die Fahrzeugodometrie auf Basis von Raddrehzahl und Lenkwinkelsensoren liefert bei niedriger Geschwindigkeit und hohen Lenkwinkelschlägen nur eine unzureichende Genauigkeit. Derartige Szenarien sind allerdings vor allem bei Einpark- und Rangiervorgängen relevant. Daher soll eine genauere Lokalisierung ermöglicht werden, indem die Sensordaten der eingesetzten Sensorik berücksichtigt werden. Für eine hochgenaue Umfeldmodellierung wird eine ebenso hochgenaue Lokalisierung benötigt. Nur so kann bei der Erstellung des Umfeldmodells eine korrekte Datenakkumulation während einer Bewegung des Fahrzeugs ermöglicht werden. Anhand der Sensordaten soll eine Bewegungsschätzung erfolgen, welche bei Bedarf mit der vorhandenen Fahrzeugodometrie fusioniert werden kann.

Erkennung dynamischer Objekte

Dynamische Objekte sollten nicht in die 3D-Modellierung der statischen Umgebung einfließen. Sie müssen daher erkannt und von der Modellierung des statischen Umfeldmodells ausgeschlossen werden. Dennoch ist es



1.5 Sensormessprinzipien zur Umfeldwahrnehmung

wichtig, dass dynamische Objekte verfolgt werden, um auf mobile Hindernisse reagieren zu können. Beispielsweise sollte ein autonomes Fahrzeug einen Sicherheitsabstand zu Fußgängern halten, während es an statischen Hindernissen wie einem parkenden Fahrzeug bis auf wenige Zentimeter vorbeifahren kann.

Bestimmung der Abstände zu Hindernissen

Die Abstände zu Hindernissen in der Umgebung sollten aus einem Umfeldmodell extrahiert werden können. Durch die Extraktion von Abstandsinformationen können dem Fahrer Informationen gegeben werden oder in einem automatischen Fahrsystem passende Trajektorien berechnet werden.

Hohe Messrate

Ein autonomes Fahrzeug muss auf Hindernisse in der Umgebung reagieren. Dies muss auch in einer dynamischen Umgebung möglich sein. Wenn sich die Umwelt verändert, dann verändert sich das vom Fahrzeug berechnete Umgebungsmodell erst durch neue Messungen der Sensoren. Daher benötigen die Sensoren eine genügend hohe Messrate, um auch Veränderungen in der Umgebung rechtzeitig zu erfassen.

Messung auch im Stand möglich

In enge Parklücken geparkte Fahrzeug haben oft kurze Abstände zu benachbarten Fahrzeugen und anderen Hindernissen. Ein System zur Unterstützung beim Ausparken benötigt auch im Stand ohne erfolgter Fahrzeugbewegung Kenntnis über die Umgebung. Daher muss die in solchen Systemen eingesetzte Sensorik die Fähigkeit besitzen auch im Stand messen zu können.

1.5 Sensormessprinzipien zur Umfeldwahrnehmung

Im Folgenden werden verschiedene Sensormessprinzipien vorgestellt, die für die Umfeldwahrnehmung von Automobilen eingesetzt werden, um Fahrerassistenzsystemen zu realisieren.



1.5.1 Ultraschallsensoren

Ultraschallsensoren bilden die Grundlage für die heutigen Einparkhilfen [2]. Üblicherweise werden sie in die hinteren und vorderen Stoßfänger integriert, um die Abstände zu Hindernissen zu bestimmen. Bei der Nutzung mehrerer Sensoren ergibt sich ein großer Erfassungswinkel, der die gesamte Fahrzeugbreite umfasst. Das Messprinzip basiert auf der Aussendung eines Ultraschallimpulses und der Auswertung des durch eine Reflexion entstandenen Echos. Im Automobilbereich sind Schallfrequenzen zwischen 40 und 50 kHz üblich [10]. Der Sensor detektiert das Zeitintervall zwischen Aussendung und Empfang eines Signals. Der Abstand zum Hindernis ergibt sich aus der Laufzeit des Signals. Ultranahbereichssensoren in der Ultraschalltechnik besitzen eine Reichweite von etwa 2,5 m [10]. Ohne Weiteres kann die Richtung des reflektierten Echos und damit die Position eines Hindernisses nicht bestimmt werden. Dieses Problem kann allerdings durch die Verwendung von mehreren Sensoren und einer Auswertung des sogenannten Kreuzechos behoben werden. Hierzu fungiert ein Sensor zugleich als Sender und Empfänger und ein weiterer Sensor ausschließlich als Empfänger des ausgesendeten Signals [1]. Dieser Sachverhalt ist in Abbildung 1.4 dargestellt. Beide Sensoren bestimmen nun die gegebenenfalls unterschiedlichen Distanzen d_1 und d_2 . Die Position der Reflexion und damit der reale Abstand zum Stoßfänger d_r kann anschließend durch Triangulation bestimmt werden.

Die Sende- und Empfangseinheit muss nach der Aussendung eines Signals zunächst zur Ruhe kommen, bevor eine Reflexion empfangen werden kann. Aus diesem Grund können reflektierte Signale unterhalb eines gewissen Mindestabstandes nicht gemessen werden.

1.5.2 Radar-Systeme

Radar-Sensoren werden in erster Linie für Abstandsregeltempomaten und für Systeme zur Kollisionswarnung eingesetzt [2]. Eine elektromagnetische Welle wird über eine oder mehrere Antennen abgestrahlt. Diese wird an Objekten reflektiert und gelangt über eine oder mehrere Empfangsantennen zurück zum Sensor. Durch die Auswertung des Empfangssignals können Distanz und Geschwindigkeit des reflektierenden Objektes sowie

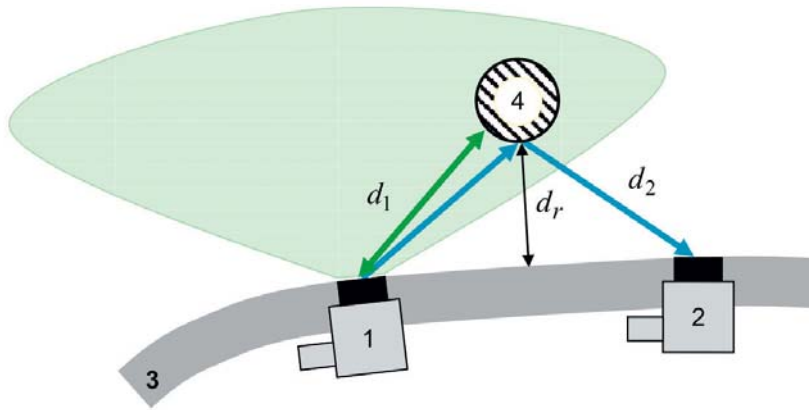


Abbildung 1.4: Abstandsmessung durch Auswertung des Kreuzechos [1]. Sensor 1 fungiert als Sender und Empfänger während Sensor 2 ausschließlich Signale empfängt. Durch Triangulation kann die Distanz. Beide Sensoren bestimmen gegebenenfalls unterschiedliche Distanzen d_1 und d_2 . Der Abstand d_r kann anschließend durch Triangulation bestimmt werden.

der Winkel, unter dem die Reflexion detektiert wurde, ermittelt werden [11].

Im Automobilbereich werden die Frequenzbereiche 24-24,25 GHz und 76-77 GHz verwendet [1]. Zur Bestimmung der Distanz kommen verschiedenen Verfahren zum Einsatz. Beim Pulsradar werden kurze Wellenpakete gesendet und ähnlich wie beim Ultraschallverfahren die Signallaufzeit gemessen. Beim Dauerstrichradar hingegen wird ein kontinuierliches Signal mit wechselnder Frequenz ausgesendet. Aus dem Vergleich des Empfangssignals mit dem Sendesignal kann auf die Distanz der Reflexion geschlossen werden. Objektgeschwindigkeiten können durch die Berücksichtigung des Dopplereffekts bestimmt werden. Um die Richtung, aus der das Signal reflektiert wurde, zu bestimmen, existieren verschiedene Verfahren. Eine Möglichkeit ist das sogenannte *Scanning*. Hier wird ein Signal mechanisch in verschiedene Richtungen gelenkt, um einen größeren Winkelbereich abzudecken. Die Amplitude des reflektierten Signals ist dann am höchsten, wenn genau in die Richtung des Objekts gestrahlt wird. Eine weitere Möglichkeit ist die parallele Aussendung und der parallele Empfang mehrerer Radarkeulen. Eine Auswertung des Verhältnisses der Amplituden lässt anschließend auf den Sichtwinkel schließen.



1 Einleitung

Moderne Radarsensoren können mehrere Ziele gleichzeitig erfassen. Im Automobil eingesetzte Sensoren bieten üblicherweise Reichweiten von bis zu 250 m [2].

1.5.3 Lidar

Lidar (*Light Detection and Ranging*) ist ein optisches Messverfahren zur Distanzbestimmung von Objekten. Im Automobilbereich werden in der Regel Laserpulse im nicht sichtbaren Infrarotbereich in die Umgebung abgestrahlt [2]. Ein Teil des abgestrahlten Lichts wird von Objekten reflektiert, gelangt zurück zum Sensor und wird von einer Fotodiode empfangen. Durch die Messung der Lichtlaufzeit kann anschließend auf die Distanz des Objektes geschlossen werden. Der Lichtstrahl ist in der Regel stark gebündelt, um eine hohe Leuchtdichte und damit eine hohe Reichweite zu erhalten. Um nun mehrere Punkte in der Umgebung zu vermessen, muss der Lichtstrahl in verschiedene Richtungen gelenkt werden. Dies kann beispielsweise durch einen rotierenden Spiegel herbeigeführt werden. Hierbei werden Winkelauflösungen von bis zu acht Messpunkten pro Grad erreicht [12]. Die Distanzgenauigkeit von Laserscannern liegt im Zentimeterbereich. Störeffekte treten allerdings bei Nebel und Gischt auf, da das Infrarotlicht auch von Wasserpartikeln in der Luft reflektiert wird.

1.5.4 Mono-Kamera

Kamerasysteme werden zunehmend in Fahrzeugen eingesetzt. Am meisten verbreitet sind hierbei zum einen die vorausschauenden Kameras, welche üblicherweise hinter der Windschutzscheibe angebracht werden. Sie erfüllen verschiedene Funktionen wie Verkehrszeichenerkennung [13], Spurerkennung [14] oder Fußgängerdetektion [15]. Zum anderen werden auch Umfeldkameras mit Weitwinkelobjektiven eingesetzt. Als Anwendung stehen hierzu die in Kapitel 1.2.3 genannten Rückfahrkameras und die in Kapitel 1.2.4 vorgestellten *Surround-View*-Systeme im Vordergrund. Während Kamerasysteme in verschiedensten Anwendungen eingesetzt werden, ist der grundlegende Aufbau aller Kameras ähnlich. Der Kern einer Kamera ist der Sensorchip. Im Automobilbereich werden in erster Linie CMOS (*Complementary Metal Oxide Semiconductor*) Sensoren eingesetzt

[16]. Der Sensorchip ist eine Matrix aus fotoempfindlichen, integrierten Schaltungen, sogenannten Pixeln. Vor dem Sensor befindet sich eine Optik. Diese hat die Aufgabe, das zur Kamera kommende Licht zu bündeln und projiziert die dreidimensionale Umgebung auf den zweidimensionalen Sensor. Auf diese Weise erhält man eine digitalisierte zweidimensionale Projektion der Umgebungshelligkeiten.

Neben einer Anzeige der Bilddaten für den Fahrer können die Bilddaten auch für unterschiedlichste Anwendungen weiterverarbeitet werden. Eine Merkmalsextraktion bietet in der Regel die Grundlage für eine Detektion und Klassifikation von Objekten wie Verkehrszeichen oder Fußgängern [2]. Durch Bildverarbeitungsverfahren ist es auch möglich, dreidimensionale Informationen aus bewegten Kamerabildern zu erzeugen. Sogenannte *Structure-from-Motion*-Verfahren (SfM) schätzen dreidimensionale Strukturen anhand von zweidimensionalen Bildfolgen, welche von unterschiedlichen Positionen aufgenommen sind [17]. Bildverarbeitungsverfahren sind aufgrund der Menge der Informationen sehr rechenaufwendig. Die zugehörigen Steuergeräte wenden oft die höchste Rechenleistung im Fahrzeug auf [2].

1.5.5 Stereo-Kamera

Das Funktionsprinzip von Stereo-Kamerasystemen ähnelt dem Prinzip des menschlichen Sehens. Eine räumliche Tiefenwahrnehmung entsteht durch die Fusion zweier Bilder. Ein Stereo-Kamerasystem besteht daher aus zwei Kameras, die einen festen Abstand zueinander besitzen.

Tiefenwerte werden durch eine nachgelagerte Bildverarbeitung von synchron aufgenommenen Bildern berechnet. Diese lassen sich aus der Distanz zwischen den Bildkoordinaten eines Punktes in einem Kamerabild und seinem korrespondierenden Punkt im zweiten Kamerabild bestimmen. Die Distanz zwischen den korrespondierenden Punkten wird Disparität genannt [18]. Um die Tiefenwerte zu bestimmen, müssen die Kameras kalibriert sein. Dies bedeutet insbesondere, dass die Positionen und Ausrichtungen (Translation und Rotation) der Kameras zueinander bekannt sein müssen. Um die Disparität zwischen den beiden Kamerabildern zu berechnen, müssen korrespondierende Punkte zwischen den Bildern gesucht werden. Hierbei kann ausgenutzt werden, dass der korrespondierende Punkt eines Bildpunktes im ersten Kamerabild auf einer Linie im zweiten



1 Einleitung

Kamerabild liegen muss. Durch eine sogenannte Rektifizierung der Kamerabilder reduziert sich die Suche zusätzlich auf eine horizontale Linie.

Tiefeninformationen können nur für den sich überlappenden Sichtbereich beider Kameras bestimmt werden. Zudem lassen sich mit dem Stereo-Verfahren keine sehr großen Öffnungswinkel realisieren. Weitere Nachteile sind die Notwendigkeit einer starren Montage und der große benötigte Bauraum durch den Einsatz von zwei versetzten Kameras.

Eingesetzt werden Stereo-Kameras in erster Linie für vorausschauende Sicherheits- und Komfortfunktionen. Zusätzlich zu den Funktionen, welche sich mit Mono-Kameras realisieren lassen, bieten Stereo-Kameras erweiterte Möglichkeiten zum vorausschauenden Fußgängerschutz [19], zur Kollisionsvermeidung [20] und zur Vermessung der Fahrbahnoberfläche [21].

1.5.6 Time-of-Flight-Kamera

Time-of-Flight-Kameras sind aktive 3D-Kameras, die auf dem Prinzip der Lichtlaufzeitmessung basieren. Eine aktive Lichtquelle strahlt moduliertes Licht in die Umgebung ab. Dieses Licht wird an Objekten reflektiert und gelangt über eine Optik auf den Bildsensor der Kamera. Jedes Pixel auf dem Sensor ist eine integrierte Schaltung zur Auswertung des reflektierten Lichtsignals. Um die Distanz zu bestimmen, existieren verschiedene Verfahren.

Bei der Messung der Phasenverschiebung wird Infrarotlicht in einem Rechteck- oder Sinussignal moduliert. Ein Pixel des Sensors bestimmt in diesem Fall die Phasenverschiebung zwischen dem ausgehenden und einfallenden Signal. Aus der Phasenverschiebung kann dann auf die Lichtlaufzeit und damit auf den vom Licht zurückgelegten Weg geschlossen werden. Die Lichtlaufzeit wird daher nicht im eigentlichen Sinne gemessen, sondern durch die Messung der Phasendifferenz bestimmt. Eine Ausprägung derartiger Time-of-Flight-Kameras ist das *Photonic Mixing Device* (PMD). PMD-Kameras basieren auf der Messung der Phasendifferenz und werden ausführlich im nachfolgenden Kapitel 2.1 vorgestellt.

Des Weiteren existieren Time-of-Flight-Kameras, welche direkt auf dem Prinzip der Messung der Lichtlaufzeit basieren. Das Messverfahren ist analog zu den lichtlaufzeitmessenden Lidar-Sensoren, die in Kapitel 1.5.3 behandelt wurden. Allerdings wird die zu vermessende Umgebung nicht

durch Ablenkung eines fokussierten Lichtstrahls abgetastet. Stattdessen wird der gesamte zu vermessende Bereich mit Lichtpulsen bestrahlt. Jedes Pixel des Sensors bestimmt nun unabhängig und zeitgleich eine dem Raumwinkel des Pixels entsprechende Distanz. Kameras dieser Art werden auch *Flash Lidar* genannt [22].

Eine weitere Art von Time-of-Flight-Kameras arbeitet nach dem *Global-Shutter*-Prinzip. Hierbei sendet die Kamera Lichtpulse in die Umgebung aus. Ein *Shutter* befindet sich vor dem Sensor und schützt den Sensor bei Verschluss vor ungewolltem Lichteinfall. Dieser *Shutter* wird nun mit dem Lichtsignal synchronisiert und öffnet nur, wenn Lichtpulse ausgesendet werden. Da das Licht eine gewisse Zeit benötigt bis es zurück zum Sensor gelangt, fällt nur ein Teil des abgestrahlten Lichts auf den Sensor. Die Länge des Zeitraums, in dem das reflektierte Licht empfangen wird, ist dabei proportional zum zurückgelegten Weg des Lichtes. Ein Beispiel einer solchen Kamera ist die *ZCam* der Firma *3DV* [23].

Time-of-Flight-Kameras können im Automobil für die vorausschauende Sicherheit verwendet werden. Hierzu werden in erster Linie PMD-Kameras eingesetzt, welche Funktionen wie den vorausschauenden Fußgängerschutz [24],[25], automatische Notbremsfunktionen [26] oder *PreCrash*-Systeme [27] ermöglichen.

1.6 Vergleich der Sensoren

In vorhergehenden Kapitel 1.5 wurden verschiedene Sensoren zur Umfeldwahrnehmung vorgestellt. Die Anforderungen der zukünftigen Fahrerassistenzsysteme an die Umfeldsensorik wurden in Abschnitt 1.4 erläutert. Im Folgenden werden die unterschiedlichen Sensoren bezüglich der Anforderungen miteinander verglichen. Abbildung 1.1 fasst den Vergleich tabellarisch zusammen.

Ultraschallsensoren, welche in Kapitel 1.5.1 vorgestellt wurden, bieten keine Möglichkeit für eine 3D-Datenverarbeitung. Somit ist auch keine Messung von Über- und Unterfahrbarkeit möglich. Auch können keine dynamischen Objekte erkannt werden. Zudem bieten sie die niedrigste Reichweite im Vergleich.

Mono-Kamerasysteme können, wie in Kapitel 1.5.4 beschrieben, 3D-Informationen durch eine nachgelagerte Bildverarbeitung liefern. Durch Weit-



	Abdeckung	Reichweite	Genauigkeit	Robustheit	Umfeldmodell	Über-/Unterfahrbarkeit	Lokalisierung	Dyn. Objekte	Abstände	Messrate	Messung im Stand	Integrierbarkeit	Kosten
Ultraschall	○	○	+	○	-	-	-	-	+	+	+	+	+
Kamera	+	+	○	○	○	+	○	+	○	+	-	+	+
Stereo	-	+	○	+	+	+	+	+	+	+	+	○	○
Laserscanner	-	+	+	+	○	-	+	○	+	+	+	○	○
3D-Laserscanner	+	+	+	+	+	+	+	+	+	+	+	-	-
Time-of-Flight-Kamera	○	+	○	+	+	+	+	+	+	+	+	○	○
Radar	○	+	○	○	-	-	-	+	+	+	+	+	○

Tabelle 1.1: Vergleich von Sensoren zur Erfüllung der Anforderungen von zukünftigen Fahrerassistenzsystemem für das Parken und Rangieren

winkelobjektive kann zudem eine hohe Abdeckung erreicht werden. Allerdings kann die Datenverarbeitung keine dichten 3D-Tiefendaten für den gesamten Abdeckungsbereich liefern. Ein weiteres Problem ist, dass ohne Bewegung der Kamera keine 3D-Daten berechnet werden können.

Diesen Nachteil besitzen Stereo-Kamerasysteme nicht. Allerdings sind bei Stereo-Kameras keine sehr großen Öffnungswinkel möglich, sodass die geforderte Abdeckung nicht erreicht werden kann. Sowohl Mono- als auch Stereo-Kameras können nur bei ausreichender Helligkeit 3D-Daten liefern. Bei einer vorausschauenden Kamera können die Frontscheinwerfer die Umgebung beleuchten. Die Bereiche seitlich des Fahrzeugs werden hingegen nicht angestrahlt, sodass passive Kamerasysteme bei Dunkelheit nur unzureichende 3D-Daten der Umgebung liefern.

Ein Laserscanner liefert sehr genaue und robuste Messdaten. Dies allerdings nur in einer oder in wenigen Ebenen, sodass keine vollständige vertikale Abdeckung erreicht werden kann. Prinzipiell kann dieses Problem durch 3D-Laserscanner gelöst werden. Diese besitzen alle gewünschten funktionalen Anforderungen. Bei einem 3D-Laserscanner sind allerdings Integrierbarkeit und Kosten ein Problem. Ein solcher Sensor müsste über dem Dach eines Fahrzeugs montiert werden. Auch die sehr hohen Kosten der Sensoren verhindern einen Serieneinsatz. Für Prototypen und als Referenzsensorik können 3D-Laserscanner dagegen eingesetzt werden.

Time-of-Flight-Kameras bieten keinen schwerwiegenden Nachteil. Heutige Time-of-Flight-Kameras besitzen zwar einen geringen Öffnungswinkel, allerdings wird in dieser Arbeit eine Möglichkeit zur Erhöhung des Öffnungswinkels der Kamera vorgestellt. Time-of-Flight-Kameras können auch bei absoluter Dunkelheit Messungen liefern. Probleme treten erst aufgrund des hohen Fremdlichtanteils bei sehr starkem Sonnenlichteinfall auf.

Radarsysteme bieten zwar sehr hohe Reichweiten und die Messung der Geschwindigkeiten, doch für eine hochgenaue dreidimensionale Vermessung des nahen Fahrzeugumfelds bieten sie keine ausreichende laterale Auflösung.

Bei der Betrachtung der Anforderungen können ausschließlich im Falle der Time-of-Flight-Kameras keine schwerwiegenden Nachteile entdeckt werden. Daher bieten Time-of-Flight-Kameras ein hohes Potenzial für einen Einsatz in Parksyste men. In dieser Arbeit wird daher die Verwendung von Time-of-Flight-Kameras als Umfeldsensor für das Parken und Rangieren untersucht. Zur Untersuchung von Time-of-Flight-Kameras für die be-



1 Einleitung

schriebenen Anwendung werden im Laufe der Arbeit PMD-Kameras eingesetzt. PMD-Kameras sind Time-of-Flight-Kameras, die auf der Messung der Phasendifferenz basieren, und decken damit nur einen Teil der existierenden Time-of-Flight-Kameras ab. Aufgrund der Ähnlichkeit der Bilddaten aller Ausprägungen von hochauflösenden Time-of-Flight-Kameras wird in dieser Arbeit das PMD-Messprinzip stellvertretend für die gesamte Breite an Time-of-Flight-Kameras eingesetzt. Nahezu alle in dieser Arbeit entwickelten Methoden können auch durch Time-of-Flight-Kameras anderer Hersteller umgesetzt werden. Es wird angenommen, dass sich lediglich die Ausprägungen der unterschiedlichen Fehlereffekte und Störfaktoren unterscheiden.

1.7 Zielsetzung

In dieser Arbeit soll ein System zur umfassenden dreidimensionalen Modellierung eines Fahrzeugumfelds durch Time-of-Flight-Kameras entwickelt werden. Eine im Automobilbereich eingesetzte Time-of-Flight-Kameratechnologie ist das *Photonic Mixing Device* (PMD). Daher soll eine PMD-Kamera als Ausprägung der Time-of-Flight-Kamera zu einer Weitwinkel-Time-of-Flight-Kamera modifiziert werden. Neben einer dreidimensionalen und omnidirektionalen Umfeldmodellierung wurden weitere notwendige Anforderungen an das Sensorsystem aus den zukünftigen Fahrerassistenzsystemen im Bereich des Parkens und Rangierens hergeleitet.

Um diese Anforderungen zu erfüllen, muss nicht nur eine entsprechende Systemauslegung und Kalibrierung für das Time-of-Flight-Kamerasystem entwickelt werden, sondern insbesondere auch neuartige Methoden zur Datenverarbeitung der Tiefendaten. Die Datenverarbeitung erstreckt sich von der Vorverarbeitung der Kameradaten bis hin zu einer dreidimensionalen Rekonstruktion der Umgebung eines Fahrzeugs.

Das dreidimensionale Umfeldmodell soll sowohl eine Darstellung von dreidimensionalen Informationen für den Fahrer als auch die Extraktion von Abstandsinformationen für automatisierte Fahrfunktionen ermöglichen.

Im nächsten Kapitel wird aus dem Stand der Technik zur Time-of-Flight-Kamera und ihrer Datenverarbeitung ein Ansatz für eine dreidimensionale Umfeldmodellierung abgeleitet. Anschließend folgt eine detaillierte Beschreibung der Struktur des eigenen Ansatzes in Kapitel 2.4.

2 Stand der Technik und Methode

2.1 PMD-Kamera

Wie in Kapitel 1.5.6 beschrieben, gehören PMD-Kameras zu den Time-of-Flight-Kameras, die auf dem Prinzip der Bestimmung der Phasendifferenz basieren. Das PMD-Prinzip wurde an der Universität Siegen entwickelt und die Weiterentwicklung wurde durch die Ausgründung der *PMD Technologies GmbH* ermöglicht [28].

Im Folgenden werden das PMD-Funktionsprinzip, die Sensordaten, verschiedene Fehlerquellen und Störfaktoren, sowie die Kalibrierung der Kameras thematisiert.

2.1.1 Funktionsprinzip

PMD-Kameras bestehen aus einem Sensor in CMOS-Technologie, einer Empfangsoptik und einer Beleuchtungseinheit [29]. Das Funktionsprinzip ist schematisch in Abbildung 2.1 dargestellt. Die Lichtquelle sendet amplitudenmodulierte Lichtpulse im nahen Infrarotlichtbereich (NIR) aus. Die Ansteuerelektronik der Lichtquelle arbeitet mit einem internen Referenzsignal $r(t)$ und moduliert die Lichtpulse mit einem Rechtecksignal [30]. Dabei kann es sich bei der Lichtquelle sowohl um eine *Light Emitting Diode* (LED) oder um eine Laserdiode handeln. Bei LEDs liegt die Modulationsfrequenz typischerweise bei 20-30 MHz, bei Laserdioden kann sie bis zu 80 MHz betragen.

Das abgestrahlte Licht wird von Objekten in der Umgebung reflektiert und gelangt über die Empfangsoptik auf den Sensor. Alle Pixel des Sensors bestimmen nun simultan die Phasendifferenz zwischen dem Referenzsignal

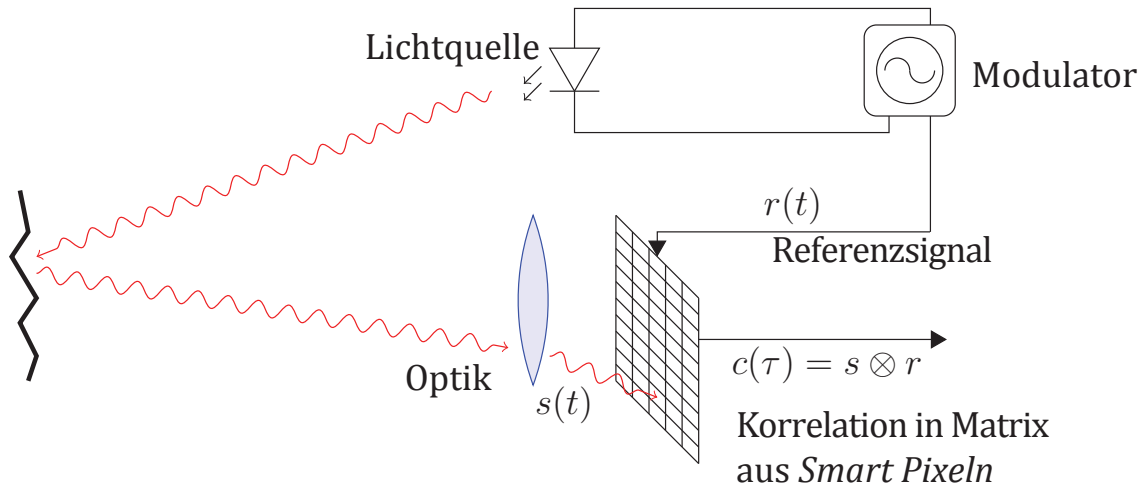


Abbildung 2.1: Funktionsprinzip eines PMD-Sensors nach [31]

$r(t)$ und dem eingehenden Lichtsignal $s(t)$. Dieser Prozess ist in Abbildung 2.2 dargestellt.

Jedes Pixel besteht aus einer integrierten Schaltung mit einer photosensitiven Fläche und einer Auswerte-Elektronik zur Auswertung des empfangenen Lichtsignals. Aufgrund ihrer Funktionalität werden die Pixel auch *Smart Pixel* genannt [33]. In Abbildung 2.3 sind der Aufbau und Seitenchnitt eines Pixels schematisch dargestellt. Es befinden sich zwei symmetrisch angeordnete lichtempfindliche Elemente, die Photogates am und bm , mit komplexen Ausleseschaltungen in jedem Pixel. Eine Auslesediode verbindet die Photogates mit der Ausleseschaltung. An die Photogates ist die Vorspannung U_0 und eine Gegentakt-Modulationsspannung $u_m(t)$ angelegt. Die Modulationsspannung verursacht ein Potentialgefälle, welches vom Licht einfallende Ladungsträger abwechselnd in die beiden Potentialtöpfe der Auslesedioden bewegt. Dieser Effekt wird auch *Elektronenschaukel* genannt und ermöglicht den Mischprozess zwischen Referenzsignal und optischem Signal. Das Prinzip zeigt Abbildung 2.4.

Auf diese Weise wird das Referenzsignal mit dem phasenverschobenen optischen Signal korreliert. Dabei entspricht die Differenz der beiden Span-

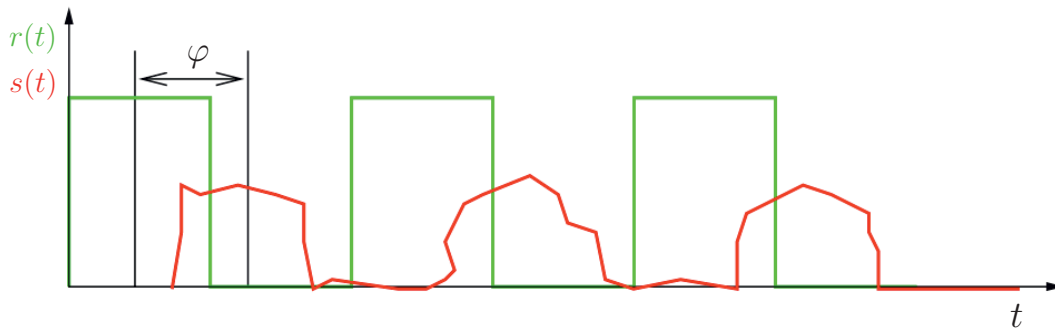


Abbildung 2.2: Phasenverschiebung φ zwischen emittiertem Lichtsignal $r(t)$ in grün und dem eintreffenden Lichtsignal $s(t)$ in rot [32]

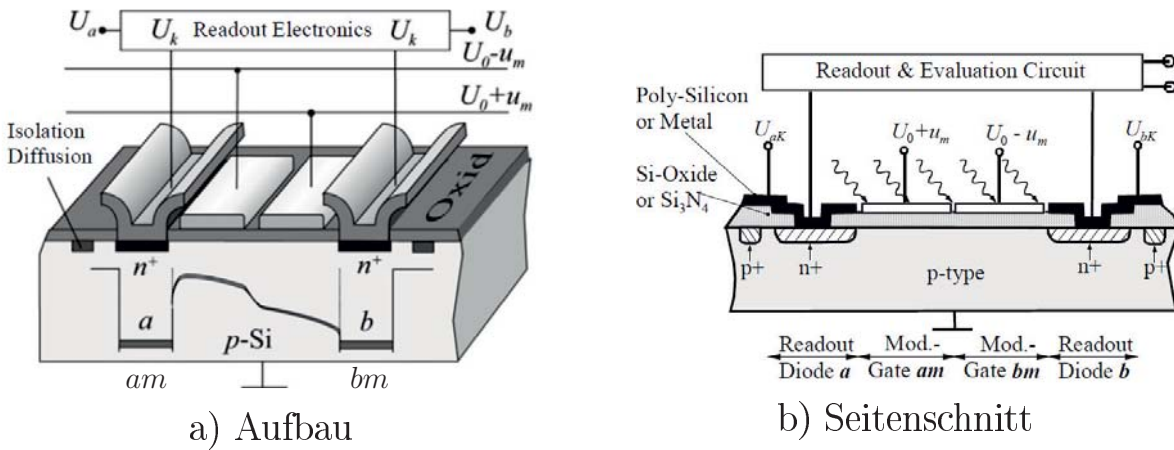


Abbildung 2.3: Aufbau (a) und Seitenschnitt (b) eines PMD-*Smart-Pixels* [34]

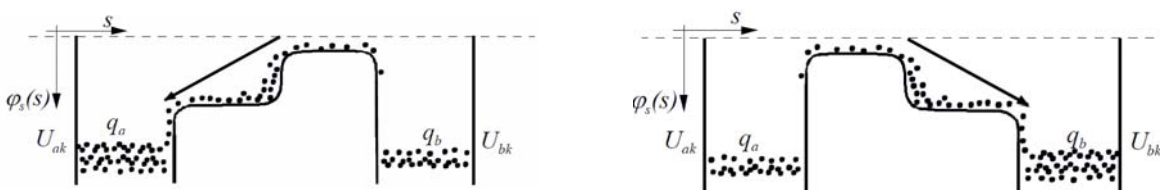


Abbildung 2.4: Funktionsprinzip der Elektronenschaukel [34]

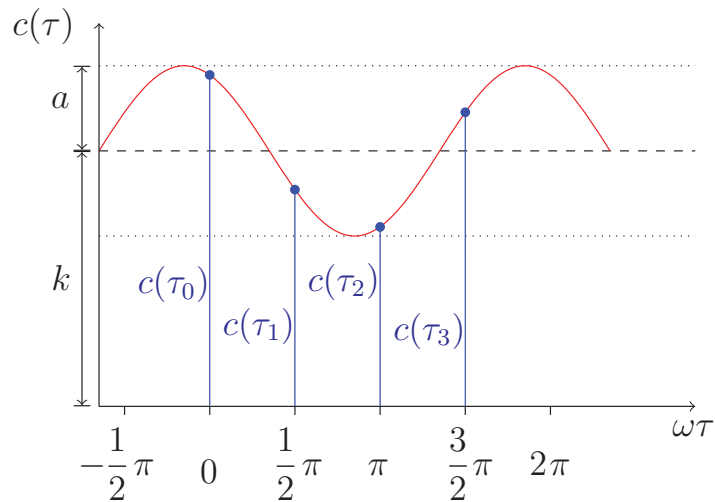


Abbildung 2.5: Abtastung der Autokorrelationsfunktion an vier diskreten Zeitpunkten

nungen einer Abtastung der Autokorrelationsfunktion $c(\tau)$, welche in Gleichung 2.1 beschrieben ist.

$$c(\tau) = s \otimes r = \lim_{T \rightarrow \infty} \int_{-T/2}^{T/2} r(t) \cdot s(t + \tau) dt \quad (2.1)$$

Unter Annahme von sinusförmigen Signalen

$$r(t) = \cos(\omega t) \quad (2.2)$$

und

$$s(t) = k + a \cos(\omega t - \tau) \quad (2.3)$$

ergibt sich nach [35]

$$c(\tau) = \frac{a}{2} \cos(\omega \tau + \varphi) \quad (2.4)$$

mit der Kreisfrequenz $\omega = 2\pi f_m$ zur Modulationsfrequenz f_m , der Amplitude a des modulierten Lichtsignals, der Intensität k und der Phasendifferenz φ . Die Demodulierung der Autokorrelationsfunktion $c(\tau)$ kann wie in Abbildung 2.5 durch vier sequentielle Abtastungen $I_i = c(\tau_i)$ mit den internen Phasenunterschieden $\tau_i = i \cdot \frac{\pi}{2} \omega$ erfolgen [35].

Anschließend kann die distanzabhängige Phasenverschiebung φ bestimmt werden durch

$$\varphi = \arctan\left(\frac{c(\tau_3) - c(\tau_1)}{c(\tau_0) - c(\tau_2)}\right). \quad (2.5)$$

Die Amplitude a des Signals $s(t)$ ergibt sich durch

$$a = \frac{\sqrt{(c(\tau_3) - c(\tau_1))^2 + (c(\tau_0) - c(\tau_2))^2}}{2} \quad (2.6)$$

und die Intensität k , vergleichbar mit dem Grauwert einer Graubildkamera, kann durch

$$k = \frac{c(\tau_0) + c(\tau_1) + c(\tau_2) + c(\tau_3)}{4} \quad (2.7)$$

ermittelt werden.

Aufgrund der konstanten Modulationsfrequenz f_m kann über die Phasenverschiebung φ die Lichtlaufzeit t durch

$$t = \frac{\varphi}{2\pi \cdot f_m} \quad (2.8)$$

bestimmt werden. Unter Berücksichtigung der Lichtgeschwindigkeit c kann die Distanz d bestimmt werden durch

$$d = \frac{c \cdot t}{2} = \frac{c \cdot \varphi}{4\pi \cdot f_m}. \quad (2.9)$$

Der Eindeutigkeitsbereich der Distanz ergibt sich aus der Modulationsfrequenz f_m . Eine Frequenz von 20 MHz ergibt beispielsweise einen Eindeutigkeitsbereich von 7,5 m. Durch Mehrfrequenzverfahren lässt sich dieser Eindeutigkeitsbereich erhöhen.

Wenn Fremdlicht, welches nicht von der Lichtquelle der Kamera stammt, auf das Pixel gelangt, hat dies zunächst den Effekt, dass die beiden Potentialtöpfe des Pixels mit Ladungsträgern gefüllt werden [30]. Da das Fremdlicht nicht moduliert ist, werden die beiden Potentialtöpfe gleichmäßig mit Ladungsträgern gefüllt. Die Abtastung der Autokorrelationsfunktion erfolgt über eine Differenzbildung der beiden Potenziale. Aus diesem Grund kann auch unter Fremdlichteinfluss eine Messung erfolgen. Dennoch gibt es zwei nachteilige Effekte. Zum einen verursacht Fremdlicht ein ungünstiges Signal-Rausch-Verhältnis, welches zu einer niedrigeren Genauigkeit

der Distanzmessung führt. Zum anderen besteht die Gefahr, dass das Pixel bei zu viel Fremdlicht sättigt und somit keine aussagekräftige Messung möglich ist. Um diesen Effekt zu vermeiden, gibt es zu jedem Pixel eine Schaltung zur Hintergrundlichtunterdrückung. Diese wird als *Suppression of Background Illumination* (SBI) bezeichnet [33]. Sie erkennt Gleichanteile in den Potentialtöpfen und entfernt sie, sodass auch unter Fremdlichteinfluss eine höhere Dynamik möglich ist.

Um ein vorteilhafteres Signal-Rausch-Verhältnis zu erreichen, muss mehr Nutzlicht zum Pixel gelangen, das heißt, die Amplitude muss steigen. Dies kann beispielsweise erreicht werden, indem die sogenannte Integrationszeit erhöht wird. Die Integrationszeit gibt die Dauer der Belichtung einer Messung an und damit die Dauer der Lichtpulse der Lichtquellen. Eine höhere Integrationszeit führt zu längeren Lichtpulsen und damit zu einer höheren Amplitude. Eine zu hohe Amplitude kann allerdings zu Bewegungsartefakten führen. Zudem verhindern lange Integrationszeiten hohe Frameraten der Kamera. Diese können bis zu 160 Hz betragen.

2.1.2 Sensordaten

Im vorherigen Abschnitt wurde erläutert, wie ein PMD-Pixel einen Wert für die Phasenverschiebung zwischen Ausgangssignal und optischem Signal bestimmt und wie dieser Wert in einen Distanzwert umgerechnet werden kann. Da ein *PMD-Imager* aus einer Matrix von PMD-Pixeln besteht, ergibt sich ein zweidimensionales Distanzbild. In Abbildung 2.6 ist das Distanzbild für einen Sensor mit einer Auflösung von 200 x 200 Pixeln abgebildet. Die Distanzwerte sind farbkodiert von rot (nah) bis blau (weit) dargestellt.

Zusätzlich zur Phasenverschiebung liefert der Sensor noch Werte für die Amplitude des Nutzsignals und für die Intensität. Amplitudenbild und Intensitätsbild sind in den Abbildungen 2.7 und 2.8 dargestellt. Die Amplitude entspricht der modulierten Lichtenergie, die von der Umgebung reflektiert wird. Ein hoher Wert für die Amplitude verbessert daher das Signal-Rausch-Verhältnis, sodass die Amplitude ein Maß für die Qualität der Distanzmessung sein kann. Die Intensität hingegen entspricht der Menge des gesamten Infrarotlichts, das auf einen Pixel trifft. Da es auch Fremdlicht enthält, das nicht von der aktiven Beleuchtung der Kamera stammt, besitzt es eine Ähnlichkeit zu einem herkömmlichen Grauwertbild. Das

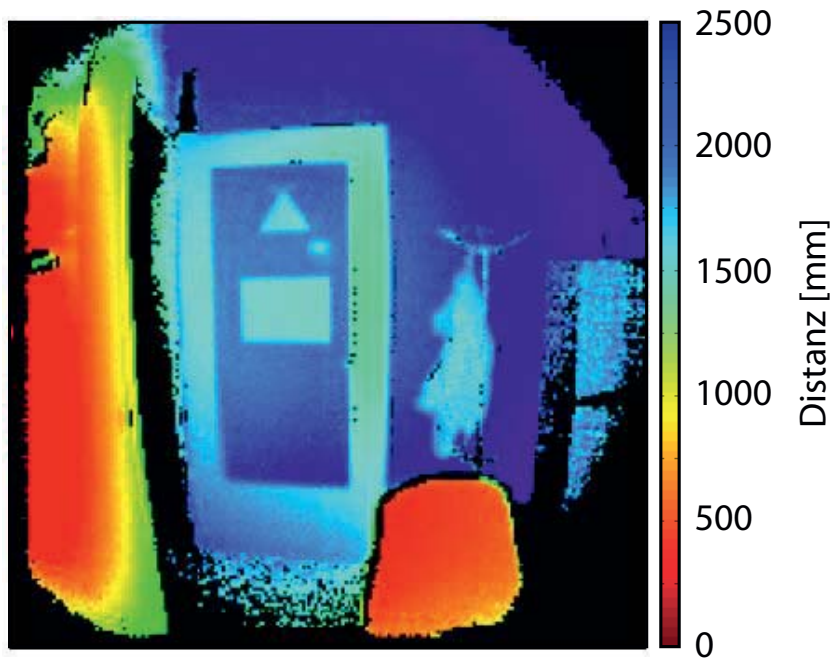


Abbildung 2.6: Distanzbild in farbkodierter Darstellung von rot (nah) bis blau (weit)

dargestellte Intensitätsbild besteht aus den unkalibrierten gemessenen Intensitätswerten. Aufgrund der unterschiedlichen Eigenschaften jedes Pixels wird die Kalibrierung des sogenannten *Fixed Pattern Noise* (FPN) benötigt. Das FPN und weitere Fehlereffekte werden ausführlich im nachfolgenden Abschnitt thematisiert.

2.1.3 Fehlerquellen und Störfaktoren

Die Messungen von PMD-Kameras leiden unter stochastischen und systematischen Fehlern.

Der stochastische Fehler äußert sich durch ein Rauschen in der Distanzmessung. Wie in Abschnitt 2.1.1 erläutert, führt ein niedriger Amplitudenwert zu einem nachteiligen Signal-Rausch-Verhältnis. Ein Pixel ist unterbelichtet, wenn eine geringe Menge an moduliertem Nutzlicht auf das Pixel trifft und sich somit ein niedriger Wert für die Amplitude ergibt. Die Menge an moduliertem Nutzlicht, das auf ein Pixel trifft, hängt von verschiedenen Faktoren ab. Klar ist, dass eine stärkere Beleuchtung zu einer höheren Amplitude führt. Die Menge an moduliertem Nutzlicht kann auch erhöht



Abbildung 2.7: Amplitudenbild



Abbildung 2.8: Intensitätsbild

werden, indem die Integrationszeit erhöht wird. Eine längere Integrationszeit hat einen positiven Effekt auf das Signal-Rausch-Verhältnis. Des Weiteren kann eine höhere Genauigkeit durch die Erhöhung der Modulationsfrequenz erreicht werden [30]. Allerdings müssen hierbei die Grenzen der Lichtquelle beachtet werden. Die Amplitude nimmt mit der Distanz des reflektierenden Objektes ab. Auch spielt die Objektreflektivität eine Rolle. Je mehr Licht vom Objekt absorbiert wird, umso geringer wird die Amplitude. Eine weitere Rolle spielt die Ausrichtung der reflektierenden Oberfläche. Die Amplitude hängt vom Winkel zwischen Lichteinfall und der Oberflächennormalen ab. Fremdlicht aus der Umgebung ist ein weiterer Störfaktor. Zwar kann die in Abschnitt 2.1.1 erläuterte SBI die Gefahr der Sättigung eines Pixels unter Fremdlichteinfluss verringern, allerdings führt Fremdlicht dennoch zu einem ungünstigeren Signal-Rausch-Verhältnis und damit zu einem größeren stochastischen Fehler.

Systematische Fehler sind Fehler, die sich bei wiederholter Messung im Mittel nicht aufheben. Durch Toleranzen in den Fertigungsprozessen und aufgrund von unterschiedlichen Oberflächeneigenschaften der Pixel besitzt jedes Pixel eine individuelle Charakteristik, welche zu einem individuellen Offset in der Distanzbestimmung führt [36]. Ein weiterer distanzabhängiger systematischer Fehler ist der sogenannte *Wiggling Error* [37]. Er entsteht durch theoretische Annahmen, die in der Praxis nicht exakt eingehalten werden. In Abschnitt 2.1 wurde beschrieben, dass bei der Bestimmung der Phasendifferenz sinusförmige Signale angenommen werden (siehe Gleichungen 2.2 und 2.3). In der Realität werden die Lichtquellen allerdings mit einem Rechtecksignal angesteuert. Dieser Unterschied zwischen Theorie und Realität führt zu einem systematischen Fehler, welcher mal positiv und mal negativ ist, abhängig von der wahren Distanz des reflektierenden Objektes. In Abbildung 2.9 ist der *Wiggling*-Fehler einer PMD-Kamera in einem Graphen dargestellt. Hier sind die Offsets gegenüber der wahren Distanz aufgetragen. Die Ausreißer in den niedrigen Distanzen werden von Lindner durch Überbelichtung erklärt [38].

Der Distanzwert, den eine PMD-Kamera liefert, repräsentiert die Entfernung, die das Licht von der Lichtquelle über das reflektierende Objekt zurück zum optischen Zentrum der Kamera zurückgelegt hat. Das Messprinzip geht im idealen Fall davon aus, dass sich die Lichtquelle am selben Punkt wie das optische Zentrum befindet. Da dies in der Praxis allerdings nicht der Fall ist, liegt in der Anordnung der Lichtquelle zur Empfangsop-

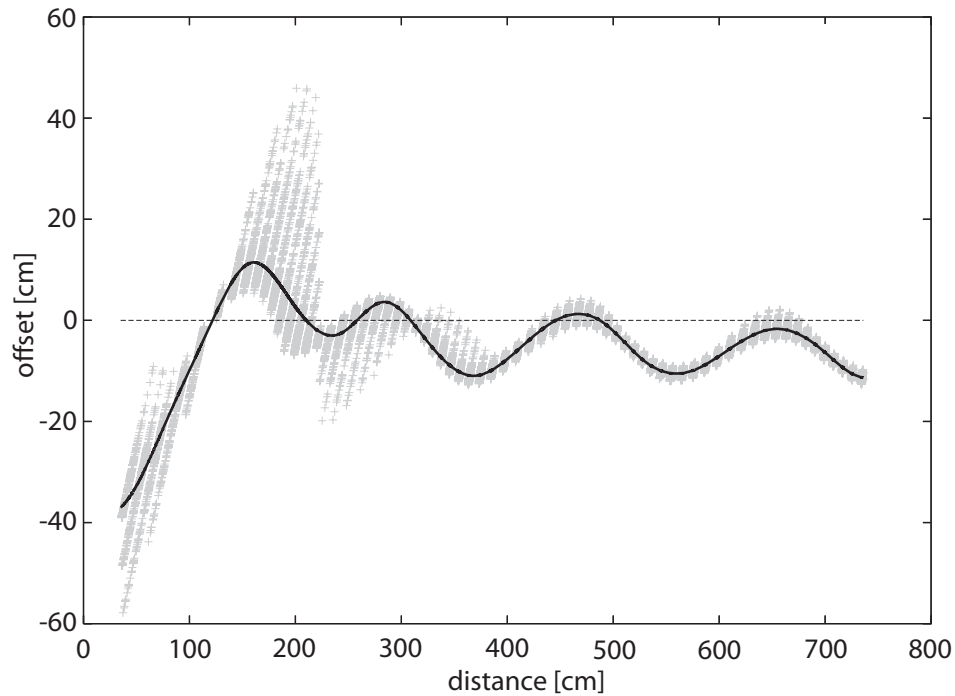


Abbildung 2.9: *Wiggling*-Fehler einer PMD-Kamera nach [38]

tik ein weiterer systematischer Fehler. Die Auswirkungen der Geometrie von Lichtquellen wurden in [39] untersucht. Insbesondere im Bereich bis zu einem Meter hat die Lichtquellengeometrie einen starken Einfluss auf die Entfernungsmessung.

Eine weitere Art von Messfehlern sind die sogenannten Mischphasen. Diese treten auf, wenn auf ein Pixel Lichtstrahlen treffen, die unterschiedliche Wege zurückgelegt haben. Ein Beispiel sind die sogenannten *Flying Pixels*. Diese treten an Objektkanten auf. Hier mischen sich in einem Pixel einerseits Lichtsignale, die von einem Objekt im Vordergrund reflektiert werden und andererseits Lichtsignale, die von einem Hintergrundobjekt reflektiert werden. Durch die Vermischung der beiden phasenverschobenen Signale wird eine Phasenverschiebung bestimmt, welche einem Distanzwert entspricht, der zwischen den Distanzen zu den beiden Objekten liegt. Abbildung 2.10 veranschaulicht diesen Effekt und Abbildung 2.11 zeigt ein Beispiel für den Effekt anhand einer 3D-Punktwolke.

Mischphasen treten auch auf, wenn Licht innerhalb der Linse reflektiert wird. Dieser Effekt wird auch *Scattering* genannt. Das Licht von Objekten, die stark reflektieren, wird im Linsensystem der Kamera gestreut und

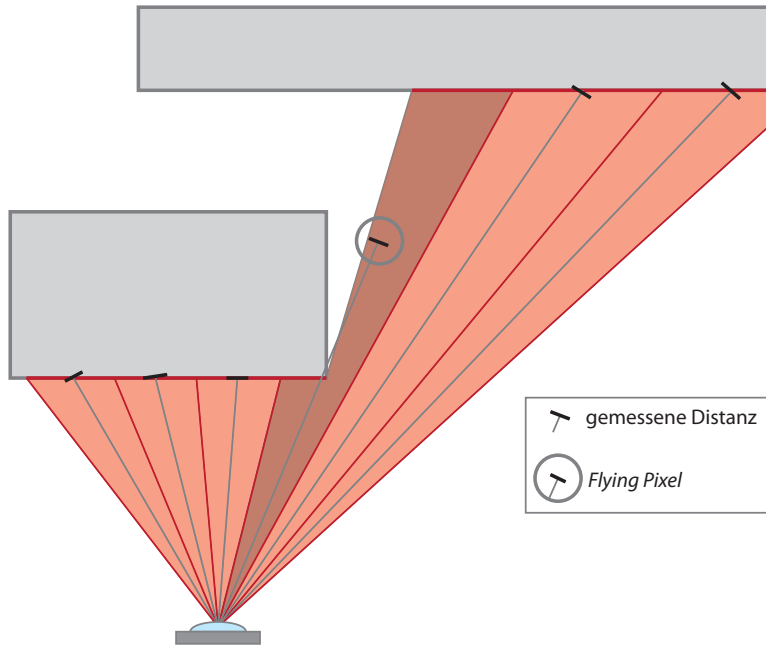


Abbildung 2.10: Entstehung des *Flying-Pixel*-Effekts

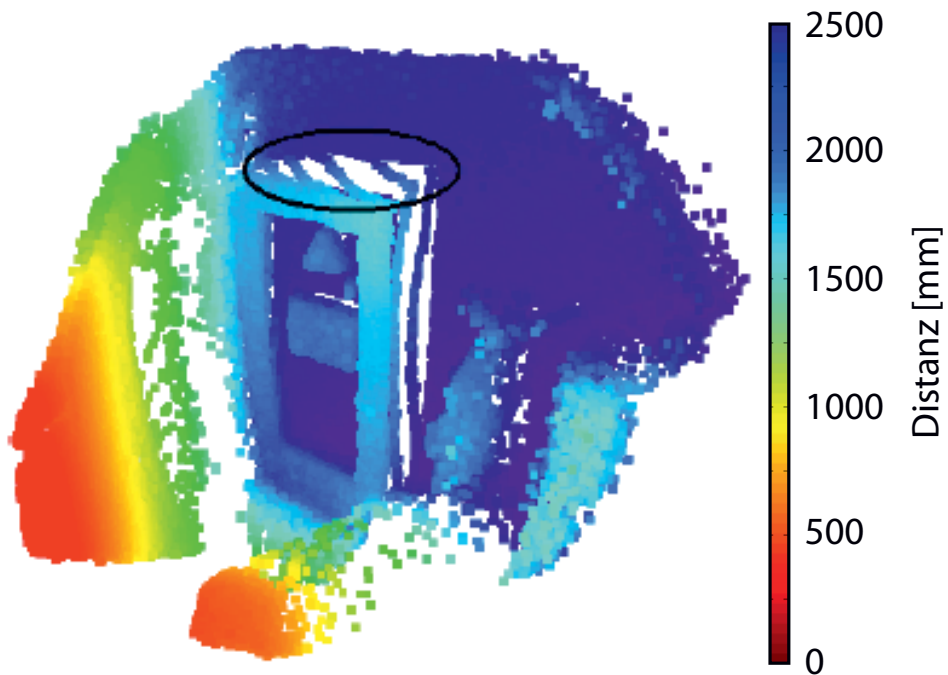
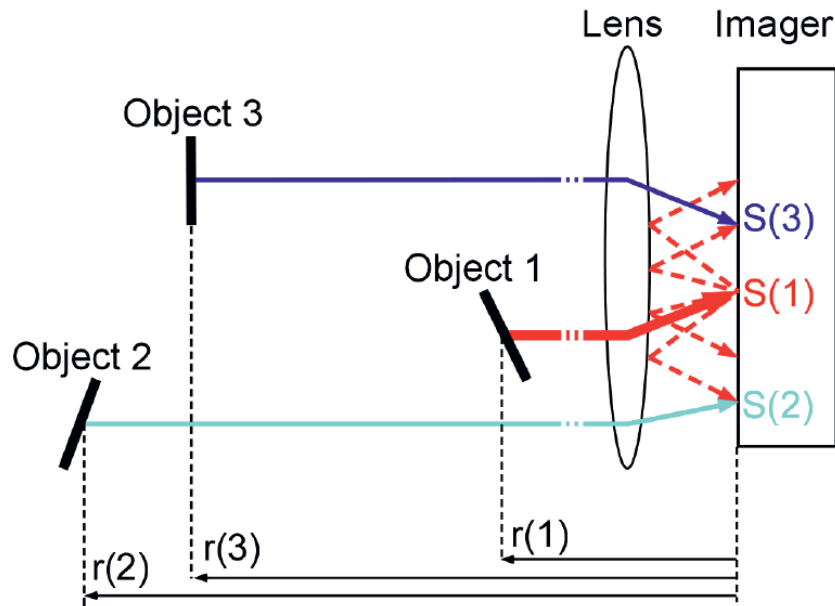


Abbildung 2.11: *Flying Pixels* in einer 3D-Punktwolke [31]

Abbildung 2.12: Entstehung des *Scattering*-Effekts [40]

gelangt auf mehrere Pixel des Sensors. In diesen Pixeln mischen sich die Lichtsignale des überstrahlenden Objektes mit denen, welche gemäß der optischen Abbildungseigenschaften der Linse auf die jeweiligen Pixel treffen. Oft verursacht ein nahes Objekt eine Verringerung der Distanzen über die gesamte Sensorfläche. Der Effekt ist in Abbildung 2.12 veranschaulicht, und in Abbildung 2.13 ist das Distanzbild einer durch *Scattering* gestörten Aufnahme zu sehen. Insbesondere in den Pixelbereichen in der Umgebung der Hand im Vordergrund werden zu geringe Distanzen gemessen.

Ein weiterer Effekt, bei dem Mischphasen entstehen, ist der sogenannte *Multipath*-Effekt [36]. Dieser tritt auf, wenn ausgesendetes Licht mehrmals von Objekten reflektiert wird, bevor es zurück zum Sensor gelangt. Das Entstehungsprinzip des Effektes ist in Abbildung 2.14 dargestellt. Dadurch, dass beim Time-of-Flight-Kameraprinzip die gesamte Szene mit aktivem Infrarotsignal beleuchtet wird, kommt es häufig zu Reflexionen zwischen Objektoberflächen, welche die Messung stören. Dabei werden die Distanzen aufgrund des längeren Weges, den das Licht zurücklegt, mit einem positiven Fehler belegt. In Abbildung 2.15 ist eine durch den *Multipath*-Effekt beeinflusste Messung zu sehen. Es kann beobachtet werden, dass die Messpunkte (rot) eine höhere Distanz aufweisen als die *Ground Truth* (grün).



Abbildung 2.13: Distanzbild einer von *Scattering* beeinflussten Aufnahme [41]. Insbesondere in dem Bereich, der sich links neben der Hand befindet, werden zu geringe Distanzen gemessen.

Auch bei Bewegung der zu messenden Objekte oder der Kamera kann es zu Mischphasen kommen. Da zur Bestimmung der Phasendifferenz mehrere konsekutive Messungen benötigt werden, kann es vorkommen, dass bei Bewegung während der Messung Phasensprünge zwischen den Lichtsignalen auftreten. Dieser Effekt wird durch hohe Integrationszeiten verstärkt.

2.1.4 Kamerakalibrierung

Bei Time-of-Flight-Kameras wird zwischen drei Formen der Kalibrierung unterschieden. Die intrinsische Kalibrierung bestimmt die internen Kameraparameter, mit denen jedem Bildpunkt in Pixelkoordinaten eine 3D-Blickrichtung in Bezug zum Kamerakoordinatensystem zugeordnet werden kann. Die Blickrichtung wird benötigt, um aus dem Distanzwert des Pixels einen 3D-Punkt zu generieren. Die extrinsische Kalibrierung bestimmt die Pose (Position und Ausrichtung) einer Kamera relativ zu einem externen Bezugskoordinatensystem. Dieses kann ein Weltkoordinatensystem oder auch ein Fahrzeugkoordinatensystem sein. Die Tiefenkalibrierung hingegen findet nur bei Tiefenkameras Anwendung und korrigiert Fehler in der

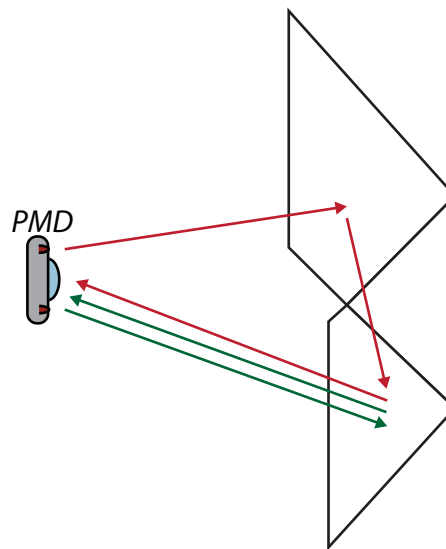


Abbildung 2.14: Entstehung des *Multipath*-Effekts

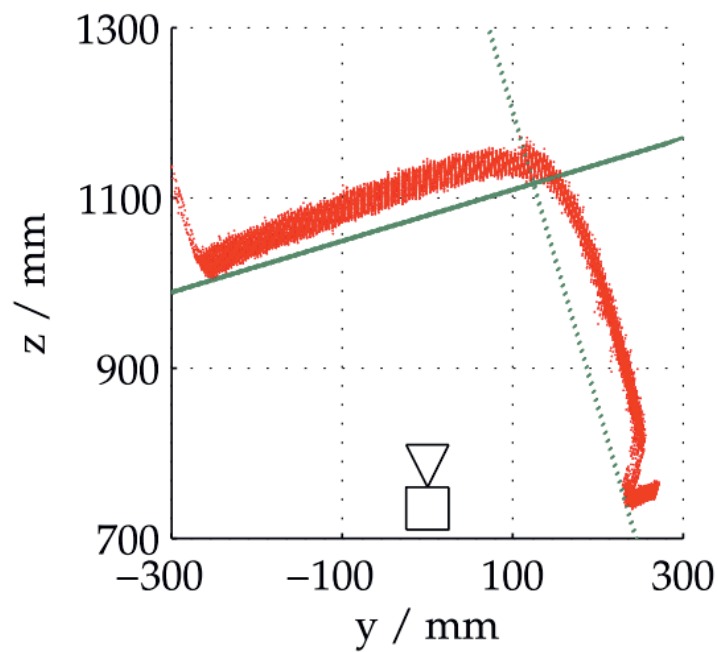


Abbildung 2.15: Messungen unter Einfluss des *Multipath*-Effekts [36]

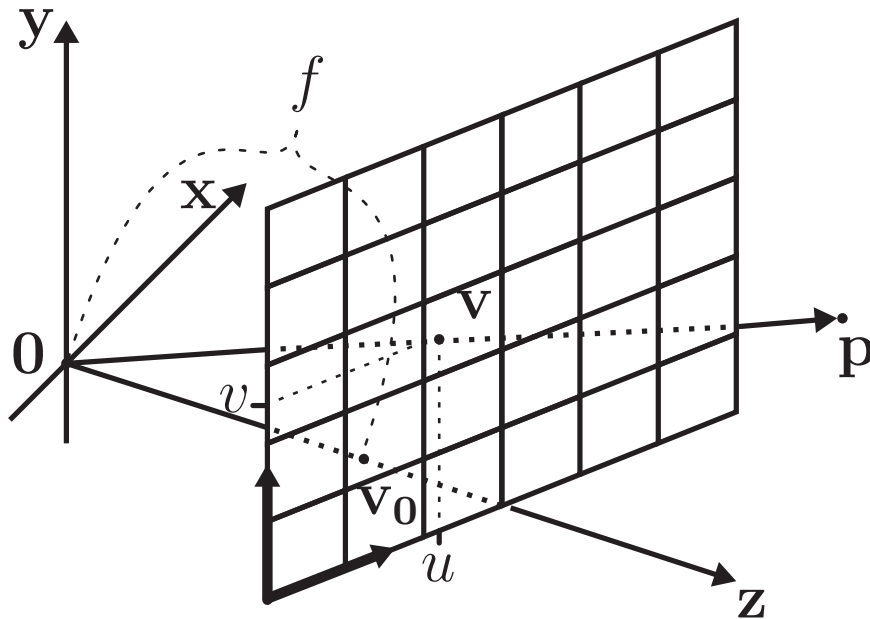


Abbildung 2.16: Das Lochkamera Modell

Distanzmessung. Der Stand der Technik für die Tiefenkalibrierung wird ausführlich in Abschnitt 2.2.1 thematisiert.

Die intrinsische und extrinsische Kalibrierung basiert auf einem mathematischen Kameramodell. Ein einfaches Modell ist das Lochkameramodell, das in Abbildung 2.16 vereinfacht dargestellt ist. Nach dem Lochkameramodell führen alle Lichtstrahlen durch einen gemeinsamen Punkt $\mathbf{0}$. Dieser Punkt wird Brennpunkt genannt und die Distanz f zur Bildebene wird als Brennweite bezeichnet. Der Brennpunkt bildet das optische Zentrum und den Ursprung des Kamerakoordinatensystems. Die zweidimensionale Projektion $\mathbf{v} = (u, v)^T$ eines Punktes $\mathbf{p} = (x, y, z)^T$ auf die Bildebene kann in homogenen Koordinaten durch

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{p} \quad (2.10)$$

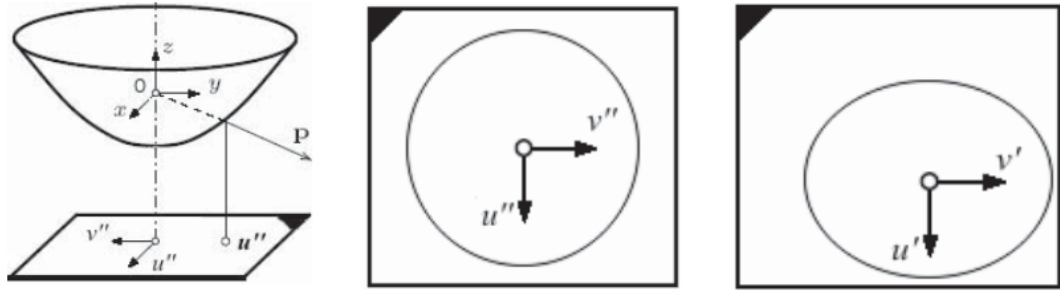


Abbildung 2.17: Das omnidirektionale Kameramodell nach Scaramuzza [43]

bestimmt werden. Dabei gilt $u = \frac{v_1}{v_3}$ und $v = \frac{v_2}{v_3}$.

Der Sensor einer Kamera entspricht der Bildebene und tastet die Helligkeitswerte in diskreten Pixeln ab. Daher wird die Brennweite f durch die Anzahl an horizontalen und vertikalen Pixeln w_p und h_p geteilt, sodass man $\alpha_u = \frac{f}{w_p}$ und $\alpha_v = \frac{f}{h_p}$ erhält. Eine zweite Adaption ist die Verschiebung des optischen Zentrums, das nicht notwendigerweise im Mittelpunkt der Bildebene liegt. Die Verschiebung wird durch den Vektor $\mathbf{v}_0 = (u_0, v_0)^T$ beschrieben. Das Kameramodell nach [42] gibt noch weitere Parameter an. Der Parameter s gibt die Scherung zwischen der x- und der y-Achse an, sodass man

$$\mathbf{v}' = \begin{pmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{p} \quad (2.11)$$

erhält. Die verzerrten Bildkoordinaten u_d und v_d können durch

$$u_d = u + (u - u_0)(k_1 r^2 + k_2 r^4) \quad (2.12)$$

und

$$v_d = v + (v - v_0)(k_1 r^2 + k_2 r^4) \quad (2.13)$$

mit dem Abstand r des Schnittpunkts zum optischen Zentrum nach [42] bestimmt werden. Die Parameter k_1 und k_2 geben die radiale Verzerrung an.

Im Fall von Weitwinkelobjektiven, die einen Öffnungswinkel von weit über 90° besitzen, bietet auch das vorgestellte Modell nach [42] keine passen-

de Annäherung. Daher existieren für Weitwinkelobjektive weitere Modelle wie das *Scaramuzza-Modell* [44]. Das Modell modelliert eine omnidirektionale Kamera mit einem paraboloiden Spiegel wie in Abbildung 2.17 dargestellt. Jedem Punkt $\mathbf{v} = (u'', v'')^T$ auf der Bildebene wird ein 3D-Vektor $\mathbf{p} = (x, y, z)$ zugewiesen, welcher der Blickrichtung des jeweiligen Pixels entspricht. Bei Übereinstimmung der Positionen von Spiegel und Bildebene kann der Zusammenhang beschrieben werden durch

$$\mathbf{p} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} u'' \\ v'' \\ f(r) \end{pmatrix}. \quad (2.14)$$

Die Funktion $f(r)$ modelliert den Zusammenhang zwischen Spiegelkrümmung und der z -Komponente von \mathbf{p} . Sie hängt ab von der Distanz eines Punktes zum Mittelpunkt der Bildebene $r = \sqrt{(u'')^2 + (v'')^2}$. Der Kern des Modells hängt von der Wahl der Funktion $f(r)$ ab. Nach Scaramuzza wird sie als Polynomfunktion modelliert mit

$$f(r) = a_0 + a_1 r + a_2 r^2 + a_3 r^3 + a_4 r^4 + \dots \quad (2.15)$$

Die Koeffizienten $a_0, a_1, a_2, a_3, a_4, \dots$ bilden die Parameter des Kameramodells. Das Modell geht bis hierhin von übereinstimmender Spiegel- und Kameraposition aus. Da dies in der Realität allerdings nicht der Fall ist, wird zusätzlich die Fehlansrichtung durch eine affine Transformation modelliert mit

$$\begin{pmatrix} u'' \\ v'' \end{pmatrix} = \begin{bmatrix} c & d \\ e & 1 \end{bmatrix} \begin{pmatrix} u' \\ v' \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \end{pmatrix}. \quad (2.16)$$

Hierdurch werden die realen verzerrten Koordinaten (u'', v'') mit den idealen Koordinaten (u', v') in Beziehung gesetzt. Als weitere Parameter entstehen die Variablen c, d, e, t_1 und t_2 der affinen Transformation.

Die intrinsische Kalibrierung einer Kamera bestimmt die Parameter des Kameramodells. Dies kann beispielsweise durch die Aufnahme eines Schachbrettmusters geschehen. Dabei wird ausgenutzt, dass die Kanten des Musters eine feste Distanz zueinander besitzen. Bei bekannten intrinsischen Kameraparametern kann die Pose des Schachbretts relativ zum Kamerakoordinatensystem aus einer einzelnen Aufnahme bestimmt werden. Wenn



die Pose des Schachbretts relativ zu einem Weltkoordinatensystem bekannt ist, können somit die extrinsischen Kameraparameter (die Kamerapose relativ zum Weltkoordinatensystem) bestimmt werden.

2.2 Datenverarbeitung von Time-of-Flight-Kameras

Im Rahmen dieser Arbeit werden neuartige Methoden zur Datenverarbeitung von Time-of-Flight-Kameras vorgestellt. Dieses Kapitel stellt den Stand der Technik bezüglich der vorgestellten Methoden vor. Es soll eine Einordnung der in den nachfolgenden Kapiteln präsentierten Methoden in den wissenschaftlichen Stand der Technik ermöglichen.

2.2.1 Tiefenkalibrierung

Eine intrinsische und extrinsische Kalibrierung einer Time-of-Flight-Kamera kann mit den bekannten Verfahren herkömmlicher Kameras erfolgen. Die Tiefendaten allerdings weisen aufgrund des Messverfahrens charakteristische Fehler auf. Systematische Fehler können durch eine Tiefenkalibrierung korrigiert werden. Nachfolgend werden bekannte Verfahren vorgestellt.

In [45] wird die Distanzkalibrierung einer PMD-Kamera beschrieben. Eine Korrekturfunktion in Form einer *B-Spline*-Funktion wird durch Referenzaufnahmen bestimmt. Dabei kann diese Funktion zum einen global bestimmt werden und zum anderen für jedes Pixel einzeln. Eine Berücksichtigung der Amplitude erfolgt nicht. Korrigiert wird in erster Linie der distanzabhängige *Wiggling*-Fehler. Dieser Ansatz wird in [38] und [46] durch die Erstellung eines zusätzlichen Fehlermodells für den amplitudenabhängigen Fehler erweitert. In [47] erfolgt die Korrektur des systematischen Distanzfehlers durch eine abweichende Demodulationstechnik. Die präsentierte Methode erreicht zwar nicht die Leistung vorheriger Methoden, benötigt allerdings deutlich weniger Referenzdaten.

[48] präsentiert eine Methode, in welcher zunächst eine Korrektur des Amplitudenbildes auf Basis des Distanzbildes erfolgt. Anschließend wird das Distanzbild für Pixel mit einem niedrigen Amplitudenwert korrigiert. Dies

erfolgt auf Basis einer durch Referenzaufnahmen erstellten distanzabhängigen Funktion. In [49] erfolgt die Tiefenkalibrierung durch die Herleitung eines Fehlermodells, das einen amplitudenabhängigen Fehler, einen distanzabhängigen Fehler und einen durch Latenz zwischen Lichtquelle und Signalgenerator verursachten Fehler enthält. Der letztgenannte Fehler wird durch einen pixelabhängigen Offset korrigiert.

Die Kalibrieremethode nach [50] verwendet *Look-Up Tables* zur Korrektur der Tiefendaten. Korrigiert wird der *Wiggling*-Effekt und ein integrationszeitabhängiger Offset. Es wird erkannt, dass die Effekte von verschiedenen Amplituden, gemessenen Distanzen und Integrationszeiten abhängen. Um dies zu berücksichtigen, werden mehrdimensionale *Look-Up Tables* generiert.

[51] stellt eine Kompensation des *Scattering*-Effekts vor. Der Effekt von *Scattering* wird zunächst analysiert und als lineare Transformation eines komplexen Bildes dargestellt. Anschließend wird ein vereinfachtes *Scattering*-Modell hergeleitet, das genutzt wird, um den Effekt zu kompensieren. Auch in [52] wird eine Kalibrieremethode zur Korrektur des *Scattering*-Effekts präsentiert. Hierbei wird eine Fehlerfunktion während des Kalibrierverfahrens um weitere Parameter zur Kompensation des *Scattering*-Effekts ergänzt.

2.2.2 Konfidenzbestimmung

Bei der Datenverarbeitung von verrauschten Sensordaten ist es in der Regel hilfreich, ein Maß für die Güte der Messung zu besitzen. Wie schon in Abschnitt 2.1.3 vorgestellt, korreliert die Amplitude mit dem Signal-Rausch-Verhältnis. Daher ist es üblich, den Amplitudenwert als Konfidenzwert für die Güte der Messung zu behandeln. Diese Vorgehensweise berücksichtigt allerdings keine systematischen Fehler oder auch Mischphaseneffekte, wie das *Scattering* oder *Flying Pixels*.

[53] präsentiert ein Modell zur Vorhersage der systematischen und stochastischen Fehler. Hier werden die Effekte von distanzabhängigen und amplitudenabhängigen Fehlern analysiert. Ein Konfidenzwert wird allerdings nicht bestimmt. Auch in [54] wird ein Modell für den statistischen Fehler der Distanzmessungen von Time-of-Flight-Kameras beschrieben. Berücksichtigt werden hierzu Amplitude, Intensität und Distanz. Dabei wird das

Signal-Rausch-Verhältnis durch einen *Maximum-Likelihood*-Schätzer geschätzt.

In [55] erfolgt eine experimentelle Untersuchung des *Scattering*-Effekts. Insbesondere werden die Auswirkungen des *Scatterings* auf die Werte für die Amplitude und Distanz geschätzt. Ein Kalibrierverfahren, das den *Scattering*-Effekt berücksichtigt, wird in [52] präsentiert. Hierzu wird ein Referenzobjekt verwendet, das aus zwei Tafeln besteht, die unterschiedliche Distanzen besitzen. Durch Referenzaufnahmen werden vier *Scattering*-Parameter für ein Fehlermodell bestimmt.

Der *Multipath*-Effekt wird in [36] modelliert. Dies geschieht durch einen am *Raytracing* angelehnten Ansatz, durch welchen für jedes Pixel getrennt zusätzliche Einflüsse durch *Multipath*-Effekte geschätzt werden können. In [56] wird eine Methode zur Erkennung von *Flying Pixels* vorgestellt. Diese werden durch die Anwendung eines adaptierten *Canny Edge Detectors* auf das Tiefenbild identifiziert. Das allgemeine Problem der Mehrwege-Effekte wird in [57] angegangen. Die Methode zur Bestimmung der Phasendifferenz wird durch ein Optimierungsproblem, welches die Daten von zusätzlichen Aufnahmen nutzt, erweitert. Auf diese Weise können durch Mehrwege-Effekte entstandene Signale von dem Signal der direkten Reflexion getrennt werden.

Ein Ansatz, welcher auf Methoden des maschinellen Lernens basiert, wird in [58] präsentiert. Um referenzierte Aufnahmen zu erhalten, wird eine PMD-Kamera mit einem Laserscanner gekoppelt. Der Laserscanner liefert *Ground-Truth*-Daten zu den Tiefenbildern der Kamera. Während des Lernvorgangs werden verschiedene Merkmale eines Pixels und die Abweichung des Distanzwerts zur wahren Distanz bestimmt. Auf Basis der extrahierten Merkmale wird ein *Random Forest* [59] trainiert. Der Konfidenzwert richtet sich nach der Abweichung der Messungen zur wahren Distanz. Nach erfolgreichem Training kann der *Random Forest* genutzt werden, um Pixeln mit bestimmten Merkmalen einen Konfidenzwert zuzuweisen.

2.2.3 Umgebungsmodellierung

Zur Modellierung der Umgebung durch verrauschte Sensordaten existieren verschiedene Datenstrukturen. Weit verbreitet sind probabilistische *Occupancy Grids*, welche die Umgebung durch Belegheitswahrscheinlichkeiten modellieren. 2.5D-Karten hingegen zielen auf die Schätzung der

Höhe von Objekten in der Umgebung ab. Ebenso existieren Methoden zur dreidimensionalen Modellierung einer Umgebung. Nachfolgend werden die verschiedenen Möglichkeiten zur Umgebungsmodellierung vorgestellt.

Occupancy Grid Maps

Occupancy Grid Maps modellieren die Umgebung als ein probabilistisches reguläres Gitter [60]. Üblicherweise werden zweidimensionale Gitter verwendet, allerdings existieren auch dreidimensionale Ansätze [61]. Jede Gitterzelle repräsentiert eine Wahrscheinlichkeit über die Belegtheit dieser Zelle. Dabei kann ein Wert von 0 beispielsweise bedeuten, dass die Zelle mit Sicherheit frei ist, und ein Wert von 1, dass die Zelle mit Sicherheit belegt ist. Die Messungen von Sensoren, welche die Umgebung wahrnehmen, werden fortlaufend in die Datenstruktur integriert. Dies geschieht im Allgemeinen über ein inverses Sensormodell, welches die Charakteristik des jeweiligen Sensors berücksichtigt [62]. Dabei wird die Belegtheitswahrscheinlichkeit einer Zelle als a-posteriori-Wahrscheinlichkeit bei gegebenen Messungen beschrieben [63]. Die Belegtheitswahrscheinlichkeiten der einzelnen Zellen werden als unabhängig voneinander betrachtet, sodass das Gesamtproblem der Kartenerstellung auf mehrere Teilprobleme der Schätzung der einzelnen Belegtheitswerte aufgeteilt werden kann. Oft wird die Belegtheitswahrscheinlichkeit als sogenannter *Log Odd* gespeichert. Dies entspricht dem Logarithmus des Quotienten aus der Wahrscheinlichkeit und der Gegenwahrscheinlichkeit und hat den Vorteil, dass bei einer Aktualisierung der Zelle Multiplikationen durch Additionen ersetzt werden können.

Occupancy Grid Maps können für die Umfelderkennung von Fahrerassistenzsystemen verwendet werden [64].

2.5D-Karten

2.5D-Karten sind reguläre zweidimensionale Gitter mit Höheninformationen. Jede Gitterzelle speichert einen Wert für die Höhe. In [65] und [66] werden 2.5D-Karten für die Wahrnehmung eines planetaren Fahrzeuges genutzt. Zunächst werden Tiefendaten aus Stereokameras erstellt. Diese werden einzeln in 2.5D-Karten umgewandelt und anschließend fusioniert. Zusätzlich zur Höheninformation wird ein Unsicherheitswert dieser Information gespeichert.



In [67] wird eine Erweiterung des *Occupancy-Grid*-Ansatzes um einen Wert für die Höhe vorgestellt. Ziel ist die Vermeidung von hohen Speicheranforderungen von *3D Occupancy Grids*. Die Verwendung einer 2.5D-Karte in einem humanoiden Roboter wird in [68] präsentiert. Die 2.5D-Karte wird aus den Daten einer Stereokamera generiert und dient zur Schrittplanung des Roboters. Diesen Ansatz erweitert [69] durch ein dreidimensionales *Occupancy Grid*. Dieses wird zur Verifizierung der Höheninformation der 2.5D-Karte genutzt. Auch [70] stellt einen humanoiden Roboter vor, dessen interne Repräsentation der Umwelt durch eine 2.5D-Karte realisiert wird. Externe Sensorik lokalisiert den Roboter. Die Wahrnehmung der Umgebung geschieht durch eine am Roboter angebrachte Time-of-Flight-Kamera. Auf Basis der Tiefendaten wird zeitgleich eine 2.5D-Karte und eine *2D Occupancy Grid Map* erstellt.

Eine Erweiterung von 2.5D-Karten wird in [71] vorgestellt. Sie hat zum Ziel, vertikale Objekte und Objekte, die einen Überhang besitzen oder gar unterfahrbar sind, zu berücksichtigen. Hierzu werden die einzelnen Gitterzellen in verschiedene Klassen eingeteilt, was durch eine Analyse der Punkte geschieht, die in eine Gitterzelle fallen. [72] stellt eine 2.5D-Karte vor, in der in jeder Gitterzelle ein Belegtheitswert, eine Höheninformation und zusätzlich die Varianz des Höhenwertes gespeichert werden.

3D-Kartierung

Eine einfache Datenstruktur zur Darstellung einer dreidimensionalen Umfeldkarte ist eine globale Punktwolke. [73] erstellt aus den Daten einer Time-of-Flight-Kamera eine unorganisierte 3D-Punktwolke. Die Registrierung der Punktwolken erfolgt durch eine Erweiterung des ICP-Algorithmus. In [74] werden herkömmliche Farbkameras mit einer Time-of-Flight-Kamera kombiniert. Aus dem Kamerabild werden 2D-Merkmalpunkte extrahiert, die anschließend durch Nutzung der Tiefendaten der Time-of-Flight-Kamera in 3D-Merkmalpunkte umgewandelt werden. Das Verfahren verwendet anschließend den *fastSLAM*-Ansatz [75]. Dieser beruht auf der Annahme, dass, wenn der wahre Pfad eines mobilen Agenten bekannt ist, das SLAM-Problem auf eine Schätzung von unabhängigen Merkmalspunkten reduziert werden kann. Die Merkmalspunkte werden zusammen mit den restlichen 3D-Punkten in einer Punktwolke gespeichert. [76] stellt ein Verfahren vor, das unter Verwendung einer *Kinect*-Kamera eine globale

Punktwolke und geometrische Objekte in einer Karte speichert. Dies geschieht durch die Extraktion von geometrischen Merkmalen mithilfe eines RANSAC-Algorithmus. Eine dichte 3D-Punktwolke wird auch in [77] aus den Daten einer Stereokamera gebildet. Hier wird ein Matching-Verfahren für Merkmalspunkte mit einem Ansatz zur visuellen Odometrie kombiniert, um eine konsistente globale Punktwolke zu generieren. Aus einer dichten Punktwolke kann ein 3D-Polygonmodell erstellt werden. [78] und [79] präsentieren verschiedene Ansätze, um aus unorganisierten Punktwolken ein 3D-Polygonmodell zu erstellen.

Auch in [80] wird die Erstellung einer globalen Punktwolke durch die Daten von Time-of-Flight-Kameras vorgeschlagen. Zusätzlich werden allerdings dreidimensionale planare Oberflächensegmente aus der Punktwolke extrahiert. Beder [81] hingegen erstellt sogenannte *Patchlets* direkt aus den Daten von Stereo- und PMD-Kameras. *Patchlets* sind orientierte Flächensegmente, die eine Schätzung für die Position und die Ausrichtung enthalten. Durch *Patchlets* kann eine dreidimensionale Umgebung modelliert werden.

Kim [82] stellt ein Verfahren vor, welches mehrere Farbkameras mit einer Time-of-Flight-Kamera kombiniert. Der Nachteil von Stereosystemen bei untexturierten Oberflächen kann somit durch die Tiefendaten der Time-of-Flight-Kamera kompensiert werden. In [83] hingegen wird die Umgebung durch die Daten einer einzelnen konventionellen Kamera dreidimensional rekonstruiert. Durch *Structure-from-Motion* werden die Kameraposen der einzelnen Aufnahmen bestimmt und zugehörige Tiefenbilder generiert. Aus den Tiefendaten wird anschließend ein Polygonmodell erstellt.

Die Erweiterung einer zweidimensionalen Gitterstruktur ins Dreidimensionale kann durch ein Voxelgitter realisiert werden. Beispielsweise wird in [84] auf Basis der Tiefendaten einer Stereokamera ein dreidimensionales Voxelgitter aufgebaut. Zusätzlich werden Geometrie-Eigenschaften der Umgebung in den einzelnen Voxeln gespeichert. Ein Polygonmodell kann zum Beispiel durch den *Marching-Cubes*-Algorithmus [85] aus einem 3D-Voxelgitter erstellt werden. Wurm [61] erweitert den *Occupancy-Grid*-Ansatz ins Dreidimensionale durch die Verwendung eines *Octrees*, die in [86] vorgestellt werden. Die resultierende Datenstruktur wird *OctoMap* genannt.

Ein alternatives volumetrisches Modell wird in [87] vorgeschlagen. Sogenannte *Truncated Signed Distance Functions* (TSDF) werden in den Vo-

xeln eines Voxel-Gitters gespeichert. Die *Signed Distance Function* gibt den Abstand eines Punktes zu einer Oberfläche an. Die aufeinanderfolgenden Tiefenbilder werden in dreidimensionale Distanzfunktionen konvertiert und mit den Daten des Voxelgitters kombiniert. Aus dem Voxelgitter kann anschließend eine Isofläche extrahiert werden. Eine Möglichkeit zur Extraktion eines Polygonmodells wird in [88] präsentiert. Zunächst wird eine erste Approximation durch das *Marching-Cubes*-Verfahren erreicht. Anschließend werden die entstehenden Polygone weiter unterteilt, bis die gewünschte Genauigkeit erreicht wird. Graber [89] beschreibt die Erstellung von 3D-Modellen durch die Nutzung von TSDF anhand einer Mono-Kamera. Dabei werden die Tiefenbilder aus *Structure from Motion* generiert. *Kinect Fusion* erstellt die TSDF aus den Daten einer *Kinect*-Kamera [90]. Zur Registrierung der Tiefenbilder wird ein ICP-Algorithmus verwendet. Eine GPU-Implementierung ermöglicht die Erstellung von 3D-Modellen in Echtzeit. In [91] wird der *Kinect-Fusion*-Ansatz um eine Anbindung von Stereo-Kameras erweitert, sodass auch Outdoor-Szenarien im Sonnenlicht brauchbare Ergebnisse liefern. Eine Erweiterung von *Kinect Fusion* wird auch in [92] vorgestellt. Hier wird ein zyklischer Puffer verwendet, um die örtliche Limitation der ursprünglichen Methode zu vermeiden. Ebenso wird der Ansatz um eine Datenreduktion durch eine Flächensegmentierung ergänzt und der ICP-Algorithmus wird durch ein Verfahren zur visuellen Odometrie ersetzt. In [93] wird die ursprüngliche Idee des *Kinect-Fusion*-Ansatzes beibehalten, allerdings wird die TSDF-Datenstruktur ausgetauscht. Das globale 3D-Modell besteht hier aus einer ungeordneten Liste von 3D-Punkten.

2.2.4 Flächenextraktion

Unter der Flächenextraktion wird die Detektion von Flächensegmenten und die Bestimmung der Flächenparameter aus den Daten von Time-of-Flight-Kameras verstanden. Eine bekannte Methode zur Bestimmung von Modellparametern aus fehlerbehafteten Daten ist das *Random-Sample-Consensus*-Verfahren (RANSAC) [94]. Das RANSAC-Verfahren kann genutzt werden, um die Parameter von Flächen in verrauschten 3D-Punktwolken zu bestimmen. Das Verfahren unterteilt die Menge der 3D-Punkte in *Inlier* und *Outlier*. Während *Inlier* eine Ebenenhypothese unterstüt-

zen, weil sie zur aufgestellten Hypothese passen, gelten *Outlier* als Punkte, die diese Hypothese nicht unterstützen. Um nun mehrere Ebenen zu erkennen, müssen die *Outlier* in Ausreißer und Pseudo-Ausreißer unterteilt werden [95]. Pseudo-Ausreißer sind dabei Punkte, die als Ausreißer für eine Ebenenhypothese gelten, gleichzeitig aber *Inlier* zu einer anderen Ebenenhypothese sind. Das *Sequential-RANSAC*-Verfahren [96] kann mit Pseudo-Ausreißern umgehen, indem der RANSAC-Algorithmus wiederholt auf denselben Datensatz ausgeführt wird, nachdem die *Inlier* der vorherigen Iteration aus dem Datensatz entfernt wurden. Eine sequentielle Durchführung führt allerdings bei fehlerhafter *Inlier*-Bestimmung zu großen Instabilitäten. Diesen Nachteil behebt das *MultiRansac*-Verfahren [97]. Hier werden die verschiedenen Hypothesen parallel aufgestellt. Allerdings besitzt dieses Verfahren den Nachteil, dass die Anzahl an Ebenen a priori bekannt sein muss. Eine weitere Variation des *Sequential-RANSAC*-Verfahrens ist das *CC-RANSAC*-Verfahren [98]. Beim *CC-RANSAC* wird davon ausgegangen, dass *Inlier* zu einer Fläche einen zusammenhängenden Bereich bilden müssen. Dies ist beim allgemeinen RANSAC-Verfahren nicht der Fall, allerdings für den Fall der Extraktion von Flächen zwingend erforderlich. Das *Preemptive-RANSAC*-Verfahren ist für Echtzeitanwendungen ausgelegt [99]. Es erweitert das RANSAC-Verfahren um eine vorherige Bewertung der aufgestellten Hypothesen und schließt Hypothesen mit geringen Konsistenzen zum Datensatz aus. Der *J-Linkage*-Algorithmus [100] kombiniert das RANSAC-Verfahren mit den Vorteilen der *Hough*-Transformation. Hier werden zunächst verschiedene Modellhypothesen aufgestellt und für jede die zugehörige Menge an *Inlier*-Punkten bestimmt. Punkte, die dieselben Modellhypothesen teilen, werden zu einem Cluster zusammengesetzt.

Ein abweichender Ansatz wird in [101] vorgeschlagen. Lineare Segmente werden entlang von vertikalen und horizontalen *Scanlines* aus dem Tiefenbild extrahiert und anschließend zu Flächensegmenten kombiniert. Anschließend wird für jedes Flächensegment eine Regressionsebene bestimmt. Ein Nachteil dieser Methode ist eine hohe Übersegmentierung während der Bildung der Flächensegmente.

In [102] wird die Extraktion von Flächensegmenten aus Tiefenbildern durch ein *Region-Growing*-Verfahren beschrieben.

Die verschiedenen Varianten des RANSAC-Verfahrens werden als globale Verfahren bezeichnet, da sie auf Basis des gesamten Datensatzes Hypo-



thesen aufstellen und diese anschließend bewerten. In lokalen Verfahren, wie dem *Scanline*-Verfahren, werden hingegen Beziehungen zwischen einzelnen Daten aufgebaut.

2.2.5 Lokalisierung

Durch Kamerasysteme kann die Lokalisierung eines Fahrzeugs unterstützt werden. Bei der visuellen Odometrie wird die Eigenbewegung des Fahrzeugs ausschließlich über die Kamerabilder geschätzt. Beim *Simultaneous Localisation and Mapping* (SLAM) wird zugleich eine Karte der Umgebung erstellt, und über einen Abgleich der Karte mit den Daten von Sensoren wird die Position des Fahrzeugs bestimmt.

Visuelle Odometrie

Visuelle Odometrie beschreibt die inkrementelle Schätzung der Pose einer Kamera oder eines Fahrzeugs durch die Beobachtungen der Änderungen in den Kamerabildern, welche durch Bewegung verursacht werden. Ein Überblick über verschiedene Methoden auf Basis von Mono- und Stereo-Kameras ist in [103] gegeben.

Verfahren zur visuellen Odometrie auf Basis von Mono-Kameras basieren in der Regel darauf, dass der optische Fluss [104] berechnet wird oder Merkmalspunkte extrahiert werden [105]. Es werden Beziehungen zwischen den Merkmalspunkten aufeinanderfolgender Kamerabilder generiert, und auf Basis der Kamerageometrie kann durch Triangulation auf die dreidimensionale Struktur der Merkmalspunkte geschlossen werden. Anschließend kann die Kamerapose bestimmt werden.

In [106] wird ein solches Verfahren um eine lokale Bündeloptimierung ergänzt und für die Lokalisierung eines Fahrzeugs verwendet. Klein [107] präsentiert eine Methode, welche die extrahierten Merkmalspunkte in einer Karte speichert und somit jedes Kamerabild mit den Merkmalspunkten der Karte vergleicht. Scaramuzza [44] beschreibt ein Verfahren zur visuellen Odometrie auf Basis einer omnidirektionalen Kamera.

Stereo-Kameras bieten zusätzliche Tiefeninformationen, welche die Bestimmung der Kamerapose unterstützen können [108]. Huang [109] stellt ein Verfahren zur visuellen Odometrie vor, welches auf den Daten einer *Kinect*-Kamera basiert. Die *Kinect*-Kamera stammt von der Firma *Micro-*

soft und wurde für die Steuerung von Videospielen entwickelt. Sie liefert neben hochauflösenden 2D-Bildern auch Tiefenbilder, die durch strukturiertes Licht bestimmt werden.

Auch Time-of-Flight-Kameras liefern Tiefendaten in Echtzeit. Dies allerdings in einer geringeren Auflösung als bei herkömmlichen Kamerasystemen. Prusak [110] kombiniert eine Time-of-Flight-Kamera mit einer Mono-Kamera für ein Verfahren zur visuellen Odometrie. Während die hochauflösende Mono-Kamera zur Extraktion und zum Matching von Merkmalspunkten herangezogen wird, liefert die Time-of-Flight-Kamera die 3D-Positionen der Merkmalspunkte. Die Extraktion von Merkmalspunkten aus den Amplitudenbildern von Time-of-Flight-Kameras wird in [111] und [112] eingesetzt. Durch die Tiefeninformationen können die zweidimensionalen Merkmalspunkte in dreidimensionale Merkmalspunkte umgerechnet werden. Während [111] Matches zwischen aufeinanderfolgenden Kamerabildern bildet, werden die dreidimensionalen Merkmalspunkte in [112] in einer globalen Karte gespeichert. Mian [113] stellt einen Merkmalsdeskriptor vor, welcher auf der dreidimensionalen Oberflächenstruktur von Tiefenbildern basiert. Durch Bildung von Korrespondenzen können die Tiefenbilder anschließend aufeinander registriert werden.

Eine weitere Möglichkeit zur Bestimmung der Kamerapose auf Basis von Tiefenbildern besteht in der dreidimensionalen Registrierung von Punktwolken. Hierzu kann der *Iterative-Closest-Point-Algorithmus* (ICP) [114] eingesetzt werden. In [115] wird eine Variante des ICP-Algorithmus verwendet, um Tiefenbilder auf eine dreidimensionale Oberfläche zu registrieren. Winkelbach [116] präsentiert einen Ansatz, welcher auf der RANSAC-Methode basiert. Es wird wiederholt zufällig eine wahrscheinliche Hypothese für die Lage der Punktwolke bestimmt, und anschließend wird die Qualität dieser Hypothese evaluiert.

Simultaneous Localisation and Mapping

Das SLAM-Problem beschäftigt sich mit der simultanen Erstellung einer Repräsentation der Umgebung eines mobilen Roboters und der Lokalisierung innerhalb dieser Umgebung [63].

Eine Möglichkeit zur Lösung des SLAM-Problems bietet der *EKF SLAM* [117]. Dieser basiert auf der Verwendung eines erweiterten Kalman-Filters (EKF). Geschätzt wird dabei nicht nur die Position des Roboters, son-

dern auch einzelne Landmarken in der Umgebung. Daher bietet das EKF-SLAM-Verfahren keine dichte Kartenrepräsentation wie im Falle einer *Occupancy Grid Map*, sondern eine Menge an Landmarken. In jedem Schritt werden die Positionen aller Landmarken durch den Kalman-Filter neu geschätzt. Das Verfahren ist gut geeignet, um Ringschlüsse bei der Kartenerstellung zu detektieren.

Das *Maximum-Likelihood*-SLAM-Verfahren ermöglicht die Repräsentation der Umgebung als dichte gitterbasierte Struktur [118]. Zu jedem Zeitpunkt, an dem eine Messung des Umfeldes geschehen ist, wird die wahrscheinlichste Position des Roboters bestimmt. Im Anschluss wird die Karte durch die neue Messung auf Basis der geschätzten Messung aktualisiert. Die Karte wird somit inkrementell aufgebaut, und es finden nur lokale Änderungen statt. Daher ist dieses Verfahren sehr rechen- und speichereffizient.

Vu [119] beschreibt ein *Maximum-Likelihood*-SLAM-Verfahren auf Basis der Daten von Laserscannern. Als Karte kommt eine *Occupancy Grid Map* zum Einsatz. Die wahrscheinlichste Pose wird bestimmt, indem um die prädizierte Pose herum Positionsstichproben (*Samples*) gestreut und anschließend bewertet werden. Bewertet wird anhand der aktuellen Messung. Jeder Messpunkt des Laserscanners wird hierzu in die Karte projiziert. Anschließend werden die Belegtheitswerte der einzelnen Zellen aufsummiert. In [120] wird eine erweiterte Methode zur Streuung von *Samples* vorgestellt. Die Methode basiert auf einem gewichteten *Resampling* analog zum Partikelfilter-Ansatz. Nach einer ersten Streuung und Bewertung von *Samples* findet eine erneute Streuung mit geringerer Varianz in der Nähe der *Samples* statt, denen eine hohe Bewertung zugewiesen wurde. Beim sogenannten *FastSLAM*-Verfahren kommt ein Partikelfilter zum Einsatz [75]. Dabei wird in jedem Partikel nicht nur eine Position gespeichert, sondern zusätzlich eine Karte. Diese Karte kann beispielsweise eine *Occupancy Grid Map* sein. Nach dem Erhalt einer Messung durch Sensoren wird zunächst die Position durch ein Bewegungsmodell prädiziert. Um die prädizierte Position werden anschließend zufällig Partikel verteilt. Auf Basis der aktuellen Karte werden die einzelnen Partikel nach einer Transformation der aktuellen Messung zur Position des Partikels bewertet. Die Karte wird anschließend durch die Berücksichtigung der aktuellen Messung für jedes Partikel einzeln aktualisiert. Im nachfolgenden *Resampling*-Schritt wird zufällig eine Teilmenge der Partikel (gewichtet nach der Bewertung

der einzelnen Partikel) ausgewählt und somit für die nachfolgende Messung weiter verfolgt.

In [121] wird eine Variante des *FastSLAM*-Algorithmus vorgestellt, welche auf die Speicherung einer gesamten Karte für jedes Partikel verzichtet. Stattdessen wird ein Vererbungsbaum erstellt, welcher die Partikel untereinander in Beziehung setzt. Der Vorteil der Methode ist eine gesteigerte Effizienz bezüglich Rechenzeit und Speicherbedarf. Nachteilig ist hingegen die Tatsache, dass nur ein binäres Belegungsgitter repräsentiert werden kann.

2.2.6 Detektion von dynamischen Objekten

Die Detektion von dynamischen Objekten durch Kameradaten wird insbesondere erschwert, wenn die Kamera an einem sich selbst bewegenden Fahrzeug angebracht ist. Im Folgenden werden Verfahren zur Erkennung von dynamischen Objekten vorgestellt. Zunächst wird auf Methoden eingegangen, die mit Mono-Kameras umgehen, und anschließend auf Verfahren, welche die Daten von Tiefenkameras nutzen.

In [122] wird ein Verfahren vorgestellt, welches auf Basis einer Farbkamera die Berechnung des optischen Flusses [104] mit einem Segmentierungsalgorithmus kombiniert. Dabei ist die Kamera auf dem sich bewegenden Fahrzeug montiert. Um Rechenzeit zu sparen, wird der optische Fluss nur für Bildbereiche berechnet, die mit hoher Wahrscheinlichkeit Objekte beinhalten. Anschließend werden Objekte durch ein *k-means-Clustering*-Verfahren [123] rekonstruiert.

Die in [124] präsentierte Methode verbindet die Berechnung des optischen Flusses mit dem Aufbau eines Hintergrund-Modells. In einem ersten Schritt wird der optische Fluss zwischen aufeinanderfolgenden Bildern berechnet. Pixel mit ähnlicher Bewegungsrichtung werden zu *Moving Blobs* zusammengefügt. Im zweiten Schritt werden mehrere Hintergrund-Modelle fusioniert, um die Umrisse der *Moving Blobs* zu verfeinern.

Die Detektion von dynamischen Objekten aus den Daten einer an einem Fahrzeug befestigten Stereo-Kamera wird in [125] und [126] beschrieben. Wenn die Eigenbewegung des Fahrzeugs bekannt ist, wird der zu erwartende optische Fluss der Kamerabilder prädiziert und mit dem gemessenen optischen Fluss verglichen. An statischen Objekten wird die Differenz zwi-



schen vorhergesagtem und berechnetem optischen Fluss nahe null sein. An den Stellen, an denen er abweicht, werden bewegende Objekte detektiert. In [80] wird eine an einem Roboter angebrachte Time-of-Flight-Kamera verwendet, um dynamische Objekte zu erkennen. Aus den 3D-Daten der Kamera werden dreidimensionale Geschwindigkeiten generiert und diese zu Clustern mit gleichmäßiger Geschwindigkeit segmentiert. Diese Cluster werden als dynamische Objekte betrachtet und zeitlich verfolgt und lokalisiert.

Auch im in [127] präsentierten Ansatz kommt eine Time-of-Flight-Kamera zum Einsatz. Diese ist hinter der Windschutzscheibe eines Kraftfahrzeugs angebracht. Hindernisse werden nach einer Segmentierung durch einen *Region-Growing*-Ansatz aus dem Tiefenbild der PMD-Kamera extrahiert. Den so ermittelten Segmenten werden Eigenschaften zugeordnet wie Schwerpunkt und Größe. Anschließend werden die Segmente in eine Objektliste eingefügt und in das Fahrzeugkoordinatensystem transformiert. Um die Objekte besser zu verfolgen, wird der Methode in [128] noch ein Kalman-Filter hinzugefügt.

Eine hinter der Windschutzscheibe angebrachte PMD-Kamera wird auch im in [129] vorgestellten Ansatz zur Hinderniserkennung eingesetzt. Ähnlich zum vorherigen Verfahren werden Segmente aus den Tiefendaten gebildet. Zusätzlich werden die ermittelten Objekte klassifiziert. Dies geschieht durch die Extraktion von Features wie dem Trägheitstensor, der Orientierung der einhüllenden Ellipse und einem *Fourier*-Deskriptor.

2.3 Diskussion des Standes der Technik

Time-of-Flight-Kameras mit Weitwinkelobjektiven werden in der Literatur nicht beschrieben. Diese Arbeit beschreibt ein **PMD-Kamerasystem mit einem Sichtbereich von 180°** und schließt damit diese Lücke im Stand der Technik. Während die Effekte von Weitwinkelobjektiven auf konventionelle Kamerasysteme ausführlich untersucht wurden, fehlt diese Untersuchung für den Fall von Time-of-Flight-Kameras. Weitwinkelobjektive erfordern aufgrund der notwendigen aktiven Beleuchtung eine neuartige Methode zur **Systemauslegung eines omnidirektionalen Time-of-Flight-Kamerasystems durch die Simulation der Lichtausbeute**. Die Systemauslegung wird in diesem Fall für ein System zur Umfeldwahr-

nehmung beim Parken und Rangieren von Automobilen durchgeführt. Allerdings ist sie auf jegliche mobile Agenten übertragbar.

Zur intrinsischen und extrinsischen Kalibrierung von Weitwinkel-Kameras existieren bekannte Verfahren, die zur Kalibrierung von Weitwinkel-PMD-Kameras verwendet werden können. Die Tiefenkalibrierung von PMD-Kameras allerdings ist noch nicht in ausreichendem Maße in der Forschung beschrieben. Berücksichtigt werden Verfahren zur Kalibrierung des amplituden- und distanzabhängigen Fehlers. Auch werden Verfahren zur Korrektur des *Scattering*-Effekts beschrieben. Diese sind allerdings sehr rechenaufwendig und können nicht in Echtzeit umgesetzt werden. Auch die Temperatur der Lichtquellen wird in der bekannten Literatur nicht berücksichtigt. Ebenso fehlt eine Untersuchung einer pixelabhängigen Korrektur im aktuellen Stand der Technik. In dieser Arbeit wird daher ein neuartiges **Verfahren zur Kalibrierung der Tiefendaten** präsentiert. Die Kalibrieremethode soll neben der Amplitude und Distanz auch die Temperatur und die Pixelposition berücksichtigen.

Messfehler, die nicht durch eine Kalibrierung korrigiert werden können, werden in einer ebenso neuartigen **Methode zur Berechnung einer Konfidenzmaske** beachtet, die jedem Pixel einen Konfidenzwert zuweist und Effekte berücksichtigt, die in der Literatur bisher unberücksichtigt bleiben. Hierzu gehören die Untersuchung des Fremdlichts auf die Genauigkeit einer Messung und die Integration von verschiedenen Messfehlern in einen einzelnen Konfidenzwert.

Zur Umfeldmodellierung existieren verschiedene Datenstrukturen. *2D Occupancy Grid Maps* modellieren die Umgebung probabilistisch, allerdings ohne Höheninformationen zu berücksichtigen. Hierzu eignen sich 2.5D-Karten, welche in jeder Gitterzelle einen Höhenwert speichern. Zusätzlich zu den gitterbasierten Umfeldrepräsentationen existieren weitere Datenstrukturen zur Speicherung von dreidimensionalen Umgebungen. Hervorzuheben ist hierbei der *Kinect-Fusion*-Ansatz, der beliebige Umgebungen dreidimensional kartieren kann. Nachteile sind allerdings der sehr hohe Speicher- und Rechenaufwand. In dieser Arbeit wird daher für die Umfeldmodellierung auf die speicher- und recheneffiziente 2.5D-Karte zurückgegriffen. Hierzu ist eine neuartige Methode zur **Integration von Time-of-Flight-Kameradaten in eine 2.5D-Karte** notwendig. Zusätzlich zur 2.5D-Karte wird der Aufbau einer *Occupancy Grid Map* durch die Daten der PMD-Kamera untersucht.



Die in der Literatur vorgestellten Methoden zur Flächenextraktion lassen sich in globale und lokale Methoden unterteilen. Globale Methoden basieren in erster Linie auf dem RANSAC-Algorithmus. Die auf RANSAC basierenden Ansätze sind zu einem großen Teil nicht echtzeitfähig, und die echtzeitfähigen Verfahren sind stark fehleranfällig im Falle von verrauschten Time-of-Flight-Kameradaten. Lokale Ansätze hingegen sind recheneffizient, neigen aber zu einer Übersegmentierung des Kamerabildes. Daher wird in dieser Arbeit ein neuartiges, echtzeitfähiges Verfahren zur **Flächenextraktion aus Time-of-Flight-Kameradaten** vorgestellt, das im Falle von Time-of-Flight-Kameradaten bessere Ergebnisse liefert als die bekannten Methoden aus dem Stand der Technik.

Bekanntes Verfahren zur visuellen Odometrie existieren in erster Linie für die Daten von Mono- und Stereo-Kameras und eignen sich für hochauflösende Bilddaten. Die vorgestellten Verfahren basieren auf der Extraktion von Merkmalspunkten und dem Bilden von Punktkorrespondenzen in aufeinanderfolgenden Bildern. Vereinzelt wurden Verfahren für die Daten von Time-of-Flight-Kameras präsentiert. Diese nutzen allerdings zur Unterstützung entweder eine weitere hochauflösende Farbkamera, eine inertielle Messeinheit oder eine Karte der Umgebung. In dieser Arbeit wird daher ein neuartiges Verfahren zur **visuellen Odometrie auf Basis von Time-of-Flight-Kameradaten** vorgestellt, das den Einsatz von mehreren Kameras berücksichtigt. Zusätzlich wird ein Verfahren zur simultanen Lokalisierung und Kartenerstellung durch die Daten der Time-of-Flight-Kameras präsentiert. Das EKF-SLAM-Verfahren besitzt den Nachteil, dass keine dichte Kartenrepräsentation aufgebaut werden kann. SLAM-Verfahren, die auf einem Partikelfilter basieren, haben einen hohen Speicher- und Rechenbedarf, da in jedem einzelnen Partikel jeweils eine gesamte Karte gespeichert wird. In dieser Arbeit wird daher auf ein *Maximum-Likelihood*-Verfahren zurückgegriffen, da dieses nur eine einzelne Karte aufbaut und somit speicher- und recheneffizienter ist.

Verfahren zur Extraktion von dynamischen Objekten, die auf den Daten von Mono- und Stereo-Kameras basieren, sind aus der Literatur bekannt. Diese beruhen meist auf der Berechnung des optischen Flusses. Da dies bei Time-of-Flight-Kamerabildern zu Problemen führt, sind diese Ansätze nicht ohne Weiteres zu übertragen. Für PMD-Kameras existieren Verfahren zur Erkennung von Hindernissen. Die bekannten Methoden basieren allerdings auf sehr niedrig auflösenden vorausschauenden PMD-Kameras.

Die Verwendung von höherauflösenden Weitwinkel-PMD-Kameras ermöglicht leistungsfähigere Algorithmen, sodass zur **Detektion von dynamischen Objekten** ein neuartiges Verfahren vorgestellt wird.

2.4 Ableitung der Methode zur 3D-Umfeldmodellierung

Diese Arbeit beschreibt ein System zur dreidimensionalen Vermessung des Fahrzeugumfelds durch den Einsatz von Time-of-Flight-Kameras. In Kapitel 1.4 wurde erläutert, dass für zukünftige Fahrerassistenzfunktionen für das Parken und Rangieren eine vollständige sensorische Abdeckung notwendig ist. Das vorgestellte System liefert eine omnidirektionale Abdeckung. Dies bedeutet, dass die Fahrzeugumgebung in allen Richtungen zu 360° Grad sensorisch erfasst wird.

Bei der Verfolgung dieses Ziels ergeben sich zwei Schwerpunkte. Zum einen beschreibt die Systemauslegung die Auslegung und Integration eines PMD-Kamerasystems, sodass die notwendige Abdeckung erreicht werden kann. Der zweite Schwerpunkt liegt in der Datenverarbeitung der PMD-Kameradaten von den Distanz- und Amplitudendaten zu einem dreidimensionalen Umfeldmodell.

Ein Überblick über die Methode ist in Abbildung 2.18 gegeben. Kapitel 3 beschreibt die Systemauslegung des PMD-Kamerasystems unter Verwendung von Weitwinkelobjektiven und die Anpassung der Beleuchtung, sodass der gewünschte Bereich im Umfeld des Fahrzeugs beleuchtet wird sowie die Integration der Sensoren in ein Fahrzeug. Zudem wird eine neuartige Methode zur Kalibrierung der Tiefendaten der Weitwinkel-PMD-Kamera vorgestellt.

Die Vorverarbeitung beinhaltet in Kapitel 4 die Ansteuerung der Sensoren durch eine Anpassung der Integrationszeit und die Glättung der Bilddaten der Kameras. Für jedes Pixel der Kamera wird ein Konfidenzwert bestimmt, welcher der Güte der Messung entspricht.

Die 3D-Umfeldmodellierung akkumuliert die vorverarbeiteten 3D-Punkt-daten der Kameras in einem Umgebungsmodell und ist in Kapitel 5 beschrieben. Als Datenstruktur wird sowohl ein *Occupancy Grid*, als auch eine 2.5D-Karte verwendet. Aus dem 3D-Umfeldmodell lassen sich Informationen für zukünftige Fahrerassistenzfunktionen ableiten.



Anschließend werden in der 3D-Bildverarbeitung, welche in Kapitel 6 erläutert wird, neuartige Methoden zur Flächenextraktion, zur Lokalisierung und zur Detektion von dynamischen Objekten angewandt. Die Flächenextraktion ermittelt planare Strukturen aus den Bilddaten der PMD-Kameras, welche als Informationsquelle für Fahrerassistenzsysteme dienen können. Zur Lokalisierung werden zwei unterschiedliche Verfahren vorgestellt. Zum einen wird ein Ansatz zur visuellen Odometrie präsentiert, der unter alleiniger Verwendung der Kamerabilder die Bewegung des Fahrzeugs bestimmt. Zum anderen wird eine Methode vorgestellt, die unter Verwendung des erstellten Umfeldmodells und der Kameradaten zu einer Verbesserung der konventionellen Fahrzeugodometrie führt. Die Extraktion dynamischer Objekte hat zum Ziel, diese Objekte von der dreidimensionalen Modellierung des statischen Umfelds auszuschließen und sie stattdessen separat zu verfolgen. Dies geschieht auf Basis von aufeinanderfolgenden Aufnahmen der PMD-Kameras.

Evaluationsverfahren

Das 3D-Umfeldmodell wird anhand eines Referenzsystems in Kapitel 7 evaluiert. Hierzu wird zum einen eine Referenztrajektorie des Fahrzeugs bestimmt und zum anderen eine Referenzumgebung hochgenau dreidimensional vermessen.

Die Referenztrajektorie wird durch ein hochgenaues Lokalisierungssystem, das auf der Abstrahlung von geformtem Laserlicht im Infrarotlichtbereich basiert, ermittelt. Die Referenzumgebung wird durch einen 3D-Laserscanner vermessen. Die Kombination der beiden Systeme ermöglicht die hochgenaue Lokalisierung des Fahrzeugs innerhalb des Referenzmodells. Für jeden Zeitpunkt kann so das erstellte Umfeldmodell mit dem Referenzmodell verglichen werden. Auch kann aufgrund der bekannten Pose des Fahrzeugs und der erfolgten extrinsischen und intrinsischen Kalibrierung für jede Kamera zu jedem Zeitpunkt ein Referenzbild generiert werden.

Daher werden sowohl das erstellte 3D-Modell als auch die korrigierten Tiefendaten der im Fahrzeug integrierten Kameras evaluiert. Anschließend wird auf die in Abschnitt 1.4 hergeleiteten Anforderungen für zukünftige Fahrerassistenzsysteme eingegangen.

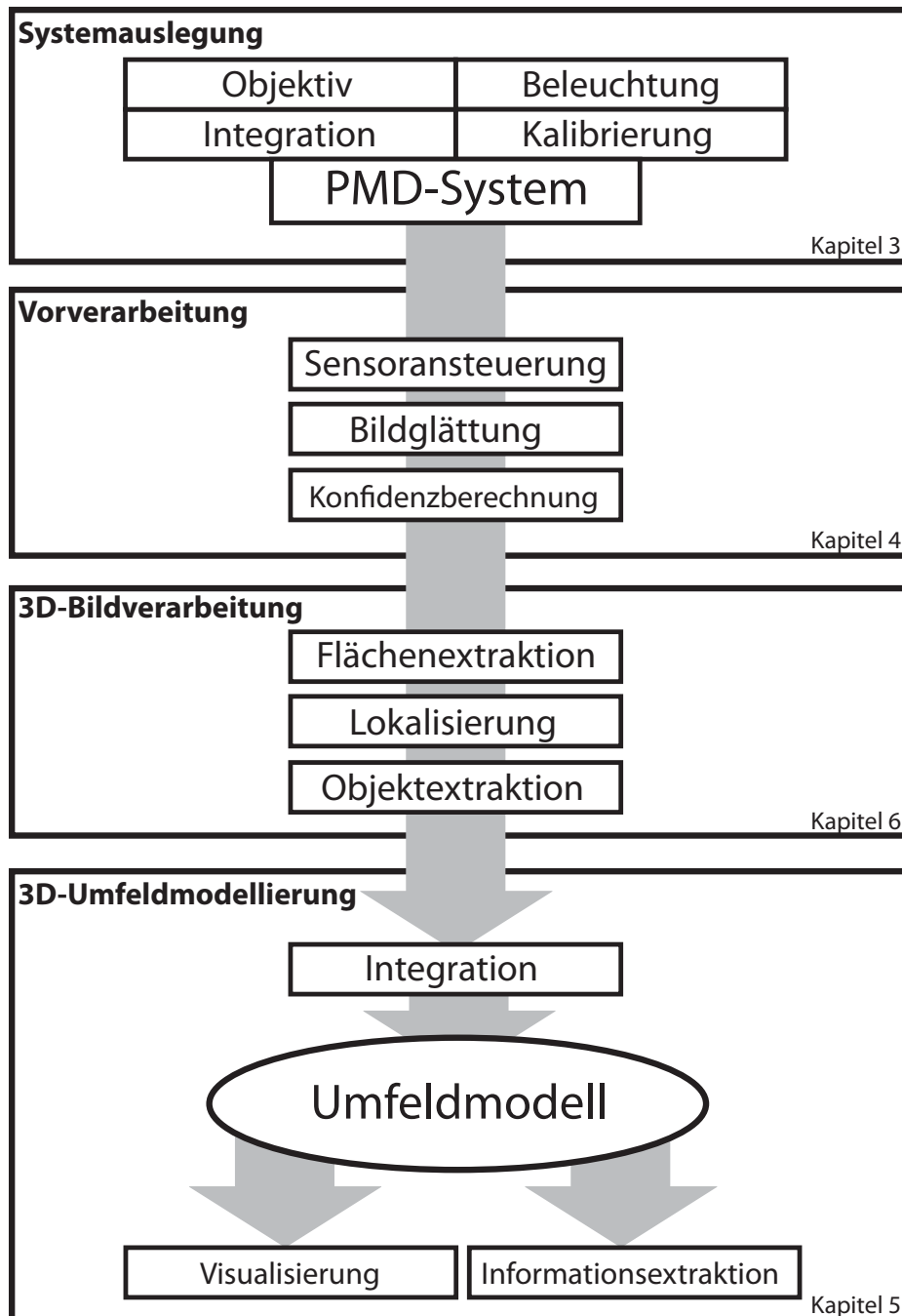


Abbildung 2.18: Überblick über die Methodik



2 Stand der Technik und Methode

Die Datenverarbeitung der Kameradaten hin zu einem Umfeldmodell beinhaltet verschiedene neuartige Verfahren, die auch unabhängig vom Gesamtziel der Erstellung des Umfeldmodells betrachtet und evaluiert werden können. Daher befindet sich eine zusätzliche Evaluierung der einzelnen Verfahren in den entsprechenden Kapiteln.

3 Systemauslegung

In diesem Kapitel wird die Auslegung eines geeigneten Weitwinkel-PMD-Sensors erläutert. Abschnitt 3.2 beschreibt, wie die Sensoren in ein Fahrzeug integriert werden können, sodass die Umgebung vollständig abgedeckt wird. Die notwendigen Anpassungen an die Beleuchtung werden in Abschnitt 3.3 thematisiert. Eine Charakterisierung der Tiefendaten erfolgt in Abschnitt 3.4. Fehler in den Tiefendaten der Kameras werden durch eine in Abschnitt 3.5 präsentierte Tiefenkalibrierung korrigiert. Eine Fusion der Sensordaten mehrerer Kameras wird in Abschnitt 3.6 vorgestellt.

3.1 Weitwinkel-PMD

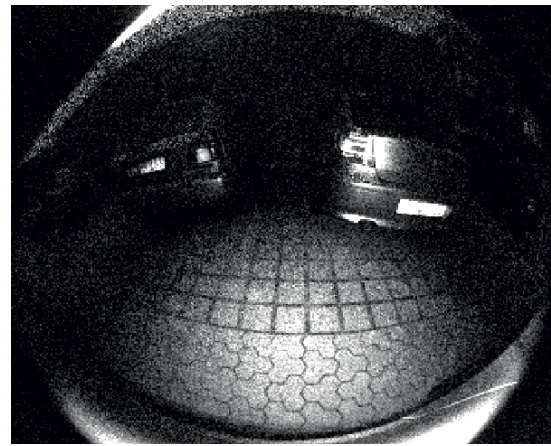
Das in Kapitel 2.4 vorgestellte Konzept sieht die Verwendung von Weitwinkel-PMD-Kameras vor. Weitwinkelobjektive werden eingesetzt, um den notwendigen horizontalen Öffnungswinkel abzudecken. In vertikaler Richtung wird hingegen ein geringerer Öffnungswinkel benötigt. Der von der Optik abgebildete Winkelbereich der eingesetzten Objektive ist demzufolge in vertikaler Richtung größer als notwendig. Neben einer Anpassung des Objektivs ist zusätzlich eine Adaption der aktiven Beleuchtung der PMD-Kamera notwendig. Ein hoher Abstrahlwinkel der LEDs würde Licht in Bereiche lenken, die nicht bestrahlt werden müssen. Daher soll das vorhandene Licht in die Bereiche gelenkt werden, die für die Vermessung der Fahrzeugumgebung interessant sind. Um dies zu realisieren, ist eine Modularisierung der Beleuchtung notwendig. Diese wird in der Entwicklung einer modularisierten Kamera berücksichtigt.

3.1.1 Weitwinkelobjektiv

Das Objektiv einer Kamera sammelt Licht aus der Umgebung und erzeugt eine optische Abbildung der Szene auf einen Sensor. Daher befindet sich



a) Standardobjektiv



b) Weitwinkelobjektiv

Abbildung 3.1: Amplitudenbilder durch Verwendung eines Standardobjektivs (a) und eines Weitwinkelobjektivs (b)

die Optik zwischen der zu vermessenen Szene und dem Sensor und besteht in der Regel aus mehreren Linsen.

Das Standardobjektiv einer PMD-Kamera kann durch ein passendes Weitwinkelobjektiv ersetzt werden. In Abbildung 4.1 ist das Amplitudenbild einer Kamera mit Standardobjektiv dem Amplitudenbild einer Kamera mit einem Weitwinkelobjektiv gegenübergestellt. Das Weitwinkelobjektiv besitzt einen Öffnungswinkel von 180° .

Weitwinkelobjektive besitzen neben dem Vorteil des größeren Sichtbereichs einige Nachteile gegenüber herkömmlichen Objektiven. Nachteile sind die allgemein geringere Lichtstärke und zum anderen der Lichtabfall an den Rändern des Sensors. Dieser Effekt wird auch Vignettierung genannt. Weitwinkelobjektive sind im Allgemeinen anfälliger gegenüber Streulichteffekten, da es oft zu Lichtreflektionen innerhalb des Linsensystems kommt. Dies führt zu einem vermehrten Auftreten des in Kapitel 2.1.3 vorgestellten *Scattering* Effekts. Ein weiterer Nachteil von Weitwinkelobjektiven ist die starke Verzerrung der Bilder.

3.1.2 Modularisierung der Beleuchtung

Die alleinige Erhöhung des Öffnungswinkels des Objektivs einer PMD-Kamera führt zwar zu einem erhöhten Sichtbereich, allerdings misst ei-

ne PMD-Kamera Distanzen nur bei ausreichender Reflexion des aktiven Lichtsignals. Das aktive Infrarotsignal muss daher auch in die Randzonen des Sichtbereichs abgestrahlt werden. Dies kann ermöglicht werden, indem der Abstrahlwinkel der LEDs auf ebenso 180° erhöht wird. Hierbei tritt allerdings der gravierende Nachteil auf, dass die Leuchtdichte der LEDs sehr stark abnimmt. Nicht in jedem Winkelbereich wird eine hohe Lichtmenge benötigt. Beispielsweise kann der vertikale Öffnungswinkel geringer sein als der horizontale. Daher sieht das Konzept vor, dass Licht nur in die Bereiche gestrahlt werden soll, in denen das System auch Distanzmessungen liefern soll. Dies kann realisiert werden, indem geringere Abstrahlwinkel für die LEDs verwendet werden wobei diese allerdings verschiedene Ausrichtungen besitzen können. Hierzu ist zwingend eine Modularisierung der Beleuchtung notwendig. Diese wird ermöglicht, indem jede LED auf eine einzelne Platine platziert wird. Die Platine kann anschließend je nach gewünschter Position und Ausrichtung der LED befestigt werden.

3.1.3 Modularisierung der PMD-Kamera

Um das zuvor vorgestellte Konzept einer modularisierten PMD-Kamera zu realisieren, wurde im Rahmen dieser Arbeit eine geeignete PMD-Kamera entwickelt.

Die entwickelte prototypische Kamera ist eine stark modularisierte Kamera. Sie besteht aus einem Kameramodul und mehreren Beleuchtungsmodulen. Das Kameramodul beinhaltet den Sensor und die Weitwinkel-Optik. An das Kameramodul können bis zu drei Verteilerplatinen angesteckt werden, mit denen wiederum jeweils drei LED-Module verbunden sind. Dabei sind die LED-Module frei beweglich und können somit beliebig ausgerichtet werden. Ein Grabbermodul fungiert als Schnittstelle zwischen Kameramodul und PC. Der Anschluss erfolgt hierbei durch USB oder Ethernet. Der verwendete Sensor stammt von der Firma *Infineon* und besitzt eine Auflösung von 352×288 Pixel. Die Wiederholrate beträgt bis zu 100 Hz. Die Kamera erlaubt es, die Modulationsfrequenz beliebig einzustellen und ermöglicht zudem, ein Mehrfrequenzverfahren zu verwenden, sodass Eindeutigkeitsfehler bei Überschreitung des Eindeutigkeitsbereichs vermieden werden. Als Lichtquellen kommen LEDs mit einem Abstrahlwinkel von 90° zum Einsatz. Eine Übersicht über die verschiedenen Kameramodule ist in Abbildung 3.2 gegeben.

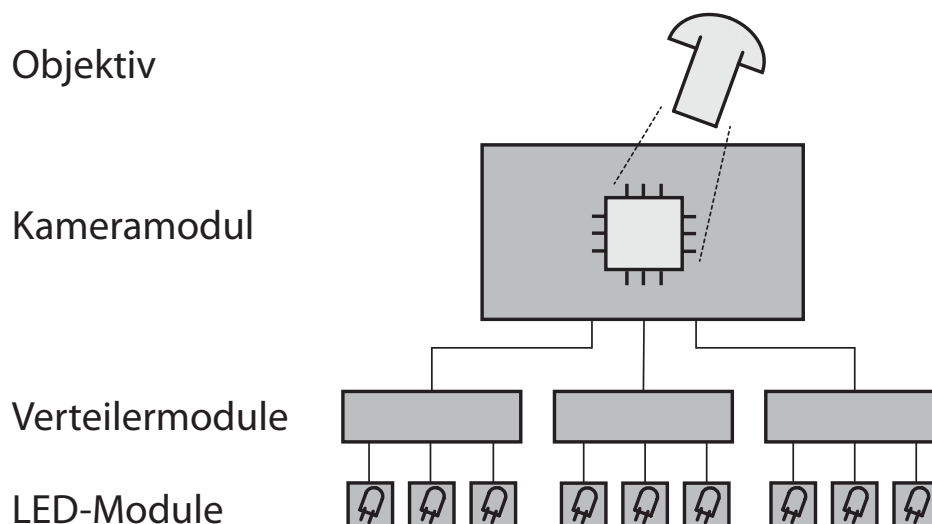


Abbildung 3.2: Übersicht über die verschiedenen Kameramodule

3.2 Integration in Versuchsträger

Als Versuchsträger für die Integration der Kameras wurde im Rahmen dieser Arbeit ein *Audi A6 Avant* aus dem Baujahr 2012 verwendet. Die Kamerapositionen werden analog zu dem *TopView*-System von *Audi* (siehe Kapitel 1.2.4) gewählt. Eine Kamera wird in jeweils ein Außenspiegelgehäuse integriert, da diese über die Fahrzeugflanken hinausragen und nur auf diese Weise die Sichtbarkeit von Objekten, die sich unmittelbar neben den Reifen befinden, sichergestellt werden kann. Die Frontkamera wird in den Kühlergrill und die Heckkamera an der Griffmulde der Heckklappe integriert. Abbildung 3.3 zeigt die prototypische Integration der Kameras.

Die Grabbermodule der Kameras und ein Fahrzeugrechner werden im Innenraum des Fahrzeugs montiert. Die optimalen Ausrichtungen der Kameraobjektive werden experimentell durch die Aufnahme von Markierungen im Fahrzeugumfeld bestimmt. Das Umfeld des Fahrzeugs wird durch reflektierende Marker abgesteckt. Nun werden Aufnahmen mit verschiedenen Neigungswinkeln der Kamera durchgeführt. Auf Basis der Sichtbarkeit der Markierungen kann ein optimaler Neigungswinkel bestimmt werden. Die Vorgehensweise ist in Abbildung 3.4 dargestellt. Es ergeben sich auf Basis der jeweiligen Höhen der integrierten Kameras als optimale Neigungswinkel 24° für die Front, 22° für das Heck und 32° für die Seitenkameras.



a) Front



b) Heck



a) Seitenspiegel

Abbildung 3.3: Die Integration der Kameras in einen Versuchsträger

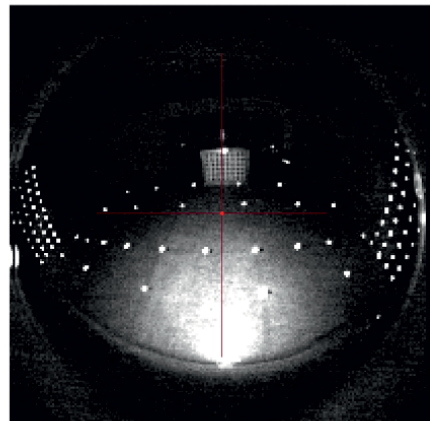
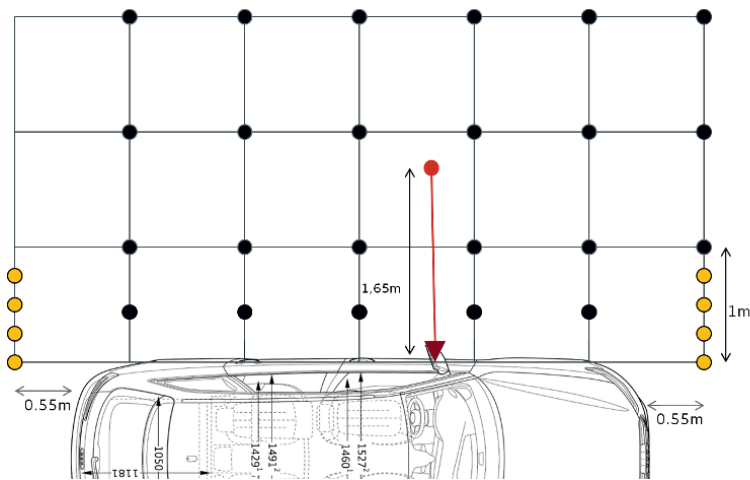


Abbildung 3.4: Die Ermittlung des optimalen Neigungswinkels der Kamera

3.3 Beleuchtung

Die aktive Infrarotbeleuchtung bildet einen wesentlichen Faktor für die Leistungsfähigkeit einer PMD-Kamera. Das Infrarotlicht ist Träger des Modulationssignal und die Modulierbarkeit der Lichtquelle entscheidet über die Güte der Distanzmessung. Zudem hängt die Lichtintensität, mit der ein Pixel des Sensors bestrahlt wird, direkt mit der Qualität der Messung zusammen. Daher muss bei der Auslegung eines PMD-Kamerasystems darauf geachtet werden, dass ausreichend Licht in die Orte gestrahlt wird, in denen gemessen werden soll.

3.3.1 Beschreibung der eingesetzten Lichtquellen

Als Lichtquellen kommen in erster Linie LED-Dioden zum Einsatz. Vermehrt werden auch in Anwendungen mit einem geringen Bedarf an optischer Leistung Laserdioden als Lichtquellen für PMD-Kamerasysteme verwendet, da sie mehrere Vorteile besitzen. Im verwendeten Kameraprototypen wird eine Infrarot-LED vom Typ *Osram Oslon SFH 4715S* [130] verwendet. Für die Simulation der Lichtausbeute einer LED sind die optische Leistung und die Abstrahlcharakteristik relevant. Für die verwendete *Osram-Oslon*-LED ergibt sich laut Datenblatt eine optische Leistung von 1070 mW. Die Abstrahlcharakteristik ist in Abbildung 3.5 dargestellt. Auf der rechten Seite ist die relative Lichtintensität gegenüber dem Abstrahlwinkel aufgetragen. Mit zunehmendem Abstrahlwinkel sinkt die Lichtintensität. Auf der linken Seite ist eine den Abstrahlwinkeln entsprechend verzerrte Darstellung gegeben wie sie bei Leuchtmittelherstellern üblich ist.

3.3.2 Simulation der Messgenauigkeit

Für die Integration in einen Versuchsträger müssen die Ausrichtungen der einzelnen LEDs bestimmt werden. Bei der Verwendung von jeweils 9 LED-Modulen pro Kamera müssen die Ausrichtungen von 36 Lichtquellen bestimmt werden. Für die Bestimmung der Parameter bietet sich eine Simulation an, da nicht jede Konfiguration manuell geprüft werden kann. Das Ziel ist eine Konfiguration, durch welche die relevanten Bereiche des Fahrzeugumfelds von den LEDs ausgeleuchtet werden.

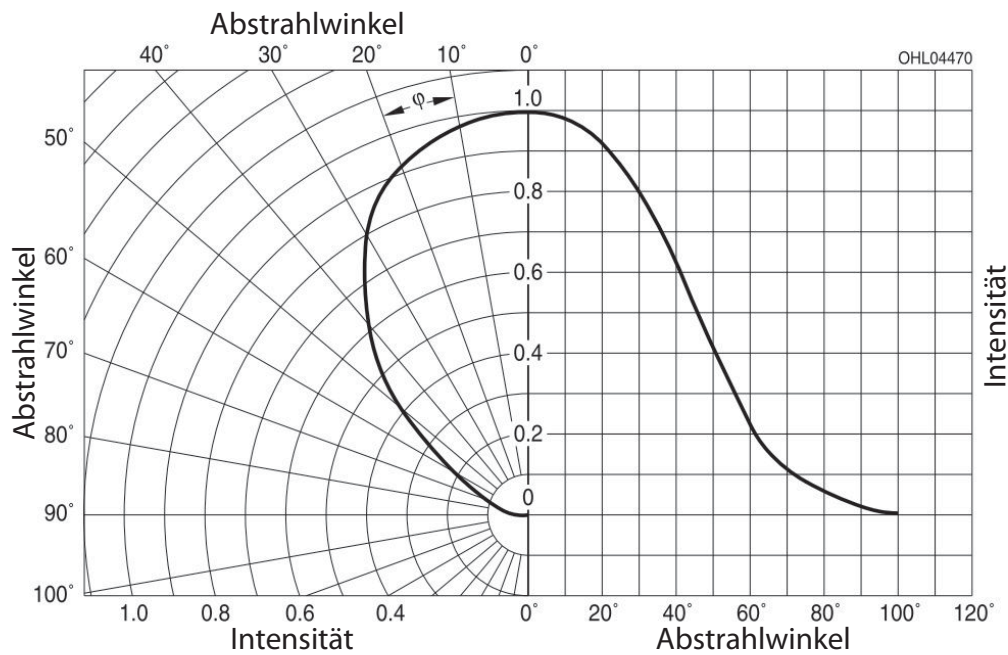


Abbildung 3.5: Abstrahlcharakteristik der verwendeten *Osram-Oslon*-Infrarot-LED [130]. Auf der rechten Seite ist die relative Lichtintensität gegenüber dem Abstrahlwinkel aufgetragen. Mit zunehmendem Abstrahlwinkel sinkt die Lichtintensität. Auf der linken Seite ist eine den Abstrahlwinkeln entsprechend verzerrte Darstellung gegeben.



3 Systemauslegung

Zur Simulation der Ausleuchtung muss auf die physikalischen Gesetze des Lichts zurückgegriffen werden. Die Lichtquellen strahlen Infrarotlicht mit einer gewissen Intensität in die Umgebung ab. Ein Teil des von Objekten reflektierten Lichts gelangt zurück zum Sensor. Die Intensität des reflektierten Lichts, das auf den photosensitiven Bereich eines Pixels trifft, bestimmt den Amplitudenwert der Messung, der wiederum einen direkten Einfluss auf die Messgüte hat. Daher wird durch eine Simulation für jedes Pixel ein Wert für die Amplitude bestimmt.

Der Zusammenhang zwischen Lichtintensität und Amplitudenwert wird anhand einer Referenzaufnahme experimentell bestimmt. Zunächst wird dazu die auf ein einzelnes Pixel treffende Lichtintensität bestimmt.

Die Lichtquellen strahlen Licht mit einer Intensität I in die Umgebung ab. Dabei hängt die abgestrahlte Lichtintensität vom Winkel zur Hauptstrahlrichtung der Lichtquelle ab. In Abbildung 3.5 ist die Abnahme der Lichtintensität nach dem Datenblatt der verwendeten LED dargestellt. Das Abstandsgesetz beschreibt die Abnahme der Lichtintensität mit der Distanz. Die an einem Objekt ankommende Lichtintensität I' beträgt

$$I' = \frac{I}{d^2} \quad (3.1)$$

wenn das Licht die Distanz d zurückgelegt hat. Es wird nun angenommen, dass die Oberfläche des reflektierenden Objektes das Licht als Lambert-scher Reflektor reflektiert. Für die reflektierte Intensität I'' gilt dann

$$I'' = L \cdot \cos(\theta) \cdot A. \quad (3.2)$$

Dabei bezeichnet θ den Winkel zwischen der Leuchtrichtung und der Normalen der Oberfläche. Der Wert A bezeichnet die Fläche der Oberfläche, welche auf ein Pixel des Sensors projiziert wird und L ist die Leuchtdichte des Reflektors. Zur Berechnung der Fläche A ist der vom jeweiligen Pixel abgedeckte Raumwinkel ω notwendig. Um diesen zu bestimmen, werden die intrinsischen Kameraparameter benötigt, die in Kapitel 2.1.4 vorgestellt wurden. Diese entsprechen den optischen Abbildungseigenschaften der Kamera und liefern für jedes Pixel einen Richtungsvektor, welcher der Blickrichtung des Pixels entspricht. Wenn die Richtungsvektoren benachbarter Pixel in Beziehung gesetzt werden, kann der vom jeweiligen Pixel abgedeckte Raumwinkel ω bestimmt werden. Die Fläche A kann anschließend durch



$$A = \frac{\omega d^2}{\cos(\theta)} \quad (3.3)$$

bestimmt werden. Die Leuchtdichte einer Oberfläche mit dem Reflexionsgrad p wird durch

$$L = \frac{\rho E}{\pi} \quad (3.4)$$

bestimmt. Dabei ist E die Bestrahlungsstärke, welche mit der Intensität I' zusammenhängt durch

$$E = \frac{I' \omega}{A}. \quad (3.5)$$

Indem die Werte in Gleichung 3.2 eingesetzt werden, kann die reflektierte Lichtintensität I'' bestimmt werden durch

$$I'' = L \cdot \cos(\theta) \cdot A = \frac{\rho I' \cos(\theta) \omega}{\pi}. \quad (3.6)$$

Berücksichtigt man nun das Abstandsgesetz aus Gleichung 3.1 erhält man durch

$$I''' = \frac{\rho I \cos(\theta) \omega}{\pi d^2} \quad (3.7)$$

die am Pixel ankommende reflektierte Lichtintensität I''' .

Die Intensität des auf das Pixel reflektierten Lichts I''' ist proportional zur Amplitude des Pixels. Da der Zusammenhang zwischen I''' und dem Amplitudenwert durch Referenzaufnahmen experimentell bestimmt wird, ist zu beachten, dass die Simulation Amplitudenwerte für Oberflächen mit derselben Objektreflektivität der Referenzoberfläche bestimmt. In Gleichung 3.7 ist allerdings zu sehen, dass die Objektreflektivität ρ proportional zur Lichtintensität I''' ist. Wenn daher die Reflektivität der Referenzoberfläche bekannt ist, kann I''' für beliebige Reflektivitäten berechnet werden.

Für die ausgehende Lichtintensität I einer Lichtquelle kann auch mit einer normalisierten Intensität im Wertebereich zwischen 0 und 1 gerechnet werden, sodass keine exakte Messung der Lichtintensitäten der Lichtquellen notwendig ist. Stattdessen werden Referenzaufnahmen einer PMD-Kamera mit einer einzelnen Lichtquelle erstellt. Eine derartige Aufnahme ist in Abbildung 3.6 auf der linken Seite dargestellt. Mit einer solchen Aufnahme kann durch Berücksichtigung der zuvor erklärten physikali-

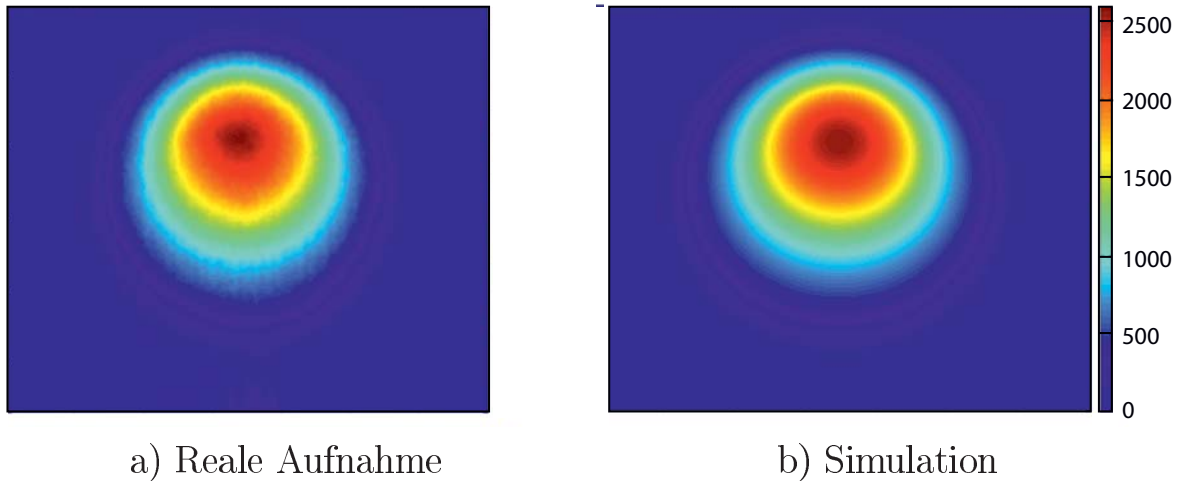


Abbildung 3.6: Gegenüberstellung der realen Aufnahme eines Amplitudenbilds mit einem simulierten Amplitudenbild (mit einer einzelnen LED als Beleuchtung)

schen Zusammenhänge sowohl die Intensitätsverteilung einer Lichtquelle als auch der Zusammenhang zwischen der reflektierten Intensität und dem Amplitudenwert eines Pixels bestimmt werden. Ein simuliertes Amplitudenbild derselben Konfiguration ist in Abbildung 3.6 rechts dargestellt.

In Abbildung 3.7 wird das simulierte Amplitudenbild einer realen Aufnahme einer Szene gegenübergestellt. Die Szene besteht aus einer horizontalen Bodenebene und einer vertikalen Wandebene. Insgesamt leuchten drei LEDs in die Szene. Es kann beobachtet werden, dass bei der realen Aufnahme im Bereich zwischen den beiden Ebenen höhere Amplitudenwerte vorhanden sind als im simulierten Bild. Dies ist darauf zurückzuführen, dass durch Mehrfachreflexionen der in Kapitel 2.1.3 vorgestellte *Multipath*-Effekt auftritt. Mehrfachreflexionen sind im Simulationsmodell nicht enthalten. Licht, das zunächst von einer ersten Ebene reflektiert wird und anschließend auf die zweite Ebene trifft, erhöht den jeweiligen Amplitudenwert. Aus diesem Grund treten insbesondere in der Nähe der Kante zwischen Boden und Wand erhöhte Amplitudenwerte auf. Das Licht, das zu der erhöhten Amplitude führt, hat aufgrund der mehrfachen Reflexion einen weiteren Weg zurückgelegt und führt daher zu keiner höheren Genauigkeit der Messung. Demzufolge ist es für die Simulation nicht problematisch, diesen Effekt unberücksichtigt zu lassen.

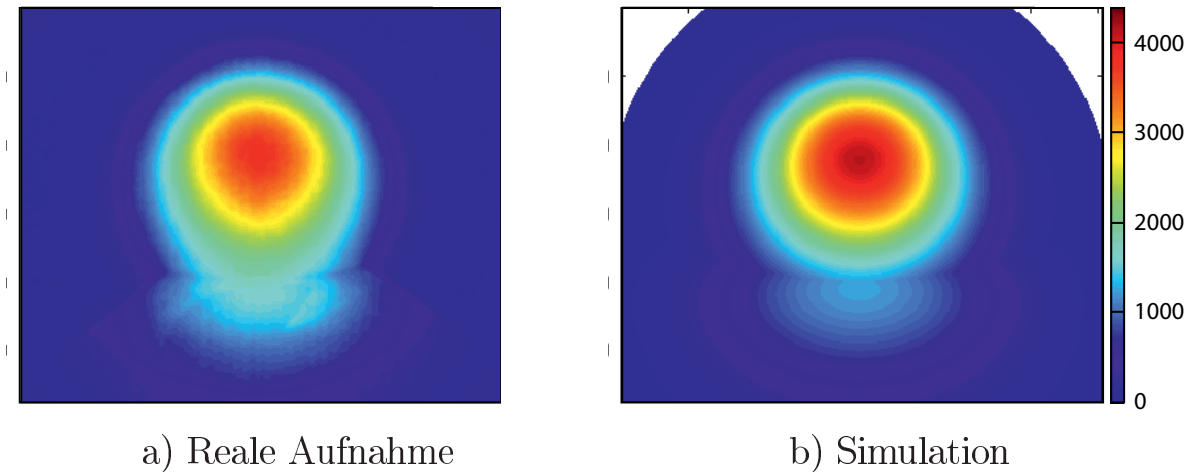


Abbildung 3.7: Gegenüberstellung einer realen Aufnahme und einem simulierten Bild. Die Szene besteht aus einer horizontalen Bodenebene und einer vertikalen Wandebene. Insgesamt leuchten drei LEDs in die Szene.

3.3.3 Positionierung der Beleuchtungsmodule

Aus den Anforderungen an die Umfeldwahrnehmung ergibt sich, dass möglichst das gesamte Umfeld des Fahrzeugs zu 360° vermessen werden soll. Daher muss ein PMD-System das Umfeld im Nahbereich über den gesamten Umfang des Fahrzeugs ausleuchten.

Pro Kamera sind neun LEDs verfügbar. Eine Konfiguration beinhaltet die Ausrichtungen der neun LEDs. Zur Evaluierung einer Konfiguration kann die im vorherigen Abschnitt vorgestellte Simulation verwendet werden. Eine Konfiguration wird anhand der Amplitudenwerte für Messungen im Umfeld des Fahrzeugs bewertet.

In Abbildung 3.8 (a) sind die simulierten Amplitudenwerte der im Versuchsträger erfolgten Integration dargestellt. Die Amplitudenwerte der einzelnen Pixel sind zur besseren Übersicht auf die korrespondierenden Oberflächen projiziert. Eine Anpassung der Ausrichtungen der LEDs kann zu den in Abbildung 3.8 (b) abgebildeten simulierten Amplitudenwerten führen.

Die verwendeten LEDs besitzen eine feste, nicht austauschbare Linse zur Bündelung des Lichts. Durch austauschbare Linsen können bei der Systemauslegung weitere Freiheitsgrade entstehen, da unterschiedliche Lin-

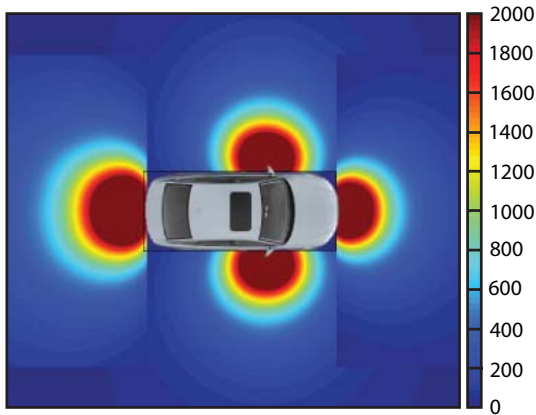
sen auch unterschiedliche Abstrahlcharakteristika der LEDs verursachen. Beispielsweise würde eine fokussierte Linse zu einer höheren Lichtintensität bei einem geringeren Öffnungswinkel führen. Bei der Verwendung von Laserdioden kann durch spezielle Strahlformoptiken eine sehr homogene Leuchtdichte über den gesamten Öffnungswinkel ermöglicht werden. Unter der Berücksichtigung unterschiedlicher Abstrahlcharakteristika mit beliebigen Öffnungswinkeln und einer homogenen Leuchtdichte der Lichtquellen können die in Abbildung 3.8(c) dargestellten Amplitudenwerte simuliert werden. Es ist zu beobachten, dass sich durch fokussierte Linsen insbesondere in den Ecken des auszuleuchtenden Bereichs höhere Amplitudenwerte erzielen lassen. Da die Laserdioden hier durch eine homogene Verteilung der Lichtenergie über einen definierten Öffnungswinkel modelliert werden und über dem definierten Öffnungswinkel hinaus kein Licht abgestrahlt wird, sind die Leuchtkegel stark abgegrenzt. Dieses Verhalten ist entsprechend im simulierten Amplitudenbild zu beobachten. Die Lichtintensität der LEDs nimmt mit zunehmendem Abstrahlwinkel ab, sodass keine abgegrenzten Leuchtkegel vorhanden sind.

3.3.4 Trennung von Lichtquelle und Empfangsoptik

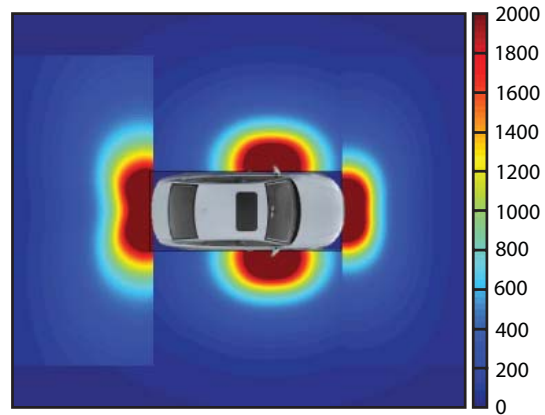
Dieser Abschnitt stellt das Konzept einer Trennung von Empfangsoptik und Lichtquellen vor. Das Ziel ist die örtliche Verteilung mehrerer Lichtquellen einer PMD-Kamera. Dies kann für eine bessere Ausleuchtung der Umgebung sorgen [39] und ist vorteilhaft für eine Integration in ein Fahrzeug. Ohne Anpassungen an die Distanzberechnung führt eine Trennung von Lichtquelle und Empfangsoptik jedoch zu Fehlern in den Tiefenwerten.

Das nachfolgend vorgestellte Verfahren ermöglicht die Positionierung mehrerer Lichtquellen an unterschiedlichen Orten. Hierzu ist es allerdings notwendig, dass die Lichtquellen sequentiell angesteuert werden. Dies bedeutet, dass in jeder Messung nur eine der Lichtquellen die Szene ausleuchtet. Die genauen Distanzdaten erhält man durch eine nachträgliche Korrektur der Tiefenwerte.

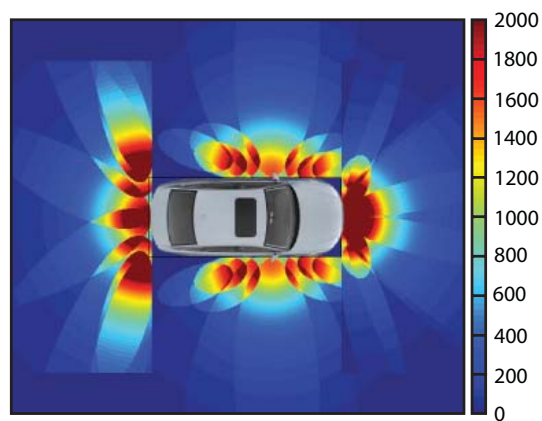
Zur Korrektur werden zum einen die optischen Abbildungsparameter und zum anderen die geometrische Anordnung der Lichtquelle und der Empfangsoptik benötigt. Dies ist in Abbildung 3.9 dargestellt.



a) Erfolgte Integration (9 LEDs)



b) Optimierung der Ausrichtungen (9 LEDs)



c) Verwendung von homogenen Lichtquellen mit beliebigen Öffnungswinkeln (9 Laserdioden)

Abbildung 3.8: Darstellung der simulierten Amplitudenwerte bei gegebenen Lichtquelleneigenschaften

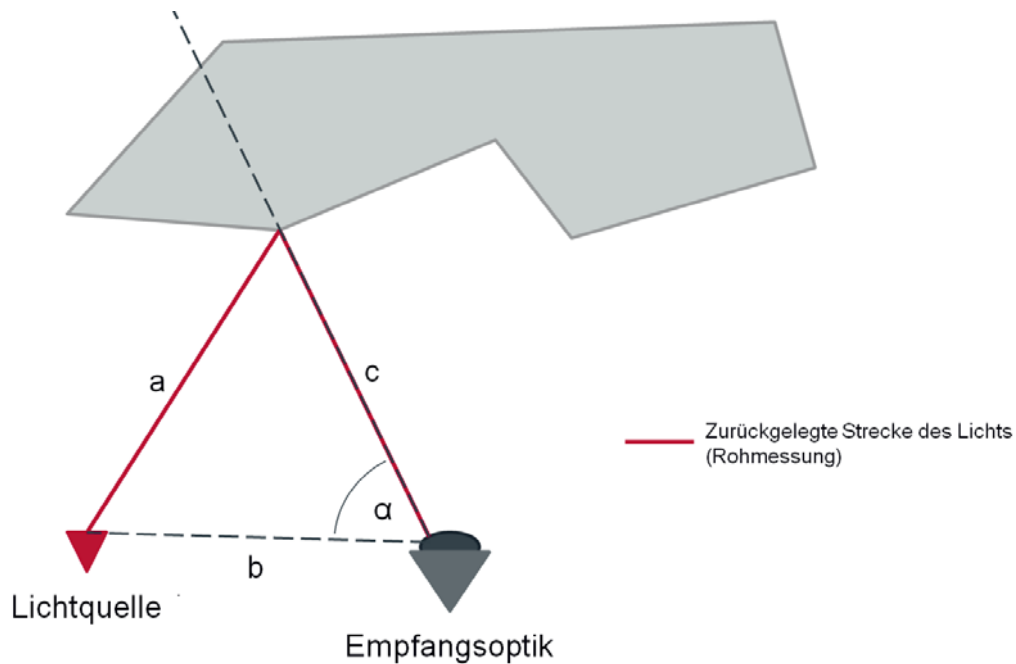


Abbildung 3.9: Die geometrische Anordnung von Lichtquelle und Empfangsoptik.

Jedem Pixel des Distanzbildes kann anhand der bekannten optischen Abbildungseigenschaften der Linse ein Raumwinkel zugeordnet werden. Da zudem die Position der Lichtquelle relativ zur Empfangsoptik bekannt ist, kann die Objektdistanz geometrisch aus der gesamten zurückgelegten Strecke des Lichtes bestimmt werden. Die Distanz b zwischen Lichtquelle und Empfangsoptik und der Winkel α zwischen dem Richtungsvektor eines Pixels (wird anhand der optischen Abbildungseigenschaften der Linse bestimmt) und dem Richtungsvektor zur Lichtquelle sind bekannt. Die Kamera misst aufgrund des Laufzeitprinzips die zurückgelegte Strecke des Lichtes, sodass der Tiefenwert d_{meas} bestimmt wird mit

$$d_{meas} = \frac{a + c}{2}. \quad (3.8)$$

Nach a aufgelöst ergibt sich

$$a = 2d_{meas} - c. \quad (3.9)$$



Gesucht wird die Entfernung c . Sie kann beispielsweise anhand des Kosinussatzes für Dreiecke bestimmt werden:

$$a^2 = b^2 + c^2 - 2 \cdot b \cdot c \cdot \cos \alpha \quad (3.10)$$

Aus Gleichung 3.9 und 3.10 ergibt sich

$$c = \frac{b^2 - 4d_{meas}^2}{2b \cos \alpha - 4d_{meas}}. \quad (3.11)$$

Auf diese Weise kann der Tiefenwert für jedes Pixel individuell korrigiert werden. Aufgrund der fehlenden Möglichkeit zur sequentiellen Ansteuerung der Lichtquellen konnte dieses Konzept nicht im Versuchsträger umgesetzt werden. Dennoch ermöglicht die theoretische Betrachtung der geometrischen Anordnung zukünftige Lösungen zur Trennung von Lichtquelle und Empfangsoptik.

3.3.5 Beleuchtung durch Laserdioden

Weiteres Potenzial zur Verbesserung der Beleuchtung bietet die Nutzung von Laserdioden. Die Verwendung von unterschiedlichen Streulinzen vor den Laserdioden ermöglicht eine nahezu beliebige Ausleuchtungscharakteristik. Allerdings sind hohe Öffnungswinkel bisher schwer umzusetzen. Ein weiterer Vorteil von Laserdioden gegenüber LEDs ist die hohe Modulierbarkeit. Durch eine Verdopplung der Modulationsfrequenz befinden sich bei gleicher Integrationszeit mehrere Phasen innerhalb einer Messung, was zu einer signifikanten Erhöhung der Messgenauigkeit führt.

Auch ist das geringere Wellenlängenspektrum des abgestrahlten Lichts ein Vorteil. Hierdurch können schmalbandigere Filter gegen die Fremdlichtbestrahlung des Sensors eingesetzt werden, sodass der Störeinfluss, der durch Fremdlicht entsteht, verringert werden kann.

Zudem sind Laserdioden in der Regel hinsichtlich ihres Energiebedarfs und der Wärmeabstrahlung effizienter als LEDs.

Nachteile sind die höheren Stückkosten der Laserdioden und die Veränderung der abgestrahlten Wellenlängencharakteristik bei Temperaturveränderungen. Diese Probleme müssen gelöst werden, um auch bei hohen optischen Leistungen die Vorteile von Laserdioden als modulierbare Lichtquellen zu nutzen.

3.4 Fehlereffekte der Tiefenmessung

Die verschiedenen Fehlerarten und Störfaktoren von PMD-Kameras wurden ausführlich in Kapitel 2.1.3 dargelegt. Im Folgenden werden die verschiedenen Fehlerarten für das entwickelte System im Fahrzeug untersucht. Dabei werden die Fehler in stochastische und systematische Fehler unterteilt. Anschließend wird ein Konzept zum Umgang mit den jeweiligen Messfehlern vorgestellt.

3.4.1 Stochastischer Fehler

In Kapitel 2.1.3 wird beschrieben, dass das Rauschen in der Distanzmessung vom Signal-Rausch-Verhältnis und damit von der Menge des empfangenen Nutzlichts abhängt. Aus diesem Grund ist der gemessene Wert für die Amplitude ein Indikator für den stochastischen Fehler der Messung. In Abbildung 3.10 ist die gemessene Standardabweichung über den gemessenen Amplitudenwert bei einer fest gewählten Integrationszeit aufgetragen. Hier zeigt sich, dass die Standardabweichung mit steigender Amplitude sinkt und durch eine Kurve beschrieben werden kann. Oft werden Pixel, die einen festgesetzten Amplitudenwert unterschreiten, als ungültig deklariert. Auf diese Weise können Tiefenmessungen, die einem starken Rauschen unterliegen, entfernt werden.

Eine weitere Quelle für ein schlechtes Signal-Rausch-Verhältnis ist die Störung der Messung durch Fremdlicht. Durch die in Kapitel 2.1.1 erläuterte Fremdlichtunterdrückung erscheinen Fremdlichtanteile nicht im Amplitudenwert. Das Intensitätsbild hingegen beinhaltet sowohl das modulierte Nutzlichtsignal als auch den unmodulierten Fremdlichtanteil. Daher kann der Intensitätswert in Zusammenhang mit dem Amplitudenwert eine Aussage über die Beeinflussung der Messung durch Fremdlicht geben.

Zur Untersuchung dieses Effekts wurde ein Sonnenlichtsimulator herangezogen. Dieser besteht aus einer starken Lichtquelle, die ein dem Sonnenlicht ähnliches Spektrum besitzt und gedimmt werden kann. Auf diese Weise können verschiedene Stärken der Fremdlichtbeeinflussung herbeigeführt werden. Die Stärke des Nutzsignals kann jeweils durch die Beleuchtung einer unterschiedlichen Anzahl von Beleuchtungsmodulen geregelt werden. Abbildung 3.11 stellt die Abhängigkeit zwischen Amplitude, Intensität und Standardabweichung dar.

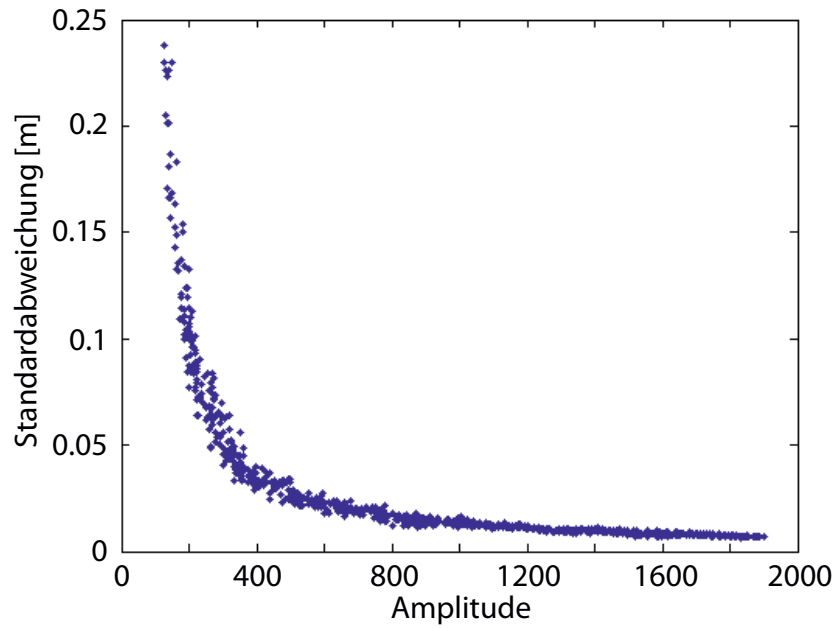


Abbildung 3.10: Zusammenhang zwischen Amplitude und Standardabweichung der Tiefenmessung

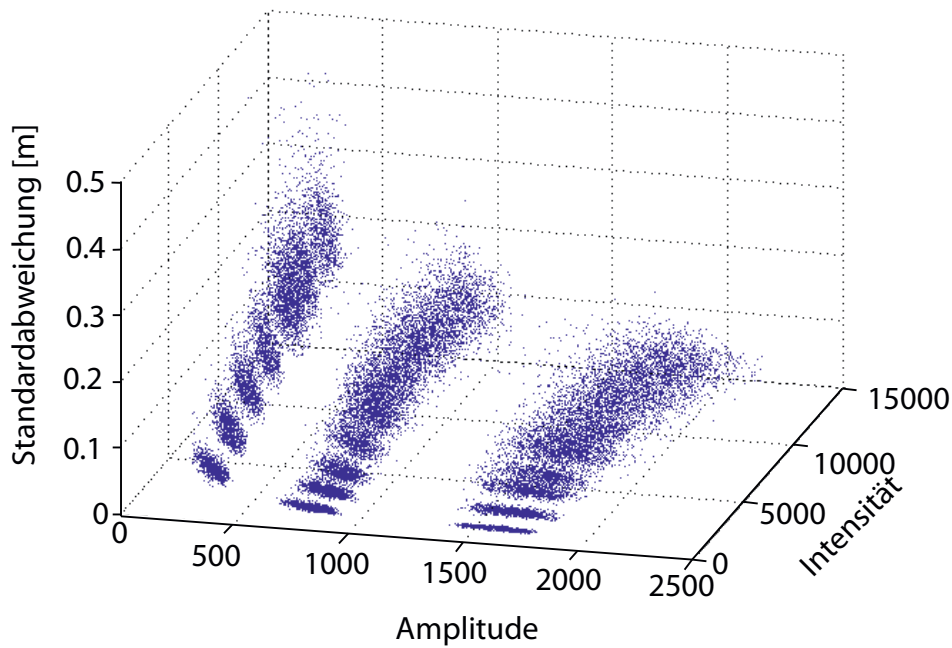


Abbildung 3.11: Zusammenhang zwischen Amplitude, Intensität und Standardabweichung der Tiefenmessung

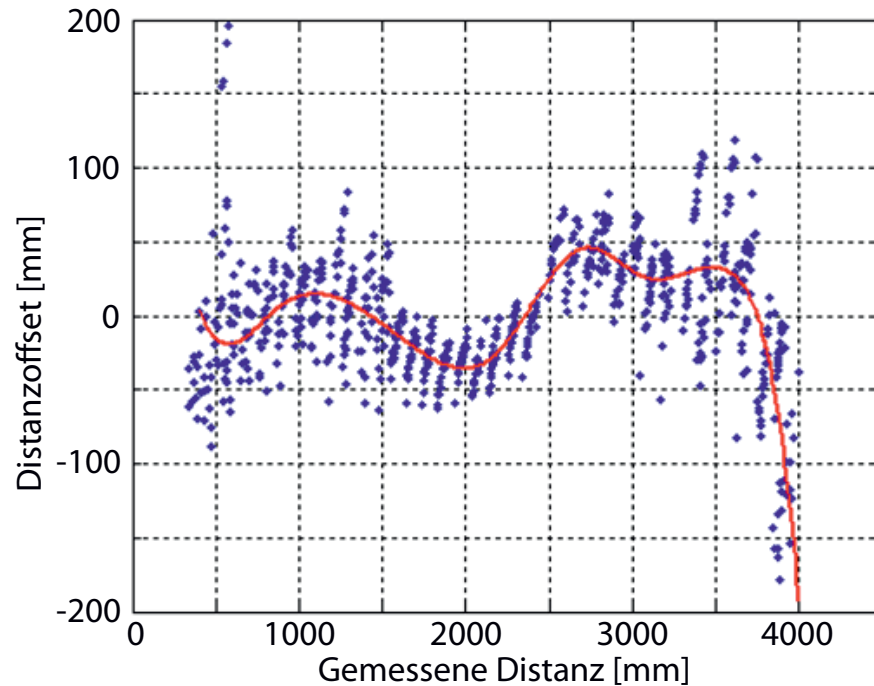


Abbildung 3.12: Zusammenhang zwischen Distanzfehler und wahrer Distanz

3.4.2 Systematische Fehler

Das im Versuchsfahrzeug integrierte Kamerasystem wird auf die in Kapitel 2.1.3 vorgestellten systematischen Fehler untersucht. Vom Hersteller wird lediglich der FPN-Fehler kalibriert.

Es zeigt sich, dass ein distanzabhängiger Fehler in den Tiefendaten existiert. Dieser wurde über verschiedene Entfernungen gegenüber der wahren Distanz in Abbildung 3.12 aufgetragen. Zum einen entsteht der distanzabhängige Fehler durch den *Wiggling*-Effekt. Zusätzliche Ursachen sind allerdings auch ein allgemeiner Offset in den Phasendaten und die in Abschnitt 3.3.4 beschriebenen Fehler, die durch einen Versatz zwischen Lichtquellen und Optik entstehen.

Auch der gemessene Amplitudenwert wirkt sich auf die Tiefenmessungen einzelner Pixel aus. Beobachtet werden kann dies insbesondere bei der Aufnahme von Schachbrettmustern, deren Felder unterschiedliche Reflektivitäten, aber die gleiche Entfernung zur Kamera besitzen. Die Tiefenwerte eines planaren Objektes mit einem Schachbrettmuster sind in Abbildung 3.13 dargestellt.

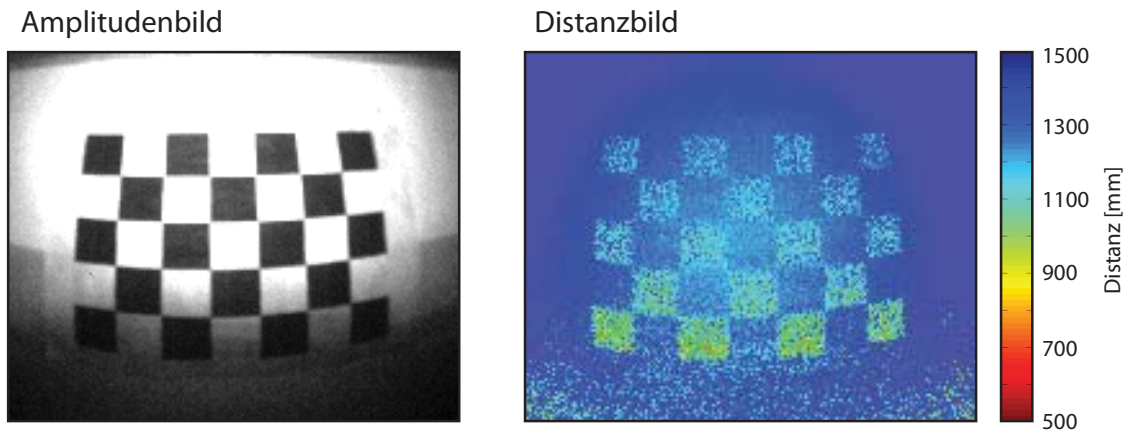


Abbildung 3.13: Die Amplituden- und Tiefenwerte einer Aufnahme eines planaren Objekts mit Schachbrettmuster

Bei der Veränderung der Temperaturen der Lichtquellen ergibt sich eine Veränderung der gemessenen Tiefenwerte. In Abbildung 3.14 sind die Temperatur und der gemessene Offset für drei verschiedene Pixel des Tiefenbildes in einem Diagramm gegeneinander aufgetragen. Dabei ergibt sich ein linearer Zusammenhang zwischen Temperatur und Offset. Auch zeigt sich, dass der lineare Zusammenhang nicht identisch ist für alle Pixel des Bildes. Die Pixel besitzen einen individuellen, linearen Zusammenhang zwischen Lichtquellentemperatur und gemessener Distanz. Dies ist darauf zurückzuführen, dass sich die LEDs einer Kamera unterschiedlich stark erhitzen und an jedes Pixel Licht in unterschiedlicher Zusammensetzung aus den unterschiedlichen Lichtquellen gelangt.

Die unterschiedlichen Mischphasen-Effekte treten im Kamerasystem in unterschiedlichen Intensitäten auf. Während der *Flying-Pixel*-Effekt analog zu herkömmlichen Time-of-Flight-Kameras auftritt, hält sich der *Multipath*-Fehler in Grenzen. Der Grund hierfür ist vor allem das starke Auftreten des *Scattering*-Effekts. Das eingesetzte Weitwinkelobjektiv führt zu einer hohen Menge an Reflexionen innerhalb des Linsensystems. Dies hat zur Folge, dass nahe Objekte, die stark beleuchtet werden, Licht auf den gesamten Sensor strahlen. Diese Überstrahlung führt dann zu einer Veränderung der Distanzwerte, die mehrere Zentimeter betragen kann.

Bewegungsartefakte können bei hohen Geschwindigkeiten auftreten. Im Geschwindigkeitsbereich des Parkens und Rangierens können diese Effekte allerdings nur selten beobachtet werden.

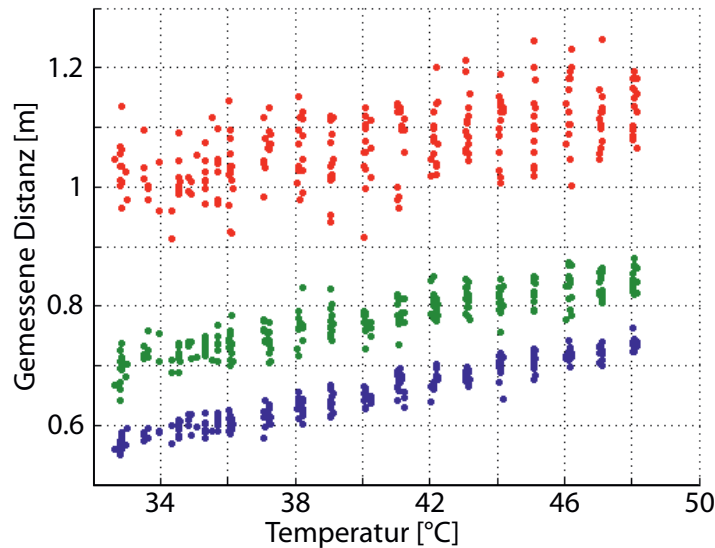


Abbildung 3.14: Zusammenhang zwischen Lichtquellentemperatur und Offset in der Tiefenmessung für drei verschiedene Pixel im Tiefenbild

3.4.3 Umgang mit Messfehlern

In den vorherigen Abschnitten wurden die verschiedenen stochastischen und systematischen Fehler des eingesetzten Kamerasystems vorgestellt. Diese verschiedenen Fehler sollten, so weit möglich, in einer Nachverarbeitung korrigiert werden. Dies kann durch eine Kalibrierung der Tiefendaten erfolgen.

Dabei erscheint auch die Korrektur der distanzabhängigen und amplitudenabhängigen Fehler möglich. Da die Temperatur der LEDs während des Betriebes der Kameras gemessen werden kann, erscheint die Korrektur des temperaturabhängigen Fehlers möglich. Auch die Fehler, welche durch die Anordnung der Lichtquellen zur Optik entstehen, können teilweise korrigiert werden. Die auftretenden *Scattering* Effekte hängen von den Objekten der Szene ab. Ein großer Teil des *Scatterings* entsteht allerdings durch Reflexionen am Fahrzeug selbst und tritt daher in jedem Zustand auf. Diese Anteile des *Scatterings* können auch in einer Kalibrierung berücksichtigt werden. In Abschnitt 3.5 wird eine Methode zur Kalibrierung und Korrektur der genannten Effekte vorgestellt.

Eine Korrektur ist allerdings nicht für alle vorgestellten Fehler möglich. Fehler, die nicht korrigiert werden können, sollen dennoch erkannt werden. Ideal ist ein aussagekräftiger Konfidenzwert pro Pixel, welcher in die nachfolgenden Verarbeitungsschritte berücksichtigt werden kann. In diesen Konfidenzwert soll der stochastische Fehler in den Tiefenmessungen einfließen. Unterbelichtung und die Störung durch Fremdlicht sollte daher erkannt werden. Auch das Auftreten von verschiedenen Mischphaseneffekten wie Scattering und *Flying Pixel* soll erkannt werden und in den pixelbasierten Konfidenzwerten einfließen. Eine Methode zur Bildung eines aussagekräftigen pixelbasierten Konfidenzwert wird in Kapitel 4.3 vorgestellt.

Multipath-Effekte und Bewegungsartefakte werden aufgrund des geringen Auftretens im verwendeten Kamerasystem vernachlässigt.

3.5 Kalibrierung

Eine PMD-Kamera muss, wie in Kapitel 2.1.4 erläutert, hinsichtlich dreier Aspekte kalibriert werden. Hierbei handelt es sich um die intrinsische, extrinsische und die Tiefenkalibrierung. Zur intrinsischen und extrinsischen Kalibrierung existieren Verfahren, welche von herkömmlichen Kameras auf eine Weitwinkel-PMD-Kamera übertragen werden können. Zur Kalibrierung der Tiefendaten wird hingegen ein neuartiges Verfahren vorgestellt, das die Eigenschaften eines Weitwinkel-PMD-Systems berücksichtigt.

3.5.1 Intrinsische und extrinsische Kalibrierung

Die optischen Abbildungseigenschaften der Kameras werden durch die intrinsische Kamerakalibrierung bestimmt. Die Kameraparameter werden mit der in Kapitel 2.1.4 vorgestellten Methode nach Scaramuzza [44] ermittelt. Das omnidirektionale Kameramodell weist jedem Pixel in Bildkoordinaten einen dreidimensionalen Vektor zu. Diese durch die intrinsische Kalibrierung bestimmten Abbildungseigenschaften werden auch für die extrinsische Kalibrierung verwendet. Bei bekannten Abbildungseigenschaften kann die relative Pose einer Landmarke mit fester Größe eindeutig bestimmt werden. Zur extrinsischen Kalibrierung werden nun Landmarken



verwendet, die eine bekannte Position relativ zum Fahrzeugkoordinatensystem besitzen. Nachdem die Pose der Landmarke relativ zum Kamera-Ursprung ermittelt wurde, kann somit die Pose der Kamera relativ zum Fahrzeugkoordinatensystem bestimmt werden.

3.5.2 Tiefenkalibrierung

Das Ziel der Tiefenkalibrierung ist die Korrektur von Fehlern in den Distanzmessungen. Hierbei sollen unterschiedliche Effekte berücksichtigt werden, welche in Abschnitt 3.4 genannt werden.

Korrigiert werden sollen die Distanz- und die Amplitudenabhängigkeit der Tiefenmessung. Auch der temperaturabhängige Fehler soll kompensiert werden. Die Distanzfehler, die durch die Lichtquellengeometrie auftreten, hängen von der Distanz eines Objektes und vom jeweiligen Pixel der Kamera ab. Daher kann auch dieser Fehler als distanzabhängiger Fehler betrachtet werden, der für jeden Pixel individuell ist. Ähnlich verhält es sich mit dem Effekt des *Scatterings*, welches durch die Oberfläche des eigenen Fahrzeugs entsteht. Das vom Fahrzeug reflektierte Licht führt in jeder Szene zu derselben Überlagerung des Lichts auf die einzelnen Pixel des Chips. Das Ausmaß des Effektes hängt vom Amplitudenwert eines jeweiligen Pixels ab und ist demzufolge individuell für jeden Pixel.

Um die genannten Effekte zu kompensieren, müssen daher Distanz, Amplitude, Temperatur und Position einzelner Pixel berücksichtigt werden. Ein erschwerender Faktor ist dabei die Tatsache, dass die jeweiligen Effekte sich gegenseitig beeinflussen, sodass sie nicht unabhängig voneinander betrachtet werden können. Lediglich der Effekt der Temperatur hat keine signifikante Auswirkung auf das Ausmaß des distanz- und amplitudenabhängigen Fehlers.

Um die genannten Effekte zu korrigieren, müsste eine mehrdimensionale Korrekturfunktion bestimmt werden mit den Dimensionen Distanz, Amplitude, Zeile, Spalte und Temperatur. Da dies nicht rechen- und speichereffizient durchzuführen ist, wird die Korrekturfunktion durch eine intelligente Datenstruktur modelliert. Im Falle des hier vorgestellten Verfahrens wird ein Suchbaum durch ein Trainingsverfahren erstellt. Das Trainingsverfahren basiert auf einem Vergleich von realen Aufnahmen mit Referenzbildern.

Die Korrektur der Tiefeninformationen erfolgt im vorgestellten Verfahren in mehreren Schritten. Zunächst wird der temperaturabhängige Fehler kompensiert. Anschließend werden die individuellen Pixel durch eine konstante Offset-Maske korrigiert. Im nachfolgenden Schritt werden die Amplituden- und Distanzwerte durch die Suche des passenden Korrekturwertes im erstellten Suchbaum berücksichtigt.

Das Verfahren wird anhand einer Referenzszene bewertet. Die Kamera befindet sich hierbei in unterschiedlichen Winkeln vor einer senkrechten Wand.

Temperaturkompensation

In Abschnitt 3.4.2 wird der temperaturabhängige Fehler der Tiefenmessungen untersucht. Es kann ein linearer Zusammenhang beobachtet werden, welcher eine individuelle Ausprägung in jedem Pixel besitzt. Der Fehler kann korrigiert werden, wenn der genaue lineare Zusammenhang für jeden Pixel bekannt ist.

Dieser kann durch Aufnahmen in einem größeren Temperaturbereich bestimmt werden. Statische Aufnahmen werden bei einer geringen Lichtquellentemperatur und einer festen Integrationszeit gestartet. Im Laufe der Zeit erhöhen sich die Temperaturen der Lichtquellen und somit auch die Tiefenwerte der einzelnen Pixel. Der lineare Zusammenhang zwischen Temperatur und gemessener Distanz kann anschließend anhand der Messdaten durch eine lineare Regression bestimmt werden. Dies kann für jeden Pixel individuell geschehen. Die resultierenden Steigungen der linearen Zusammenhänge sind für die Pixel der Frontkamera des Fahrzeugs in Abbildung 3.15 dargestellt. Die aufzunehmende Szene besteht aus einer horizontalen Bodenebene und einer vertikalen Wand und ist in Abbildung 3.16 dargestellt.

Der lineare Zusammenhang kann durch die Angabe einer Basistemperatur t_{basis} und der Steigung m definiert werden. Der korrigierte Distanzwert kann durch

$$d_{korrigiert} = d_{aktuell} + (t_{basis} - t_{aktuell}) \cdot m \quad (3.12)$$

mit Berücksichtigung der aktuellen Temperatur $t_{aktuell}$ aus dem aktuellen Distanzwert $d_{aktuell}$ bestimmt werden.

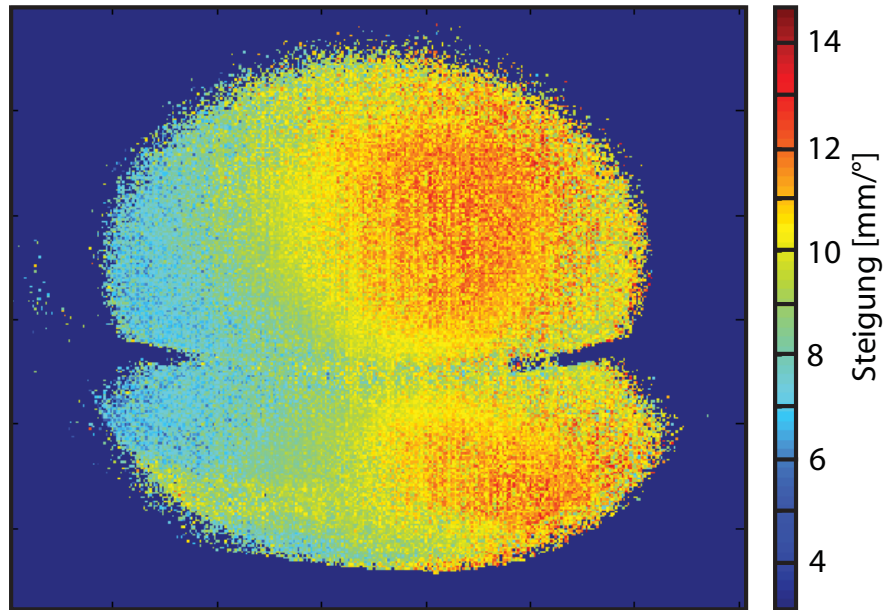


Abbildung 3.15: Darstellung der Temperaturabhängigkeit der Tiefenmessung. Abgebildet sind die individuellen Steigungen des linearen Zusammenhangs zwischen Temperatur und Distanzwert eines Pixels.

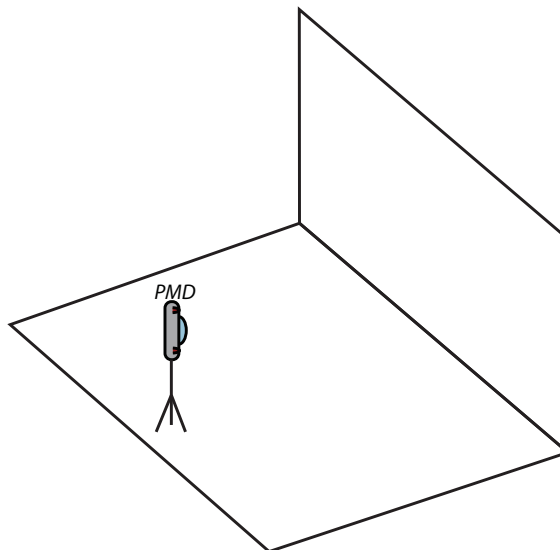


Abbildung 3.16: Die aufzunehmende Szene besteht aus einer horizontalen Bodenebene und einer vertikalen Wand.

Offset-Maske

Die Offset-Werte werden durch einen Vergleich zwischen den Tiefenbildern von realen Aufnahmen und Referenziefenbildern generiert. Die Erstellung der Referenziefenbilder erfolgt durch das zur Evaluierung aufgebaute Referenzsystem und wird ausführlich in Kapitel 7.2.3 thematisiert. In jedem Vergleich werden nur die Pixel betrachtet, deren Amplitude einen Schwellwert überschreitet. Auf diese Weise werden Tiefenwerte, die einem sehr starken Rauschen unterliegen, nicht bei der Berechnung des Offsets mit einbezogen.

Vor der Ermittlung der Offset-Werte werden die aufgenommenen Tiefenwerte durch das im vorherigen Abschnitt vorgestellte Verfahren der Kompensation der Lichtquellentemperatur korrigiert.

In einem ersten Versuch wird ein globaler Offset bestimmt, welcher auf alle Pixel angewendet wird. Dies geschieht, indem der durchschnittliche Offset über alle gültigen Pixel und alle betrachteten Aufnahmen ermittelt wird. Das Ergebnis der Korrektur einer beispielhaften Aufnahme durch einen einzelnen globalen Offset ist in Abbildung 3.17 zu sehen. Dargestellt sind die nach der Korrektur verbleibenden Fehler pro Pixel.

Es können hohe Abweichungen beobachtet werden, da die Unterschiede zwischen den einzelnen Pixeln nicht berücksichtigt werden. Dieser Effekt macht daher die Bestimmung eines individuellen Offsets für jeden Pixel notwendig.

Die einzelnen Offset-Werte werden analog zum globalen Offset aus dem Durchschnittswert der Differenzen zwischen den realen Tiefenwerten und den Tiefenwerten der Referenzbilder ermittelt. Für jedes Pixel wird ein individueller Offset bestimmt, sodass eine zweidimensionale Offset-Maske entsteht. Die Offset-Werte können anschließend noch in die Bereiche der Maske extrapoliert werden, für die in den Trainingsdaten keine gültigen Messungen existieren. Anschließend wird eine Glättung des Bildes durchgeführt. Die verbleibenden Fehler nach der Korrektur des Tiefenbildes durch eine Offsetmaske sind in Abbildung 3.18 dargestellt.

Es kann eine deutliche Verbesserung gegenüber der Korrektur durch einen globalen Offset beobachtet werden. Dennoch existieren einige Bereiche, in denen große Fehler auftreten. Diese treten insbesondere in Bereichen mit niedriger Amplitude auf. Da der amplitudenabhängige Fehler nicht berücksichtigt wird, können bei niedrigen Amplituden auch nach der Korrektur durch die Offset-Maske hohe Fehler auftreten.

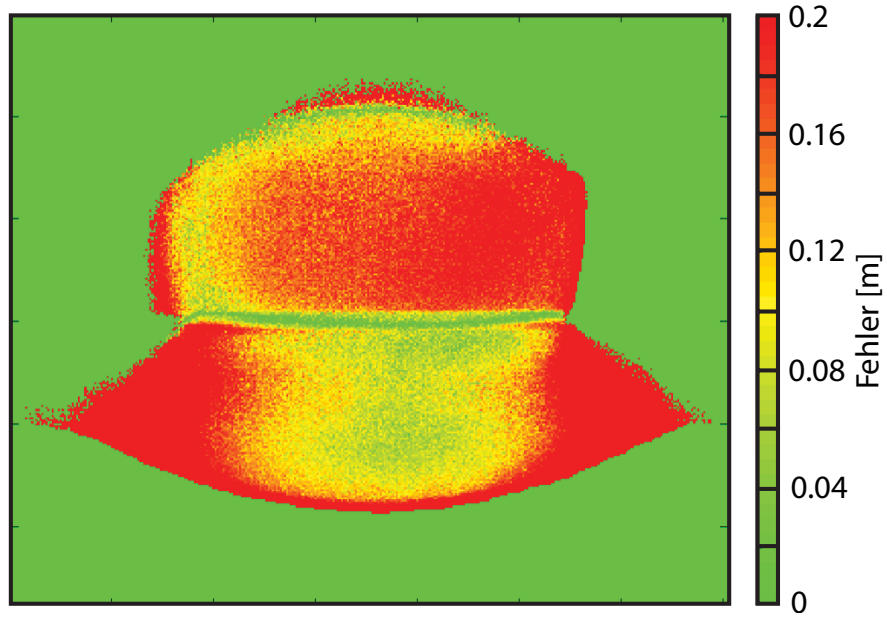


Abbildung 3.17: Verbleibende Fehler nach der Korrektur durch einen globalen Offset

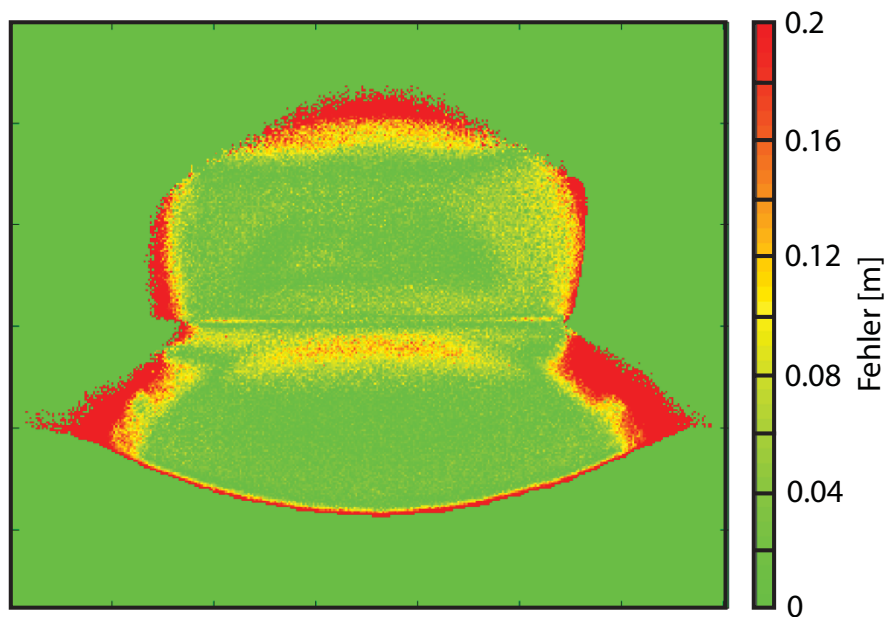


Abbildung 3.18: Verbleibende Fehler nach der Korrektur durch eine Offset-Maske

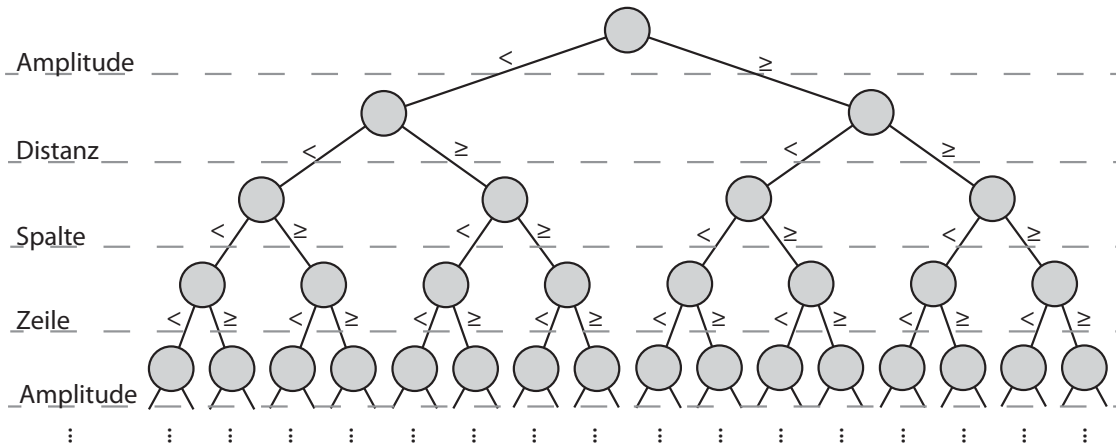


Abbildung 3.19: K-d-Baum mit den vier Dimensionen Amplitude, Distanz, Zeile und Spalte.

Suchbaum-basierte Korrektur

Um die distanz- und amplitudenabhängigen Fehler zu kompensieren, wird eine mehrdimensionale Korrekturfunktion durch einen mehrdimensionalen Suchbaum modelliert. Im Falle des vorgestellten Verfahrens wird auf einen k-d-Baum zurückgegriffen. K-d-Bäume werden in [131] ausführlich behandelt. Ein k-d-Baum besitzt k Dimensionen. Jeder Knoten des Baumes entspricht einem k-dimensionalen Punkt. Jeder Knoten, welcher kein Blatt des Baumes ist, teilt den Suchraum durch eine Hyperebene in zwei Bereiche. Die beiden Kinder eines Knotens repräsentieren dabei die Punkte, die auf den jeweiligen Seiten der teilenden Hyperebene liegen. Jede Ebene des Baumes ist einer Dimension zugeordnet. Beim Durchlaufen des Baumes von der Wurzel bis zu einem beliebigen Blatt gelangt man an Knoten, die abwechselnd den unterschiedlichen Dimensionen zugeordnet sind.

Da im Falle der Korrekturfunktion die Dimensionen Amplitude, Distanz, Zeile und Spalte berücksichtigt werden sollen, wird ein vierdimensionaler Baum eingesetzt. Abbildung 3.19 stellt einen solchen vierdimensionalen Baum schematisch dar. Der Baum wird dabei automatisiert durch Trainingsaufnahmen generiert. Herangezogen werden erneut Paare aus Tiefen-



3 Systemauslegung

bildern von realen Aufnahmen und Referenztiefenbildern. Pixel, die eine niedrige Amplitude besitzen, werden analog zur Erstellung der Offset-Maske aussortiert. Die übrigen Punkte werden nacheinander in den k-d-Baum eingetragen. Als Schlüssel werden dabei Tupel aus Amplitude, Distanz, Zeile und Spalte gebildet, und als Korrekturwert wird die Differenz zwischen dem Tiefenwert der Referenz und dem realen Tiefenwert gewählt. Auf diese Weise wird ein vierdimensionaler k-d-Baum erstellt, mit welchem durch Angabe eines beliebigen Tupels aus Amplitude, Distanz, Zeile und Spalte nach passenden Korrekturwerten gesucht werden kann.

Hierdurch entsteht ein sehr großer Baum, da jeder gültige Pixel aus den Trainingsdaten zu einem Knoten des Baumes wird. Dies macht die Reduktion der Datengröße des Suchbaumes notwendig. Die Knoten des Baumes werden daher zu kombinierten Knoten zusammengefasst.

Hierzu wird ein iterativer Prozess angewendet, der nacheinander zufällig einen noch nicht zusammengefassten Knoten bestimmt. Im k-d-Baum wird anschließend eine Suche nach den N nächsten Nachbarn auf Basis der euklidischen Distanz zwischen den normierten Tupeln aus Amplitude, Distanz, Zeile und Spalte durchgeführt. Die einzelnen Dimensionen werden auf Werte zwischen 0 und 1 normiert, um eine passende Gewichtung zwischen den Dimensionen bei der Bestimmung der euklidischen Distanz zu erhalten. Ein neues zusammengefasstes Tupel wird anschließend aus dem gewichteten Durchschnitt der insgesamt N Punkte gebildet. Dabei erfolgt zum einen die Gewichtung auf Basis der euklidischen Distanz. Zum anderen erhalten die neuen schon kombinierten Punkte ein höheres Gewicht, da sie die Daten von mehreren Knoten beinhalten. Das Ergebnis der Kombination der benachbarten Knoten wird als neuer kombinierter Knoten in den Baum eingefügt. Die benachbarten Knoten selbst werden gelöscht, es sei denn es handelt sich um schon zuvor kombinierte Knoten.

Auf diese Weise ist es möglich, die Datenmenge des Suchbaums zu reduzieren und zu glätten, ohne einen großen Verlust an Informationen zu erhalten. Im Suchbaum kann anschließend für beliebige Pixel nach einem Korrekturwert gesucht werden. Abbildung 3.20 stellt die verbleibenden Fehler nach der Korrektur des beispielhaften Tiefenbildes durch das vorgestellte Verfahren dar. Es ist zu beobachten, dass auch an den Rändern des sichtbaren Bereiches geringe Fehler auftreten, da nun verschiedene Amplitudenwerte berücksichtigt werden. Dennoch treten an einzelnen Positionen

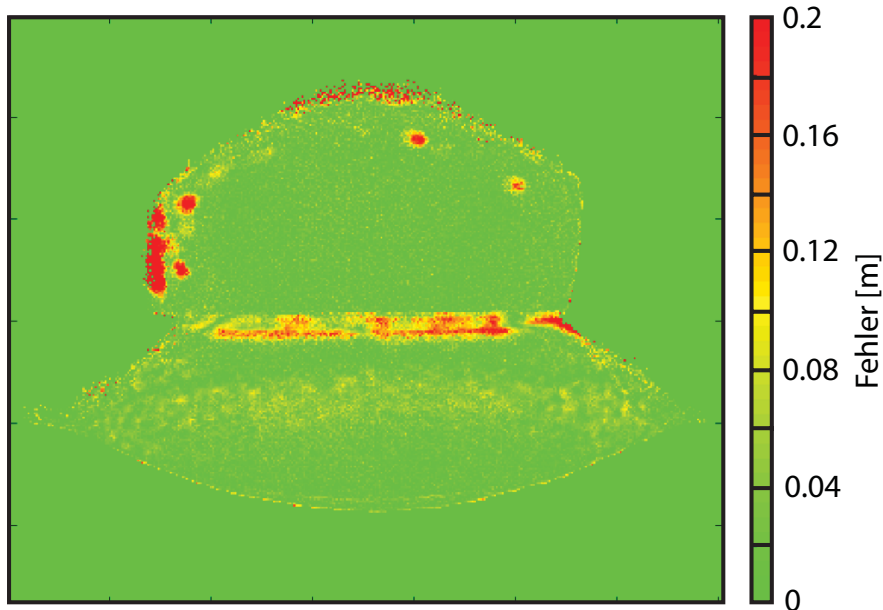


Abbildung 3.20: Verbleibende Fehler nach der Korrektur durch einen auf Basis von Trainingsdaten erstellten Suchbaum

im Bild lokal begrenzte Fehler auf. Diese sind darauf zurückzuführen, dass an diesen Pixelpositionen keine ausreichende Trainingsdatenmenge vorhanden war. Da auch die Pixelposition bei der Korrektur berücksichtigt wird, führen Lücken einzelner Pixelbereiche in der Trainingsdatenmenge zu Fehlern in der Korrektur des Tiefenbildes. Eine sinnvolle Möglichkeit diese Lücken zu verhindern, wäre eine intelligente, automatisierte Bestimmung der Trainingsdatenmenge durch das in 7.2 vorgestellte Referenzsystem.

3.6 Fusion der Sensordaten

Die Tiefendaten der Kameras müssen in 3D-Punkte umgerechnet und in ein gemeinsames Koordinatensystem transformiert werden. Dies wird im nächsten Abschnitt vorgestellt. Anschließend wird thematisiert, wie eine gegenseitige Beeinflussung der Kameras vermieden werden kann.

3.6.1 Transformation der Messpunkte

Jeder Pixel einer Kamera liefert eine eindimensionale Distanzinformation. Um zu einer gemeinsamen Repräsentation der Tiefenmessungen aller Kameras zu gelangen, sind mehrere Koordinatentransformationen notwendig. Die eindimensionalen Distanzwerte werden zunächst zu dreidimensionalen Punkten im Kamerakoordinatensystem transformiert. Dies geschieht unter Berücksichtigung der jeweiligen Pixelkoordinaten durch die optischen Abbildungseigenschaften der Kamera. Hierzu wird aufgrund der verwendeten Weitwinkelobjektive die in Kapitel 2.1.4 vorgestellte intrinsische Kalibrierung nach Scaramuzza [44] verwendet.

Das Ergebnis der in Kapitel 2.1.4 präsentierten extrinsischen Kalibrierung ist eine Transformationsmatrix, die aus einer Rotationsmatrix und einem Translationsvektor besteht. Diese berücksichtigt die Kamerapose und transformiert Messpunkte vom Kamerakoordinatensystem in ein übergeordnetes Fahrzeugkoordinatensystem. Für jede Kamera existiert hierbei eine eigene Transformationsmatrix für die Transformation in das Fahrzeugkoordinatensystem. Der Bezugspunkt des Fahrzeugkoordinatensystems ist der von der Mitte der Hinterachse des Fahrzeugs auf den Boden projizierte Punkt. Nach der Transformation vom Kamera- zum Fahrzeugkoordinatensystem beziehen sich die 3D-Punkte aller Kameras auf ein gemeinsames Koordinatensystem.

Bei Bewegung des Fahrzeugs ist noch eine weitere Transformation notwendig. Die Transformation der 3D-Punkte vom Fahrzeugkoordinatensystem in das Weltkoordinatensystem wird aus der Fahrzeugpose bestimmt und besteht ebenfalls aus einer Rotation und einer Translation. Der Bezugspunkt befindet sich an einem beliebigen Startpunkt auf der Bodenebene. Die verschiedenen Ursprünge der Koordinatensysteme sind in Abbildung 3.21 dargestellt.

3.6.2 Gegenseitige Beeinflussung

Wenn mehrere PMD-Kameras simultan dieselbe Oberfläche vermessen, können Fehlereffekte auftreten. Da in jeder Kamera sowohl das eigene als auch das Lichtsignal der jeweils anderen Kamera reflektiert wird, entsteht ein Mischeffekt von unterschiedlichen Lichtsignalen. Dieser Effekt führt zu Fehlern in den Tiefenmessungen, da die Phasendifferenzen zwischen ausge-

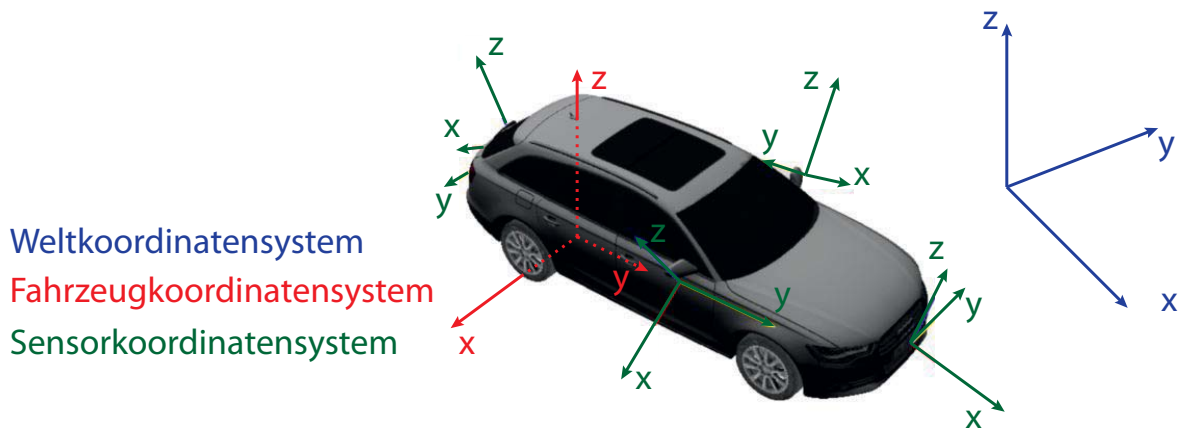


Abbildung 3.21: Darstellung der verschiedenen Koordinatensysteme

henden und eingehenden Lichtsignalen nicht mehr korrekt bestimmt werden können. In der in Abschnitt 3.2 definierten Integration befinden sich an den Ecken des Fahrzeugs Bereiche, die von mehreren PMD-Kameras erfasst werden. Um hier eine gegenseitige Beeinflussung zu vermeiden, werden benachbarte Kameras jeweils nacheinander zu einer Messung getriggert. Gegenüberliegende Kameras können eine gleichzeitige Messung durchführen, da sie keinen überlappenden Bereich besitzen.

Eine weitere Möglichkeit, diesen Fehler zu vermeiden, ist die Verwendung von unterschiedlichen Modulationsfrequenzen. Hierdurch wird verhindert, dass eine fehlerhafte Phasendifferenz bestimmt wird. Allerdings wirkt das Lichtsignal einer zweiten Kamera als Fremdlicht und verschlechtert somit das Signal-Rausch-Verhältnis. Ebenso könnten die Lichtsignale durch ein Codemultiplexverfahren unterschieden werden [132]. Diese Möglichkeit bietet die eingesetzte PMD-Technologie allerdings nicht.



4 Vorverarbeitung

Die Vorverarbeitung ist eine erste Bearbeitung der Kamerabilder. Hier wird detektiert, ob Überbelichtung besteht, um hierdurch die Integrationszeit zu steuern. Die Bilddaten werden geglättet, und ein Konfidenzbild mit einem Konfidenzwert für jeden Pixel wird berechnet.

4.1 Dynamische Sensoransteuerung

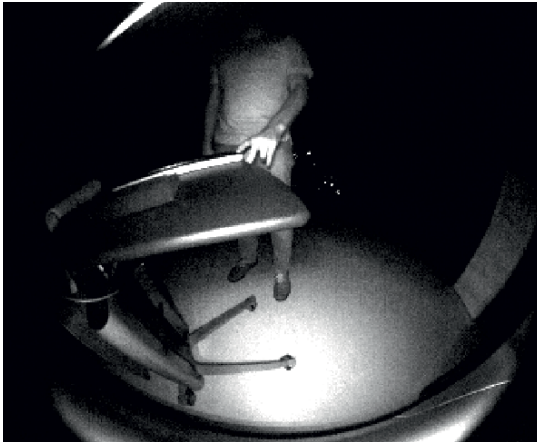
Eine dynamische Steuerung der Integrationszeit kann Effekte wie eine Unter- oder Überbelichtung vermeiden. Die Amplituden- und Distanzbilder werden analysiert, um die Integrationszeit an die Szene anzupassen.

4.1.1 Überbelichtung und Unterbelichtung

In Kapitel 2.1.3 wurde beschrieben, dass bei einem geringen Nutzsignal das Rauschen in der Distanzmessung zunimmt. Durch ein Erhöhen der Integrationszeit kann die Menge des nutzbaren Signals und damit die Qualität der Messung erhöht werden. Zu hohe Integrationszeiten können allerdings zu einer Überbelichtung der Szene führen. Bei einem zu starken Nutzsignal kann ein Pixel sättigen, sodass keine korrekte Distanz mehr bestimmt werden kann. Zudem können stark belichtete, nahe Objekte zu dem in Kapitel 2.1.3 dargestellten *Scattering*-Effekt führen.

Daher sollte das Auftreten von Überbelichtung detektiert werden, um die Integrationszeit bei Bedarf auf eine kürzere Länge zu setzen. Wenn keine Überbelichtung vorhanden ist, sollte die Integrationszeit auf einen möglichst hohen Wert gesetzt werden, da höhere Integrationszeiten im Allgemeinen zu einem besseren Signal-Rausch-Verhältnis führen und die Genauigkeit und die Reichweite der Distanzmessung erhöhen.

Limitiert wird die Integrationszeit durch verschiedene Faktoren. Eine zu lange Integrationszeit verlängert die Dauer einer einzelnen Aufnahme und



a) Keine Überbelichtung

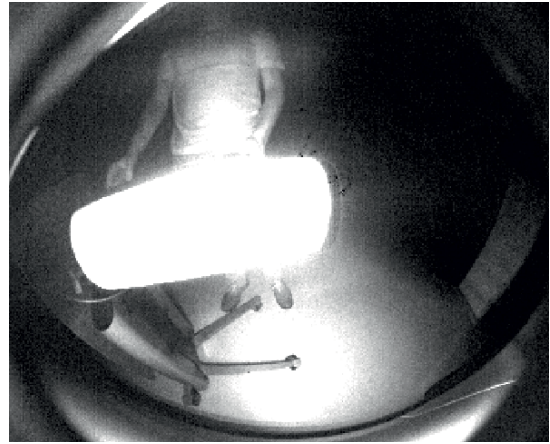
b) Überbelichtung durch *Scattering*

Abbildung 4.1: Amplitudenbilder ohne (a) und mit (b) Auftreten von Überbelichtung

schränkt daher die Framerate der Kamera ein. Zudem führen lange Integrationszeiten bei Bewegung zu Bewegungsartefakten. Auch generieren zu lange Integrationszeiten eine erhöhte Hitzeentwicklung der Lichtquellen. Bei der in Kapitel 3.2 vorgestellten Integration wird die maximale Integrationszeit auf 5 ms definiert.

4.1.2 Anpassung der Integrationszeit

Zur dynamischen Anpassung der Integrationszeit wurde eine neuartige Methode entwickelt, welche die Integrationszeit abhängig von der Szene auf eine von drei verschiedenen Integrationszeiten setzt. Das System läuft so lange mit der höchsten Integrationszeit, bis eine Überbelichtung detektiert wird. Dann wechselt das System zu einer niedrigeren Integrationszeit. Wird nun erneut oder weiterhin eine Überbelichtung detektiert, wird auf die niedrigste Integrationszeit gesetzt. Sobald anschließend erkannt wird, dass die Gefahr der Überbelichtung nicht mehr vorhanden ist, wird wieder auf die höhere Stufe gesetzt.

Der Kern der Methode ist die Erkennung von Überbelichtung in den Amplitudenbildern. Bei der Überbelichtung von nahen Objekten kommt es zum *Scattering*-Effekt. Licht wird innerhalb des Linsensystems reflektiert und streut über den gesamten Chip. In Abbildung 4.1 (a) ist zu sehen,

dass die Optik die Szene nicht in die Eckbereiche des Sensors abbildet. Bei Aufnahmen, bei denen keine Streulichteffekte auftreten, bleiben diese Pixel dunkel. Treten hingegen Streulichteffekte auf, gelangt Licht auch in diese Bereiche und die Amplitudenwerte der Pixel steigen. Dies zeigt Abbildung 4.1 (b). Durch diesen Effekt kann Streulicht, das bei Überbelichtung auftritt, an erhöhten Amplitudenwerten erkannt werden. Hierzu wird der Durchschnittswert der Amplitudenwerte aller Pixel bestimmt, die eine Maximaldistanz zum Mittelpunkt des Sensors überschreiten. Überschreitet dieser Durchschnittswert der Amplitude einen Schwellwert, dann wird davon ausgegangen, dass die Szene überbelichtet ist, und die Integrationszeit wird auf die nächstniedrigere Stufe gesetzt.

Das System sollte die Integrationszeit wieder auf die höhere Stufe zurücksetzen, wenn davon ausgegangen werden kann, dass die Gefahr der Überbelichtung nicht mehr besteht. Um dies umzusetzen, wird, unmittelbar nachdem die Integrationszeit herabgesetzt wurde, die Durchschnittsamplitude über das gesamte Bild bestimmt. Dieser Durchschnittswert wird um 10% erhöht und gilt dann als Schwellwert zum Zurückschalten zur nächsthöheren Integrationszeit. Auf diese Weise kann detektiert werden, sobald die Objekte in der Szene weniger stark beleuchtet werden und somit die Gefahr der Überbelichtung reduziert ist.

Als Stufen für die verschiedenen Integrationszeiten werden für die erfolgte Systemauslegung die Werte $5000 \mu s$, $2000 \mu s$ und $500 \mu s$ festgelegt. Für das Herunterschalten zur nächstniedrigeren Integrationszeit konnte beim *Camboard ATV* mit dem verwendeten 180° Weitwinkelobjektiv für die durchschnittliche Amplitude in den Eckbereichen der Wert 200 als Schwellwert experimentell bestimmt werden.

4.2 Bildglättung

Sowohl die Distanzdaten als auch die Amplitudendaten unterliegen einem Rauschen. Das Rauschen kann durch eine Glättung der Daten reduziert werden. Da die Amplituden und Tiefendaten als 2D-Bilddaten verfügbar sind, können bekannte Filter der Bildverarbeitung genutzt werden. Hierbei macht es keinen Unterschied, ob es sich bei den Werten um Distanzwerte oder um Intensitätswerte handelt. Bei der Glättung der Daten ist darauf zu achten, dass das Rauschen entfernt wird ohne dass relevante Strukturen des Bildes verloren gehen.



Abbildung 4.2: Bilateraler Filter nach [136]

Ein solcher kantenerhaltender Filter ist der nichtlineare Median-Filter [133], welcher schon in [134] vorgeschlagen wurde. Der Wert eines Pixels wird hierbei auf den Median seiner Nachbarschaft gesetzt. Auf diese Weise bleiben Kanten im Bild erhalten.

In [135] wird ein signaladaptiver Glättungsoperator vorgestellt, welcher auf einer lokalen Kantenschätzung basiert. Auf diese Weise können Kanten nach der Glättung erhalten werden.

Eine Alternative ist der bilaterale Filter nach [136]. Auch dieser ist nicht-linear und kantenerhaltend. Durch eine Gewichtungsfunktion werden die Pixelwerte der Nachbarschaft eines Pixels unterschiedlich stark gewichtet. Dabei richtet sich die Gewichtung nicht nur nach der Entfernung des Pixels, sondern auch nach der photogrammetrischen Ähnlichkeit. Das Prinzip des Filters ist in Abbildung 4.2 dargestellt.

4.3 Konfidenzberechnung

In diesem Abschnitt wird eine Methode zur Erstellung einer Konfidenzmaske mit einem Konfidenzwert pro Pixel vorgestellt. Der Konfidenzwert gibt an, inwiefern der Messung eines Pixels vertraut werden kann, und besitzt einen Wertebereich zwischen 0 und 1. In Kapitel 2.2.2 wird der Stand der Technik bezüglich der Konfidenzbestimmung von Time-of-Flight-Kameras beschrieben.

Im nachfolgend vorgestellten Verfahren sollen verschiedene auftretende Fehler berücksichtigt werden, die in Kapitel 3.4 untersucht werden. Für

einen Teil der Fehler kann eine Korrektur stattfinden. Die hierzu umgesetzte Tiefenkalibrierung wird in Kapitel 3.5.2 vorgestellt. In den Konfidenzwert sollen die Fehlerarten einfließen, die nicht durch eine Kalibrierung korrigiert werden können. Hierzu zählt der stochastische Fehler, der von der Menge an reflektiertem Nutzlicht und vom einfallenden Fremdlicht abhängt. Auch Mischphasen-Effekte, wie der *Scattering*-Effekt, der durch nahe Objekte in der Szene entsteht, soll in einem Konfidenzwert berücksichtigt werden. Zu den zu berücksichtigenden Mischphaseneffekten gehört auch der *Flying-Pixel*-Effekt.

Das nachfolgend vorgestellte Verfahren bestimmt jeweils einen Konfidenzwert pro Pixel für den stochastischen Fehler, den *Scattering*-Effekt und für den *Flying-Pixel*-Effekt. Anschließend werden die Konfidenzwerte zu einem gemeinsamen Wert kombiniert.

4.3.1 Stochastischer Fehler

Eine Untersuchung des stochastischen Fehlers wird in Kapitel 3.4.1 präsentiert. Die Stärke des Rauschens hängt von der Menge des auf einen Pixel reflektierten Nutzlichts und der Menge des Fremdlichts ab. Dabei liefert die Amplitude ein Maß für die Menge des reflektierten Nutzlichts, und die Intensität ist ein Maß für die Menge an Fremdlicht. Durch einen Sonnenlichtsimulator konnten die Standardabweichungen der gemessenen Tiefenwerte unter verschiedenen Bedingungen bestimmt werden.

Mit den gemessenen Werten für die Standardabweichungen wird ein künstliches neuronales Netz trainiert. Hierbei handelt es sich um eine dem menschlichen Gehirn nachempfundene Datenstruktur, die aus vielen miteinander vernetzten Neuronen besteht. Jedes Neuron reagiert dabei auf Eingangsgrößen und hat wiederum einen Einfluss auf nachfolgende Neuronen. Künstliche neuronale Netze werden ausführlich in [137] behandelt. Als Eingangsgrößen kommen die gemessenen Amplituden- und Intensitätswerte zum Einsatz. Die gemessene Standardabweichung der Tiefenwerte gilt als Ausgangsgröße des künstlichen neuronalen Netzes. Das auf diese Weise durch überwachtetes Lernen trainierte künstliche neuronale Netz gibt für eine beliebige Kombination von Amplitude und Intensität eine Standardabweichung aus. In Abbildung 4.3 sind die gemessenen Standardabweichungen in blau den Standardabweichungen des künstlichen neuronalen Netzes in rot gegenübergestellt.

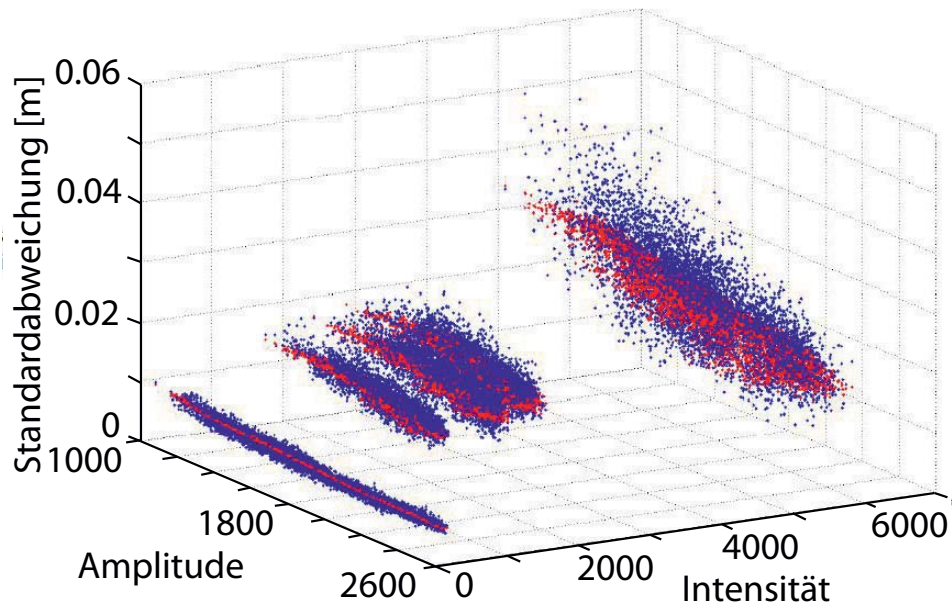


Abbildung 4.3: Gegenüberstellung der gemessenen Standardabweichungen (blau) und der Standardabweichungen des künstlichen neuronalen Netzes (rot)

Der durch das künstliche neuronale Netz bestimmte Wert für die geschätzte Standardabweichung wird durch

$$c_{std} = \frac{2}{\sqrt{\pi}} \int_0^{\frac{v}{\sigma\sqrt{2}}} e^{-\tau^2} d\tau \quad (4.1)$$

auf den Wertebereich des Konfidenzwertes gebracht. Durch den Parameter v kann der Zusammenhang zwischen Standardabweichung und Konfidenzwert beeinflusst werden.

In den Abbildungen 4.4 (a)-(d) sind beispielhafte Amplituden- und Intensitätswerte zusammen mit der geschätzten und der gemessenen Standardabweichung dargestellt. Es kann beobachtet werden, dass die Standardabweichungen auf der am Boden beleuchteten Fläche korrekt vorhergesagt werden. In den übrigen Bereichen wird kaum Nutzlicht zurück auf die Pixel reflektiert, sodass keine korrekten Messungen möglich sind und nur Rauschen produziert wird. Die Abbildungen 4.4 (e)-(f) stellen die geschätzte Standardabweichung der gemessenen Standardabweichung einer

Indoor-Aufnahme gegenüber. Auch hier kann beobachtet werden, dass die geschätzten Standardabweichungen in den Bereichen des Bildes, in denen auch Messungen durchgeführt werden können, mit den gemessenen Standardabweichungen übereinstimmen.

4.3.2 *Scattering*-Effekt

Der *Scattering*-Effekt entsteht bei Reflexionen innerhalb des Linsensystems und wird ausführlich in Kapitel 2.1.3 erläutert. Insbesondere das von nahen Objekten reflektierte Licht kann auf größere Bereiche des Chips streuen, sodass in den betroffenen Pixeln Mischphasen-Effekte entstehen. Das nachfolgend vorgestellte Konzept basiert auf der Erkennung eines *Scattering* verursachenden Objekts. Um ein solches Objekts zu identifizieren, wird der Einfluss auf die benachbarten Pixel bestimmt. Der Einfluss unterschiedlicher Objekte wird auf Basis von Referenzaufnahmen ermittelt. Um den Einfluss durch *Scattering* darzustellen, wird die Darstellung einer Messung als Phasenvektor verwendet, die auch in [138] genutzt wird. Dabei fließt die gemessene Phase und damit die Distanz als Orientierung des Phasenvektors ein. Die gemessene Amplitude entspricht der Länge des Vektors. Mischphasen-Effekte, die sich aus der Überlagerung verschiedener Signale bilden, können somit durch Vektoradditionen dargestellt werden. In Abbildung 4.5 ist der *Scattering*-Effekt auf Basis der Phasenvektordarstellung illustriert. Ein Objekt (rot) verursacht durch seine Nähe zur Kamera einen signifikanten *Scattering*-Effekt, sodass das von diesem Objekt reflektierte Signal auf benachbarte Pixel gelangt. In der Phasenvektordarstellung kann der Effekt als die Addition eines Anteils des roten Phasenvektors zu den korrekten Phasenvektoren benachbarter Pixel betrachtet werden.

Der gemessene Phasenvektor eines benachbarten Pixels (blau und grün) besteht somit aus der Addition des korrekten Phasenvektors und eines Anteils des *Scattering* verursachenden Phasenvektors.

Die hier vorgestellte Methode basiert daher auf den folgenden Schritten: Zunächst wird im Bild nach Objekten gesucht, die *Scattering* verursachen. Anschließend wird für jedes dieser Objekte ein Phasenvektor bestimmt. Dieser beschreibt das Signal, welches auf die benachbarten Pixel strahlt. Zuletzt wird für jedes einzelne Pixel individuell die Stärke des Einflusses

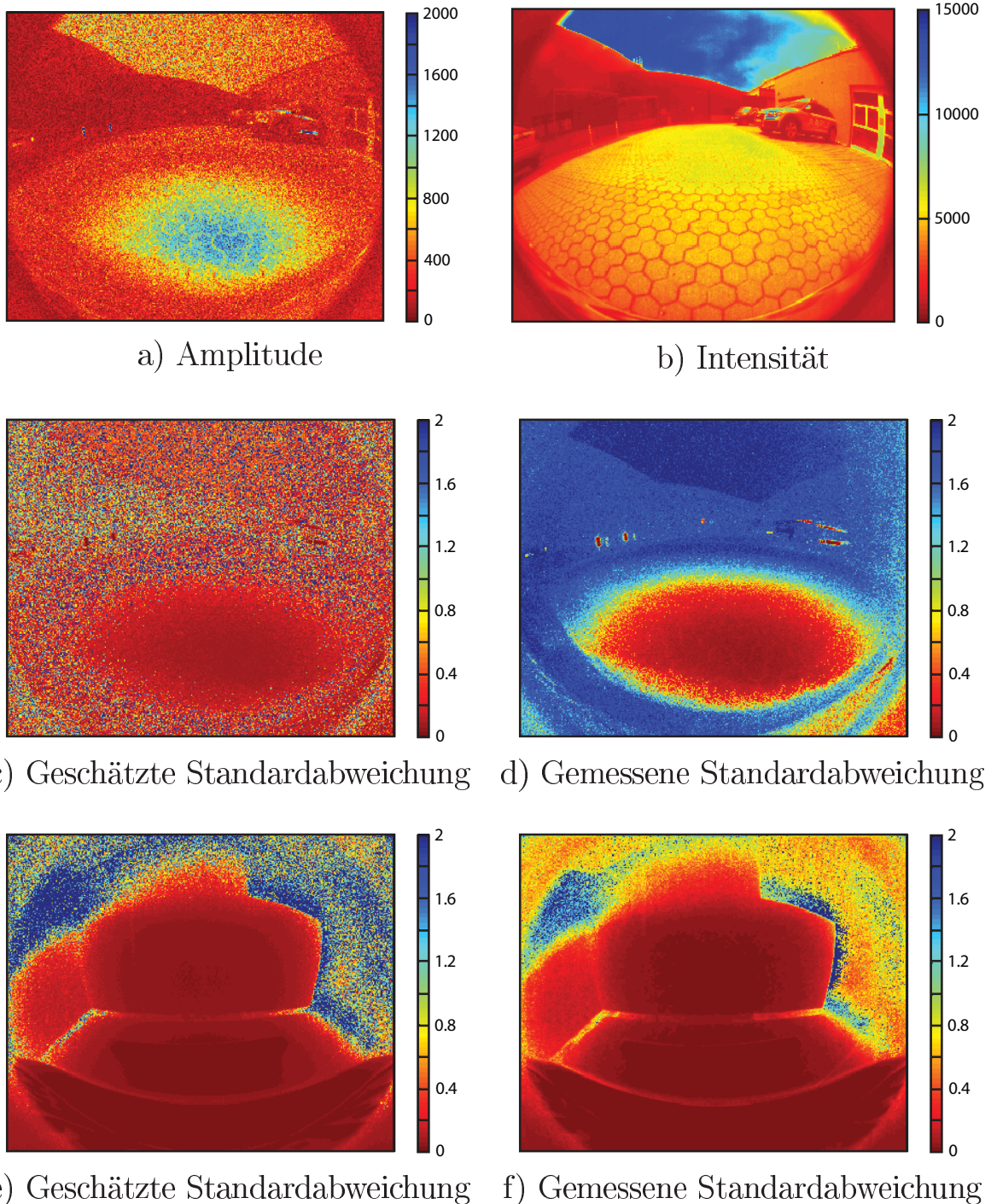


Abbildung 4.4: Ergebnisse der Bestimmung der Standardabweichung aus den Amplituden- und Intensitätswerten. In (a) ist das Amplitudenbild und in (b) das Intensitätsbild einer Outdoor-Aufnahme abgebildet. Die geschätzte Standardabweichung der einzelnen Pixel sind in (c) und die gemessenen Standardabweichungen in (d) dargestellt. In (e) und (f) sind die geschätzten und gemessenen Standardabweichungen einer Indoor-Szene abgebildet.

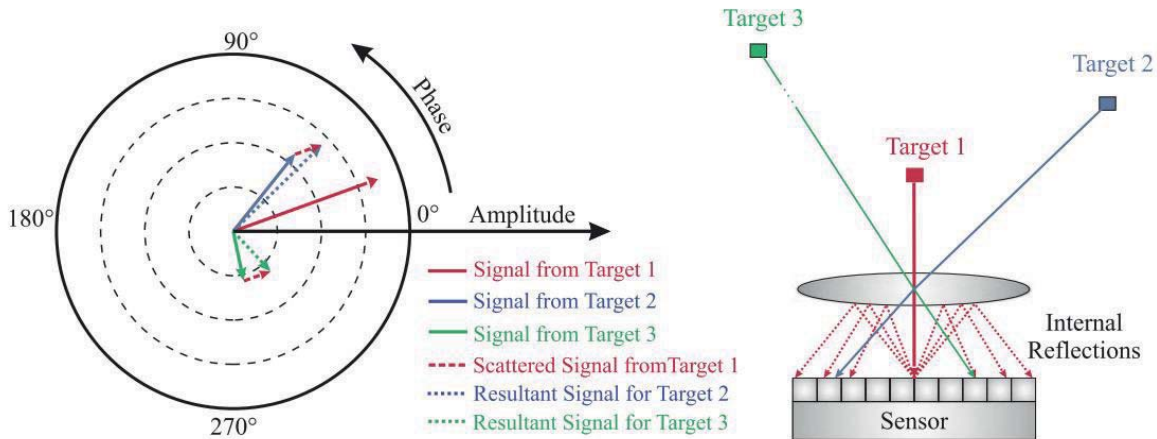


Abbildung 4.5: Illustration des *Scattering*-Effekts [138]. Das reflektierte Signal eines nahen Objekts (rot) wird innerhalb des Linsensystems reflektiert und strahlt auf benachbarte Pixel. Der Effekt kann in der Phasenvektordarstellung als Addition von Vektoren dargestellt werden.

des Phasenvektors bestimmt. Dieser Wert wird dann in einen Konfidenzwert umgerechnet.

Zur Suche eines störenden Objekts werden nur die Pixel betrachtet, die eine ausreichend hohe Amplitude besitzen und nicht ausmaskiert sind. Zunächst werden im Tiefenbild diejenigen Pixel bestimmt, die im Vergleich zu ihren Nachbarpixeln eine minimale Distanz besitzen. Wenn diese Distanz geringer ist als 1 m, dann wird der Pixel als Saatpunkt eines *Region-Growing*-Verfahrens genutzt. Durch das *Region-Growing*-Verfahren wird das zum Pixel zugehörige Segment extrahiert. Wenn dieses Segment aus mindestens 100 Pixeln besteht, dann wird es als Objekt betrachtet, das zu einem *Scattering*-Effekt führt.

Die durchschnittliche Amplitude und der durchschnittliche Tiefenwert der zu einem Objekt zugehörigen Pixel werden ermittelt. Aus diesen wird der Phasenvektor des Objekts bestimmt. Dieser wiederum beeinflusst die Messungen der benachbarten Pixel. Um den Einfluss auf benachbarte Pixel zu ermitteln, werden Referenzaufnahmen mit und ohne nahes Objekt im Vordergrund gemacht. Zunächst wird eine Szene, die sich im Hintergrund befindet, aufgenommen. Anschließend wird die Aufnahme mit einem nahen

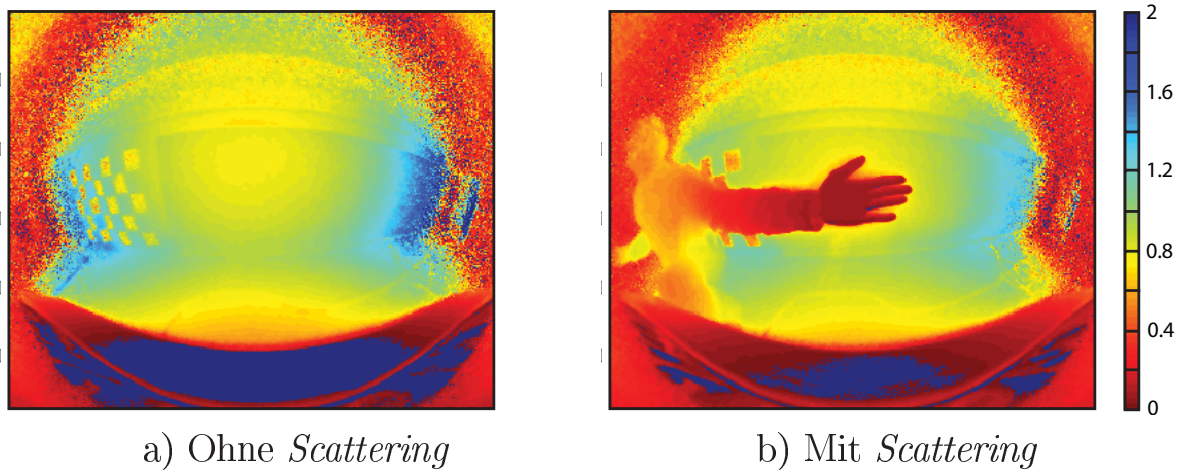


Abbildung 4.6: Beispielhafte Tiefenbilder einer Szene einmal ohne (a) und einmal mit (b) Vordergrundobjekt.

Objekt, das *Scattering* verursacht, wiederholt. In Abbildung 4.6 (a) und (b) sind beispielhafte Aufnahmen einer Szene mit und ohne Vordergrundobjekt abgebildet. Das Differenzbild der beiden Tiefenbilder entspricht dabei dem Effekt des *Scatterings* und kann als Referenz für den Einfluss des Objekts betrachtet werden.

Es kann beobachtet werden, dass die Stärke des Einflusses von der Distanz eines Pixels zu dem Objekt auf der Bildebene abhängt. Aus den Daten von mehreren Differenzbildern wird daher ein Zusammenhang zwischen der Größe eines Objekts, der Distanz eines Pixels zu dem Objekt und dem Einfluss des Phasenvektors auf den jeweiligen Pixel extrahiert. Der Einfluss des Phasenvektors wird als Verhältnis zwischen dem Phasenvektor des Objekts und dem Einfluss-Phasenvektor, der auf den wahren Vektor des jeweiligen Pixels addiert wird, angegeben. Der Sachverhalt ist in Abbildung 4.7 illustriert.

Anhand des aus den Referenzdaten bestimmten Zusammenhangs kann für ein beliebiges Objekt der Einfluss auf einen beliebigen Pixel eines Tiefenbildes geschätzt werden. Entscheidend ist dabei die geschätzte Veränderung der Distanz. Hierzu wird der geschätzte Distanzwert eines Pixels durch Subtraktion des Einfluss-Phasenvektors vom gemessenen Phasenvektor bestimmt. Die Orientierung des resultierenden Vektors entspricht dabei der geschätzten Distanz. Aus der Differenz zwischen gemessener

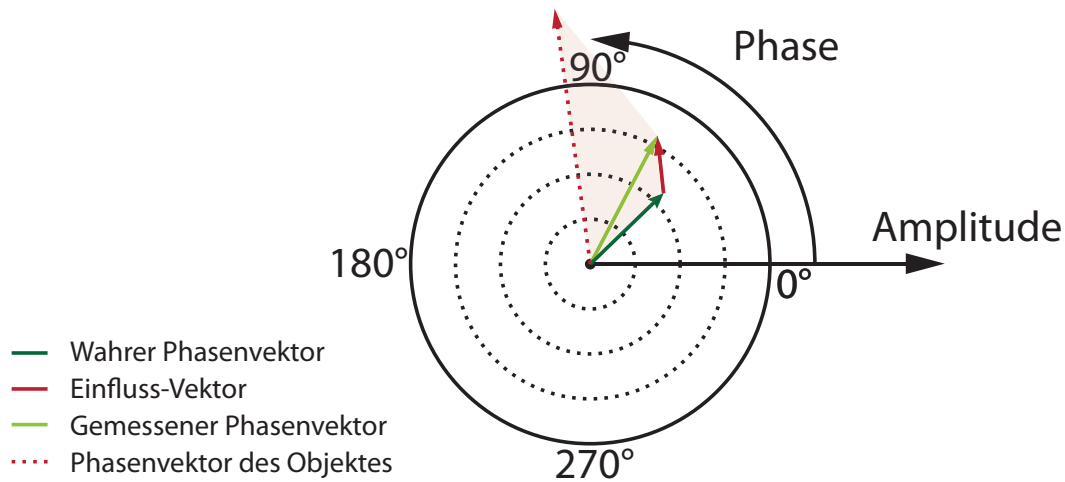


Abbildung 4.7: Die Länge des Einfluss-Phasenvektors hängt von der Distanz eines Pixels zum Objekt ab. Der gemessene Phasenvektor setzt sich zusammen aus dem korrekten Phasenvektor eines Pixels und dem Einfluss-Phasenvektor.

Distanz d_{meas} und geschätzter Distanz d_{est} wird ein Konfidenzwert c_{scat} nach

$$c_{scat} = 1 - \frac{2}{\pi} \arctan(|d_{est} - d_{meas}|) \quad (4.2)$$

ermittelt.

In Abbildung 4.8 (a) ist das nach der vorgestellten Methode geschätzte Konfidenzbild dargestellt. Gegenübergestellt wird das Bild einer *Ground-Truth*-Konfidenz (b). Diese wurde analog zu Abbildung 4.6 aus dem Vergleich zwischen zwei Aufnahmen einer Szene mit und ohne Vordergrundobjekt erstellt. Um die *Ground-Truth*-Konfidenz zu bestimmen, wird Gleichung 4.2 herangezogen. Es kann beobachtet werden, dass die geschätzten Konfidenzwerte über weite Bereiche übereinstimmen. Lediglich für die Pixel, die unmittelbar an das Vordergrundobjekt angrenzen, wird eine deutlich geringere Abweichung geschätzt. Dies ist auf den *Flying-Pixel*-Effekt zurückzuführen, der zusätzlich zum *Scattering* zu Fehlern in der Tiefenmessung führt. Die Berücksichtigung des *Flying-Pixel*-Effekts wird im nachfolgenden Kapitel thematisiert.

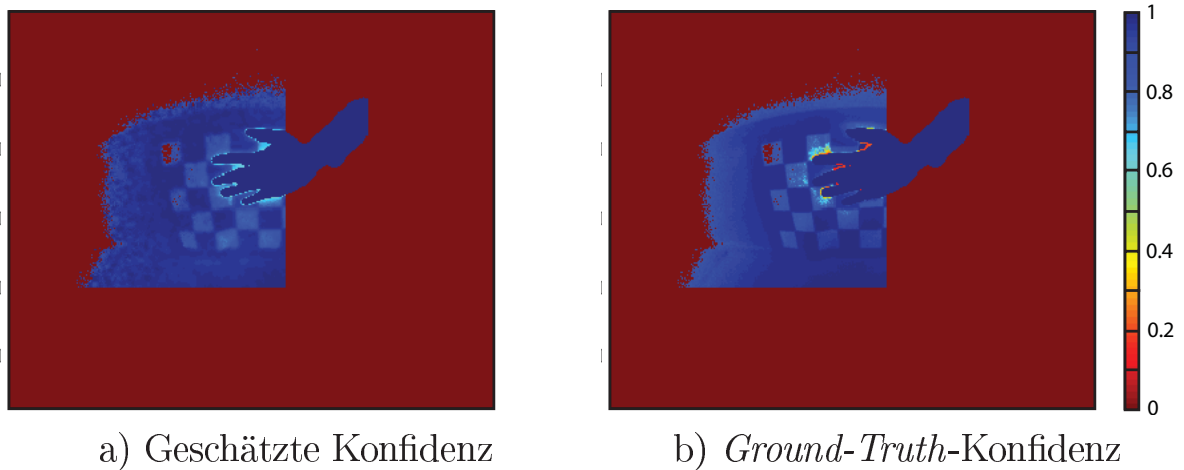


Abbildung 4.8: Vergleich zwischen dem geschätzten Konfidenzbild nach der vorgestellten Methode (a) und einer *Ground-Truth*-Konfidenz (b). Die *Ground-Truth*-Konfidenz wird berechnet aus dem Vergleich zweier Tiefenbilder, die einmal mit und einmal ohne Vordergrundobjekt aufgenommen werden.

4.3.3 *Flying-Pixel*-Effekt

Flying Pixel sind wie das *Scattering* ein Mischphaseneffekt und werden in Kapitel 2.1.3 erläutert. Sie entstehen bei der Überlagerung von verschiedenen Signalen auf einen einzigen Pixel. Dabei stammt ein Signal von der Reflexion an einem Vordergrund- und ein anderes Signal von der Reflexion an einem Hintergrundobjekt. Die resultierende gemessene Distanz entspricht dann einem Punkt zwischen den beiden Objekten.

Vor der Verarbeitung der Tiefendaten sollten *Flying Pixel* im Bild erkannt werden und dann als ungültig markiert werden. Da *Flying Pixel* in der Regel sehr große Abweichungen besitzen, wird der Konfidenzwert eines gefundenen *Flying Pixels* direkt auf einen Wert von 0 gesetzt.

Im vorgestellten Verfahren wird die Methode nach [56] verwendet. Hier wird ein *Canny*-Kantendetektor mit einem adaptiven Schwellwert auf dem Tiefenbild angewendet. Da die Messgenauigkeit der Tiefenmessung von der gemessenen Amplitude abhängt, wird der Schwellwert abhängig vom jeweiligen Amplitudenwert gesetzt. In Kapitel 3.4.1 wurde festgestellt, dass

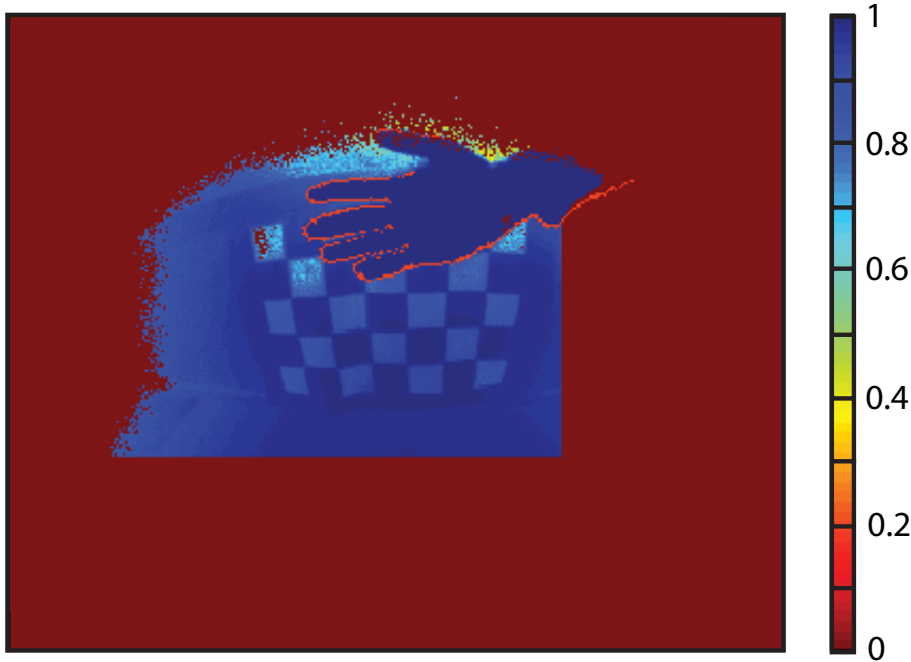


Abbildung 4.9: Detektierte *Flying Pixel* sind im dargestellten Konfidenzbild orange markiert

nicht nur die Amplitude, sondern auch die Intensität einen Einfluss auf die Messgenauigkeit hat. Aus diesem Grund wird das Verfahren nach [56] erweitert, indem der Fremdlichteinfluss berücksichtigt wird. Hierzu fließt bei der Bildung des adaptiven Schwellwerts nicht mehr ausschließlich die Amplitude ein. Stattdessen wird der Wert für die geschätzte Standardabweichung nach der in Abschnitt 4.3.1 vorgestellten Methode in die Bestimmung des adaptiven Schwellwerts mit einbezogen. Auf diese Weise wird auch die Anfälligkeit des Sensors gegenüber Fremdlicht bei der Detektion von *Flying Pixel* berücksichtigt.

In Abbildung 4.9 ist ein Konfidenzbild dargestellt, in dem die detektierten *Flying Pixel* orange markiert sind.

4.3.4 Kombination

Die zuvor vorgestellten Verfahren bestimmen jeweils einen Konfidenzwert abhängig vom stochastischen Fehler, vom *Scattering* und vom *Flying-Pixel*-Effekt. Die Werte werden zu einem einzelnen Konfidenzwert zu-



Abbildung 4.10: Kombiniertes Konfidenzbild. Die Helligkeit gibt die Höhe des Konfidenzwertes an. *Flying Pixel* sind in rot dargestellt.

sammengefasst. Bei der Kombination der Werte ist zu beachten, dass die Effekte eines einzelnen Fehlers nicht durch andere Effekte wieder ausgeglichen werden können. Wenn beispielsweise ein Tiefenwert durch *Scattering* einen hohen Fehler aufweist, dann kann ein geringer stochastischer Fehler diesen Effekt nicht ausgleichen.

Aus diesem Grund können die einzelnen Konfidenzwerte nicht durch die Bildung eines Durchschnittswertes zusammengefasst werden. Stattdessen soll der Fehler mit der größten Auswirkung den kombinierten Konfidenzwert bestimmen. Daher resultiert der kombinierte Konfidenzwert aus dem Minimum der einzelnen Konfidenzwerte.

In Abbildung 4.10 ist ein kombiniertes Konfidenzbild dargestellt. *Flying Pixel* sind in rot dargestellt.

5 3D-Umfeldmodellierung

5.1 Datenstruktur

Zur Umfeldmodellierung wird eine gitterbasierte Repräsentation der Umgebung verwendet. Hierbei handelt es sich um eine 2.5D-Höhenkarte, die mit einer probabilistischen Belegtheitskarte (*Occupancy Grid Map*) kombiniert wird. Der Stand der Technik bezüglich 2.5D-*Grids* und *Occupancy Grids* wird in Kapitel 2.2.3 vorgestellt.

Das umgesetzte 2.5D-*Grid* ist ein reguläres Gitter. In jeder Gitterzelle g_i wird ein Höhenwert z_i gespeichert. Zusätzlich wird ein Konfidenzwert c_i in jeder Zelle gespeichert. Dieser gibt an, inwiefern der berechneten Höhe vertraut werden kann.

Zur Implementierung der Belegtheitskarte wird in jeder Zelle zusätzlich zum Höhen- und Konfidenzwert ein Wahrscheinlichkeitswert angegeben. Dieser gibt an, mit welcher Wahrscheinlichkeit die Zelle belegt ist. \mathbf{m}_i bezeichnet dabei die Zufallsvariable, welche die beiden Zustände *belegt* oder *nicht belegt* annehmen kann. Die Wahrscheinlichkeit, dass eine Zelle belegt ist, hängt dabei von der Menge der Messungen $\mathbf{d}_{1:k}$ ab und wird durch

$$p(\mathbf{m}_i | \mathbf{d}_{i:k}) \quad (5.1)$$

beschrieben. Nach [63] kann der Wert als Logarithmus des Quotienten aus der Wahrscheinlichkeit und ihrer Gegenwahrscheinlichkeit, dem sogenannten *LogOdd* ausgedrückt werden. Hierdurch ergibt sich der Ausdruck

$$l_{k,i} = \ln \left(\frac{p(\mathbf{m}_i | \mathbf{d}_{i:k})}{p(\bar{\mathbf{m}}_i | \mathbf{d}_{i:k})} \right). \quad (5.2)$$

Die Repräsentation als *LogOdds* birgt Vorteile bei der inkrementellen Aktualisierung der Werte bei neuen Messungen. Multiplikationen können so durch Additionen ersetzt werden.

Alle Werte werden relativ zur Initialposition und Ausrichtung des Fahrzeugs gespeichert. Das Gitter bewegt sich nicht mit dem Fahrzeug, sondern bleibt statisch. Die Zellen werden zu Beginn mit der Höhe $z_i = 0$, der Konfidenz $c_i = 0$ und einer Belegtheitswahrscheinlichkeit von 0,5 initialisiert, was einen *LogOdd*-Wert von $l_i = 0$ zur Folge hat.

5.2 Integration von PMD-Messdaten

Die Punktwolken der im Fahrzeug integrierten Kameras werden fortlaufend in die Datenstruktur des Umfeldmodells integriert. Zum einen werden die Höhenwerte des 2.5D-*Grids* angepasst und zum anderen die Belegtheitswerte des *Occupancy Grids* aktualisiert, und zwar unabhängig voneinander. Betrachtet werden nur die Punkte der Kameras, die nach der in Kapitel 4.3 vorgestellten Methode einen ausreichend hohen Konfidenzwert besitzen. Die 3D-Punkte werden, wie in Kapitel 3.6.1 beschrieben, in das Weltkoordinatensystem, das identisch ist mit dem Koordinatensystem des Umfeldmodells, transformiert. Anschließend werden die einzelnen Punkte den einzelnen Gitterzellen zugeordnet.

Jeder Gitterzelle g_i ist nun eine Menge von Punkten zugewiesen. Da nach der Zuordnung nur noch der Höhenwert und Konfidenzwert des Punktes relevant ist, kann die Menge der Punkte $P_i \subseteq \mathbb{R}^2$ auf die z -Koordinaten und Konfidenzwerte der Punkte reduziert werden. Jedes Punktelement $(z_j, c_j) \in P_i$ enthält einen Höhenwert z_j und einen Konfidenzwert c_j . Der maximale Höhenwert wird mit z_{max} bezeichnet und der durchschnittliche Höhenwert und Konfidenzwert mit z_{avg} und c_{avg} . Der Höhenwert z_i einer Gitterzelle g_i wird aktualisiert, wenn $c_{avg} > 0,9 \cdot c_i$.

Zur Aktualisierung der Zelle wird eines von drei unterschiedlichen Verfahren angewandt. Die *Average*-Variante bestimmt zunächst den Mittelwert z_{avg} der Höhenwerte aller in eine Zelle fallenden Punkte durch

$$z_{avg} = \frac{1}{n_{P_i}} \sum_{(z_j, c_j) \in P_i} z_j. \quad (5.3)$$

Dabei entspricht n_{P_i} der Anzahl der in P_i enthaltenen Elemente. Die *Maximum*-Variante aktualisiert den Wert der Zelle mit dem größten Höhen-

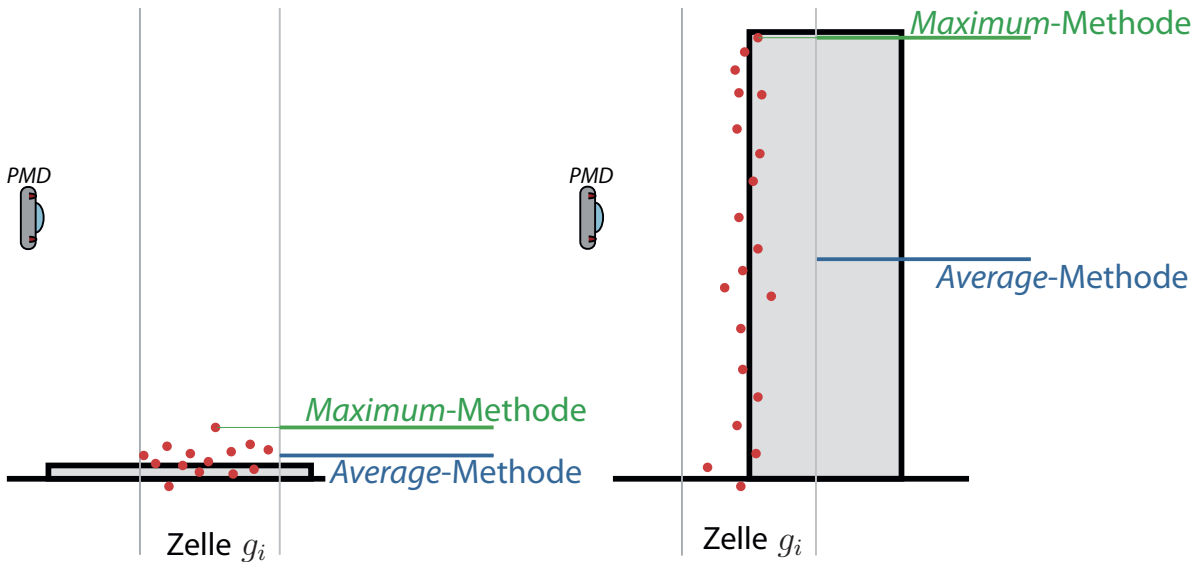


Abbildung 5.1: Die unterschiedlichen Varianten zur Bestimmung des Höhenwertes für die Aktualisierung einer Zelle g_i

wert z_{max} aller Elemente aus P_i . Eine Kombination der beiden Ansätze bildet die *Mixed*-Variante. Hier wird zunächst überprüft, ob die Höhe des höchsten Punktes z_{max} einen Schwellwert z_{thresh} übersteigt. Wenn dies der Fall ist, dann wird die *Maximum*-Variante gewählt und anderenfalls die *Average*-Variante. Dies kann ausgedrückt werden durch

$$z_{mixed} = \begin{cases} z_{max}, & \text{falls } z_{max} > z_{thresh} \\ z_{avg}, & \text{sonst} \end{cases} . \quad (5.4)$$

Der bestimmte Höhenwert (z_{avg} , z_{max} oder z_{mixed}) wird anschließend zur Aktualisierung der Zelle g_i eingesetzt. In Abbildung 5.1 sind die verschiedenen Werte schematisch dargestellt, wobei auf der linken Seite zu beobachten ist, dass sich der Mittelwert der in die Zelle fallenden Punkte als Höhenwert zur Aktualisierung der Zelle eignet. Die abgetasteten Oberflächenpunkte befinden sich alle auf derselben Höhe, und die Unterschiede zwischen den gemessenen Werten basieren auf dem Messrauschen des Sensors. Durch die Bestimmung von z_{avg} kann das Messrauschen ausgeglichen werden. Der Wert für z_{max} würde fälschlicherweise die Höhe eines Punktes mit einem hohen Messrauschen übernehmen und somit im Allgemeinen zu



hohe Höhenwerte bestimmen. In Abbildung 5.1 auf der rechten Seite wird ein senkrechtes Objekt vermessen. Hier besitzen die abgetasteten Punkte unterschiedliche Höhenwerte, und die Höhenwerte der in die Zelle fallenden Punkte strecken sich korrekterweise vom Boden bis zur Höhe des Objektes. Der Wert für z_{avg} bestimmt hier einen Wert, der etwa der halben Höhe des Objektes entspricht. In diesem Fall ist z_{max} die bessere Wahl. Der Höhenwert der höchsten Messung entspricht der Höhe des Objekts. Die *Mixed*-Variante würde mit z_{mixed} in beiden Fällen den passenderen Höhenwert bestimmen. Im Nachfolgenden wird daher die *Mixed*-Variante zur Ermittlung des Höhenwertes angenommen und daher z_{mixed} stellvertretend für alle Varianten in den Formeln genutzt.

Der Höhenwert z_i^k der Zelle g_i wird nun zum Zeitpunkt k der aktuellen Messung durch Berücksichtigung von z_{mixed} und dem vorherigen Höhenwert z_i^{k-1} aktualisiert:

$$z_i^k = \begin{cases} \frac{n_i^{k-1}}{n_i^k} \cdot z_i^{k-1} & + \frac{n_{P_i}}{n_i^k} \cdot z_{mixed}, & \text{falls } n_i^{k-1} < n_{max} \\ \frac{n_{max} - n_{P_i}}{n_{max}} \cdot z_i^{k-1} & + \frac{n_{P_i}}{n_{max}} \cdot z_{mixed}, & \text{sonst} \end{cases} \quad (5.5)$$

Dabei bezeichnet n_i^{k-1} die Anzahl der bisher in der *Grid*-Zelle berücksichtigten Messungen. n_i^k wird durch Berücksichtigung der n_{P_i} neuen Messungen aktualisiert nach

$$n_i^k = n_i^{k-1} + n_{P_i}. \quad (5.6)$$

Das Gewicht von neuen Messungen hängt daher von der Gesamtanzahl der bisher berücksichtigten Messungen ab, die in eine Zelle fallen. Um zu verhindern, dass das Gewicht einer neuen Messung bei einer hohen Anzahl an vorherigen Messungen zu gering wird, wird der Schwellwert n_{max} eingeführt. Wenn die Gesamtanzahl der Messungen einer Zelle n_{max} überschreitet, dann hängt das Gewicht neuer Messungen nicht mehr von der Gesamtanzahl der Messungen n_i^{k-1} ab. Der Konfidenzwert einer Zelle c_i^k wird analog zum Höhenwert aus dem vorherigen Konfidenzwert c_i^{k-1} und der durchschnittlichen Konfidenz der Messungen c_{avg} nach

$$c_i^k = \begin{cases} \frac{n_i^{k-1}}{n_i^k} \cdot c_i^{k-1} & + \frac{n_{P_i}}{n_i^k} \cdot c_{avg}, & \text{falls } n_i^{k-1} < n_{max} \\ \frac{n_{max} - n_{P_i}}{n_{max}} \cdot c_i^{k-1} & + \frac{n_{P_i}}{n_{max}} \cdot c_{avg}, & \text{sonst} \end{cases} \quad (5.7)$$

aktualisiert.

Die Belegtheitswerte werden parallel zu den Höhenwerten aktualisiert. Es wird auch auf die Einteilung der Messwerte in die einzelnen Zellen zurückgegriffen. In Abschnitt 5.1 wurde die Speicherung der Belegtheitswahrscheinlichkeiten als *LogOdds* erläutert. Nach [63] kann die Integration einer neuen Messung auf Basis der *LogOdd*-Repräsentation durch

$$l_{k,i} = \ln \left(\frac{p(\mathbf{m}_i | \mathbf{d}_k)}{p(\bar{\mathbf{m}}_i | \mathbf{d}_k)} \right) + l_{k-1,i} + l_{0,i} \quad (5.8)$$

mit

$$l_{k-1,i} = \ln \left(\frac{p(\mathbf{m}_i | \mathbf{d}_{1:k-1})}{p(\bar{\mathbf{m}}_i | \mathbf{d}_{1:k-1})} \right) \quad (5.9)$$

und

$$l_{0,i} = \ln \left(\frac{p(\bar{\mathbf{m}}_i)}{p(\mathbf{m}_i)} \right) \quad (5.10)$$

erfolgen. Der Ausdruck $p(\mathbf{m}_i | \mathbf{d}_k)$ gibt die Wahrscheinlichkeit an, dass die Zelle g_i belegt ist, gegeben die aktuelle Messung \mathbf{d}_k , und wird durch das sogenannte inverse Sensormodell I modelliert [63]. Zur Aktualisierung der einzelnen Zellen, wird demzufolge eine Modellierung des inversen Sensormodells benötigt.

Im Falle eines *Occupancy Grids* wird nicht nur der Wert der Zelle aktualisiert, in welche eine Messung fällt, sondern auch die Zellen, die im Sichtstrahl des Sensors zwischen Sensor und Messpunkt liegen. Da der Sensor in diesen Zellen nicht gemessen hat, sollte sich die Wahrscheinlichkeit erhöhen, dass diese Zellen frei sind. Die Bestimmung der Zellen zwischen Messpunkt und Sensor erfolgt durch den recheneffizienten *Bresenham*-Algorithmus [139].

Das inverse Sensormodell I berücksichtigt bei der Aktualisierung der Belegtheitswerte die Konfidenz c_i , den Höhenwert z_i und die Distanz d_c eines Punktes zu einer zu aktualisierenden Zelle. Dabei wird d_c nach

$$d_c = d_m - d_i \quad (5.11)$$

aus der Differenz zwischen der Distanz d_i des Punktes zur Kamera und der Distanz d_m der Zelle zur Kamera berechnet und kann demzufolge

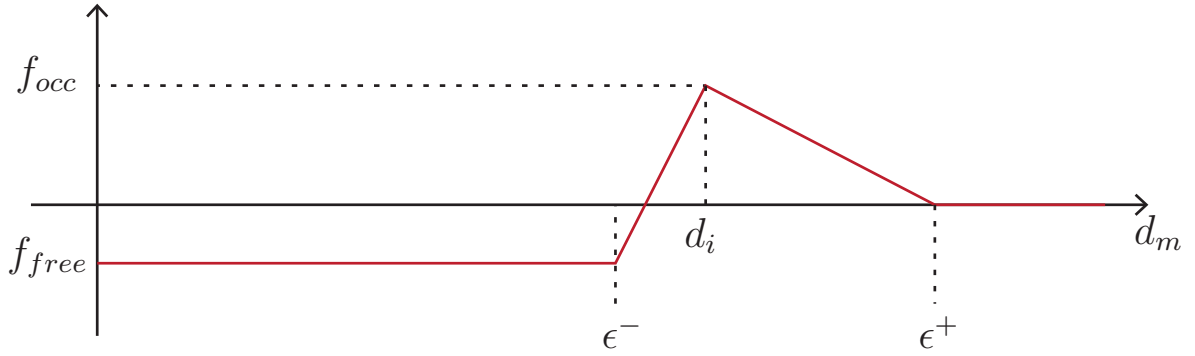


Abbildung 5.2: Das inverse Sensormodell

auch negative Werte annehmen, wenn sich die Zelle zwischen Punkt und Kamera befindet. Das inverse Sensormodell ist durch

$$I(c_i, z_i, d_c) = \begin{cases} f_{free}(c_i, z_i) & , \text{wenn } d_c < \epsilon^- \\ \frac{f_{occ}(c_i, z_i) - f_{free}(c_i, z_i)}{\epsilon^-} d_c + f_{occ}(c_i, z_i) & , \text{wenn } \epsilon^- < d_c \leq 0 \\ f_{occ}(c_i, z_i) & , \text{wenn } d_c = 0 \\ \frac{0,5 - f_{occ}(c_i, z_i)}{\epsilon^+} d_c + f_{occ}(c_i, z_i) & , \text{wenn } 0 < d_c \leq \epsilon^+ \\ 0,5 & , \text{sonst} \end{cases} \quad (5.12)$$

gegeben. ϵ^- und ϵ^+ definieren die Breite des Einflusses einer Messung auf die benachbarten Zellen. f_{free} und f_{occ} beschreiben durch die Parameter w_{free} und w_{occ} die Stärke der Senkung oder Erhöhung der Belegheitswahrscheinlichkeiten und sind durch

$$f_{free}(c_i, z_i) = w_{free} \frac{c_i + z_i}{2} \quad (5.13)$$

und

$$f_{occ}(c_i, z_i) = w_{occ} \frac{c_i + z_i}{2} \quad (5.14)$$

gegeben.

Das inverse Sensormodell zeigt Abbildung 5.2. Die Belegheitswahrscheinlichkeiten der Zellen, die sich zwischen Kamera und Messpunkt befinden, werden abhängig von der Konfidenz und dem Höhenwert des Messpunktes gesenkt, während die Belegheitswahrscheinlichkeit der Zelle, in der sich der Messpunkt befindet, erhöht wird. Über die Zellen hinter dem Mess-

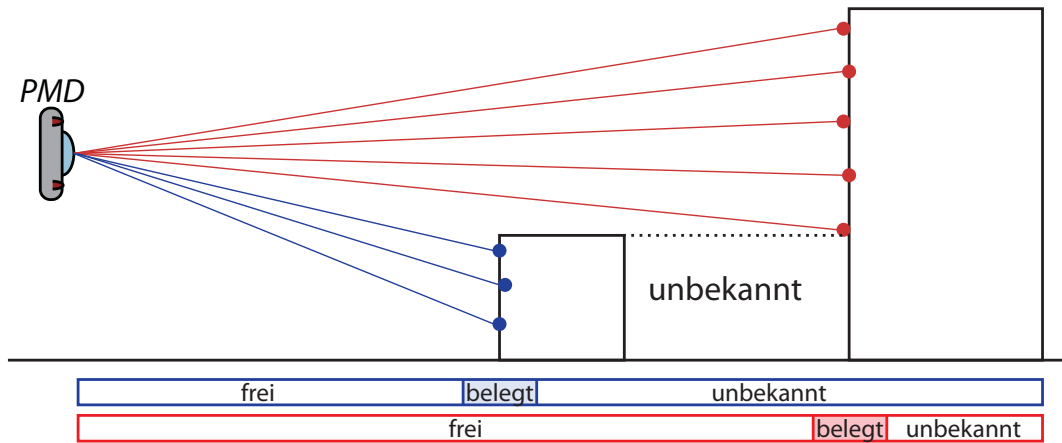


Abbildung 5.3: Ein niedriges Vordergrundobjekt kann durch die Messung eines Hintergrundobjekts fälschlicherweise als frei deklariert werden.

punkt kann keine Aussage hinsichtlich ihrer Belegtheit getroffen werden. In unmittelbarer Nachbarschaft der Zelle des Messpunktes existiert durch den von ϵ^- und ϵ^+ eingegrenzten Bereich ein fließender Übergang. Durch diesen wird die Unsicherheit der gemessenen Distanz in das Modell mit einbezogen.

5.3 3D-Occupancy-Grid

Im Falle der *2D-Occupancy-Grid-Map* werden die dreidimensionalen Kameradaten auf ein zweidimensionales *Grid* projiziert. Bei der Integration von Messpunkten werden die Zellen zwischen Sensor und Messpunkt als frei angenommen. Dies muss, wie in Abbildung 5.3 dargestellt, bei hohen Objekten nicht immer der Fall sein. Ein kleines Vordergrundobjekt kann durch die Messungen eines Hintergrundobjektes fälschlicherweise als frei deklariert werden. Eine Lösung zur Vermeidung dieses Effektes bietet die Verwendung eines *3D-Occupancy Grids*.

Durch *3D-Occupancy-Grids* kann der Vorteil der dreidimensionalen Messfähigkeit der Time-of-Flight-Kameras genutzt werden. Ein *3D-Occupancy-Grid* besteht aus einem dreidimensionalen regulären Gitter. Aus den zweidimensionalen Zellen des *2D-Grids* entstehen somit dreidimensionale *Voxel*. Daher kann ein *3D-Occupancy-Grid* auch, wie in Abbildung 5.4 darge-

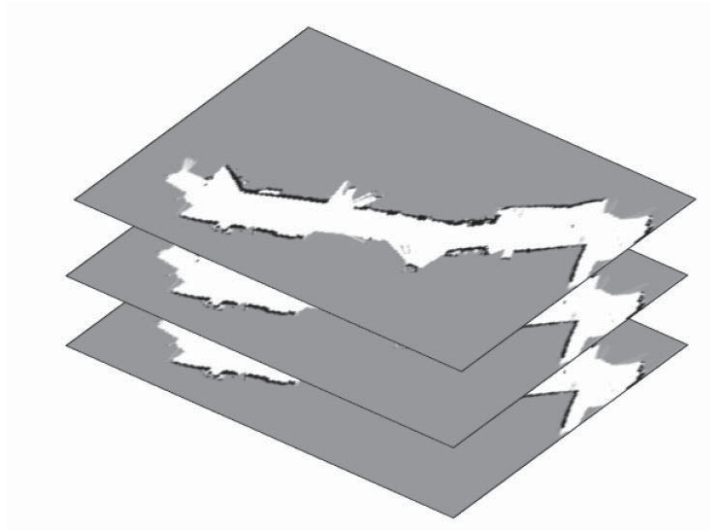


Abbildung 5.4: Ein 3D-*Occupancy-Grid* kann als mehrschichtiges 2D-*Occupancy-Grid* betrachtet werden.

stellt, als ein mehrschichtiges 2D-*Occupancy-Grid* betrachtet werden. Die Auflösung in der Höhe entspricht dabei der Anzahl an Schichten.

Die Integration von neuen Messungen in ein dreidimensionales *Grid* muss gegenüber dem 2D-*Grid* angepasst werden. Der im 2D-*Grid* eingesetzte *Bresenham*-Algorithmus kann auf die dreidimensionale Struktur erweitert werden. Die Verwendung des zweidimensionalen Algorithmus führt allerdings nur zu vernachlässigbaren Fehlern und ist auf der anderen Seite aufgrund der geringeren Komplexität deutlich recheneffizienter. Daher kann auch beim 3D-*Grid* auf den zweidimensionalen *Bresenham*-Algorithmus zurückgegriffen werden.

Ein weiterer Unterschied bei der Integration von neuen Messungen besteht im inversen Sensormodell. Um Messpunkte in die korrekte Schicht des 3D-*Grids* einzutragen, wird der Höhenwert des Messpunktes betrachtet. Der Höhenwert fließt nicht wie im Falle des 2D-*Grids* in die Aktualisierung des Belegtheitswertes ein, sondern legt ausschließlich die Höhe des entsprechenden *Voxels* fest. Auf diese Weise fällt bei der Verwendung des 3D-*Occupancy-Grids* kaum zusätzlicher Rechenaufwand an. Lediglich der Speicherbedarf vervielfacht sich um die Anzahl der Schichten.

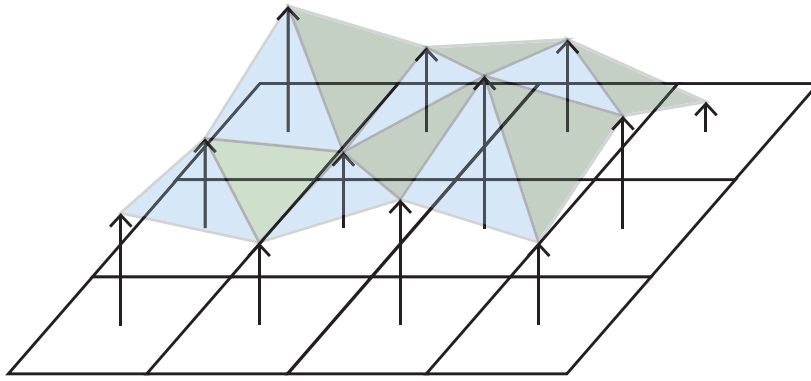


Abbildung 5.5: Triangulierung der Höhenwerte des 2.5D-Grids

5.4 Visualisierung

Zur Visualisierung des Umfeldmodells wird ein 3D-Mesh aus den Daten der 2.5D-Karte erstellt. Aus den Höhenwerten benachbarter Gitterzellen werden Dreiecke generiert. Jedes Dreieck wird aus drei Höhenwerten bestimmt. Dies ist in Abbildung 5.5 dargestellt.

Ein Eckpunkt eines Dreiecks wird als Vertex bezeichnet. Die laterale Position eines Vertex ist aufgrund der statischen Gittergeometrie bekannt. Die vertikale Position ergibt sich aus dem Höhenwert der jeweiligen Gitterzelle.

Zur Visualisierung der geometrischen Struktur des Modells werden die Dreiecke schattiert. Dies geschieht auf Basis der Normalen der Dreiecke. Je größer die Abweichung der Normalen zur Lichteinfallrichtung, umso dunkler wird das Dreieck gezeichnet. Die Schattierung eines beispielhaften Umfeldmodells ist in Abbildung 5.6 dargestellt. Das Fahrzeug wird als 3D-Modell entsprechend der Position im Umgebungsmodell eingefügt.

Eine weitere Möglichkeit zur Visualisierung ist die Texturierung des 3D-Modells durch Kamerabilder. Wenn die intrinsischen und extrinsischen Kameraparameter bekannt sind, kann jeder Bildpixel einer Kamera auf das 3D-Modell projiziert werden. Die eingesetzten PMD-Kameras liefern ein Amplitudenbild. Auch dieses kann in die Szene projiziert werden. Allerdings werden in den Amplitudenbildern nur die beleuchteten Bereiche hell dargestellt. Eine Alternative ist die Nutzung von Intensitätsbildern, die eine Ähnlichkeit zu Graustufenbildern besitzen. Zusätzliche Kameras können bei bekannten extrinsischen und intrinsischen Parametern glei-

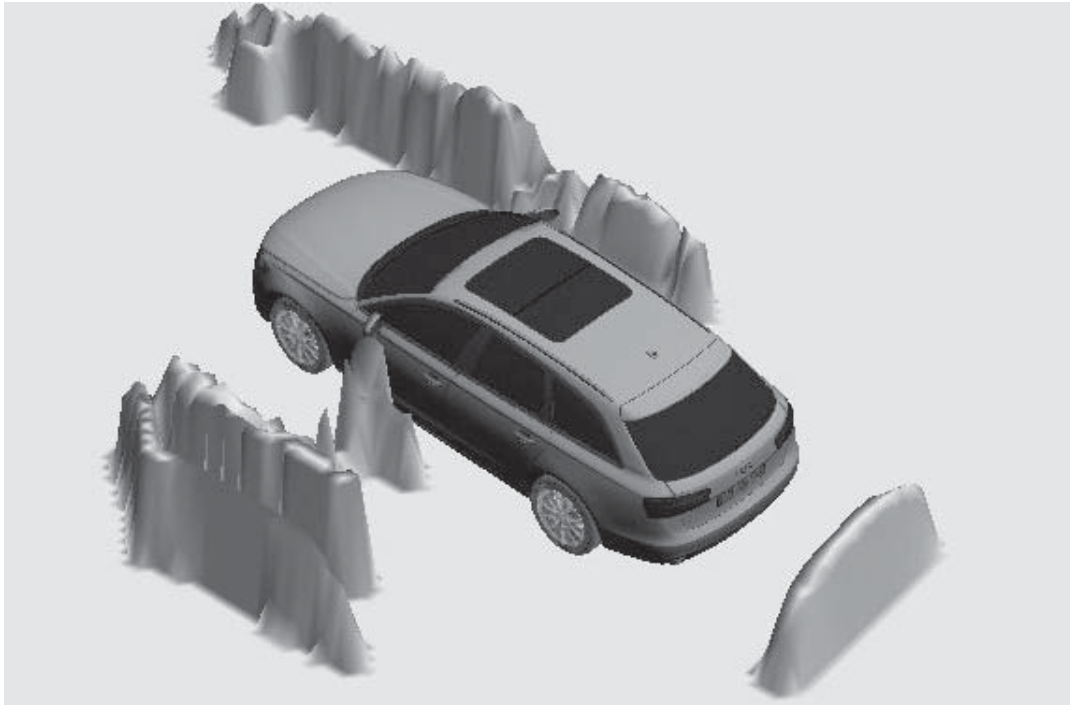


Abbildung 5.6: Schattierung des 3D-Modells abhängig von den Normalen der Dreiecke

chermaßen auf das Modell projiziert werden. Beispielsweise kann die Texturierung auch durch hochauflösende Farbkameras erfolgen.

5.5 Informationsextraktion

Aus dem *2.5D-Grid* werden Hindernisse extrahiert. Dabei werden Zellen, die einen Höhenwert besitzen, der eine gewisse Höhe überschreitet, als Hindernis betrachtet. Zusammenhängende Bereiche von Gitterzellen, die als Hindernis betrachtet werden, können zu Hindernisobjekten zusammengefügt werden. Eine relevante Information ist die Distanz eines Hindernisobjektes zum Fahrzeug. Zur Bestimmung der Distanz wird das Fahrzeug als Rechteck modelliert. Die Distanz einer Gitterzelle kann dann im Zweidimensionalen geometrisch bestimmt werden. In Abbildung 5.7 sind die Hindernisse abhängig von der Distanz zum Fahrzeug rot eingefärbt.



Abbildung 5.7: Einfärbung der nahen Hindernisse abhängig von der Distanz zum Fahrzeug



6 3D-Bildverarbeitung

6.1 Flächenextraktion

In diesem Kapitel wird eine Methode beschrieben, welche die Parameter von Flächen aus den Kameradaten einer PMD-Kamera extrahiert. Die Methode wurde schon in [140] präsentiert. Extrahierte Flächen können wertvolle Informationen für Fahrerassistenzsysteme bilden oder auch das generierte dreidimensionale Umgebungsmodell verbessern.

6.1.1 Kombination von lokaler und globaler Flächenextraktion

Im Kapitel 2.2.4 wird ausführlich der Stand der Technik bezüglich der Extraktion von Flächen aus Tiefenbildern und 3D-Punktwolken beschrieben. Es existieren sowohl globale Methoden, die ein oder mehrere Flächenmodelle in die gesamte Punktwolke einpassen, als auch lokale Methoden, die aus den Nachbarschaftsbeziehungen der Pixel zusammenhängende Bereiche extrahieren. Der nachfolgend vorgestellte Ansatz ist eine Kombination aus einer lokalen und einer globalen Herangehensweise.

Zunächst werden nach einer Vorverarbeitung Liniensegmente gebildet, die anschließend zu Flächensegmenten kombiniert werden. Es folgt eine globale Flächenextraktion ausgehend von den zuvor extrahierten Flächensegmenten. Im letzten Schritt werden zusammengehörige Flächensegmente fusioniert.

Die Ergebnisse der verschiedenen Verarbeitungsschritte werden anhand einer Beispielszene dargestellt. Hierbei handelt es sich um eine Modellumgebung, welche in Abbildung 6.1 abgebildet ist.

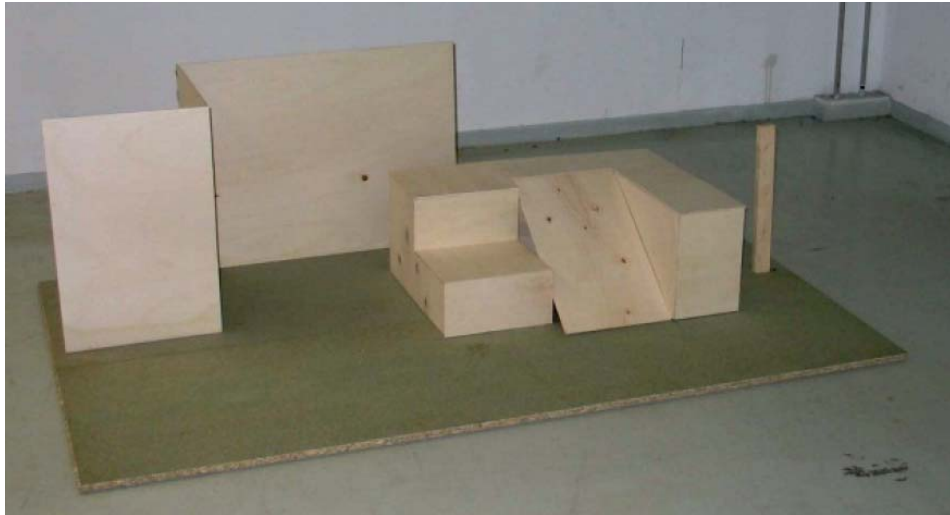


Abbildung 6.1: Die verschiedenen Verarbeitungsschritte werden anhand der abgebildeten Modellumgebung dargestellt und bewertet.

6.1.2 Vorverarbeitung

Die Vorverarbeitung hat das Ziel, die Tiefendaten für die nachfolgenden Verarbeitungsschritte aufzubereiten und führt somit zu einer robusteren Flächenextraktion. Zunächst werden ungültige Messwerte entfernt. Dies geschieht auf Basis der in Kapitel 4.3 vorgestellten Konfidenzwerte. Tiefenwerte mit einem niedrigen Konfidenzwert werden nicht weiter verarbeitet.

Anschließend werden die Tiefendaten durch die in Kapitel 4.2 genannten Filter geglättet. Es werden konsekutiv zunächst ein Median-Filter und anschließend der bilaterale Filter angewandt.

Die Berechnung der 3D-Punktwolke ist auch ein Teil der Vorverarbeitung und wurde bereits in Kapitel 3.6.1 vorgestellt.

6.1.3 Liniensegmentierung

Vor der Liniensegmentierung wird die Umgebung jedes Pixels im Tiefenbild auf Planarität überprüft. Die Menge der Pixel in der näheren Umgebung eines Pixels wird als *Patch* bezeichnet und beinhaltet den betreffenden Pixel und seine Nachbarpixel. Für jeden dieser *Patches* wird nun eine Hauptkomponentenanalyse [141] durchgeführt. Die relative Varianz kann

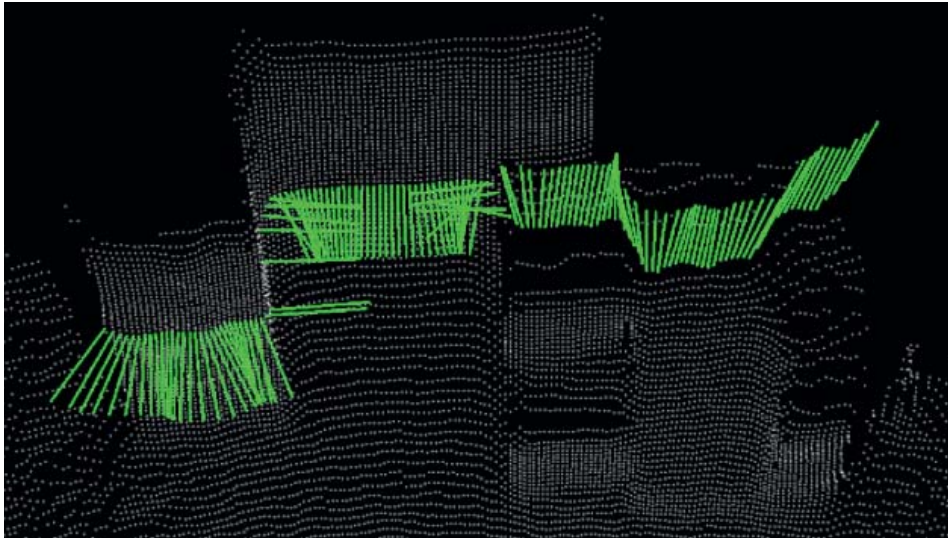


Abbildung 6.2: Extraktion von Normalen entlang einer horizontalen *Scanline* [140]

bestimmt werden, indem die ersten beiden Eigenwerte durch die Summe aller Eigenwerte dividiert werden. Wenn die relative Varianz einen spezifischen Schwellwert überschreitet, dann wird davon ausgegangen, dass keine Planarität gegeben ist, und der Pixel wird verworfen.

Die dritte Hauptkomponente, die den kleinsten Eigenwert besitzt, entspricht der Normalen des *Patches*. Die Normalen der *Patches* entlang einer horizontalen *Scanline* sind in Abbildung 6.2 dargestellt. Diese Normalen werden für die Bestimmung der Liniensegmente herangezogen. Um Normalen miteinander zu vergleichen, werden sie in sphärische Koordinaten mit einem Polar- und einem Azimutwinkel umgerechnet. Das Verfahren für die Segmentierung startet mit einem ersten Pixel. Nun werden nacheinander Pixel entlang einer *Scanline* hinzugefügt, wenn die zugehörigen Normalen nicht stark voneinander abweichen. Wenn die beiden Normalen nicht ähnlich sind, dann endet das Liniensegment und ein neues beginnt. Diese Segmentierung wird sowohl für alle horizontalen als auch für alle vertikalen *Scanlines* durchgeführt. Die vertikale Liniensegmentierung ist in Abbildung 6.3 abgebildet. Zur besseren Übersicht ist nur jede zehnte Linie farblich dargestellt. An dem rot eingekreisten Bereich kann eine Übersegmentierung beobachtet werden und an dem grün markierten Bereich schlägt die Segmentierung fehl.

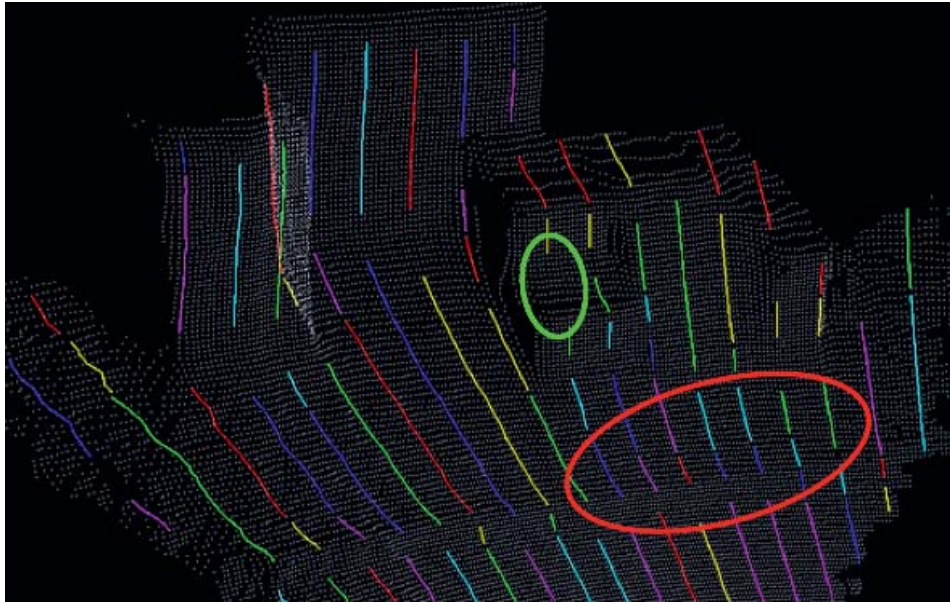


Abbildung 6.3: Die Liniensegmentierung wird durchgeführt durch einen Vergleich von Normalen entlang vertikaler und horizontaler *Scanlines* [140]

6.1.4 Flächensegmentierung

Für die Bestimmung von Flächensegmenten werden die zuvor extrahierten horizontalen und vertikalen Liniensegmente berücksichtigt. Zwei benachbarte Punkte werden demselben Flächensegment zugeordnet, wenn zwei Bedingungen erfüllt sind. Beide Punkte müssen demselben Liniensegment zugeordnet sein. Ist diese Bedingung erfüllt, werden die zum gemeinsamen Liniensegment orthogonalen Liniensegmente der beiden Punkte betrachtet. Wenn der Unterschied der Orientierungen der beiden orthogonalen Liniensegmente nicht einen Schwellwert überschreitet, dann werden beide Punkte demselben Flächensegment zugeordnet.

Um diese Überprüfungen für alle benachbarten Punkte effizient durchzuführen, wird der *Connected-Component-Labeling-Algorithmus* [142] verwendet. Das Ergebnis der Flächensegmentierung ist in Abbildung 6.4 dargestellt.

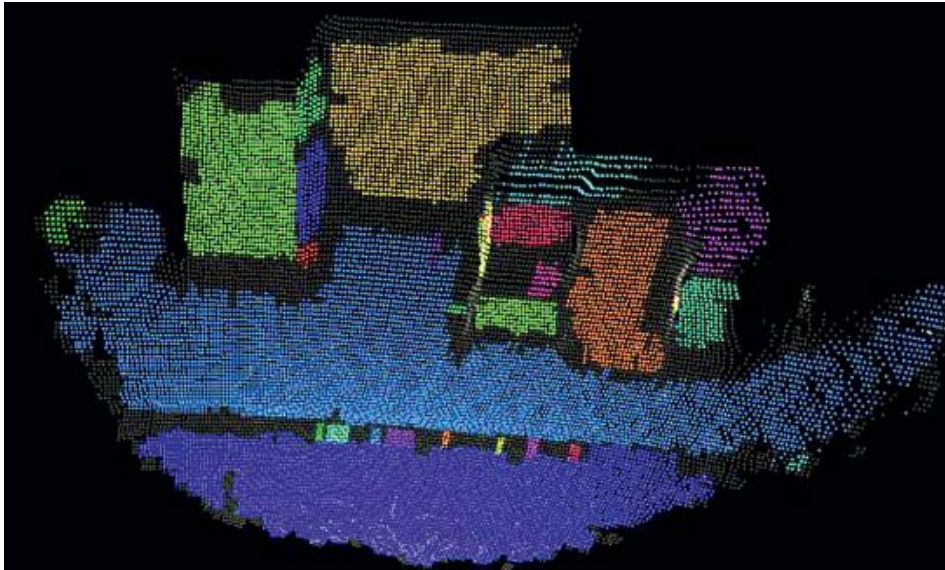


Abbildung 6.4: Ergebnis der Flächensegmentierung [140]

6.1.5 Globale Flächenextraktion

Um die Robustheit der Flächenextraktion zu erhöhen, folgt auf die Flächensegmentierung eine globale Flächenextraktion. Hierzu werden in einem ersten Schritt auf Basis der Hauptkomponentenanalyse [141] Regressionsebenen aus den 3D-Punkten der Flächensegmente berechnet. Treten viele Ausreißer auf, kann eine RANSAC-Methode [94] für eine robustere Berechnung der Regressionsebenen eingesetzt werden.

Nach der Berechnung einer Regressionsebene wird für die Punkte der gesamten Punktwolke überprüft, ob sie zu dieser Ebene hinzugefügt werden. Hierbei werden die folgenden drei Kriterien überprüft:

- Die orthogonale Distanz des Punktes zur Ebene ist kleiner als ein definierter Schwellwert
- Die Normale des zum 3D-Punkt korrespondierenden *Patches* ist ähnlich zur Normalen der Ebene
- Der 3D-Punkt liegt in einem zusammenhängenden Bereich mit den Punkten der Regressionsebene

Da dieses Verfahren wiederholt für die verschiedenen Regressionsebenen auf die gesamte Punktwolke angewendet wird, kann es vorkommen, dass

Punkte mehreren Regressionsebenen zugeordnet werden. Dieses Problem wird im nachfolgenden Schritt gelöst.

6.1.6 Fusion von Regressionsflächen

Die Fusion von Regressionsflächen hat zwei Ziele. Zum einen soll gesichert werden, dass Punkte nur zu einer einzigen Fläche zugeordnet sind, und zum anderen sollen Flächen, die durch eine Übersegmentierung in verschiedene Flächen aufgeteilt wurden, zusammengefügt werden.

Zunächst werden Paare von Flächen bestimmt, die einen überlappenden Bereich besitzen. Für jedes dieser Flächenpaare werden die folgenden Bedingungen überprüft:

$$\alpha_n = \arccos \left(\frac{n_{P_i} \cdot n_{P_n}}{\|n_{P_i}\| \|n_{P_n}\|} \right) \leq C_1 \quad (6.1)$$

$$d = \left| \frac{n_{P_i} + n_{P_n}}{2} \cdot \frac{z_{P_i} - z_{P_n}}{\|z_{P_i} - z_{P_n}\|} \right| \leq C_2 \quad (6.2)$$

Gleichung 6.1 überprüft, ob die Differenz der Orientierung α_n zwischen den Normalen n_{P_i} und n_{P_n} der Flächen P_i und P_n den Schwellwert C_1 überschreitet. Gleichung 6.2 bestimmt, ob der Vektor, der die beiden Schwerpunkte z_{P_i} und z_{P_n} der Flächen verbindet, orthogonal zu den Normalen n_{P_i} und n_{P_n} ist. Wenn einer der beiden Schwellwerte C_1 oder C_2 überschritten wird, wird angenommen, dass es sich um verschiedene Flächen handelt. Anderenfalls werden die Flächen fusioniert.

Nach der Fusionierung werden alle Punkte betrachtet, die weiterhin mehr als einer Fläche zugeordnet sind. Auf Basis der Ähnlichkeit zwischen der Normalen des zum Punkt zugehörigen *Patches* und der Normalen der Flächen wird der Punkt der passenderen Fläche zugeordnet. Zuletzt wird für alle Flächen geprüft, ob sie aus einer zusammenhängenden Fläche bestehen. Wenn dies für eine Fläche nicht der Fall ist, dann wird sie entsprechend der zusammenhängenden Bereiche aufgeteilt.

Abbildung 6.5 stellt die Flächenbereiche dar, die sich nach der globalen Flächenextraktion und dem Fusionierungsschritt ergeben. Es kann beobachtet werden, dass alle planaren Strukturen der Modellszene erkannt werden.

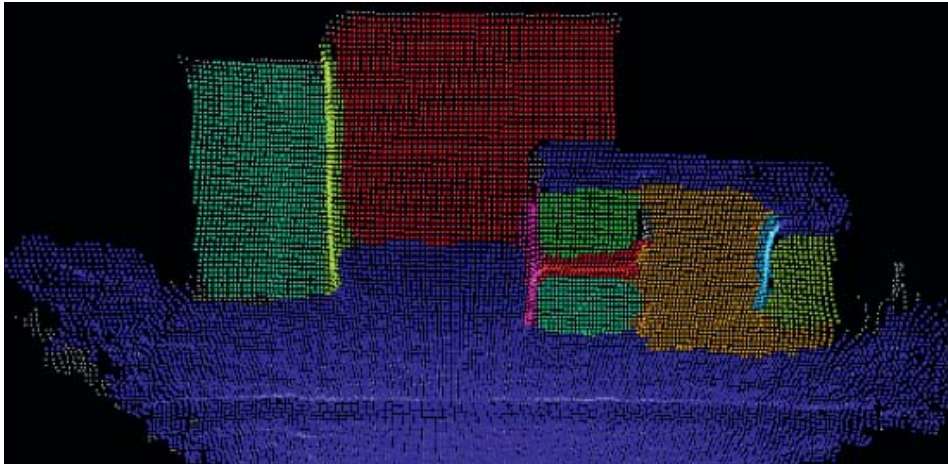
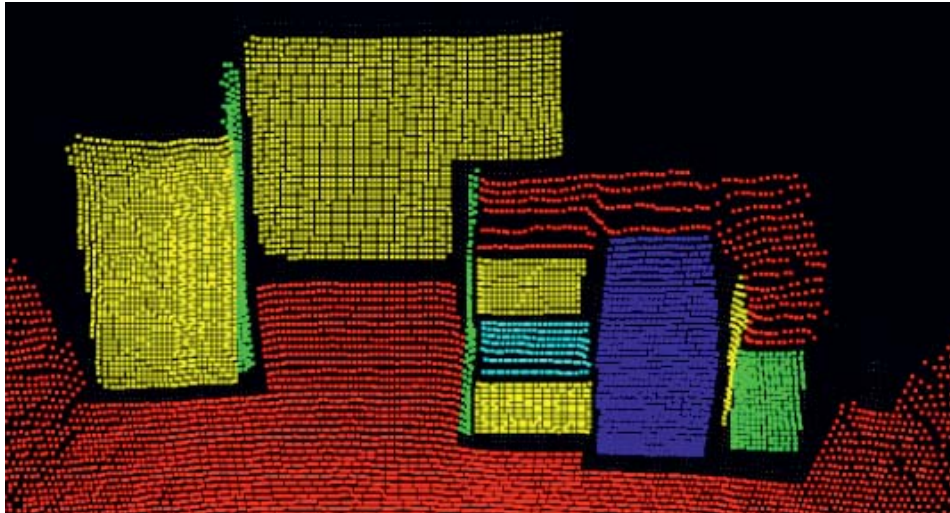


Abbildung 6.5: Das Ergebnis nach der globalen Flächenextraktion und dem Fusionierungsschritt [140]

6.1.7 Evaluierung

Die Evaluierung der Methode erfolgt auf Basis von *Ground-Truth*-Daten eines Referenzmodells. Mit dieser Referenz wird die Methode auch mit verschiedenen Methoden aus dem Stand der Technik (siehe Kapitel 2.2.4) verglichen. Eine Einfärbung der Punkte nach der Zugehörigkeit in den *Ground-Truth*-Daten ist in Abbildung 6.6 dargestellt.

Um die Leistungsfähigkeit des entwickelten Algorithmus zu bewerten, werden verschiedene Bewertungsmaße definiert. Den vom Algorithmus erkannten Anteil der Flächen aus den *Ground-Truth*-Daten gibt die Detektionsrate r_{detect_planes} an. Die Übersegmentierungsrate $r_{overseg_planes}$ gibt das Verhältnis zwischen den insgesamt extrahierten Flächen und den Flächen in den *Ground-Truth*-Daten an. Der mittlere Orientierungsfehler α_{diff} gibt die mittlere Abweichung zwischen der Normalen der extrahierten Fläche und der Normalen in den *Ground-Truth*-Daten an. Die Zuordnung der Punkte zu den Flächen wird bewertet über die Detektionsrate r_{detect_points} und der Übersegmentierungsrate $r_{overseg_points}$ der Punkte. r_{detect_points} gibt den Anteil der aus den *Ground-Truth*-Daten korrekt zugeordneten Punkte an und $r_{overseg_points}$ beschreibt das Verhältnis zwischen der gesamten Menge der Punkte, die einer Fläche zugeordnet wurden, und den Punkten, die in der *Ground-Truth* der Fläche zugeordnet sind. Um die Werte zu bestimmen, wurde eine Modellumgebung aus verschiedenen Blickwinkeln

Abbildung 6.6: Einfärbung der Punkte nach der *Ground-Truth* [140]

Bewertungsmaß	$r_{detect_{planes}}$	$r_{overseg_{planes}}$	α_{diff}	$r_{detect_{points}}$	$r_{overseg_{points}}$
Wert	0,9883	1,3117	5,3771°	0.9010	0,1351

Tabelle 6.1: Übersicht über die Bewertungsmaße der Flächenextraktion

aufgenommen. Die *Ground-Truth* wurde manuell bestimmt und anschließend in einem automatisierten Verfahren mit den Ergebnissen des Algorithmus verglichen. Die verschiedenen Bewertungsmaße sind in Tabelle 6.1 zusammengefasst. Die Laufzeit des vorgestellten Algorithmus wurde auf einem Desktop PC evaluiert, mit den Komponenten *Intel Sandy Bridge i5-2320* Prozessor (3 GHz), 4 GByte RAM und *Microsoft Windows 7 x64*. Die Methode wurde mehrfach auf verschiedenen Aufnahmen der PMD-Kamera durchgeführt. Es befanden sich durchschnittlich 18500 als gültig markierte Pixel in den Aufnahmen. Die Laufzeiten der verschiedenen Bearbeitungsschritte sind in Tabelle 6.2 gegeben.

Die entwickelte Methode zur Flächenextraktion aus Tiefendaten wurde mit Methoden aus dem Stand der Technik verglichen. Das Ergebnis der vorgestellten Methode ist in Abbildung 6.7 dargestellt. Rot eingekreist ist ein Bereich, an dem die Methode fehlschlägt. Aufgrund von hohem Messrauschen kann an dieser Stelle nicht zwischen den verschiedenen Flächen unterschieden werden. Die Ergebnisse der verschiedenen Methoden aus

Verarbeitungsschritt	Laufzeit [ms]
Vorverarbeitung	9,8
Normalenextraktion	50,3
Liniensegmentierung	6,5
Flächensegmentierung	2,1
Globale Flächenextraktion	9,5
Fusion	8,8
Gesamtlaufzeit	87,0

Tabelle 6.2: Laufzeiten der verschiedenen Verarbeitungsschritte

dem Stand der Technik sind in den Abbildungen 6.8, 6.9 und 6.10 dargestellt. Es kann beobachtet werden, dass der *Scanline*-Algorithmus [101], wie erwartet, zu einer starken Übersegmentierung führt. Die *Sequential-RANSAC*-Methode [96] schlägt fehl bei der Unterscheidung zwischen verschiedenen Flächen. Nur die Grundfläche kann korrekt erkannt werden. Der *J-Linkage*-Algorithmus [100] besitzt ähnliche Probleme, wie die *Sequential RANSAC* Methode, da er einerseits Flächen übersegmentiert und andererseits Grenzen zwischen unterschiedlichen Flächensegmenten unberücksichtigt lässt.

6.2 Lokalisierung

Die Erstellung eines hochgenauen Umfeldmodells ist, wie in Kapitel 5 beschrieben, darauf angewiesen, die Position des Fahrzeugs zu den Zeitpunkten der Tiefenmessungen zu kennen. Ungenauigkeiten in der Eigenlokalisierung des Fahrzeugs führen direkt zu Ungenauigkeiten in der Kartenerstellung.

Dieses Kapitel stellt drei verschiedene Methoden zur Eigenbewegungsschätzung vor. Zunächst wird in Abschnitt 6.2.1 auf die etablierte Methode der radgetriebenen Odometrie und der inertialen Messeinheiten eingegangen. Da diese für die Ziele dieser Arbeit keine ausreichende Genauigkeit besitzt, werden zwei weitere Methoden präsentiert. Die visuelle Odometrie (Abschnitt 6.2.2) basiert auf der Schätzung der Eigenbewegung ausschließ-

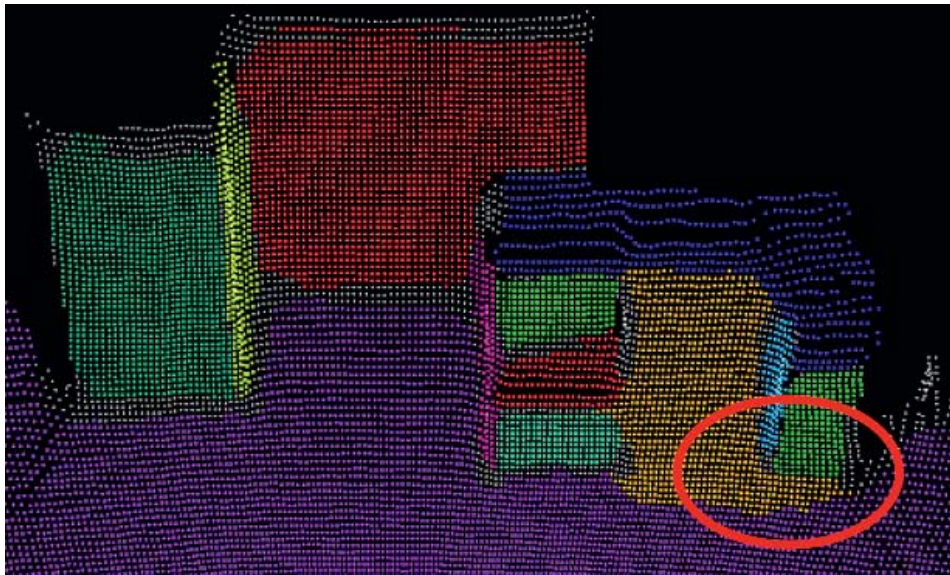


Abbildung 6.7: Ergebnis der vorgestellten Methode zur Flächenextraktion [140]

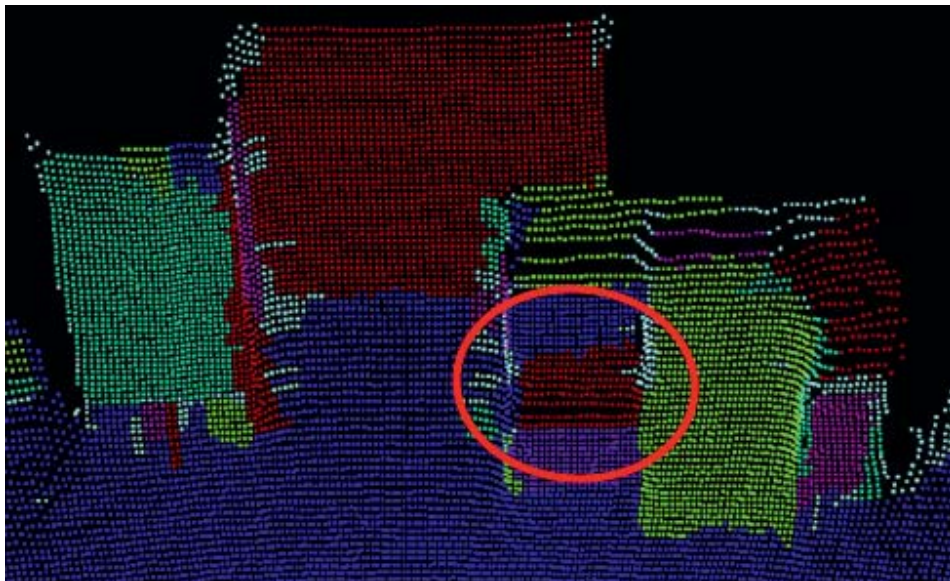


Abbildung 6.8: Ergebnis der *Scanline*-Methode [101] zur Flächenextraktion [140]

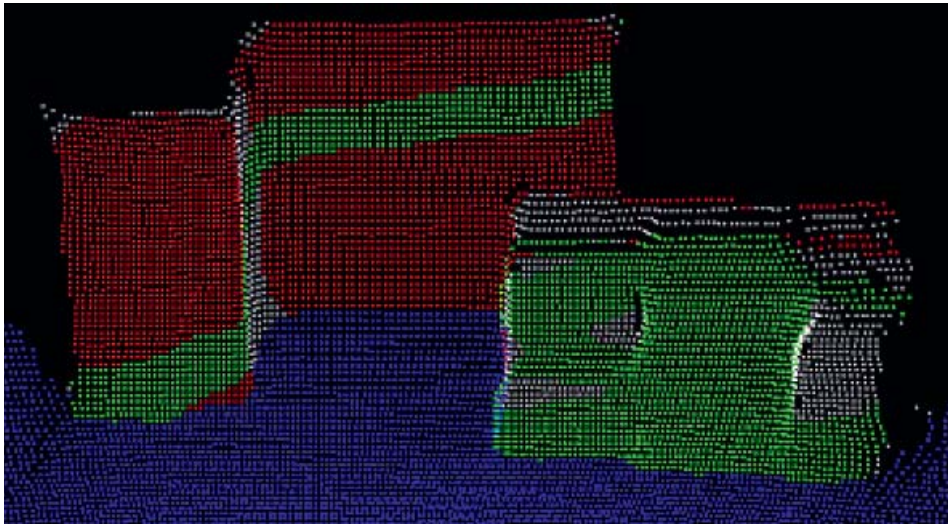


Abbildung 6.9: Ergebnis der *Sequential-RANSAC*-Methode [96] zur Flächenextraktion [140]

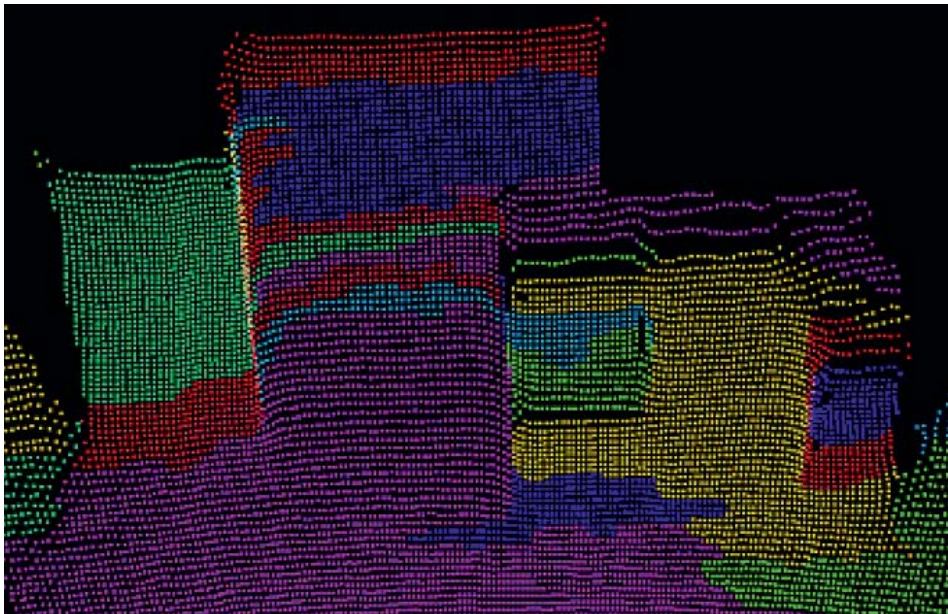


Abbildung 6.10: Ergebnis der *J-Linkage*-Methode [100] zur Flächenextraktion [140]



lich aus den Kameradaten. Ebenso wird in Abschnitt 6.2.3 ein Verfahren zur simultanen Lokalisierung und Kartenerstellung vorgestellt. Die Lokalisierung basiert dabei auf einem Abgleich der Kameradaten mit dem parallel erstellten Umgebungsmodell.

6.2.1 Radgetriebene Odometrie und inertielle Messeinheit

Eine herkömmliche Methode zur Bestimmung der Fahrzeugeigenbewegung kann durch die radgetriebene Odometrie erfolgen. Hier werden die Radumdrehungen der Räder eines Fahrzeugs durch Raddrehzahlsensoren und der Lenkwinkel des Fahrzeugs durch einen Lenkwinkelsensor gemessen. Die Bewegung des Fahrzeugs kann anschließend anhand eines Bewegungsmodells geschätzt werden. In der Regel wird die nichtangetriebene Achse des Fahrzeugs zur Messung herangezogen, um Schlupf der Räder zu minimieren. Bei Schlupf entstehen Fehler, da Radumdrehungen gemessen werden, die nicht zu einer entsprechenden Bewegung des Fahrzeugs führen.

Diese wird über den Radumfang der Räder bestimmt. Da der Radumfang eines Fahrzeugs nicht konstant ist und von verschiedenen Faktoren wie Abrieb oder Luftdruck abhängt, treten zusätzliche Ungenauigkeiten in der Eigenbewegungsschätzung auf.

Zusätzlich kann zur Eigenbewegungsschätzung eine inertielle Messeinheit verwendet werden, die Beschleunigungen und Drehraten in sechs Freiheitsgraden misst. Die Messungen einer inertialen Messeinheit werden vom Schlupf der Räder nicht beeinträchtigt und können diesen daher detektieren und zu einem gewissen Grad korrigieren. Die *Audi AG* setzt zur Entwicklung von Fahrerassistenzsystemen eine leistungsfähige inertielle Messeinheit im sogenannten *EgoMaster*-System [143] ein. Die drei Beschleunigungen und Drehraten in drei Achsen können mit einer Wiederholrate von 100 Hz gemessen werden.

Der Nachteil von radgetriebener Odometrie und inertialen Messeinheiten ist die Akkumulation von Fehlern über die Zeit. Da die Fahrzeugpose nur inkrementell über neue Messungen berechnet werden kann, führt die Akkumulation von Messfehlern zu einem Drift in der Schätzung der Eigenbewegung. Große Fehler treten insbesondere bei niedrigen Fahrzeuggeschwindigkeiten und starkem Lenkwinkelschlag auf. Da diese Szenarien

gerade für das Parken und Rangieren relevant sind, wird eine erweiterte Eigenbewegungsschätzung benötigt, um eine korrekte Karte für die Umgebung des Fahrzeugs aufzubauen.

6.2.2 Visuelle Odometrie

Der Stand der Technik zur visuellen Odometrie wird in Kapitel 2.2.5 vorgestellt. Keine der vorgestellten Methoden ist in der Lage, auf Basis der PMD-Kameradaten zufriedenstellende Ergebnisse zu liefern. Aus diesem Grund wird nachfolgend ein neuartiges Verfahren zur visuellen Odometrie auf Basis der Daten von PMD-Kameras vorgestellt. Aspekte dieses Verfahrens wurden zuvor in [144] und [31] veröffentlicht.

Methode zur visuellen Odometrie

Visuelle Odometrie beschreibt die inkrementelle Schätzung der Pose einer Kamera oder eines Fahrzeugs durch die Beobachtungen der Änderungen in den Kamerabildern, die durch Bewegung verursacht werden [103]. Die vorgestellte Methode basiert auf der Kombination einer 2D-Bildverarbeitung des Amplitudenbildes mit einer 3D-Datenverarbeitung der Tiefendaten. Die Verarbeitungsschritte des Verfahrens sind in Abbildung 6.11 dargestellt. Nach einer Vorverarbeitung der Kameradaten werden Merkmalspunkte in den Amplitudenbildern extrahiert. Zwischen den Merkmalspunkten aufeinanderfolgender Aufnahmen werden Korrespondenzen gebildet. Diese zweidimensionalen Korrespondenzpunkte können durch das Tiefenbild der Kamera unmittelbar in 3D-Punktkorrespondenzen umgewandelt werden. Dies geschieht durch eine effiziente, auf Quaternionen basierende Registrieremethode. Eine Ausreißerdetektion erhöht die Robustheit der Registrierung, indem ungültige Korrespondenzen detektiert werden. Auch durch die Verwendung von mehreren Kameras kann die Zuverlässigkeit des Verfahrens erhöht werden.

Vorverarbeitung

Die Amplitudenbilder werden vorverarbeitet, um die Robustheit der Merkmalsextraktion zu steigern. Hierzu wird, wie in Abschnitt 4.2 vorgestellt, ein bilateraler Filter nach [136] auf das Amplitudenbild angewandt.

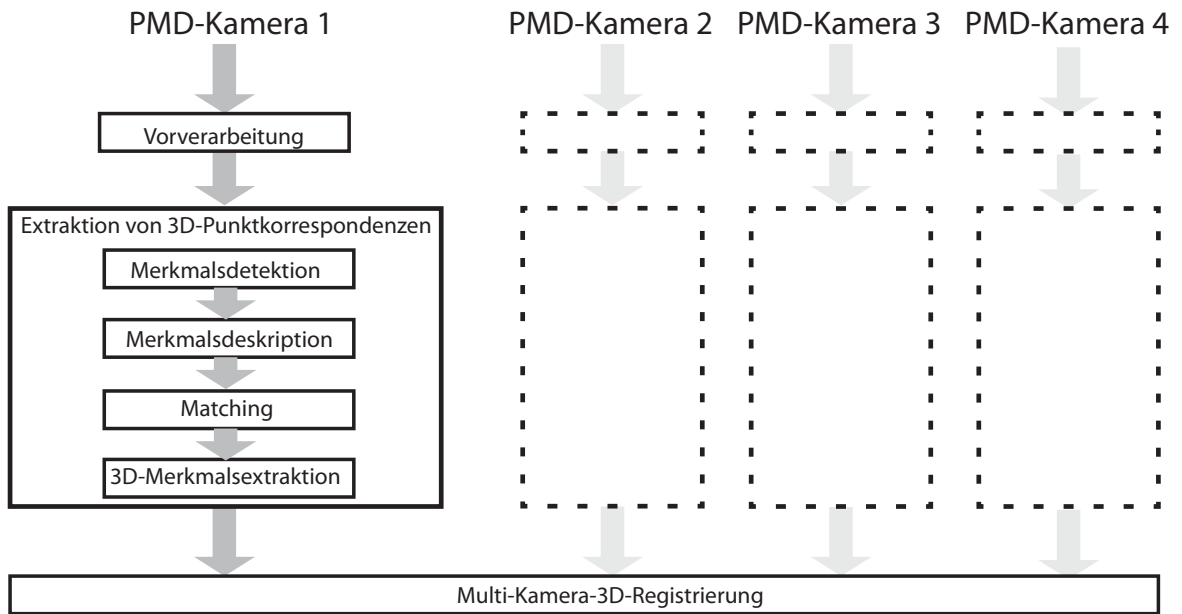


Abbildung 6.11: Überblick über die Methode zur visuellen Odometrie

Ungültige Pixel werden über die Berechnung der Konfidenz nach Kapitel 4.3 entfernt und von der weiteren Verarbeitung ausgeschlossen.

Extraktion von Punktkorrespondenzen

Um die Änderung der Posen zwischen zwei Aufnahmen zu schätzen, müssen Beziehungen zwischen den beiden Aufnahmen bestehen. Die vorgestellte Methode basiert auf der Bildung von Korrespondenzen zwischen charakteristischen Merkmalspunkten aufeinanderfolgender Amplitudenbilder. Wie in Abbildung 6.11 dargestellt, ist der erste Schritt die Detektion von charakteristischen Merkmalspunkten. Für Grauwertbilder existieren verschiedene Verfahren zur Extraktion von Merkmalspunkten. Weit verbreitet sind beispielsweise *Good Features to Track* (GFTT) [145], *Features from Accelerated Segment Test* (FAST) [146], *Scale Invariant Feature Transform* (SIFT) [147] und *Speeded Up Robust Features* (SURF) [148]. Die Verfahren unterscheiden sich beispielsweise hinsichtlich der Berechnungseffizienz oder der Invarianz gegenüber Rotation, Skalierung oder affinen Transformationen.

Nach der Extraktion von Merkmalspunkten werden diese beschrieben und mit einer eindeutigen Signatur versehen. Dies geschieht durch einen Merk-

malsdeskriptor. Die Deskriptoren SIFT und SURF besitzen eine hohe Klassifikationsleistung und Robustheit [105] und sind weit verbreitet. Allerdings benötigt der Vergleich der gleitpunktbasierten Vektordescriptoren von SIFT und SURF eine hohe Rechenleistung insbesondere bei der Verwendung mehrerer Kameras. Aus diesem Grund wird auch der *Fast Retina Keypoint Descriptor* (FREAK) [149] verwendet. FREAK beschreibt einen Merkmalspunkt durch binäre Werte, sodass verschiedene Signaturen sehr effizient über die Bestimmung der Hamming-Distanz verglichen werden können.

[105] evaluiert Leistung und Eigenschaften der verschiedenen genannten Detektoren und Deskriptoren im Kontext von visueller Odometrie. Allerdings ist noch keine Evaluierung unterschiedlicher Detektoren und Deskriptoren unter Berücksichtigung der speziellen Charakteristiken von Time-of-Flight-Kameras (niedrige Auflösung, aktive Beleuchtung) durchgeführt worden. Aus diesem Grund werden die verschiedenen Detektor-Deskriptor-Kombinationen individuell ausgewertet.

Um die Anzahl der verfügbaren Korrespondenzen zu erhöhen, kommt *Multi Frame Matching* zum Einsatz. Dies bedeutet, dass Korrespondenzen nicht nur zwischen einem Bild I_i und dem direkten Vorgänger I_{i-1} bestimmt werden, sondern auch zwischen I_i und weiteren früheren Bildern $I_{i-2}, I_{i-3}, \dots, I_{i-n}$. Dabei entspricht n der Anzahl der vergangenen Bilder, in denen nach korrespondierenden Merkmalspunkten gesucht wird. Durch *Multi Frame Matching* steht für die nachfolgende 3D-Registrierung eine höhere Anzahl an Korrespondenzen zur Verfügung.

3D-Registrierung

Der vorhergehende Prozessschritt liefert korrespondierende 2D-Punkte zwischen Paaren von Amplitudenbildern. Jeder Pixel des Amplitudenbildes besitzt einen Tiefenwert im Distanzbild. Durch Berücksichtigung des Tiefenwertes können die 2D-Punktkorrespondenzen unmittelbar in 3D-Punktkorrespondenzen transformiert werden. Das Ergebnis sind zwei Punktwolken, in denen jeder Punkt \mathbf{p}_k^{i-1} , ($k = 1, \dots, K$) der ersten Punktwolke mit genau einem Punkt \mathbf{p}_k^i der zweiten Punktwolke korrespondiert. Die Bewegungsänderung zwischen den beiden Aufnahmen kann durch eine Registrierung der beiden Punktwolken ermittelt werden. Das Problem besteht



darin, dass eine Rotationsmatrix \mathbf{R}_i^{i-1} und ein Translationsvektor \mathbf{t}_i^{i-1} bestimmt werden müssen, welche die euklidische Distanz zwischen den beiden 3D-Punktwolken minimieren durch

$$\arg \min_{\mathbf{R}_i^{i-1}, \mathbf{t}_i^{i-1}} \sum_k \|\mathbf{R}_i^{i-1} \cdot \mathbf{p}_k^i + \mathbf{t}_i^{i-1} - \mathbf{p}_k^j\|^2. \quad (6.3)$$

Die optimale Transformation kann durch eine sehr effiziente, auf Quaternionen basierende Methode nach [150] bestimmt werden. Sie ist auch Teil des weit verbreiteten *Iterative-Closest-Point-Algorithmus* (ICP) [114]. In dem hier vorgestellten Verfahren kann allerdings aufgrund der bestehenden Korrespondenzen auf eine iterative Verfeinerung verzichtet werden, und die Lösung kann sehr effizient in geschlossener Form berechnet werden.

In Kapitel 2.1.3 werden die verschiedenen Fehlerquellen und Störfaktoren der PMD-Kameras vorgestellt. Diese Fehler führen logischerweise zu Fehlern in der Bestimmung der optimalen Transformation. Ebenso führen falsche Korrespondenzen zu einer fehlerhaften Transformation. Aus diesem Grund werden Ausreißer durch eine *Random-Sample-Consensus-Methode* (RANSAC) entfernt. RANSAC ist ein nichtdeterministisches Verfahren und liefert verschiedene Ergebnisse bei jeder Durchführung. Um ein wiederholbares Ergebnis mit einer Wahrscheinlichkeit p zu erreichen, wird eine bestimmte Anzahl an Iterationen N benötigt. N kann durch

$$N = \frac{\log(1 - p)}{\log(1 - (1 - r_{outlier})^S)} \quad (6.4)$$

berechnet werden. Dabei bezeichnet $r_{outlier}$ den relativen Anteil an Ausreißern und S gibt die Anzahl an verwendeten Beobachtungen zur Bestimmung einer Hypothese an. Im Falle der Registrierung von 3D-Punktwolken werden mindestens $S_{min} = 3$ Beobachtungen benötigt, um die Transformation zu berechnen.

Multi-Kamera-System

Durch die Verwendung von mehreren Kameras zur Bestimmung der Eigenbewegung kann zum einen eine höhere Genauigkeit und zum anderen eine höhere Robustheit gegenüber dynamischen Objekten erwartet wer-

den. Bei einer Kombination der Tiefendaten von vier Kameras, die in unterschiedliche Richtungen gerichtet sind, werden die 3D-Punktwolken in einem gemeinsamen Koordinatensystem betrachtet. Die Extraktion von Punktkorrespondenzen wird für jede Kamera separat durchgeführt. Die 3D-Registrierung der dreidimensionalen Korrespondenzen wird hingegen in einem Schritt für die Daten aller Kameras kombiniert durchgeführt. Dieses Verfahren verspricht die folgenden Vorteile:

- **Verschiedene Arten von Kamerabewegungen:** Da jede Kamera in eine andere Richtung orientiert ist, führt eine Bewegung des Fahrzeugs zu unterschiedlichen Bewegungen der einzelnen Kameras. Beispielsweise bewegt sich bei einer Fahrt des Fahrzeugs nach vorne auch die Frontkamera nach vorne, während sich die seitlichen Kameras seitwärts bewegen und die hintere Kamera rückwärts. Eine Bewegung entlang der Hauptsichtachse der Kamera führt in der Regel zu einer Skalierung der Merkmale, eine seitliche Bewegung zu einer Translation. Diese Effekte werden von den unterschiedlichen Deskriptoren in verschiedener Weise berücksichtigt.
- **Kompensation von Fehlern im Tiefenbild:** Fehler in den Tiefenbildern, die in der Aufnahme einer Kamera auftreten, können durch die Aufnahmen der übrigen Kameras kompensiert werden. Wenn beispielsweise das Tiefenbild einer Kamera aufgrund schlecht reflektierender Objekte hohe Fehler aufweist, besteht die Möglichkeit, dass hochwertige Daten der übrigen Kameras diesen Effekt durch die Ausreißerentfernung ausgleichen können.
- **Robustheit gegenüber dynamischen Objekten:** Dynamische Objekte führen in der Regel zu einer fehlerhaften Bestimmung der Eigenbewegung der Kameras. Werden mehrere Kameras mit einem deutlich größerem Sichtfeld verwendet, werden dynamische Objekte nur einem kleinen Anteil der Punktkorrespondenzen entsprechen. Durch die Verwendung der RANSAC Methode zur Ausreißerentfernung ist die Wahrscheinlichkeit hoch, dass dynamische Anteile, die nur in einem kleinen Teil der Daten auftreten, als Ausreißer deklariert werden und somit nicht im Registrierungsprozess einbezogen werden.



		Detektor			
		GFTT	FAST	SIFT	SURF
Deskriptor	SIFT	++	++	++	+
	SURF	-	-	-	○
	FREAK	○	+	○	-

++	Sehr gute Leistung
+	Gute Leistung
○	Mittlere Leistung
-	Schlechte Leistung

Tabelle 6.3: Qualitative Bewertung der verschiedenen Detektor-Deskriptor-Kombinationen

Ergebnisse

Das vorgestellte Verfahren zur visuellen Odometrie wird evaluiert anhand verschiedener Versuchsfahrten. Als PMD-Kameras kamen *ConceptCams* von der Firma *PMDTechnologies GmbH* [28] zum Einsatz. Diese besitzen einen Öffnungswinkel von 90° und besitzen eine geringere Leuchtstärke. Auch die Auflösung ist mit 200×200 Pixel niedriger als die des *Camboard ATV*. Die Versuchsfahrten wurden mit niedrigen Geschwindigkeiten und hohen Lenkwinkelschlägen durchgeführt.

Um die Leistung des inkrementellen Ansatzes zur Schätzung der Eigenbewegung zu bewerten, wird eine inkrementelle Metrik gewählt. [151] stellt ein inkrementelles Fehlermaß für Translationen und Rotationen vor. Im Gegensatz zu einem absoluten Fehlermaß werden bei einem inkrementellen Fehlermaß die einzelnen Fehler zwischen zwei Posenbestimmungen akkumuliert. Dies vermeidet den Effekt, dass sich Fehler im Laufe einer Messfahrt wieder ausgleichen können, und macht das inkrementelle Fehlermaß zu einem aussagekräftigen Wert, der die Leistung eines inkrementellen Schätzverfahrens angibt.

Eine qualitative Bewertung ist in Abbildung 6.3 gegeben. Die Evaluierung der vorgestellten Detektor-Deskriptor-Kombinationen zeigt, dass die Kombination aus dem SIFT-Deskriptor und dem FAST- oder dem SIFT-Detektor (SIFT/FAST und SIFT/SIFT) bei der Anwendung auf Amplitudenbilder der PMD-Kamera zu den niedrigsten Fehlern in der Ei-

genbewegungsschätzung führt. Die Verwendung der RANSAC-basierten Ausreißerdetektion konnte zu einer Verminderung des Fehlers um 20% bei FAST/SIFT und 10% bei SIFT/SIFT führen. Für SIFT/SIFT ist es vorteilhaft, die Anzahl an Beobachtungen von $S_{min} = 3$ auf $S = 5$ zu erhöhen, um die Hypothese zu bilden (Siehe Gleichung 6.4). Bei der Verwendung des FAST-Detektors reicht ein Wert von $S_{min} = 3$ aus. Während die bilaterale Filterung bei SIFT/SIFT den Fehler um bis zu 10% reduzierte, konnte kein positiver Effekt bei FAST/SIFT beobachtet werden. Dies ist darauf zurückzuführen, dass die Filterung zu einer Glättung des Bildes und damit zu weniger charakteristischen Ecken und Kanten führt, auf die der FAST-Detektor abzielt. Hinsichtlich des *Multi Frame Matching* stellt sich eine Korrespondenzsuche in $n = 2$ vergangenen Bildern als optimal heraus.

Die Evaluierung zeigt, dass sich bei der Verwendung mehrerer PMD-Kameras der Fehler in der Eigenbewegungsschätzung signifikant reduzieren lässt und eine höhere Robustheit erzielt werden kann. Die Hinzunahme einer zweiten Kamera konnte zu einer Reduktion des Fehlers in statischen Situationen um 25% führen. Bei vier Kameras verringert sich der Fehler sogar um 40%. Dieser Effekt kann in dynamischen Situation noch verstärkt werden. Wenn im Kamerabild einer einzelnen Kamera großflächig dynamische Objekte aufgenommen werden, treten trotz Ausreißerdetektion sehr hohe Fehler in der Berechnung der Eigenbewegung auf. Die Berücksichtigung von einer zweiten oder von insgesamt vier Kameras konnte in den dargestellten Szenarien zu einer Reduktion von 82% bzw. 95% führen. Daher kann ein Multi-Kamera-System empfohlen werden, um den Schätzfehler zu verringern, und dies insbesondere in dynamischen Umgebungen.

In Abbildung 6.12 ist sowohl der inkrementelle, als auch der absolute Fehler der Eigenbewegungsschätzung über mehrere Frames dargestellt. Zum Vergleich sind die Ergebnisse der Methode *foviss* [109] aus dem Stand der Technik gegenübergestellt. Es ist zu erkennen, dass die vorgestellte Methode bei der Verwendung von FAST/SIFT und vier Kameras signifikant bessere Ergebnisse liefert.

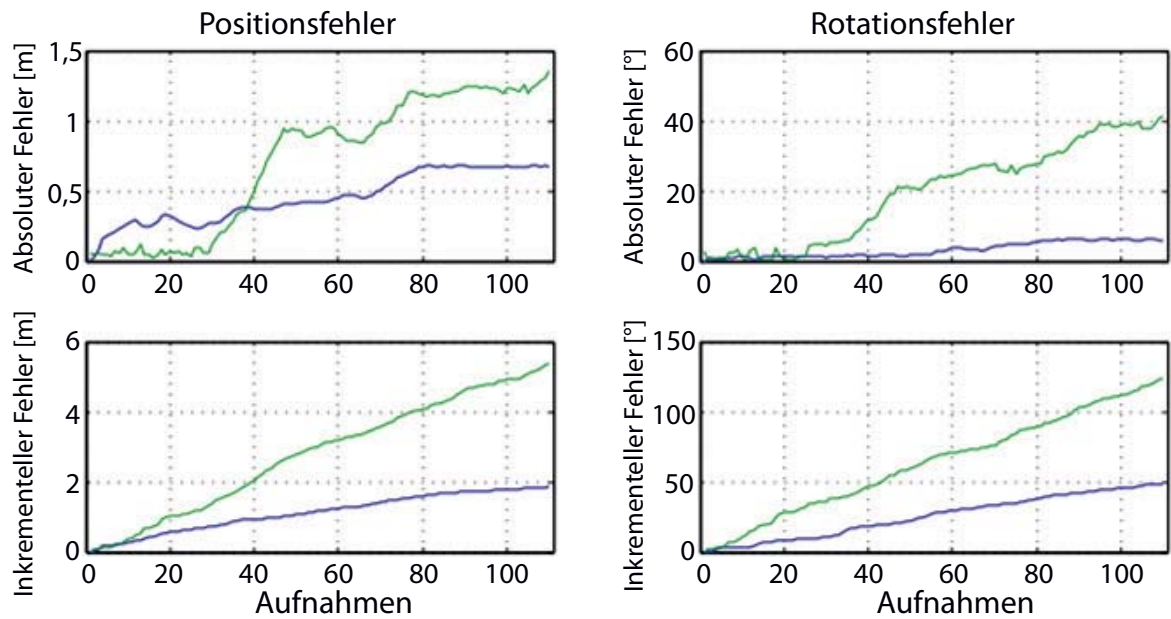


Abbildung 6.12: Vergleich der absoluten und der inkrementellen Fehler der Eigenbewegungsschätzung über mehrere Aufnahmen hinweg. In blau dargestellt ist die vorgestellte Methode mit FAST/SIFT als Detektor-Deskriptor-Kombination bei der Verwendung von vier Kameras. Grün markiert ist die Methode *fovis* [109] aus dem Stand der Technik.

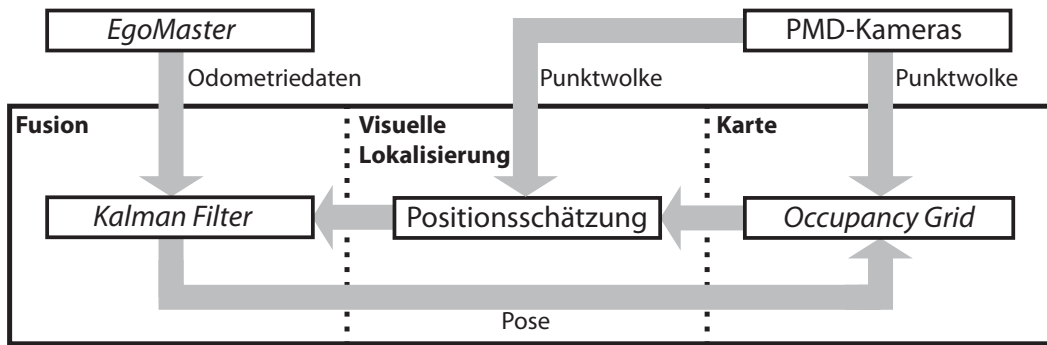


Abbildung 6.13: Überblick über das vorgestellte SLAM-Verfahren

6.2.3 *Simultaneous Localisation and Mapping*

Sowohl die radgetriebene Odometrie in Kombination mit einer inertialen Messeinheit als auch das vorgestellte Verfahren zur visuellen Odometrie unterliegen einem Drift, der dazu führt, dass bei längerem Rangieren große Fehler in der Positionsschätzung des Fahrzeugs auftreten. Dieses Problem kann ein Ansatz zum *Simultaneous Localisation and Mapping* (SLAM) lösen. Das SLAM-Problem adressiert die simultane Lokalisierung und Kartenerstellung. Zur Lokalisierung werden die Sensordaten mit der erstellten Karte abgeglichen. Dabei wird die Karte abhängig von der geschätzten Pose erstellt. Der Stand der Technik bezüglich SLAM ist in Kapitel 2.2.5 vorgestellt worden.

Ein Überblick über das nachfolgend vorgestellte Verfahren ist in Abbildung 6.13 gegeben. Die Positionsschätzung beruht auf einem Abgleich der Punktwolken mit dem erstellten Umgebungsmodell. Durch einen Kalman-Filter wird die visuelle Lokalisierung auf Basis der Kameradaten mit den Odometriedaten des Fahrzeugs fusioniert.

Visuelle Lokalisierung

Das Ziel der visuellen Lokalisierung ist es, die Fahrzeugpose durch einen Abgleich der Sensordaten mit der aufgebauten Karte zu bestimmen. Der Aufbau der Umfeldkarte wird in Kapitel 5 thematisiert. Im vorgestellten Verfahren wird das probabilistische *Occupancy Grid* genutzt. Dabei werden zunächst verschiedene Positionshypothesen aufgestellt. Diese werden anschließend anhand des bereits aufgebauten *Occupancy Grids* be-

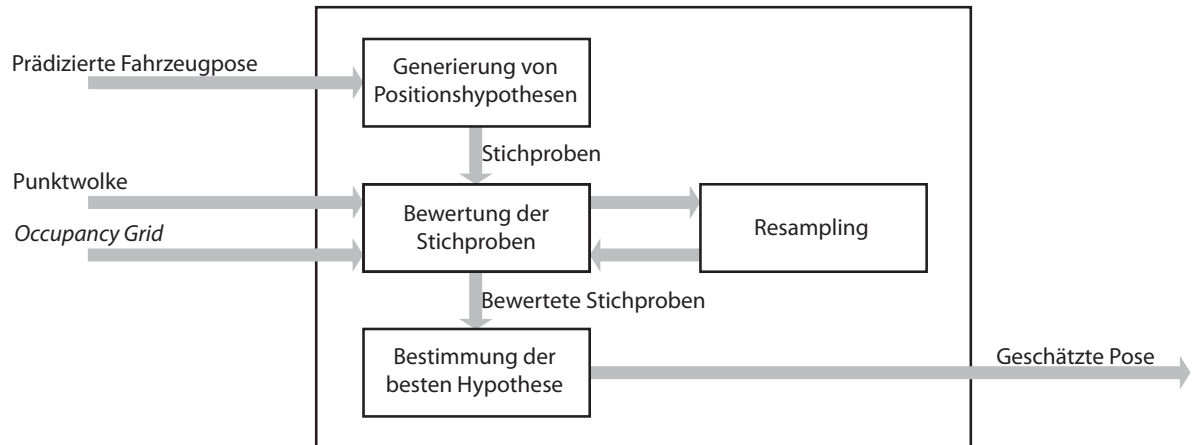
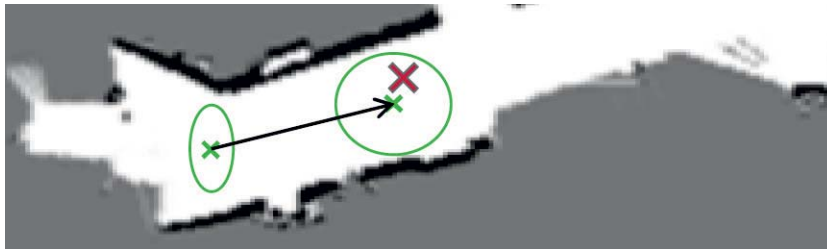


Abbildung 6.14: Überblick über die Methode zur visuellen Lokalisierung durch einen Abgleich der Sensordaten mit der aufgebauten Umfeldkarte

wertet. Die Bewertung der Positionshypothesen führt schließlich zu einer Schätzung der eigenen Position, die als Messung für den nachfolgenden Kalman-Filter dient. Einen Überblick über das Verfahren zur visuellen Lokalisierung zeigt Abbildung 6.14.

Zunächst werden Positionshypothesen zufällig erstellt. Als Basis dient die Prädiktion des Kalman-Filters. Eine Positionshypothese besteht aus den 2D-Punktkoordinaten x und y , sowie einer Orientierung φ . Es werden nun eine Anzahl $n_{samples}$ von Positionshypothesen in der Nähe der prädizierten Position randomisiert generiert. Die Verteilung der Stichproben erfolgt durch eine Normalverteilung mit der prädizierten Position als Zentrum. Der Prozess ist in Abbildung 6.15 dargestellt. Die prädizierte Position ist mit einem grünen Kreuz und die reale Position mit einem roten Kreuz dargestellt. Die Positionshypothesen, die normalverteilt in der Nähe der prädizierten Position generiert werden, sind in schwarz dargestellt.

Aus der Menge der Positionshypothesen soll eine einzelne Hypothese bestimmt werden. Hierzu ist eine Bewertung der einzelnen Hypothesen notwendig, bei der die aktuelle Punktwolke der Kameras herangezogen wird. Nur Punkte, die sich oberhalb der Bodenebene befinden und damit ein Hindernis darstellen, werden berücksichtigt. Die gesamte Punktwolke in Bezug zum Fahrzeugkoordinatensystem wird zur Position der Positionshypothesen



a) Prädiktion des Messwertes



b) Streuung um die prädizierte Position

Abbildung 6.15: Die Erzeugung von Positionshypothesen erfolgt durch eine normalverteilte Streuung um eine prädizierte Position.

hypothese transformiert. Die Hypothese soll eine hohe Bewertung erhalten, wenn sich die Punktwolke mit dem bereits berechneten *Occupancy Grid* deckt. Dies ist in Abbildung 6.16 dargestellt. Eine Positionshypothese wird schlecht bewertet, wenn die korrespondierende Punktwolke auf als frei deklarierte Zellen fällt, und sie wird gut bewertet, wenn die korrespondierende Punktwolke auf als belegt deklarierte Zellen fällt.

Das Gewicht w_i einer Positionshypothese wird aus den N Punkten der aktuellen Punktwolke durch

$$w_i = \frac{1}{2N} \sum_{j=1}^N \begin{cases} p(\mathbf{m}(\mathbf{x}_j)) + c_j & , \text{ wenn } p(\mathbf{m}(\mathbf{x}_j)) > 0,5 \\ 0 & , \text{ sonst} \end{cases} \quad (6.5)$$

bestimmt. Dabei bezeichnet $\mathbf{m}(\mathbf{x}_j)$ die Zelle, welcher der Punkt \mathbf{x}_j nach der Transformation zugeordnet ist, und folglich $p(\mathbf{m}(\mathbf{x}_j))$ die Belegheitswahrscheinlichkeit der Zelle. Indem die Konfidenz c_j der einzelnen Punkte berücksichtigt wird, wird Punkten mit hoher Konfidenz ein höheres Gewicht zugeordnet.



a) Beispiel einer schlecht bewerteten Positionshypothese



b) Beispiel einer gut bewerteten Positionshypothese

Abbildung 6.16: Eine Positionshypothese wird schlecht bewertet, wenn die korrespondierende Punktwolke auf als frei deklarierte Bereiche fällt (a), und gut bewertet, wenn die Punkte auf als belegt deklarierte Zellen fällt (b).

Die Erhöhung der Anzahl an Positionshypothesen hat zur Folge, dass die Wahrscheinlichkeit, dass eine zur Realität passende Hypothese generiert wird, erhöht wird. Gleichzeitig führt eine Zunahme der Anzahl an Positionshypothesen zu einem Anstieg der Rechenzeit. In [120] wird eine Methode vorgestellt, welche die Dichte der Positionshypothesen durch ein gewichtetes Resampling erhöht. Diese Methode wird im vorgestellten Verfahren übernommen. Nach der ersten Streuung und Bewertung der Positionshypothesen findet eine randomisierte Selektion von Positionshypothesen statt. Die Hypothesen werden probabilistisch und in Abhängigkeit von ihrem Gewicht bestimmt. Je höher das Gewicht, desto wahrscheinlicher ist es, dass die Positionshypothese ausgewählt wird. Hypothesen können auch mehrfach ausgewählt werden. Wenn eine Positionshypothese gezogen wird, dann wird die Position und Ausrichtung künstlich durch eine Normalverteilung mit der Standardabweichung σ_r einem Rauschprozess unterzogen. Auf diese Weise entstehen neue Positionshypothesen, die eine höhere Dichte in der Nähe der gut bewerteten Hypothesen besitzen. Der Resampling-Schritt kann iterativ angewendet werden. In jedem weiteren Resampling-Schritt wird die Standardabweichung σ_r mit einem Faktor f multipliziert, um die Dichte mit jedem Schritt zu erhöhen. In Abbildung 6.17 ist das Prinzip des Resamplings anhand eines Beispiels dargestellt. Nach einer Bewertung der Positionshypothesen erfolgt der erste Resampling-Schritt. Dieser führt dazu, dass die Dichte der neuen Hypothesen in der Umgebung der gut bewerteten Samples zunimmt. Dasselbe geschieht nach der Anwendung des zweiten Resampling-Schritts. Als guter Kompromiss zwischen Rechenzeit und Qualität der bestimmten Positionshypothese hat sich die Verwendung von zwei Resampling-Schritten mit 200 Positionshypothesen und einem Faktor $f = 0,4$ erwiesen.

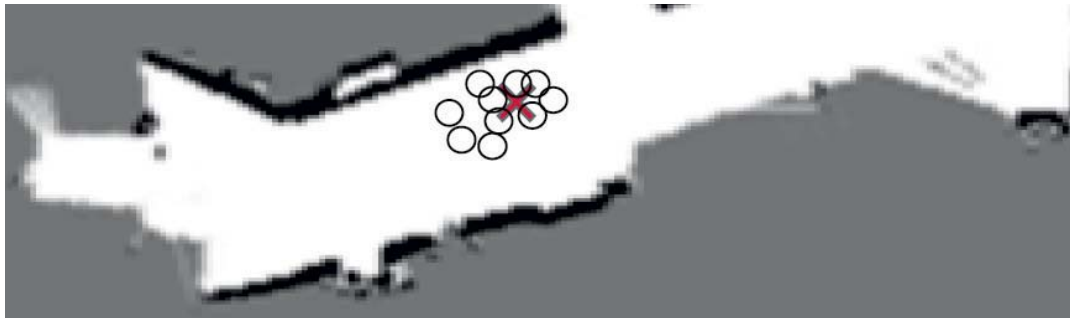
Aus der gesamten Menge der Positionshypothesen nach dem letzten Resampling-Schritt wird die Hypothese ausgewählt, die das höchste Gewicht besitzt. Die Position und die Ausrichtung dieser besten Hypothesen werden anschließend als Ergebnis der visuellen Lokalisierung an den nachfolgenden Kalman-Filter übergeben.

Fusion mit konventioneller Odometrie

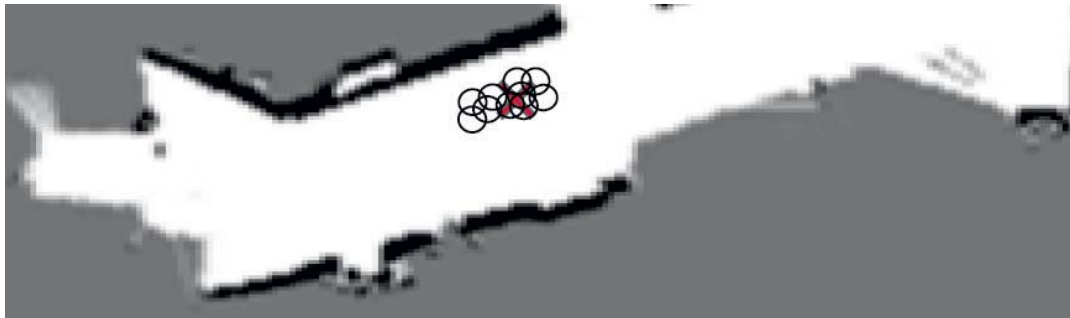
Die konventionelle Odometrie auf Basis von Raddrehzahlsensoren und einer inertialen Messeinheit wird durch einen erweiterten Kalman-Filter



a) Bewertung der Positionshypothesen



b) Erstes Resampling



c) Zweites Resampling

Abbildung 6.17: Nach einer Bewertung der Positionshypothesen (a) erfolgt der erste Resampling-Schritt (b). Dieser führt dazu, dass die Dichte der neuen Hypothesen in der Umgebung der gut bewerteten Samples zunimmt. Nach der Anwendung des zweiten Resampling-Schrittes erhöht sich wiederum die Dichte um die gut bewerteten Hypothesen (c).

(EKF) mit der vorgestellten Methode zur visuellen Lokalisierung fusioniert. Der erweiterte Kalman-Filter wird ausführlich in [152] thematisiert. Die Methode der konventionellen Odometrie unterliegt, wie in Abschnitt 6.2.1 erläutert, einem Drift und damit einem niederfrequenten Rauschen, während die vorgestellte Methode zur visuellen Lokalisierung einem hochfrequentem Rauschen ausgesetzt ist. Ziel der Fusion ist es daher, die Vorteile der beiden Verfahren zu kombinieren und somit in Summe eine genauere Lokalisierung zu erreichen.

Die Messungen der konventionellen Odometrie (Geschwindigkeit und Gierrate) fließen über ein Bewegungsmodell in das Systemmodell des Kalman-Filters ein. Als Fahrzeugmodell wird das 4-Rad-Ackermann-Lenkmodell [153] unter der Annahme von konstanter Gierrate und konstanter Geschwindigkeit (Constant Turn Rate and Velocity) [154] verwendet.

Die Eigenbewegungsschätzung der visuellen Odometrie wird als Messung betrachtet und fließt somit für den Innovationsschritt in das Messmodell des Kalman-Filters ein. Das Prozessrauschen, das im Systemmodell zur Anwendung kommt, wird als konstant definiert. Das Messrauschen wird hingegen in Abhängigkeit von der Qualität der Messung der visuellen Lokalisierung modelliert. Im vorhergehenden Abschnitt wurde erläutert, dass die Dichte der Positionshypothesen nach den Resampling-Schritten um die Positionen der als gut bewerteten Hypothesen zunimmt. Aus diesem Grund kann durch die Streuung der Positionshypothesen eine Aussage über die Konfidenz der visuellen Lokalisierung getroffen werden. Somit wird die Varianz in der Streuung der Positionshypothesen für die Modellierung des Messrauschens verwendet. Auf diese Weise wird in Situationen, in denen die visuelle Lokalisierung eine hohe Sicherheit besitzt, stärker auf die geschätzte Position vertraut, als in Situationen, in denen eine hohe Unsicherheit in der visuellen Lokalisierung vorliegt. Dies ist beispielsweise der Fall, wenn sich das Fahrzeug in bisher unerkundete Bereiche bewegt und daher noch keine Karte zum Abgleich mit den Kameradaten existiert.

Die Daten der konventionellen Odometrie ermöglichen einen Prädiktionschritt des Kalman-Filters. Da die Daten der konventionellen Odometrie mit einer höheren Frequenz bereitgestellt werden, finden mehrere Prädiktionschritte zwischen zwei Innovationsschritten statt. Auf diese Weise kann die Frequenz der Eigenbewegungsschätzung deutlich höher sein, als die Frequenz der visuellen Lokalisierung.

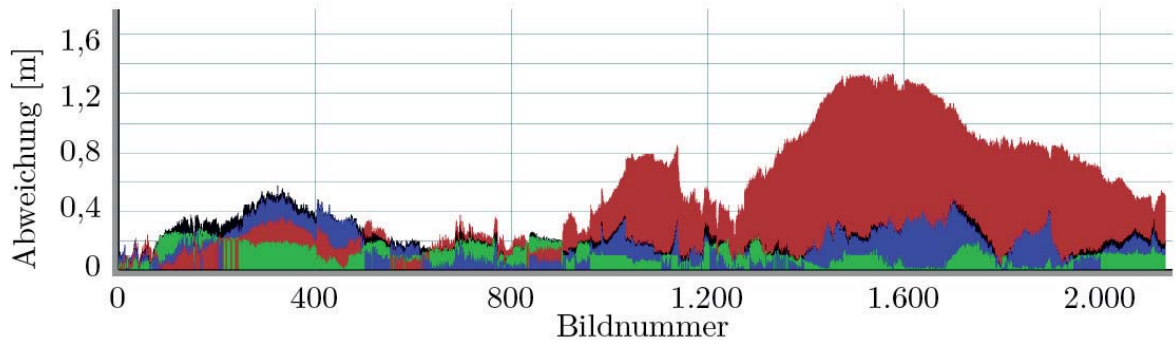


Abbildung 6.18: Die Abweichungen des *EgoMasters* (rot) und der vorgestellten Lokalisierungsmethode (schwarz) zu einer Referenztrajektorie. Der Fehler der Lokalisierungsmethode wird unterteilt in einen longitudinalen Fehler (blau) und einen lateralen Fehler (grün).

Evaluierung

Die vorgestellte Methode und das Verfahren der konventionellen Odometrie werden mit einer Referenztrajektorie verglichen. Als konventionelle Odometrie wird das in Abschnitt 6.2.1 vorgestellte *EgoMaster*-System genutzt. Insgesamt werden vier verschiedene Einparkszenarien mit mehreren Richtungswechseln ausgewertet. Als Vergleichswert wird der euklidische Abstand zwischen den bestimmten Fahrzeugpositionen und der Referenzposition eingesetzt.

Die Auswertung einer beispielhaften Sequenz ist in Abbildung 6.18 dargestellt. Die Abweichung des *EgoMasters* ist in rot dargestellt und die entwickelte Lokalisierungsmethode in schwarz. Dabei ist der Fehler der eigenen Methode unterteilt in einen longitudinalen (blau) und einen lateralen Fehler (grün). Es ist zu beobachten, dass die entwickelte Methode in der abgebildeten Sequenz zunächst stärker von der Referenz abweicht als der *EgoMaster*. Dies ist darauf zurückzuführen, dass zu Beginn noch keine Karte existiert. Die visuelle Lokalisierung vergleicht die Messdaten der Kameras mit der bereits erstellten Karte. Der Vergleich der Daten mit einer noch nicht aufgebauten Karte führt daher zu den zu beobachtenden Fehlern. Nach einer kurzen Zeit kann sich die visuelle Lokalisierung allerdings dann anhand der aufgebauten Karte orientieren, sodass sich der

	Sequenz 1	Sequenz 2	Sequenz 3	Sequenz 4	Gesamt
Ø-Abweichung <i>Egomaster</i> [m]	1,7492	0,5389	0,3804	0,4154	0,7840
Ø-Abweichung Methode [m]	0,1363	0,2913	0,0784	0,0601	0,1634
Verbesserung [%]	92,2	45,9	79,4	85,5	79,2

Tabelle 6.4: Die durchschnittlichen Abweichungen der letzten 50 Aufnahmen für alle vier Sequenzen.

Fehler wieder verringert. Ab etwa der Bildnummer 1000 und insbesondere bei Bildnummer 1300 kann ein hoher Drift des *Egomasters* beobachtet werden. Auch die vorgestellte Lokalisierungsmethode weist hier eine Spitze in der Abweichung auf, da die Bewegungsschätzung des *EgoMasters* im Kalman Filter berücksichtigt wird. Dennoch kann sich die visuelle Lokalisierung zusätzlich an der aufgebauten Karte orientieren, sodass der Fehler wieder ausgeglichen werden kann. Zum Ende der Sequenz weist der *Egomaster* eine Abweichung von 0,3804m auf, während die vorgestellte Lokalisierungsmethode bei einem Fehler von 0,0784m endet, was einer Verringerung des Fehlers um 79,4% entspricht.

Neben der abgebildeten Sequenz wurden drei weitere Sequenzen ausgewertet. In allen Szenarien endet die Sequenz mit einer Fahrzeugposition, welche ähnlich ist zur Position zu Beginn der Sequenz. Das Ziel der Lokalisierung ist der Aufbau einer konsistenten Karte. Daher ist die Genauigkeit der Lokalisierung besonders wichtig, wenn sich das Fahrzeug in der Umgebung einer zuvor erstellten Karte bewegt. Um diese Szenarien zu bewerten, werden die jeweils letzten 50 Sequenzen jeder Sequenz analysiert. Die durchschnittlichen Abweichungen der letzten 50 Aufnahmen für alle vier Sequenzen sind in Tabelle 6.4 dargestellt. Insgesamt ergibt sich eine durchschnittliche Abweichung von 16 cm, welche einer Verringerung des Fehlers um 79,2% entspricht.

6.3 Detektion dynamischer Objekte

Der Stand der Technik bezüglich der Extraktion von dynamischen Objekten aus Kameradaten wurde in Kapitel 2.2.6 präsentiert. Keine der vor-

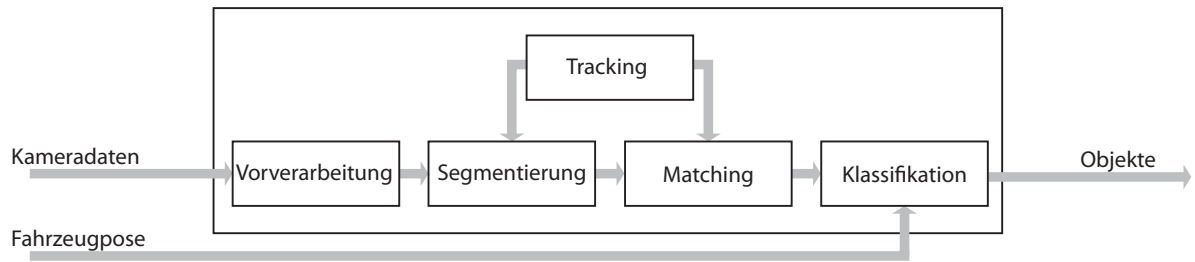


Abbildung 6.19: Überblick über die Methode zur Detektion von dynamischen Objekten

gestellten Methoden ermöglicht eine robuste Detektion von dynamischen Objekten aus den verrauschten Tiefendaten einer sich bewegenden Weitwinkel PMD-Kamera. Aus diesem Grund wird nachfolgend ein neuartiges Verfahren zur Detektion dynamischer Objekte aus den Daten von Weitwinkel PMD-Kameras vorgestellt. Inhalte dieser Methode wurden bereits in [155] veröffentlicht.

6.3.1 Überblick über die Methode

Abbildung 6.19 gibt einen Überblick über die verschiedenen Verarbeitungsschritte der Methode. Zunächst werden die verrauschten Daten vorverarbeitet und ungültige Pixel entfernt. Anschließend wird das Tiefenbild segmentiert, und die verschiedenen Segmente werden charakterisiert. Auf Basis der extrahierten Segmente werden im Objekt-Matching-Schritt Verknüpfungen zwischen den Bildern aufeinanderfolgender Segmente gebildet. Die verschiedenen Objekte werden über mehrere Bilder hinweg verfolgt. Eine anschließende Klassifizierung bestimmt, ob es sich um ein statisches oder dynamisches Objekt handelt. Hierzu wird die Eigenbewegung des Fahrzeugs berücksichtigt.

6.3.2 Vorverarbeitung

Die Vorverarbeitung hat das Ziel, die Daten für die nachfolgenden Verarbeitungsschritte aufzubereiten. Zunächst wird das Bild mit dem in Kapitel 4.2 vorgestellten bilateralen Filter [136] geglättet. Anschließend werden

ungültige Pixel nach der in Kapitel 4.3 präsentierten Methode zur Konfidenzberechnung entfernt. Nach der Berechnung der 3D-Punktwolke aus den Tiefendaten werden zudem Punkte, die sich auf oder unter der Bodenebene befinden, herausgefiltert. Dynamische Objekte befinden sich stets über der Bodenebene.

6.3.3 Segmentierung

Ziel der Segmentierung ist die Extraktion zusammenhängender Bereiche aus den Bilddaten, die Objekten in der Umgebung entsprechen. Die Segmentierung basiert auf den Tiefendaten und extrahiert Segmente, die konsistent hinsichtlich ihrer Distanzwerte sind. Um dies umzusetzen, wird ein *Seeded-Region-Growing*-Algorithmus angewendet. Ausgehend von einem Startpunkt s wird ein zusammenhängender Bereich markiert. Nacheinander werden die Nachbarn eines Pixels betrachtet, und wenn ein Nachbarpixel als zugehörig markiert wird, dann wird das Verfahren rekursiv auf diesen Pixel angewandt. Ein Pixel wird hinzugefügt, wenn er sich in direkter Nachbarschaft eines bereits hinzugefügten Pixels befindet und wenn der Unterschied zwischen den beiden Distanzwerten geringer als ein Schwellwert T ist. Die Qualität der Segmentierung hängt von der Wahl des Schwellwertes T und von der Wahl des Startpunktes s ab.

In Abschnitt 6.3.5 ist beschrieben, dass Objekte über längere Sequenzen in aufeinanderfolgenden Bildern verfolgt werden. Auf Basis der Vorhersage der Objektposition im Bild werden Startpunkte zur Segmentierung gesetzt. Auf jedes Objekt, für das eine Vorhersage für das aktuelle Bild möglich ist, wird der Algorithmus angewandt. Somit werden die entsprechenden Segmente bestimmt. Um neu im Bild erscheinende Objekte berücksichtigen zu können, werden weitere Startpunkte gesetzt. Diese werden anhand eines *Scanline*-Algorithmus bestimmt [101]. Der Algorithmus wird nur auf die Bereiche angewandt, die einerseits als gültige Pixel markiert und andererseits noch nicht einem Segment zugeordnet wurden. Über das *Scanline*-Verfahren werden nun Punkte gesucht, die bezüglich ihres Distanzwertes ein lokales Minimum bilden. Diese Punkte werden als Startpunkte für weitere Durchläufe des *Seeded-Region-Growing*-Algorithmus gewählt. Da dieses Vorgehen eine starke Übersegmentierung des Bildes zur Folge hat, werden nur Segmente weiter betrachtet, die aus einer Mindestanzahl an Pixeln bestehen. In Abbildung 6.20 sind die extrahierten Seg-

mente einer Aufnahme durch unterschiedliche Einfärbungen dargestellt. Das entsprechende Amplitudenbild ist in Abbildung 6.21 abgebildet. Nach der Segmentierung werden verschiedene Eigenschaften der einzelnen Segmente bestimmt. Diese Eigenschaften beinhalten den Schwerpunkt sowohl in zweidimensionalen Bild-, als auch in dreidimensionalen Weltkoordinaten, den durchschnittlichen Tiefenwert und die umfassende *Bounding Box*. Diese Eigenschaften werden im nachfolgenden Matching-Schritt verwendet.

6.3.4 Matching

Das Objekt-Matching hat zum Ziel, Verbindungen zwischen den Segmenten der aktuellen Aufnahme und den Segmenten vorheriger Aufnahmen herzustellen. Hierzu wird ein Matching-Verfahren eingesetzt, das nachfolgend erläutert wird.

Zusätzlich zu den im vorherigen Abschnitt aufgeführten Eigenschaften werden zu jedem Segment Merkmalspunkte extrahiert. Hierzu wird das Amplitudenbild der Kamera in Betracht gezogen. Nach [31] liefert eine Kombination des *Features-from-Accelerated-Segment-Test*-Detektor (FAST) [146] mit dem *Scale-Invariant-Feature-Transform*-Deskriptor (SIFT) [147] gute Ergebnisse auf Basis des Amplitudenbildes einer PMD-Kamera. Zur Einsparung von Rechenzeit werden detektierte Merkmalspunkte nur durch den Deskriptor beschrieben, wenn dieser einem Segment zugeordnet werden konnte. Hierzu ist die euklidische Distanz des Merkmalspunkt zu einem Segment auf der Bildebene maßgeblich. Das Ergebnis der Extraktion von Merkmalspunkten und ihrer Zuordnung zu den Segmenten ist in Abbildung 6.21 zu sehen.

Segmente, denen Merkmalspunkte zugewiesen wurden, werden im Folgenden als Objekte bezeichnet. Auf Basis der extrahierten Merkmalspunkte werden nun Beziehungen zwischen den im aktuellen Bild extrahierten Objekten und den im vorherigen Bild extrahierten Objekten gebildet. Sei $\mathbf{O}_t = \{O_t^1, \dots, O_t^n\}$ die Menge der in der aktuellen Aufnahme zum Zeitpunkt t detektierten Objekte und entsprechend \mathbf{O}_{t-1} und \mathbf{O}_{t-2} die Menge der aus den vorhergehenden Aufnahmen zu den Zeitpunkten $t-1$ und $t-2$ extrahierten Objekte. Ein Objekt $O_t^k \in \mathbf{O}_t$ wird nun mit den Objekten in \mathbf{O}_{t-1} in Beziehung gesetzt. Dies geschieht durch einen Vergleich der Merkmalspunkte. Es werden Korrespondenzen zwischen den Merkmalspunkten



Abbildung 6.20: Darstellung der aus einer Aufnahme extrahierten Segmente [155].

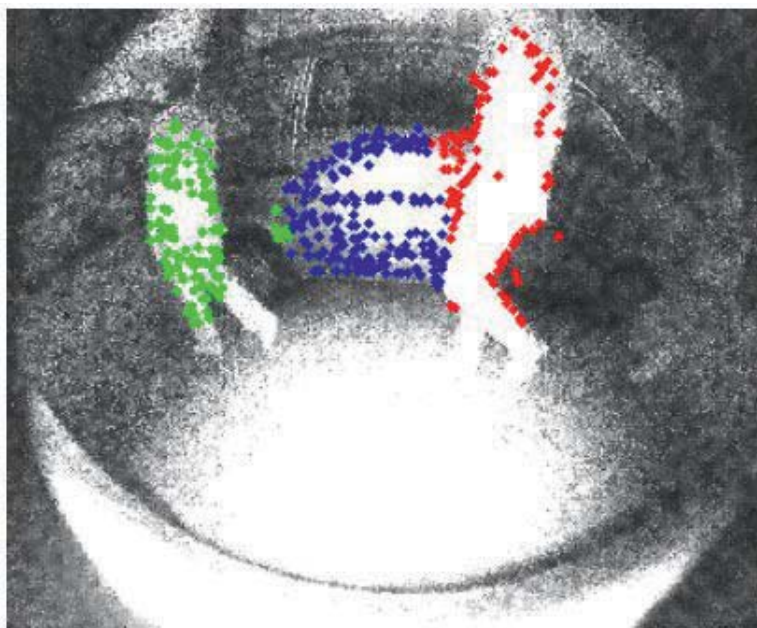


Abbildung 6.21: Extraktion und Zuordnung von Merkmalspunkten auf Basis von FAST und SIFT [155].



von O_t^k und einem Objekt $O_{t-1}^l \in \mathbf{O}_{t-1}$ gesucht. Wenn die Anzahl der korrespondierenden Merkmalspunkte einen Schwellwert überschreitet, dann wird davon ausgegangen, dass es sich um dasselbe Objekt handelt, und die Suche kann beendet werden. Wenn kein passendes Objekt in \mathbf{O}_{t-1} gefunden wird, dann wird anschließend in den Objekten der vorherigen Aufnahme \mathbf{O}_{t-2} gesucht. Sollte auch hier kein passendes Objekt gefunden werden, wird ein neues Objekt erstellt.

In der Objektliste sind Informationen über die Trajektorien der einzelnen Objekte gespeichert sowie die extrahierten Eigenschaften und Merkmalspunkte.

6.3.5 Tracking

Die Verfolgung der Objekte basiert auf einem Kalman-Filter [156]. Dieser erlaubt die Prädiktion eines Prozesszustandes auf Basis eines Prozessmodells und des aktuellen Zustandes. Zu jeder neuen Messung wird die Prädiktion verfeinert. Wie auch in [157] wird als Prozessmodell ein *Constant-Velocity*-Modell verwendet.

Die Prädiktion kommt an zwei Stellen zum Einsatz. Zum einen werden die Schwerpunkte der Objekte fortlaufend verfolgt, und zum anderen wird die Prädiktion eingesetzt, um passende Startpunkte für den in Abschnitt 6.3.3 vorgestellten *Seeded-Region-Growing*-Algorithmus zu definieren.

6.3.6 Klassifizierung

Das Ziel der Klassifizierung ist eine Unterscheidung der extrahierten und verfolgten Objekte in statische und dynamische Objekte. Diese Unterscheidung beruht auf der Verschiebung der Objekte im Weltkoordinatensystem seit der ersten Beobachtung des Objekts und der entsprechenden Eigenbewegung des Fahrzeugs.

Die Berücksichtigung der Eigenbewegung ist notwendig, um die Erkennung von dynamischen Objekten robust zu gestalten. Insbesondere Szenarien, in denen sich das Fahrzeug entlang eines lang ausgedehnten Objektes bewegt können Probleme verursachen, wenn ausschließlich die Bewegung der extrahierten Objekte berücksichtigt wird. Wenn sich das Fahrzeug beispielsweise entlang einer Wand bewegt, ist zu jedem Zeitpunkt nur ein kleiner Teil der Wand im Kamerabild als gültige Messung sichtbar. Da sich das

Fahrzeug bewegt, verschiebt sich ebenso der sichtbare Bereich dieser Wand analog zur Fahrzeugbewegung. Das zuvor vorgestellte Verfahren erkennt, dass es sich in den unterschiedlichen Kamerabildern um dasselbe Objekt handelt, bestimmt aber eine Bewegung des Objekts entlang der Fahrzeugtrajektorie. Um dies zu vermeiden, wird die folgende Methode umgesetzt. Zunächst wird ein Vektor $\Delta\vec{p}$ bestimmt, welcher der Verschiebung des Objekts seit der ersten Beobachtung entspricht. Hierzu wird nur die Bewegung in x und y -Richtung berücksichtigt. Entsprechend wird ein Vektor $\Delta\vec{v}$ ermittelt, welcher der Verschiebung der Fahrzeugposition während des selben Zeitraumes entspricht. Anschließend wird ein Differenzvektor $\Delta\vec{d}$ nach Gleichung 6.6 bestimmt.

$$\Delta\vec{d} = |\Delta\vec{v} - \Delta\vec{p}| \quad (6.6)$$

$\Delta\vec{d}$ wird herangezogen, um einen *Score*-Wert h zu bestimmen nach

$$h = \alpha \|\Delta\vec{d}\|_2 + \beta(n) \|\Delta\vec{p}\|_2. \quad (6.7)$$

Dabei ist α ein konstanter Parameter und $\beta(n)$ skaliert den Differenzvektor $\Delta\vec{p}$ abhängig von der Anzahl an Beobachtungen n .

Die Klassifikation erfolgt anschließend durch einen Vergleich von h mit den vordefinierten Schwellwerten T_{Low} und T_{High} nach

$$\text{Zustand} = \begin{cases} \text{Statisch} & , \text{ wenn } h \leq T_{Low} \\ \text{Dynamisch} & , \text{ wenn } h \geq T_{High} \\ \text{Unbekannt} & , \text{ sonst.} \end{cases} \quad (6.8)$$

Auf diese Weise kann sichergestellt werden, dass der Zustand von Objekten nach wenigen Beobachtungen korrekt klassifiziert werden kann.

6.3.7 Evaluierung

Die Extraktion von dynamischen Objekten wird evaluiert, indem die Ergebnisse des Verfahrens mit einer Referenz verglichen werden. Als Referenz wurden Aufnahmen generiert, in denen die dynamischen Pixel manuell markiert wurden. Zur Evaluierung wird nur ein zur Seite gerichteter Sensor herangezogen. Die Aufnahmen wurden in verschiedenen Rangierszenarien erzeugt mit verschiedenen dynamischen Objekten in der Umgebung

	Kein Hintergrund	Hintergrund
Sensitivität	0.9390	0.7142
Spezifität	0.9038	0.9256
Präzision	0.8850	0.7534

Tabelle 6.5: Effizienz der Methode mit und ohne statischen Hintergrund

(Fußgänger, Fahrzeuge, Fahrräder). In Abbildung 6.22 sind beispielhafte Szenarien dargestellt.

Zur Bewertung werden die folgenden Kriterien definiert:

- Sensitivität (*true positive rate*): Die Sensitivität entspricht dem Anteil an positiven Messungen, die korrekterweise als positiv gemessen wurden.
- Spezifität (*true negative rate*): Die Spezifität entspricht dem Anteil an negativen Messungen, die korrekterweise als negativ gemessen wurden.
- Präzision: Die Präzision ist das Verhältnis zwischen den *true positives* und allen als positiv bestimmten Messungen.

Zur Bestimmung der optimalen Parameter werden ROC-Kurven aufgestellt. In Abbildung 6.23 ist beispielsweise die ROC-Kurve für den Parameter β aus Gleichung 6.7 dargestellt.

Die Ergebnisse der Evaluierung sind in Tabelle 6.5 aufgelistet. Es wird unterschieden zwischen Szenarien, in denen sich die dynamischen Objekte vor einem statischen Hintergrundobjekt bewegen, und Szenarien, in denen dies nicht der Fall ist. Es kann beobachtet werden, dass das Verfahren eine höhere Leistung erzielt, wenn kein Hintergrund vorhanden ist. In diesem Fall ist die Anzahl an *true positives* deutlich höher als *false positives*, sodass ein höherer Wert für die Sensitivität entsteht. Ein statischer Hintergrund führt dazu, dass eine höhere Anzahl an *false positives* bestimmt werden und somit die Sensitivität sinkt. Gleichzeitig steigt bei einer höheren Anzahl an statischen Objekten die Anzahl an *false positives*, sodass die Präzision sinkt.

In Abbildung 6.24 sind die Ergebnisse als Graphen in Abhängigkeit von der Integrationszeit und dem Einfluss von Fremdlicht dargestellt. Eine



a) Dynamische Objekte vor einem statischen Hintergrund



b) Erkennung von statischen Hindernissen

Abbildung 6.22: Beispielhafte Szenarien zur Evaluierung der Methode

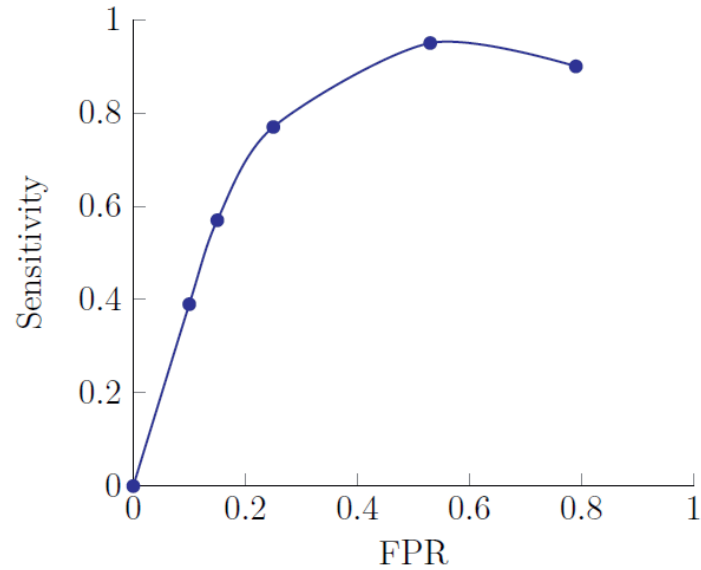
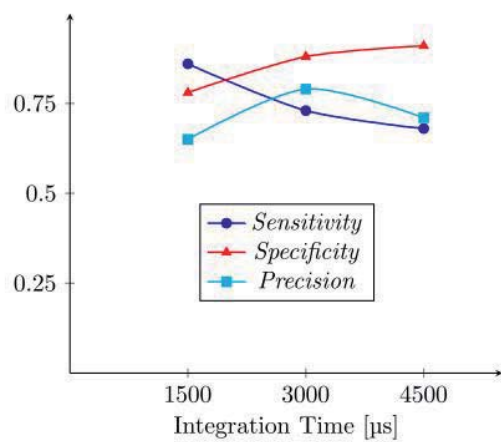
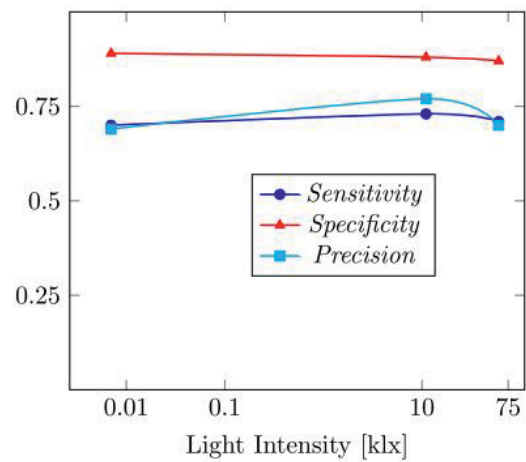


Abbildung 6.23: Die ROC-Kurve zur Bestimmung des optimalen Wertes für den Parameter β . Hierzu werden Sensitivität und Spezifität für verschiedene Werte von β gegeneinander aufgetragen.

geringe Integrationszeit führt dazu, dass eine geringere Anzahl an dynamischen Objekten erkannt wird, da sie einer geringeren Sichtbarkeit entspricht. Andererseits wird eine geringere Anzahl an statischen Objekten fälschlicherweise als dynamisch klassifiziert, sodass die Spezifität steigt. In Abbildung 6.24 (b) ist dargestellt, dass ein hoher Einfluss von Fremdlicht zu einem Sinken der Präzision führt, während die Sensitivität und die Spezifität auf einem gleichbleibenden Niveau verbleiben.



(a) Integrationszeit



(b) Fremdlicht

Abbildung 6.24: Ergebnisse in Abhängigkeit von der Integrationszeit und dem Einfluss von Fremdlicht



7 Evaluierung

Das entwickelte System wird anhand einer experimentellen Untersuchung evaluiert. Hierzu werden die Ergebnisse des Systems mit einer Referenz verglichen. Ein Fahrzeug rangiert innerhalb einer bekannten Modellumgebung. Ein Referenzsystem bestimmt die genauen Abmaße der Fahrzeugumgebung und die Pose des Fahrzeugs zu jedem Zeitpunkt. Sowohl das Umfeldmodell als auch die Tiefenbilder werden durch einen Vergleich mit der Referenz evaluiert. Das System wird hinsichtlich der Anforderungen zukünftiger Fahrerassistenzsysteme für das Parken und Rangieren bewertet.

7.1 Versuchsaufbau

Für den Versuch wurde eine Messhalle mit einem Referenzsystem zur Verfolgung des Fahrzeugs eingerichtet (siehe Abbildung 7.1). Das Fahrzeug bewegt sich dabei innerhalb einer Modellumgebung. Diese besteht aus verschiedenen Objekten und ist in Abbildung 7.2 dargestellt. Mehrere Stellwände formen längere Begrenzungen, die das Fahrzeug überragen. Einzelne Säulen bilden hohe Hindernisse. Auch kleinere Objekte, die vom Fahrer schnell übersehen werden, befinden sich in der Modellumgebung. Bordsteine werden durch längere Holzbalken simuliert.

Zur Erstellung der Referenzaufnahmen wird das Fahrzeug nun innerhalb dieser Modellumgebung rangiert. Dabei bewegt es sich sowohl vorwärts als auch rückwärts. Die Geschwindigkeiten variieren dabei zwischen -5 bis 7 km/h.

7.2 Referenzsystem

Das Referenzsystem besteht zum einen aus der hochgenauen Vermessung einer Referenzumgebung und zum anderen aus der Bestimmung der Fahr-



Abbildung 7.1: Die Messhalle, die mit einem Referenzsystem zur Bestimmung der Fahrzeugposition ausgestattet ist.



Abbildung 7.2: Die Modellumgebung besteht aus verschiedenen Stellwänden und Objekten



zeugtrajektorie. Im Nachfolgenden werden die Bestimmung des Referenzmodells und der Fahrzeugtrajektorie vorgestellt.

7.2.1 Referenzumgebung

Der Versuchsaufbau wird in Abschnitt 7.1 erläutert. Ein wesentlicher Bestandteil des Referenzsystems ist die hochgenaue Vermessung der Modellumgebung. Dies geschieht durch einen 3D-Laserscanner vom Typ *Z+F Imager 5010C* [158].

Eine Laserdiode sendet kurze, fokussierte Lichtpulse in die Umgebung aus. Der Lichtstrahl trifft auf ein Objekt und wird an der Objekt Oberfläche reflektiert. Das reflektierte Licht wird vom Laserscanner empfangen, und die Lichtlaufzeit wird bestimmt. Auf Basis der Lichtlaufzeit kann analog zu den Time-of-Flight-Kameras eine Distanz bestimmt werden.

Der ausgesandte Laserstrahl wird durch einen um eine Achse rotierenden Spiegel mehrfach umgelenkt, sodass der austretende Laserstrahl um 360° verschwenkt wird. Zusätzlich wird das Gehäuse um eine weitere Achse zu 360° rotiert, sodass eine vollständige dreidimensionale Vermessung ermöglicht wird.

Die Winkelauflösung des Laserscanners beträgt nach Datenblatt bis zu 100 000 Schritte in einer vollständigen Umdrehung [158]. Die Genauigkeit der Messung liegt bei unter 1 mm. Zu jeder Distanzmessung werden die zugehörigen Horizontal- und Vertikalwinkel gespeichert. Aus diesen Daten wird ein zugehöriger 3D-Punkt im Koordinatensystem des Scanners bestimmt. Auf diese Weise entstehen hochgenaue 3D-Punktwolken der Umgebung mit mehreren Millionen Messpunkten. Eine beispielhafte Messung zeigt Abbildung 7.3. Messungen aus verschiedenen Positionen des Laserscanners können durch eine Punktwolkenregistrierung zu einer gemeinsamen Punktwolke zusammengesetzt werden. Auf diese Weise können Lücken, die durch Verdeckung entstehen, geschlossen werden.

Die so generierte Punktwolke kann anschließend durch Triangulierungsverfahren in ein 3D-Mesh umgewandelt werden. Dies führt zu Vorteilen bei der Berechnung von Schnittpunkten zwischen dem Modell und beliebigen Linien.



Abbildung 7.3: 3D-Punktwolke der Modellumgebung aus der Vermessung durch einen 3D-Laserscanner

7.2.2 Referenztrajektorie

Zur Bestimmung der dynamischen Fahrzeugpose wurde das sogenannte *iGPS*-System von der Firma *Nikon* verwendet [159]. Es handelt sich hierbei um ein modulares Messsystem, das die Pose von Objekten im dreidimensionalen Raum bestimmen kann. In [160] wird beschrieben, dass es sich auch zur Verfolgung von sich bewegenden Objekten eignet.

Das *iGPS*-System besteht aus mehreren Laser-Transmittern und optischen Sensoren als Empfängern. Die Bestandteile des Systems sind in Abbildung 7.4 wiedergegeben. Jeder Transmitter sendet zwei verschiedene Signale aus. Zu Beginn jedes Zyklus wird ein kurzer Lichtpuls in die gesamte Umgebung abgestrahlt. Das zweite Signal stammt von einem rotierenden Gehäuse, aus dem zwei Laserdioden Lichtstreifen aussenden. Die Lichtstreifen besitzen eine Strahlweite von etwa 60° und sind so angeordnet, dass sie um etwa 30° gegeneinander verdreht sind. Jeder Transmitter besitzt eine individuelle Charakteristik. Die Lichtstreifen sind um unterschiedliche Winkel verdreht und jeder Transmitter besitzt eine eigene Rotationsgeschwindigkeit. Aus den bekannten geometrischen Eigenschaften können, wenn das Signal am Empfänger empfangen wurde, hochgenau Azimuth- und Elevationswinkel zwischen Empfänger und Transmitter be-



Abbildung 7.4: Empfänger (links) und Transmitter (rechts) eines *iGPS*-Systems [160]

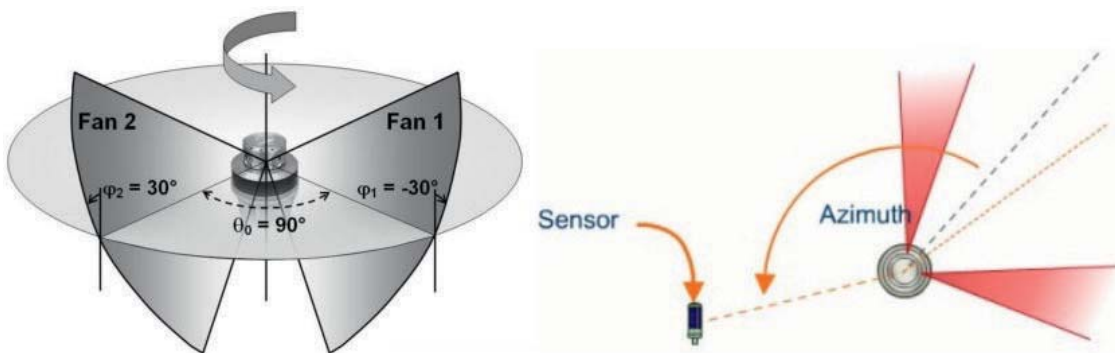


Abbildung 7.5: Bestimmung des Azimuth- und Elevationswinkels durch die Abstrahlung von zwei verdrehten Lichtstreifen [160]

stimmt werden. Dies ist in Abbildung 7.5 dargestellt. Die Distanz zum Transmitter bleibt zunächst unbekannt. Da das System allerdings aus mehreren Transmittern besteht, können die Winkel zu jedem Transmitter berechnet werden, und analog zum GPS-System kann dann eine hochgenaue Position des Empfängers berechnet werden. Werden mindestens zwei zueinander fest angebrachte Empfänger verwendet, ergibt sich eine 6D-Pose.

In der Messhalle, in der die Evaluierung durchgeführt wurde, befinden sich insgesamt 6 Transmitter. Zu jedem Zeitpunkt sollten mindestens zwei Transmitter vom Empfänger sichtbar sein. Vor der Verwendung muss das System kalibriert werden. Dies geschieht durch einen starren Stab, an dem mehrere Sender befestigt sind. Zudem wird während der Kalibrierung das Koordinatensystem bestimmt.



Abbildung 7.6: Befestigung der *iGPS*-Empfänger auf dem Dach des Fahrzeugs durch einen starren Körper

Das *iGPS* liefert Messungen mit einer Wiederholrate von 40 Hz. Die Genauigkeit des Systems beträgt laut Datenblatt auch bei sich bewegenden Empfängern < 1 mm und eignet sich daher als Referenz für die Fahrzeugbewegung. Auf dem Dach des Fahrzeugs wird ein starres Objekt mit zwei Empfängern befestigt (siehe Abbildung 7.6). Die Pose der Empfänger relativ zum Fahrzeugkoordinatensystem kann anschließend durch das *iGPS*-System bestimmt werden.

Um die einzelnen Aufnahmen der Kameras den gemessenen Posen des *iGPS*-System zuordnen zu können, müssen die *iGPS*-Messungen mit dem Fahrzeugrechner synchronisiert werden. Hierzu wird ein Modul eingesetzt, das in einem festen Takt mit 10 Hz ein Triggersignal aussendet. Zum einen startet das Signal eine Messung des *iGPS*-Systems, zum anderen wird eine CAN-Botschaft an das Fahrzeug gesendet. Diese Botschaft wird durch den Car-PC empfangen. Auf diese Weise wird die *iGPS*-Messung mit der Systemzeit des Car-PCs synchronisiert. Die Messung des *iGPS*-Systems wird zunächst vom Empfänger auf dem Fahrzeugdach drahtlos an den zentralen Messrechner gesendet und dort ausgewertet. Die berechnete Position wird anschließend über WLAN zum Fahrzeugrechner gesendet. Die Latenz zwischen Messung und Empfang beträgt etwa 300 bis 600 ms.

7.2.3 Referenziefenbilder

Das beschriebene Referenzsystem ermöglicht es, Referenziefenbilder der PMD-Kameras zu erstellen. Diese werden zum einen zur Bewertung der Tiefendaten herangezogen und zum anderen für die Tiefenkalibrierung verwendet, die in Kapitel 3.5.2 vorgestellt wird.

Voraussetzung für die Bestimmung der Referenziefenbilder ist eine erfolgte intrinsische und extrinsische Kalibrierung. Das Referenzsystem liefert sowohl eine hochgenaue Fahrzeugpose als auch eine hochgenaue Vermessung der Modellumgebung. Die gemessene Pose des Fahrzeugs wird hierbei relativ zum Weltkoordinatensystem angegeben. Da die Kameraposen relativ zum Fahrzeugkoordinatensystem durch die extrinsische Kalibrierung bekannt sind, kann die Pose einer Kamera durch eine Transformation relativ zum Weltkoordinatensystem bestimmt werden. Werden nun auch die durch die intrinsische Kalibrierung bekannten optischen Abbildungseigenschaften berücksichtigt, kann für jeden Pixel der Kamera ein Vektor bestimmt werden, der dem Sehstrahl des Pixels im Weltkoordinatensystem entspricht.

Zu jedem Pixel existiert nun mit dem Brennpunkt der Kamera ein Punkt und mit der Blickrichtung ein Vektor im Weltkoordinatensystem. Wird der entstehende Strahl nun mit dem 3D-Modell der Modellumgebung geschnitten, kann per *Raytracing* der Schnittpunkt zwischen Strahl und Modellumgebung bestimmt werden. Die Distanz zwischen diesem Schnittpunkt und dem Kamera-Ursprung entspricht der Referenzdistanz für den Tiefenwert des jeweiligen Pixels. Wird für jedes Pixel auf diese Weise die Referenzdistanz bestimmt, entsteht ein Referenziefenbild, das mit den realen Messungen verglichen werden kann. In Abbildung 7.7 ist ein reales Tiefenbild dem entsprechenden Referenziefenbild gegenübergestellt.

7.3 Evaluationskriterien

Das vorgestellte Referenzsystem erlaubt einerseits die Evaluierung der Tiefenbilder der Kameras und andererseits die Bewertung des entstehenden Umfeldmodells.

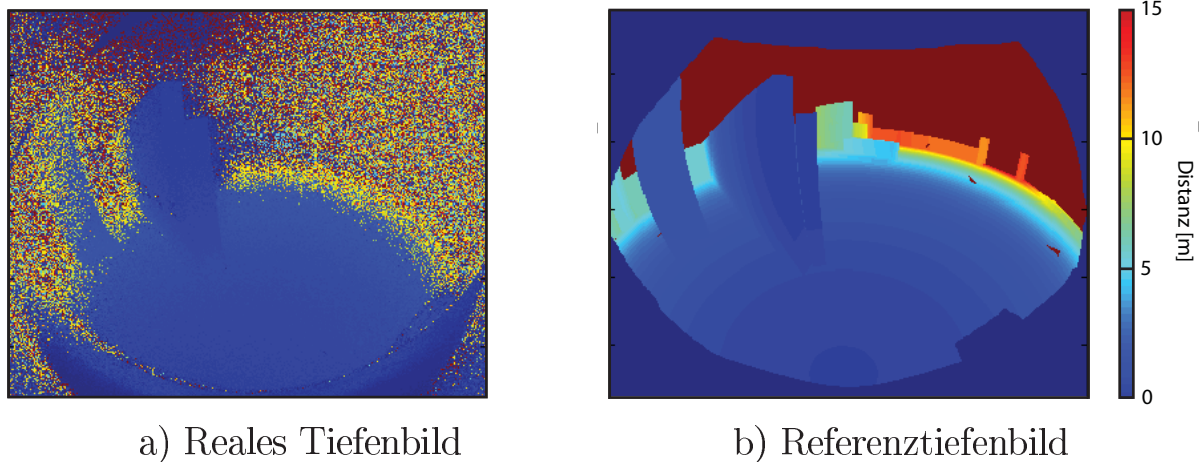


Abbildung 7.7: Gegenüberstellung eines realen Tiefenbildes (a) mit dem entsprechenden Referenz-tiefenbild (b)

7.3.1 Tiefenbilder

Zu jeder beliebigen Aufnahme einer der Kameras kann, wie in Abschnitt 7.2.3 erläutert, das zugehörige Referenz-tiefenbild generiert werden. Die Evaluierung der Tiefenbilder basiert auf einem Vergleich der realen Aufnahmen mit der *Ground Truth*. Dabei wird der absolute Fehler in der Distanzmessung bestimmt. Dieser soll zunächst für jede Kamera separat, abhängig von der Pixelposition bestimmt werden. Auf diese Weise erhält man für jede Kamera ein Fehlerbild, das für jeden Pixel den durchschnittlichen absoluten Fehler enthält. Die Tiefenbilder werden bewusst nicht anhand der zu evaluierenden Aufnahmen kalibriert. Es ist zu erwarten, dass die bestimmten Fehlerwerte sehr hoch ausfallen, da der Fehler in diesem Falle unabhängig von der Amplitude betrachtet wird. Dennoch bildet die Verteilung des Fehlers innerhalb eines Bildes eine wertvolle Information.

Die Höhe des durchschnittlichen absoluten Fehlers kann anschließend abhängig von der Position in der Umgebung des Fahrzeugs betrachtet werden. Dies unterstützt die Bewertung der erfolgten Systemauslegung. Zur Berücksichtigung der Amplitude soll der Zusammenhang zwischen Amplitude und absolutem Fehler bestimmt werden. Hierzu werden die Fehlerwerte aller Kameras und aller Pixel zusammengefasst, um einen möglichst allgemeingültigen Zusammenhang zu bestimmen.



7.3.2 Umfeldmodell

Das Umfeldmodell wird mit dem aus dem 3D-Laserscan erstellten Referenzmodell verglichen. Hierzu wird die gesamte Datenstruktur des Umfeldmodells zu beliebigen Zeitpunkten gespeichert. Nach einer Transformation des Referenzmodells in das Koordinatensystem des Umfeldmodells kann das vom System erstellte Umfeldmodell mit der Referenz verglichen werden.

Hierzu werden die Fehler in der Höhenbestimmung der einzelnen Gitterzellen bestimmt. Dabei wird der absolute Fehler abhängig von der Position in der Fahrzeugumgebung bestimmt. Dadurch ist es möglich, die Qualität des Umfeldmodells abhängig von den verschiedenen Bereichen um das Fahrzeug zu bewerten.

Zudem werden die Fehler in der Bestimmung der Distanzen zu Objekten berechnet. Für Fahrerassistenzsysteme ist in erster Linie die Distanz zu Hindernissen relevant. Hierzu werden Distanzen zu Objekten in der Umgebung anhand des Umfeldmodells und des Referenzmodells bestimmt. Der durchschnittliche Fehler in der Distanzbestimmung wird anschließend abhängig von der wahren Distanz und der Position des Hindernisses relativ zum Fahrzeug ermittelt.

7.4 Ergebnisse

Nachfolgend werden die Ergebnisse der Evaluierung vorgestellt. Zunächst wird erläutert, inwiefern dem entwickelten Referenzsystem vertraut werden kann. Anschließend werden die Tiefenbilder und das Umfeldmodell nach den im vorherigen Kapitel definierten Evaluierungskriterien bewertet.

7.4.1 Referenzsystem

Das für die Evaluierung entwickelte Referenzsystem wurde in Abschnitt 7.2 erläutert. Es basiert auf mehreren hochgenauen Messsystemen, die zu einem gemeinsamen Referenzsystem zusammengefügt werden. Insbesondere bei der Berechnung der Referenztiefenbilder können verschiedene

Faktoren zu Ungenauigkeiten in der Erstellung der *Ground Truth* führen. Für die Berechnung der Referenztiefenbilder wird auf die in Kapitel 2.1.4 vorgestellte Kamerakalibrierung nach [43] zurückgegriffen. Die resultierenden optischen Abbildungseigenschaften weisen dabei einen geringen Fehler auf, der durch Rückprojektion auf die Sensorebene bestimmt werden kann. Im Falle der eingesetzten Kameras liegt dieser Wert bei unter einem Pixel. Im Allgemeinen führt dies zu einem sehr geringen Fehler. Lediglich an Objektkanten kann es vorkommen, dass der zum Pixel zugehörige Sichtstrahl ein anderes Objekt trifft und somit eine fehlerhafte Distanz bestimmt wird.

Eine weitere Fehlerquelle liegt in einer ungenauen extrinsischen Kalibrierung der Kameras. Der Fehler bei der Bestimmung der Kameraposen besitzt einen ähnlichen Wert wie der Fehler in der Bestimmung der intrinsischen Kameraparameter, da dieselbe Methode genutzt wird. Allerdings kann sich die Position und Ausrichtung der einzelnen Kameras minimal während des Betriebs aufgrund von Materialeigenschaften der Fahrzeugbauteile verändern. Zusätzlich liegt eine weitere Fehlerquelle in der Fahrwerkseinstellung des Fahrzeugs. Die Höhe der Kameras kann daher in einem Bereich von mehreren Millimetern schwanken.

Einen vergleichbaren Effekt hat die Kalibrierung der *iGPS*-Empfänger, die auf dem Fahrzeugdach befestigt werden. Zur Kalibrierung wird die relative Pose der Empfänger zum Ursprung des Fahrzeugkoordinatensystems ermittelt. Die Genauigkeit liegt dabei maximal im Bereich der Messgenauigkeit des *iGPS*-Systems und wird zusätzlich von weiteren Unsicherheiten beeinflusst. Hierzu gehört insbesondere die manuelle Ausrichtung eines Messstabes auf die auf den Boden projizierte Lage der Hinterachse. Hierbei können Fehler im Bereich von mehreren Millimetern auftreten.

Die Messgenauigkeit des *iGPS*-Systems wird mit unter einem Millimeter angegeben. Dies konnte sich durch die durchgeführten Messungen allerdings nicht bestätigen lassen. Teilweise treten sehr starke Ausreißer auf, die eine Abweichung von bis zu 20 cm aufweisen. Dies ist auf Fehlereffekte des *iGPS*-Systems zurückzuführen, wie beispielsweise auftretende *Multipath*-Effekte. Diese Effekte können allerdings als Ausreißer erkannt werden und als ungültig deklariert werden. Nach einer Ausreißerentfernung treten dennoch Fehler im Bereich von bis zu 2 cm auf. Eine Analyse

der Positionsmessungen ergibt eine Standardabweichung von 4-7 mm. Es wird vermutet, dass die Höhe der Standardabweichung von der Anzahl an sichtbaren Transmittern abhängt. Zudem sind die Transmitter teilweise an den Wänden der Messhalle befestigt. Da sich diese im Laufe eines Tages aufgrund von Temperaturveränderungen ausdehnen kann, verschieben sich die Positionen der Transmitter um mehrere Millimeter. Auch dies kann zu einer zusätzlichen Ungenauigkeit führen.

Eine weitere Quelle für einen Fehler in den Referenzdaten liegt in der Synchronisierung der *iGPS*-Messdaten mit den Bilddaten der Kameras. Zur Synchronisierung wird gleichzeitig eine CAN-Botschaft an das Fahrzeug und ein Trigger-Signal an das *iGPS*-System gesendet. Somit erhält die erfolgte Positionsmessung einen Zeitstempel im Fahrzeugrechner. Die Genauigkeit des Zeitstempels kann mit einer Latenz von bis zu 0,3 ms angegeben werden. Bei den niedrigen Geschwindigkeiten des Rangierens kann hierdurch lediglich ein Positionsfehler von bis zu 0,5 mm verursacht werden. Allerdings besteht eine weitere unbekannte Latenz zwischen den Aufnahmen der Kamerabilder und ihrer Zeitstempelung im Fahrzeugrechner. Die Zeitstempelung erfolgt, sobald die Bilddaten vom Fahrzeugrechner empfangen werden. Dies führt zu einer weiteren Störung der Synchronisierung. Die Latenz wird im Bereich weniger Millisekunden angenommen, sodass sie, abhängig von der Geschwindigkeit des Fahrzeugs, zu einem Fehler von wenigen Millimetern führt.

Die Vermessung des 3D-Modells anhand eines 3D-Laserscanners und die Erstellung des Flächenmodells aus den verschiedenen Punktwolken ist ebenfalls fehlerbehaftet. Die tatsächliche Genauigkeit des 3D-Modells konnte nicht bestimmt werden. Allerdings kann das 3D-Modell durch einen Abgleich von Messpunkten mit den Positionsmessungen des *iGPS*-Systems verifiziert werden. Dies ergab eine maximale Abweichung von etwa 20 mm zwischen den Punkten des 3D-Modells und den *iGPS*-Messungen. Dabei wird angenommen, dass die Genauigkeit des 3D-Modells höher ist als die Genauigkeit des *iGPS*-Systems.

Zusammenfassend kann angenommen werden, dass sowohl in den Referenztiefenbildern als auch in der Referenzposition Fehler im Bereich von wenigen Zentimetern auftreten können. Dies sollte in der Bewertung der Tiefenbilder und des erstellten Umfeldmodells berücksichtigt werden.

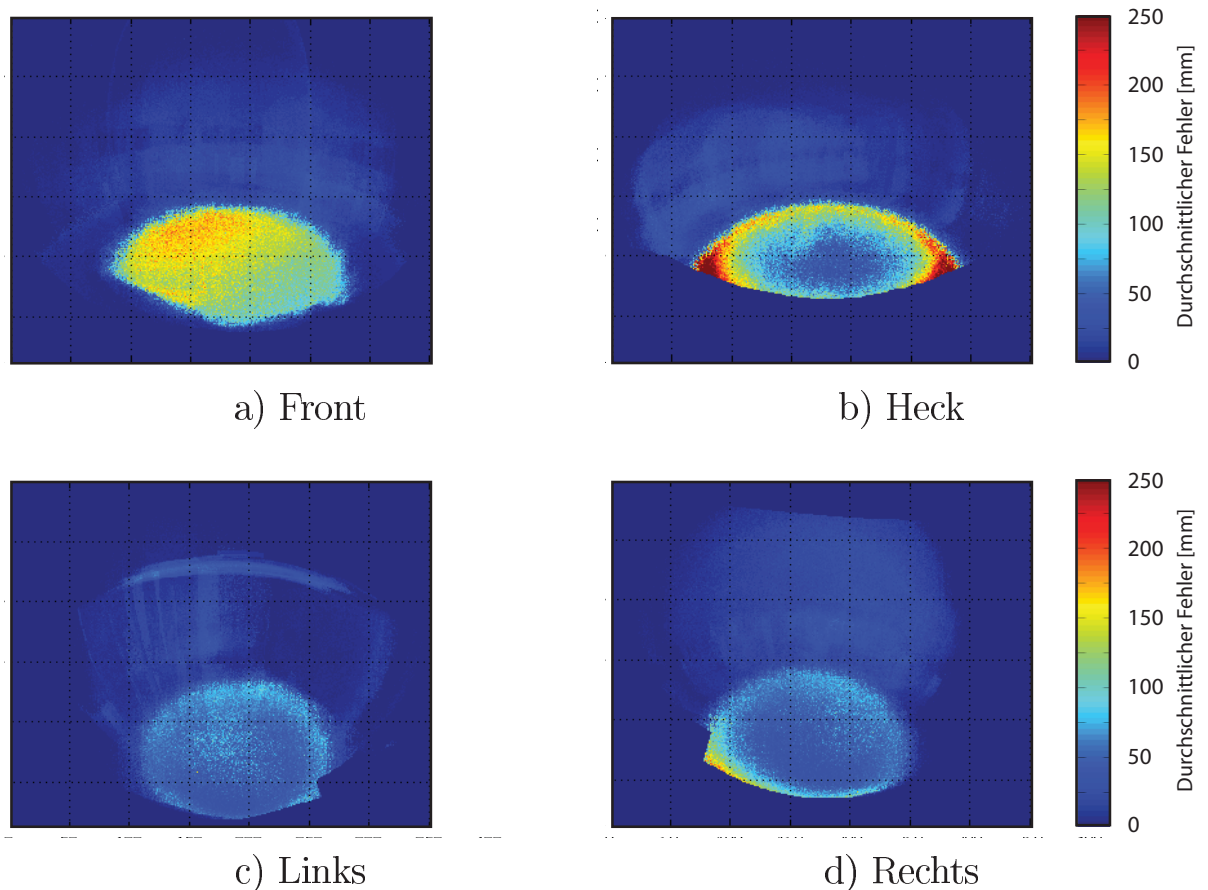


Abbildung 7.8: Die durchschnittlichen Fehler in den Tiefenwerten der einzelnen Pixel aller vier Kameras

7.4.2 Tiefenbilder

Aus dem Vergleich zwischen Referenztiefenbildern und den realen Tiefenbildern können für jedes Pixel einzeln Fehlerwerte bestimmt werden. Die Paare aus realen Tiefenbildern und Referenztiefenbildern wurden zufällig an unterschiedlichen Positionen der gesamten Versuchsfahrten innerhalb der Referenzumgebung gebildet. Es werden nur die Pixel betrachtet, die einen Mindestwert für die Amplitude überschreiten. Der gewählte Wert beeinflusst dabei die Ergebnisse stark, da geringe Amplituden in der Regel zu höheren Fehlerwerten führen. In den in Abbildung 7.8 dargestellten Bildern wurde eine Mindestamplitude von 500 gewählt.

Es wird deutlich, dass die Fehlerwerte der auf den Boden zeigenden Pixel in der Frontkamera deutlich höher ausfallen, als in den beiden Seitenkame-

ras. An der Front werden durchschnittliche Fehler im Bereich zwischen 120 und 170 mm ermittelt, während bei den Seitenkameras Fehler im Bereich zwischen 20 und 80 mm auftreten. Im Vergleich zu den Seitenkameras ist die Frontkamera sehr niedrig angebracht. Daher haben Änderungen im Richtungsvektor eines Pixels aufgrund des geringen Winkels zwischen Boden und Sichtstrahl eine starke Auswirkung auf die Distanz zum Boden. Dieser Effekt fällt bei den Kameras in den Außenspiegeln geringer aus. Daher spricht der hohe Fehler der Frontkamera für einen Fehler in der Neigung des Fahrzeugs. Schon ein geringer Orientierungsfehler in der Referenz kann zu hohen Fehlern der Frontkamera bei der Vermessung des Bodens führen, während sich der Fehler auf die Vermessung des Bodens durch die Seitenkameras weitaus geringer auswirkt.

In der Aufnahme der Heckkamera in Abbildung 7.8 (b) ist ein weiterer Effekt bei der Vermessung des Bodens zu beobachten. Die Lichtquellen der Kameras erzeugen einen Leuchtkegel, der den Boden beleuchtet. In der Mitte des beleuchteten Bereichs treten nur geringe Fehler im Bereich von etwa 40 mm auf, während der Fehler zum Rand des Lichtkegels hin zunimmt auf einen Wert von bis zu 250 mm. Dies ist durch den in Kapitel 2.1.3 erläuterten *Scattering*-Effekt zu erklären. Die Kamera im Heck des Fahrzeugs beleuchtet große Bereich des eigenen Fahrzeuges, unter anderem das Nummernschild und die Heckschürze. Der weiße Lack des Versuchsträgers führt dazu, dass eine große Menge Licht an der Oberfläche des Fahrzeugs reflektiert wird und zu einem *Scattering* führt, das die Tiefenmessungen stört. Die Stärke der Beeinflussung hängt dabei von der Amplitude eines Pixels ab. Die Pixel im Außenbereich des Leuchtkegels weisen eine geringe Amplitude auf und werden daher stark durch das *Scattering* beeinflusst, sodass die hohen Fehler im Randbereich entstehen.

Es wird deutlich, dass hohe Fehler bei der Vermessung des Bodens auftreten, die im Falle der Frontkamera durch Fehler des Referenzsystems hervorgerufen werden. Das Ziel der Umfeldwahrnehmung ist es, Hindernisse in der Umgebung des Fahrzeugs zu erkennen. Daher wird die Auswertung der Tiefenbilder ohne Berücksichtigung der Punkte durchgeführt, die sich nahe der Bodenebene befinden. Die Ergebnisse sind in Abbildung 7.9 dargestellt. Der durchschnittliche Tiefenfehler in der Messung von Objekten in der Umgebung liegt zwischen 0 und 40 mm. Dies spricht für die Tatsache, dass das Referenzsystem einen geringen Fehler im Neigungswinkel

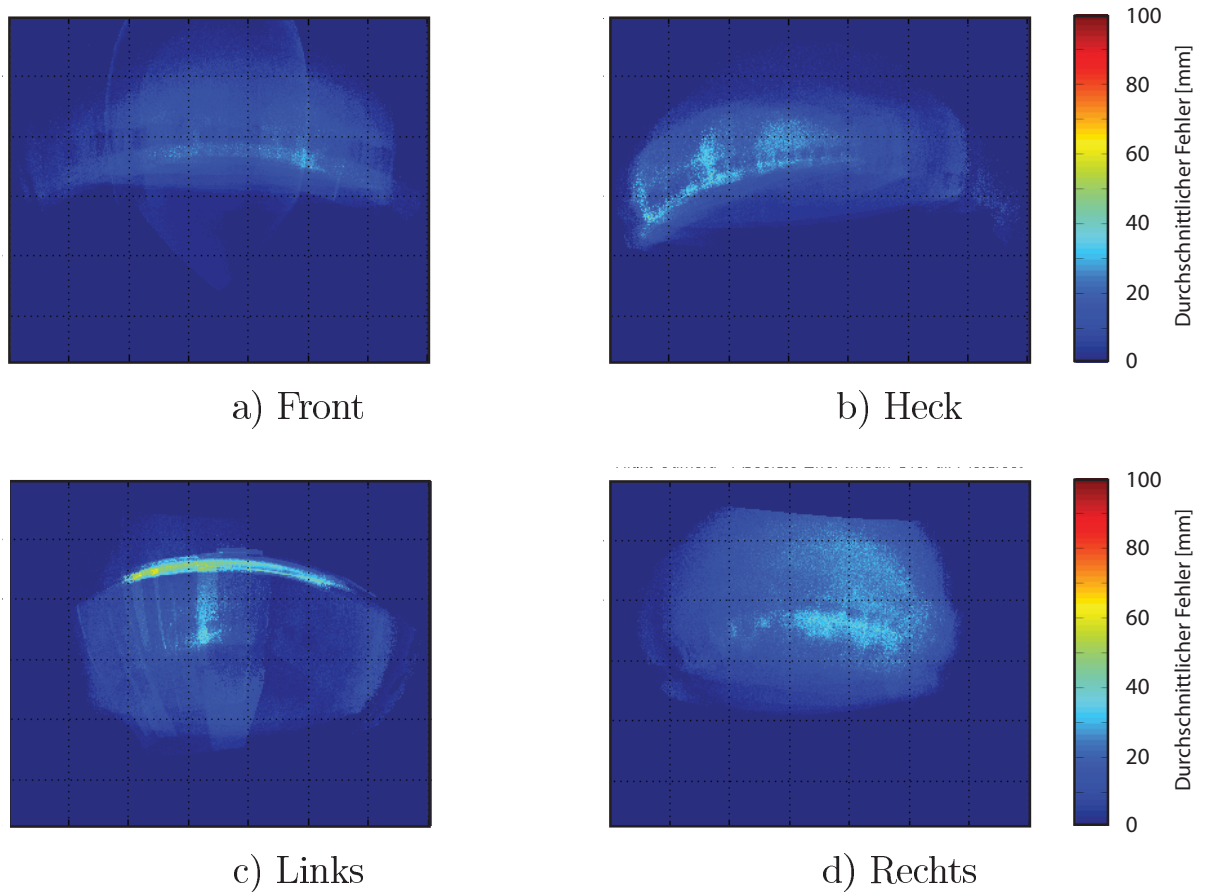


Abbildung 7.9: Die durchschnittlichen Fehler in den Tiefenwerten der einzelnen Pixel aller vier Kameras ohne Berücksichtigung von Tiefenwerten, die den Boden vermessen

des Fahrzeugs bestimmt. Die Vermessung von Objekten wird hierdurch nur wenig beeinflusst.

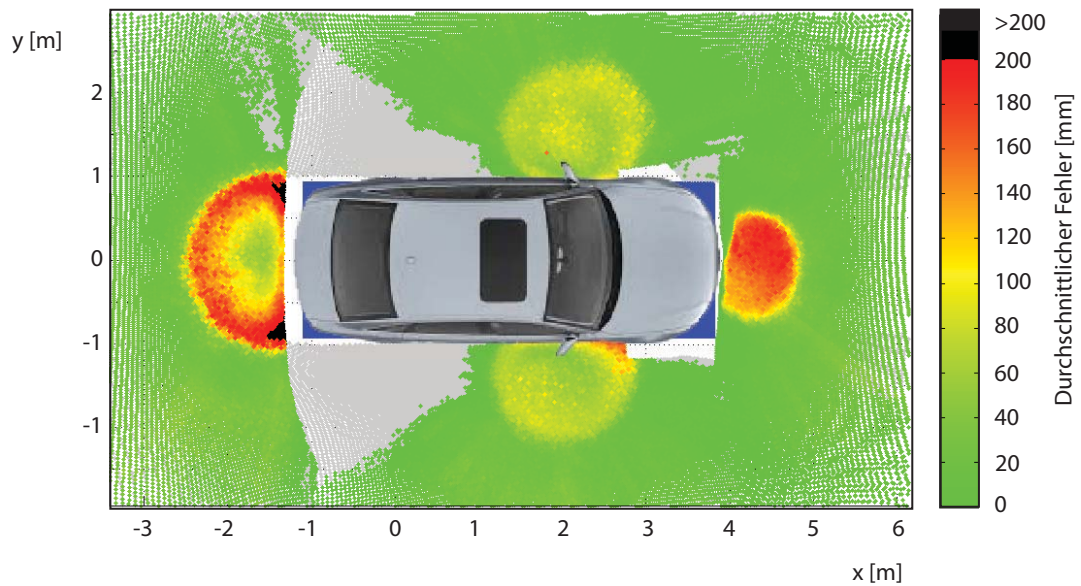
Die dargestellten, durchschnittlichen Fehler der Tiefenwerte der einzelnen Pixel werden in Abbildung 7.10 (a) als Ansicht aus der Vogelperspektive dargestellt. Die Bilder der vier Kameras werden zusammengefügt und durch eine Projektion der einzelnen Pixel auf den Boden abgebildet. Der Wert eines Pixels wird dabei an die Stelle im Bild eingetragen, die dem Auftreffpunkt des Sichtstrahls des Pixels auf den Boden entspricht. In der Umgebung der Kameras befinden sich jeweils die Pixel, die den Boden beleuchten. Diese weisen, wie zuvor erläutert, hohe Fehler auf. In den umliegenden Bereichen treten nur geringe Fehler auf, da hier die Messungen der Hindernisse ausschlaggebend sind. Hierbei ist allerdings zu beachten, dass die Objekte sich nicht an den im Bild eingetragenen Positionen befunden haben müssen, sondern nur der korrespondierende Pixel ein Hindernis vermessen hat, das sich zwischen Kamera und Auftreffpunkt befindet. Auch wird deutlich, dass die Kameras die hinteren Bereiche an den Flanken des Fahrzeugs nicht ausreichend abdecken.

Analog stellt Abbildung 7.10 (b) den durchschnittlichen Fehler in der Tiefenmessung aus der Vogelperspektive ohne Berücksichtigung der Bodenebene dar. Wie schon bei der Analyse der Tiefenbilder treten deutlich geringere Fehler bei der Vermessung von Hindernissen auf, als bei der Vermessung des Bodens.

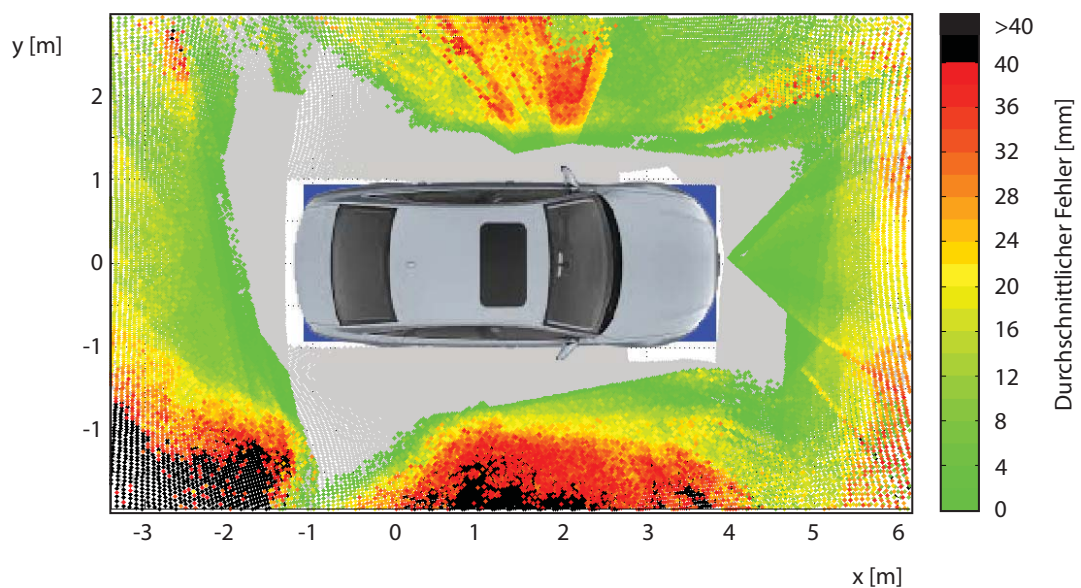
7.4.3 Umfeldmodell

Das erstellte Umfeldmodell wird durch einen Vergleich mit dem Referenzmodell bewertet. Hierzu werden die Abweichungen der ermittelten Höhenwerte zu den *Ground-Truth*-Höhenwerten gebildet. Dabei wird zu verschiedenen Zeitpunkten das bis zu diesem Zeitpunkt erstellte Umfeldmodell mit der Referenz in Bezug gesetzt. Für die möglichen Positionen im Umfeld des Fahrzeugs werden die durchschnittlichen Abweichungen der Höhenwerte berechnet. Das Ergebnis ist in Abbildung 7.11 dargestellt.

Die Darstellung gibt einen Überblick über die Güte der Vermessung der verschiedenen Bereiche in der Umgebung des Fahrzeugs. Es wird deutlich, dass in der Umgebung der Sensoren sehr geringe Abweichungen gemessen werden. In der Nähe der Sensoren beträgt der durchschnittliche Höhenfehler nur wenige Zentimeter. Die größten Fehler ergeben sich an den hinteren



a) Berücksichtigung des Bodens



b) Keine Berücksichtigung des Bodens

Abbildung 7.10: Darstellung der durchschnittlichen Fehler in der Tiefenmessung aus der Vogelperspektive durch die Projektion der einzelnen Pixel aller Kameras auf die Bodenebene

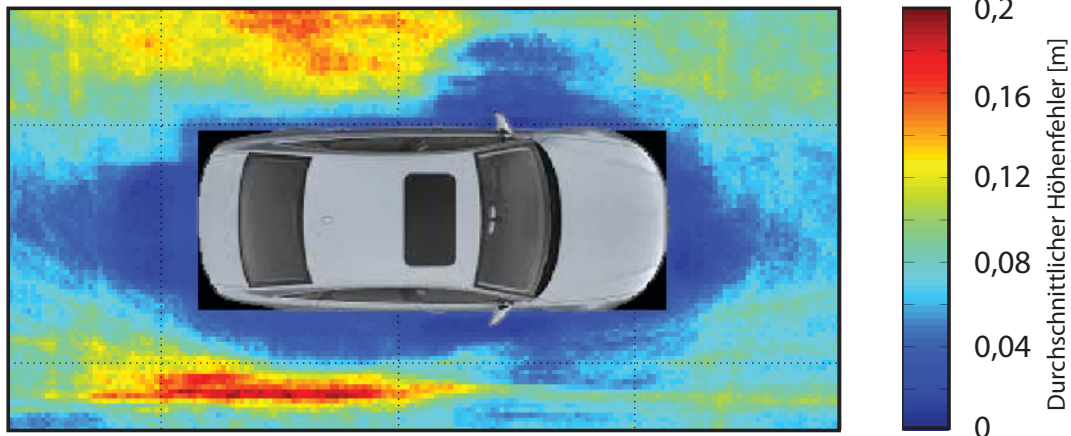


Abbildung 7.11: Die durchschnittlichen Abweichungen zwischen den ermittelten Höhenwerten des Umfeldmodells und der *Ground Truth* abhängig von der Position relativ zum Fahrzeug

Bereichen der Flanken des Fahrzeugs. Hier beträgt der durchschnittliche Höhenfehler in Fahrzeugnähe bis zu 16 cm. Wie schon aus der Analyse der Tiefenbilder im vorherigen Abschnitt deutlich wird, besitzen die Sensoren in diesen Bereichen nur eine geringe Abdeckung. Auch an den Fahrzeugecken treten höhere Abweichungen auf als unmittelbar vor den Kameras. Dies ist darauf zurückzuführen, dass sich diese Bereiche im Kamerabild an den Rändern befinden und durch die Beleuchtung nicht gleichmäßig ausgeleuchtet werden.

Der Höhenfehler berücksichtigt nicht die absolute Höhe von Objekten. Beispielsweise besteht keine Gefahr, wenn ein 50 cm hohes Hindernis mit einer Höhe von 70 cm vermessen wird. Andererseits wäre es ein Problem, wenn ein Hindernis nicht als solches erkannt oder eine fehlerhafte Distanz zum Hindernis bestimmt würde. Daher werden Objekte aus den verschiedenen Umfeldmodellen, die zu unterschiedlichen Zeitpunkten gebildet wurden, extrahiert. Die Distanz der auf diese Weise ermittelten Objekte zum Fahrzeug wird bestimmt und mit der wahren Distanz aus der *Ground Truth* verglichen. Das Umfeld des Fahrzeugs wird in verschiedene Bereiche eingeteilt, und für jeden Bereich wird der durchschnittliche Fehler zwischen der aus dem Umfeldmodell ermittelten Distanz und der *Ground-Truth*-Distanz bestimmt. Die resultierende Darstellung ist in Abbildung 7.12 gegeben.

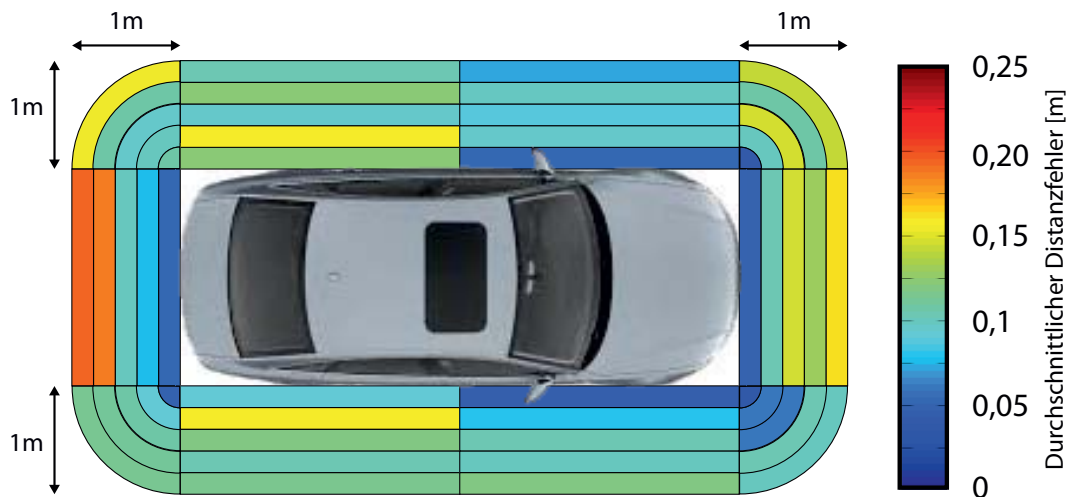


Abbildung 7.12: Die durchschnittlichen Abweichungen zwischen den ermittelten Distanzen zu Objekten und den *Ground-Truth*-Distanzen für verschiedene Bereiche im Fahrzeugumfeld

An den Bereichen der beiden Seitensensoren werden sehr geringe durchschnittliche Fehler bestimmt. In den Bereichen bis zu 80 cm Entfernung wird eine durchschnittliche Abweichung von maximal 11 cm ermittelt. Anders verhält es sich bei der Frontkamera. Hier werden im Nahbereich des Fahrzeugs durchschnittliche Fehler von unter 5 cm ermittelt. Allerdings nimmt der Fehler mit zunehmender Entfernung auf bis zu 16 cm zu. In den Bereichen an den Fahrzeugecken treten bei einer Entfernung von bis zu 40 cm durchschnittliche Fehler von maximal 10 cm und bei Entfernungen über 40 cm Fehler von bis zu 16 cm auf. Die Heckkamera liefert im Nahbereich durchschnittliche Fehler von unter 5 cm und im Bereich bis etwa 60 cm durchschnittliche Fehler von unter 11 cm. Zwischen 60 und 100 cm Entfernung steigt der Fehler allerdings auf bis zu 19 cm. Die Analyse der Tiefendaten hat ergeben, dass bei der Heckkamera starkes *Scattering* auftritt. In höheren Entfernungen werden vergleichsweise niedrige Amplituden gemessen, sodass diese Messungen vom *Scattering* beeinflusst werden. An der Front, an den vorderen Seitenbereichen und am Heck werden im Bereich bis zu 40 cm Genauigkeiten von etwa 10 cm erreicht. Dies ist nicht der Fall für die hinteren Seitenbereiche. Hier ist kein deutlicher Verlauf der Messgenauigkeit abhängig von der Distanz zu beobachten. Dies ist darauf zurückzuführen, dass sich keine Kamera in unmittelbarer Nähe befindet,

sondern der gesamte Bereich von den Kameras in den Seitenspiegeln vermessen wird.

7.5 Diskussion

Das in Abschnitt 7.2 vorgestellte Referenzsystem erlaubt eine Evaluierung sowohl der Tiefenbilder der einzelnen Kameras als auch des aufgebauten Umfeldmodells. Bei der Evaluierung der Tiefenbilder wurde zunächst die Vermessung des Bodens zur Bestimmung von durchschnittlichen Fehlern berücksichtigt. Hier ergab sich für die Seitenkameras eine geringe Abweichung, während bei der Frontkamera sehr hohe Fehlerwerte auftraten. Diese hohen Fehler konnten durch einen Fehler in der Bestimmung der Fahrzeugneigung durch das Referenzsystem erklärt werden. Wenn bei der Bestimmung der Tiefenfehler nur Hindernisse berücksichtigt werden, sind die durchschnittlichen Fehler bei allen Kameras deutlich geringer. Auch bei der Bewertung der Höhenwerte des Umfeldmodells konnten im Frontbereich nur sehr geringe Fehler ermittelt werden.

In den Bildern der hinteren Kamera sind an den Randbereichen hohe Fehlerwerte zu beobachten. Diese treten aufgrund des in Kapitel 2.1.3 erläuterten *Scattering*-Effekts auf. Da eine Korrektur des Effekts nicht problemlos möglich ist, wird das Auftreten von *Scattering* in dieser Arbeit in einem per-Pixel-Konfidenzwert berücksichtigt. Die geringen Abweichungen bei der Bestimmung des Höhenfehlers im Umfeldmodell zeigen, dass die Fehlereffekte in den Tiefenbildern durch die Bestimmung eines Konfidenzwertes und durch eine gewichtete Akkumulation der gemessenen 3D-Punkte in den Gitterzellen erfolgreich kompensiert werden können. Andererseits treten dennoch bei der Bestimmung der Distanzen zu Objekten in höheren Entfernungen aufgrund der niedrigen Amplitudenwerte vergleichsweise hohe Abweichungen auf. Der *Scattering*-Effekt, der durch Reflexionen am eigenen Fahrzeug auftritt, kann durch einen geschickten Einbau der Kameras und der Beleuchtungsmodule verhindert werden. Dies ist in jedem Fall für weitere Aufbauten zu empfehlen.

Die Bestimmung der durchschnittlichen Abweichungen zwischen den ermittelten Höhenwerten des Umfeldmodells und der *Ground Truth* konnte aufzeigen, welche Bereiche der Fahrzeugumgebung das Kamerasystem zuverlässig abdeckt. Hier konnte gezeigt werden, dass die Abdeckung an



den hinteren Bereichen der Fahrzeugflanken zu gering für eine exakte Vermessung der Fahrbahnoberfläche ist. Dies wurde auch schon bei der Systemauslegung in Kapitel 3.3.3 durch eine Simulation der Ausleuchtung vorhergesagt. Hier wurde vorgeschlagen, fokussierte und homogene Lichtquellen zu verwenden, um die Lichtverteilung beliebig zu definieren. Die hinteren Bereiche der Fahrzeugflanken sind jedoch die am wenigsten kritischen Bereiche im Umfeld des Fahrzeugs, da statische Objekte, die in diese Bereiche gelangen, sich zunächst in einem anderen Bereich befinden, der die Objekte messen kann. Dies liegt daran, dass das Fahrzeug durch die hintere Achse nicht lenken kann und das Heck des Fahrzeugs sich somit nicht zur Seite bewegt. Da die Umfeldmodellierung durch den Aufbau einer Karte umgesetzt wird, sind Objekte, die zunächst sichtbar sind und anschließend in einen wenig abgedeckten Bereich gelangen, dennoch dem System bekannt. Die durchschnittlichen Abweichungen zwischen den ermittelten Distanzen zu Objekten und den wahren Distanzen fallen daher auch für die hinteren Bereiche der Fahrzeugflanken nicht deutlich höher aus, als in den gut abgedeckten Bereichen. Für automatisierte Systeme kann dies allerdings dennoch ein Problem darstellen, wenn sich dynamische Hindernisse in der Umgebung befinden. Ein dynamisches Objekt könnte sich über den hinteren Seitenbereich dem Fahrzeug nähern. Daher wird für automatisierte Systeme empfohlen, die Systemauslegung gemäß der Vorschläge aus Kapitel 3.3.3 anzupassen. Im Allgemeinen führt eine Erhöhung der Lichtleistung ebenso zu einer verbesserten Abdeckung.

Auch die Bereiche an den vorderen Fahrzeugecken weisen eine geringe Abdeckung auf. Verglichen mit den hinteren Bereichen der Fahrzeugflanken treten hier geringere Messfehler auf. Allerdings sind diese Bereiche für die Erkennung von Hindernissen als kritischer zu bewerten, da sich Objekte bei einem starken Lenkwinkelschlag auf direktem Kollisionsweg zum Fahrzeug befinden können. Daher ist für die Wahrnehmung dieser Bereiche ebenfalls eine Anpassung der Lichtquellen gemäß Kapitel 3.3.3 zu empfehlen. Zwar konnte während der experimentellen Versuche keine Situation hervorgerufen werden, in der ein Objekt, das sich der Fahrzeugecke näherte, nicht erkannt werden konnte. Dennoch könnte dies ein Problem sein für kleine Objekte, die eine sehr niedrige Reflektivität besitzen.

Mit Ausnahme der hinteren Fahrzeugflanken konnte für alle Bereiche in der Umgebung des Fahrzeugs eine durchschnittliche Abweichung zwischen ermittelter und wahrer Distanz zu Objekten zwischen 4 und 11 cm bei ei-

ner Entfernung von bis zu 40 cm ermittelt werden. Im Bereich von bis zu 100 cm liegt die Genauigkeit im schlimmsten Fall bei 19 cm. Da die Genauigkeit bei Annäherung an das Fahrzeug steigt, reicht dies für eine zuverlässige Bestimmung von Abständen zu Hindernissen aus.

7.6 Eignung für zukünftige Fahrerassistenzfunktionen

In Kapitel 1.4 werden die Anforderungen an die Umfeldwahrnehmungen für zukünftige Fahrerassistenzfunktionen beim Parken und Rangieren erläutert. Auf Basis der Evaluierung des Gesamtsystems werden diese zur Bewertung des Systems wieder aufgegriffen.

Die in Abschnitt 7.4 vorgestellte Evaluierung ergibt, dass eine **360°-Abdeckung** durch die PMD-Kameras prinzipiell gegeben ist. Im sehr nahen Bereich des Fahrzeugs werden Objekte je nach Position mit einer Genauigkeit zwischen 4 und 11 cm erkannt. Dennoch ist zu beobachten, dass die Genauigkeit und Reichweite an den Fahrzeugecken und an den hinteren Bereichen der Fahrzeugflanken abnimmt. Dabei sind die vorderen Seitenbereiche als kritischer zu betrachten als die hinteren, da der vordere Bereich beim rückwärtsfahren ausscheren kann. Auf der anderen Seite konnte gezeigt werden, dass der Aufbau des Umfeldmodells verhindert, dass Hindernisse in den kritischen Bereichen gar nicht erkannt werden. In Kapitel 3.3.2 wird gezeigt, dass die Verwendung von Lichtquellen mit geringerem Öffnungswinkel zu einer höheren Abdeckung der Fahrzeugecken und an den hinteren Bereichen der Fahrzeugflanken führen kann. Für eine vollständige 360°-Abdeckung sollte daher die Systemauslegung gemäß Kapitel 3.3.3 angepasst werden.

Die **Reichweite** hängt stark von der Position und der Reflektivität eines Hindernisses ab. Aus diesem Grund kann für das vorgestellte System keine allgemeingültige maximale Reichweite angegeben werden. Für eine Reichweite von 1 m um das Fahrzeug herum wurde die Genauigkeit bei der Bestimmung der Distanzen zu Objekten ermittelt. Diese liegt im Nahbereich im günstigsten Fall bei 4 cm und im schlechtesten Fall bei 19 cm. Objekte in einer größeren Entfernung werden im Allgemeinen mit einer geringeren Genauigkeit wahrgenommen.



Eine weitere Anforderung ist die **Robustheit gegenüber Störeffekten**. Die eingesetzten PMD-Kameras liefern aufgrund des aktiven Messprinzips auch in absoluter Dunkelheit Tiefenmessungen. Niederschlag wie Regen oder Schnee verursachen nur leichte Störungen. Die größte Schwäche des Systems ist die Anfälligkeit gegenüber Sonnenlicht. Zwar besitzen die Pixel der PMD-Kamera eine Fremdlichtunterdrückung, allerdings führt diese dazu, dass das Rauschen in den Tiefenmessungen signifikant zunimmt. Für einen Einsatz im Automobil sollte die Fremdlichtunterdrückung ausgebaut werden. Dies kann beispielsweise durch den Einsatz von Laserdioden als Beleuchtung oder durch eine Erweiterung der Schaltung für die Fremdlichtunterdrückung auf den einzelnen Pixeln des Sensors geschehen. **Absolute Messgenauigkeit** kann durch eine Kalibrierung der Kameras erreicht werden. Eine neuartige Kalibriermethode wird in Kapitel 3.5.2 vorgestellt. Die Messgenauigkeit des Systems im Fahrzeug wird in Abschnitt 7.4 evaluiert und liegt im Nahbereich des Fahrzeugs zwischen 4 und 10 cm. Dies ist ausreichend, um eine Hinderniserkennung für Park- und Rangierfunktionen umzusetzen.

Das in dieser Arbeit entwickelte Verfahren zur **3D-Umfeldmodellierung** wird in Kapitel 5 vorgestellt. Die Ergebnisse der Evaluierung der Methode werden in Abschnitt 7.4 präsentiert. Diese zeigen, dass durch eine 3D-Umfeldmodellierung mithilfe der eingesetzten Weitwinkel-PMD-Kameras die Erkennung von Hindernissen durchgeführt werden kann. Bereiche, die vom Kamerasystem ausgeleuchtet werden, können mit einer hohen Genauigkeit vermessen werden.

Für zukünftige Fahrerassistenzfunktionen muss beim Rangieren eines Fahrzeugs eine **Verbesserung der Lokalisierung durch Sensordaten** erfolgen. Die Lokalisierung durch den Einsatz der Weitwinkel-PMD-Kameras wird in Kapitel 6.2 vorgestellt. Ein neuartiges, rein auf den Kameradaten basiertes Verfahren zur visuellen Odometrie wird in Kapitel 6.2.2 vorgestellt. Dieses führt allerdings zu einem Drift in der Eigenbewegungsschätzung, sodass eine Kombination mit der Koppelnavigation des Fahrzeugs nötig ist. Das anschließend in Kapitel 6.2.3 vorgestellte Verfahren zur simultanen Lokalisierung und Kartierung führt zu einer deutlichen Verbesserung der Eigenbewegungsschätzung.

Die **Erkennung dynamischer Objekte** wird in Kapitel 6.3 thematisiert. Es wird ein neuartiges Verfahren zur Erkennung dynamischer Objekte aus den Daten der im Fahrzeug integrierten Weitwinkel-PMD-Kameras vor-

gestellt. Die Evaluierung des Verfahrens zeigt, dass eine zuverlässige Erkennung dynamischer Objekte ermöglicht werden kann.

Eine **Bestimmung der Abstände zu Hindernissen** erfolgt auf Basis des erstellten Umfeldmodells. Evaluiert wird dies in Abschnitt 7.4. Diese zeigt, dass mit dem vorgestellten Verfahren Hindernisse in den gut ausgeleuchteten Bereichen der Umgebung des Fahrzeugs mit ausreichender Genauigkeit erkannt werden können. Probleme bestehen in Bereichen, die nicht direkt ausgeleuchtet werden. Dies ist in den hinteren Seitenbereichen der Fall. Zwar ermöglicht der Aufbau einer Karte die Bestimmung der Distanzen zu Hindernissen auch in ungünstigen Bereichen, allerdings kann dies ein Problem für dynamische Hindernisse darstellen. Um im gesamten Umfeld auch dynamische Hindernisse wahrnehmen zu können, muss eine Anpassung der Beleuchtung gemäß Kapitel 3.3.3 vorgenommen werden.

Eine **hohe Messrate** kann aufgrund der hohen Bildwiederholrate der PMD-Kameras von bis zu 100 Hz ermöglicht werden. Aufgrund der hohen Integrationszeit der Kameras und der Vermeidung von Störeffekten durch eine sequentielle Triggerung reduziert sich die Wiederholrate im vorgestellten System auf 15 Hz. Auch diese Messrate ist ausreichend, um während des Parkens und Rangierens auf dynamische Hindernisse reagieren zu können.

Aufgrund des aktiven Messprinzips ist bei der Verwendung von Time-of-Flight-Kameras eine **Messung auch im Stand möglich**. Anders als bei Verfahren, die 3D-Daten aus aufeinanderfolgenden Bildern einer Monokamera berechnen, ist keine Bewegung der Kamera zur Bestimmung von Tiefeninformationen notwendig.



8 Zusammenfassung

8.1 Fazit

Moderne Fahrerassistenzsysteme basieren auf der Wahrnehmung der Umgebung des Fahrzeugs. Zukünftige Fahrerassistenzsysteme richten sich unter anderem auf das assistierte oder automatisierte Parken und Rangieren eines Fahrzeugs. Die heute eingesetzten Sensoren zur Umfeldwahrnehmung können die Anforderungen zukünftiger Fahrerassistenzsysteme für das Parken und Rangieren nicht vollständig erfüllen. Durch die Auswertung verschiedener Sensoren wurde erkannt, dass Time-of-Flight-Kameras ein hohes Potenzial für die dreidimensionale Umfeldwahrnehmung besitzen.

Das Ziel dieser Arbeit war es daher, eine omnidirektionale 3D-Umfeldmodellierung zu 360° durch entsprechend ausgelegte Time-of-Flight Kameras zu ermöglichen. Die Arbeit beinhaltet als zwei Schwerpunkte die Systemauslegung eines PMD-Kamerasystems durch Weitwinkelobjektive und die Datenverarbeitung der Kameradaten hin zu einem dreidimensionalen Umfeldmodell.

Die Realisierung einer Weitwinkel-PMD-Kamera konnte durch eine Modularisierung der Kamera erreicht werden. Einzelne Lichtquellen können somit in unterschiedliche Richtungen geschwenkt werden und die gewünschten Bereiche innerhalb des Sichtfelds des Weitwinkelobjektivs ausleuchten. Zur Integration in ein Fahrzeug ist die Simulation der Lichtausbeute notwendig. Hierzu wurde ein Simulationsverfahren vorgestellt, mit dem die Ausrichtungen der einzelnen LEDs bestimmt werden können. Ein neuartiges Verfahren zur Tiefenkalibrierung konnte die absolute Messgenauigkeit der Kameras signifikant erhöhen.

Die Datenverarbeitung der Kameras beginnt bei der Vorverarbeitung der Amplituden- und Distanzbilder. Teil der Vorverarbeitung ist eine neuartige Methode zur pixelbasierten Konfidenzbestimmung. Diese ermöglicht

eine zuverlässige Aussage über die Messgenauigkeit eines einzelnen Tiefenwertes.

Zur Modellierung des Umfelds wurde auf eine Kombination einer 2.5D-Karte und eines *Occupancy Grids* zurückgegriffen. Die in 3D-Weltkoordinaten transformierten Tiefenwerte können somit fortlaufend in eine dreidimensionale Datenstruktur eingefügt werden.

Um die Umgebungsrekonstruktion zu verbessern, wurden die Probleme der Lokalisierung und der dynamischen Objekte in den Blick genommen. Zur Lokalisierung wurde ein neuartiges Verfahren zur visuellen Odometrie und ein Verfahren zur simultanen Lokalisierung und Kartierung vorgestellt. Ebenso konnte ein neues Verfahren zur Detektion von dynamischen Objekten entwickelt werden. Zusätzlich wurde ein neuartiges Verfahren zur Flächenextraktion aus den Kameradaten vorgestellt.

Die Evaluierung der 3D-Umfeldmodellierung konnte die Stärken und Schwächen der vorgestellten Methode und der Kamertechnologie aufzeigen. Es konnten Schwächen bei der Messgenauigkeit unter starkem Sonnenlicht festgestellt werden. Sie zeigte aber vor allem, dass der Einsatz der entwickelten Weitwinkel-PMD-Kameras Objekte im Umfeld des Fahrzeugs mit einer Genauigkeit von 4-11 cm im Nahbereich des Fahrzeugs bis zu einer Entfernung von 40 cm erkennen kann. Im Bereich bis zu einem Meter liegt der durchschnittliche Fehler im äußersten Fall bei 19 cm. Probleme bestehen bei der Abdeckung der Fahrzeugecken und der hinteren Bereiche der Fahrzeugflanken. Zwar ermöglicht die vorgestellte Methode zur Umfeldmodellierung den Aufbau einer Karte, sodass Objekte auch in den wenig abgedeckten Bereichen im Umfeldmodell vorhanden sind. Allerdings ist dies keine Lösung für dynamische Hindernisse. Für Systeme, welche die Verantwortung für das Rangieren vollständig vom Fahrer übernehmen, wird eine Erweiterung der Ausleuchtung notwendig werden. Es wurde dennoch gezeigt, dass der Einsatz von Weitwinkel-Time-of-Flight-Kameras die Anforderungen zukünftiger Fahrerassistenzsysteme für das Parken und Rangieren erfüllen kann.

8.2 Ausblick

In zukünftigen Arbeiten kann eine Weiterentwicklung der eingesetzten Kameras erfolgen. Insbesondere die Beleuchtung sollte weiterentwickelt wer-

den. Beispielsweise können die einzelnen LEDs mit unterschiedlichen Linsen versehen werden, sodass eine noch größere Freiheit bei der Ausleuchtung des Umfelds gewährt werden kann. Stark fokussierte LEDs können somit die kritischen Bereiche ausleuchten. Auch die Verwendung von Laserdioden mit strahlformenden Optiken kann dazu beitragen, noch spezifischere Ausleuchtungscharakteristika zu ermöglichen. Laserdioden könnten zudem aufgrund der geringeren Wellenlängenbandbreite gegenüber LEDs die Fremdlichtrobustheit erhöhen.

Weitere Forschung kann sich mit der Trennung von Lichtquelle und Empfangsoptik beschäftigen. Hierzu wurden in dieser Arbeit erste Ergebnisse präsentiert. Vorstellbar wäre ein Kamerasystem, das durch mehrere Beleuchtungscluster an unterschiedlichen Positionen durch eine sequentielle Ansteuerung der Lichtquellen große Bereiche ausleuchten kann. Beispielsweise könnten sich mehrere Lichtquellen entlang der Fahrzeugflanken befinden, die nacheinander angesteuert werden.

Die Time-of-Flight-Kameratechnologie entwickelt sich stetig weiter. Inzwischen existieren Kameras mit einer Auflösung von 640×480 Pixeln und auch kombinierte RGB-Z-Kameras, die Tiefenwerte und Farbwerte durch einen einzigen Sensor liefern [161]. Auch die Verbesserung der Messgenauigkeit von Time-of-Flight-Kameras ist ein aktuelles Forschungsthema. Eine höhere Genauigkeit kann zum einen durch eine stärkere Beleuchtung und zum anderen durch leistungsfähigere Sensoren ermöglicht werden.

Höhere Genauigkeiten würden zu einer genaueren 3D-Modellierung der Umgebung führen. Insbesondere die in Kapitel 2.2.3 vorgestellten Methoden zur 3D-Kartierung, wie das *KinectFusion*-Verfahren, werden erst durch eine Erhöhung der Genauigkeit der Tiefendaten ermöglicht.

Die vorgestellte Methode zur Umfeldmodellierung durch ein 2.5D-Grid ermöglicht bisher keine Modellierung von Unterfahrbarkeit. Die Integration der Unterfahrbarkeitsmodellierung in der Datenstruktur stellt einen Ansatzpunkt für weitere Arbeiten dar. Auch die farbliche Texturierung des extrahierten 3D-Modells der Umgebung kann insbesondere durch RGB-Z-Time-of-Flight-Kameras aber auch durch eine Kombination aus Time-of-Flight- und Mono-Kamera als sinnvolle Erweiterung der vorgestellten Methode betrachtet werden.

Die Verbesserung der 3D-Modellierung kann zudem durch die Erkennung von geometrischen Strukturen in den Tiefenbildern erfolgen. Beispielsweise bietet es sich an, die in Kapitel 6.1 vorgestellte Methode zur Flächenex-



8 Zusammenfassung

traktion für eine Verbesserung der 3D-Modellierung zu nutzen. Ebenso kann die Methode zur Extraktion von Flächen auf weitere geometrische Objekte erweitert werden.



Literaturverzeichnis

- [1] K. Reif, *Automobilelektronik*. Wiesbaden: Vieweg+Teubner Verlag, 2012.
- [2] H. Winner, S. Hakuli, und G. Wolf, *Handbuch Fahrerassistenzsysteme*. Springer, 2011.
- [3] H. Rudolph, *Audi Q3: Entwicklung und Technik*. Springer, 2013.
- [4] D. Ahrens und T. Bertram, „Parkassistent mit automatisierter Längs- und Querführung,” 8. Dortmunder Autotag, 2013.
- [5] Audi AG, „Driver assistance systems of tomorrow,” 2012. [Online]. Verfügbar: https://www.audi-mediaservices.com/publish/ms/content/en/public/hintergrundberichte/2012/03/08/networked_mobility/driver_assistance0.html
- [6] Volkswagen AG, „Volkswagen zeigt neueste Entwicklungen zu Parkassistenzsystemen,” 2013. [Online]. Verfügbar: https://www.volkswagen-media-services.com/detailpage/-/detail/Volkswagen-zeigt-neueste-Entwicklungen-zu-Parkassistenzsystemen/view/440891/6e1e015af7bda8f2a4b42b43d2dcc9b5?p_p_auth=liUiIBf9
- [7] BMW AG, „Ready for takeover!” 2011. [Online]. Verfügbar: https://www.press.bmwgroup.com/global/pressDetail.html?title=ready-for-takeover&outputChannelId=6&id=T0119745EN&left_menu_item=node__5236
- [8] B. Gutjahr und M. Werling, „Automatic collision avoidance during parking and maneuvering - An optimal control approach,” *IEEE Intelligent Vehicles Symposium 2014*, S. 636–641, 2014.



- [9] H. Vorobieva, S. Glaser, N. Minoiu-Enache, und S. Mammar, „Automatic parallel parking with geometric continuous-curvature path planning,” *IEEE Intelligent Vehicles Symposium 2014*, S. 465–471, 2014.
- [10] K. Reif, *Fahrstabilisierungssysteme und Fahrerassistenzsysteme*. Springer, 2010.
- [11] M. I. Skolnik, *Radar handbook*. McGraw-Hill, 1970.
- [12] Ibeo Automotive Systems GmbH, „Ibeo Web Page,” 2013. [Online]. Verfügbar: <http://www.ibeo-as.com>
- [13] A. De La Escalera, L. E. Moreno, M. A. Salichs, und J. M. Armingol, „Road traffic sign detection and classification,” *IEEE Transactions on Industrial Electronics* 1997, S. 848–859, 1997.
- [14] Z. Kim, „Robust Lane Detection and Tracking in Challenging Scenarios,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 9, S. 16–26, Mar. 2008.
- [15] P. Dollár, C. Wojek, B. Schiele, und P. Perona, „Pedestrian detection: an evaluation of the state of the art.” *IEEE transactions on pattern analysis and machine intelligence*, Vol. 34, Nr. 4, S. 743–61, Apr. 2012.
- [16] B. Hosticka, „CMOS imaging for automotive applications,” *IEEE Transactions on Electron Devices*, Vol. 50, Nr. 1, S. 173–183, 2003.
- [17] F. Dellaert, S. Seitz, C. Thorpe, und S. Thrun, „Structure from motion without correspondence,” *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, S. 557–564.
- [18] O. Faugeras, *Three-dimensional computer vision: a geometric viewpoint*. MIT Press, 1993.
- [19] S. Nedeveschi, S. Bota, und C. Tomiuc, „Stereo-based pedestrian detection for collision-avoidance applications,” *IEEE Transactions on Intelligent Transportation Systems*, Vol. 10, Nr. 3, S. 380–391, 2009.



- [20] C. Rabe, U. Franke, und S. Gehrig, „*Fast detection of moving objects in complex scenarios,*” *IEEE Intelligent Vehicles Symposium*, S. 398–403, Juni 2007.
- [21] F. Oniga und S. Nedeveschi, „*Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection,*” *IEEE Transactions on Vehicular Technology*, Vol. 59, Nr. 3, S. 1172–1182, 2010.
- [22] R. Stettner, „*Compact 3D flash lidar video cameras and applications,*” *SPIE Defense, Security, and Sensing*, 2010.
- [23] G. J. Iddan und G. Yahav, „*3D imaging in the studio (and elsewhere...),*” *Proceedings SPIE*, Vol. 4298, S. 48–55, 2001.
- [24] M. Roehder, „*Ein situativ entscheidendes Fahrzeugsystem für den vorausschauenden Fußgängerschutz,*” Dissertation, Technische Universität Kaiserslautern, 2011.
- [25] B. Elias, „*Vorausschauende Fahrzeugsensorik mit Photonic Mixer Device und Videokamera für den aktiven Fußgängerschutz,*” Dissertation, RWTH Aachen, 2009.
- [26] M. Roehder, S. Humphrey, B. Giesler, und K. Berns, „*Improving pedestrian safety in urban scenarios through autonomous collision avoidance,*” *Advanced Microsystems for Automotive Applications*, S. 165–175, 2010.
- [27] T. Ringbeck, B. Hagebeuker, H. Kraft, und M. Paintner, „*PMD-basierte 3D-Optosensoren zur Fahrzeugumfelderfassung,*” *Sensoren im Automobil II*, 2007.
- [28] PMDTechnologies GmbH, „*Website PMDTechnologies GmbH,*” 2014. [Online]. Verfügbar: <http://www.pmdtec.com>
- [29] A. Kolb, E. Barth, R. Koch, und R. Larsen, „*Time-of-flight cameras in computer graphics,*” *Computer Graphics Forum*, Vol. 29, S. 141–159, 2010.



- [30] T. Möller, H. Kraft, J. Frey, M. Albrecht, und R. Lange, „*Robust 3d measurement with pmd sensors*,“ *Range Imaging Day, Zürich*, 2005.
- [31] Y. Dalbah, N. Dingeldey, und F. M. Wahl, „*Visual odometry with high resolution time-of-flight cameras*,“ *International Conference on Machine Vision*, 2013.
- [32] C. Ege, „*Applikationsorientierte 3D-Bilddatenverarbeitung für ein PMD-Kamerasystem*,“ Dissertation, Hochschule Konstanz, 2006.
- [33] H. Kraft, J. Frey, T. Moeller, und M. Albrecht, „*3D-camera of high 3D-frame rate, depth-resolution and background light elimination based on improved PMD (photonic mixer device)-technologies*,“ *OPTO 2004, AMA Fachverband*, 2004.
- [34] X. Luan, „*Experimental investigation of photonic mixer device and development of TOF 3D ranging systems based on PMD technology*,“ Dissertation, Universität Siegen, 2006.
- [35] R. Lange, „*3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*,“ Dissertation, Universität Siegen, 2000.
- [36] S. Fuchs, „*Calibration and multipath mitigation for increased accuracy of time-of-flight camera measurements in robotic applications*,“ Dissertation, Technische Universität Berlin, 2012.
- [37] M. Frank, M. Plaue, H. Rapp, und F. A. Hamprecht, „*Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras*,“ *Optical Engineering*, 2009.
- [38] M. Lindner, „*Calibration and real-time processing of time-of-flight range data*,“ Dissertation, Universität Siegen, 2010.
- [39] S. Nesper und A. Seyfahrt, „*Der Einfluss der Lichtquellengeometrie auf die Entfernungsmessung von PMD-Kameras*,“ *Beiträge der Oldenburger 3D-Tage*, S. 389–396, 2010.
- [40] J. Mure-Dubois und H. Hügli, „*Optimized scattering compensation for time-of-flight camera*,“ *Optics East 2007*, Sep. 2007.



- [41] P. Westfeld, „*Geometrische und stochastische Modelle zur Verarbeitung von 3D-Kameradaten am Beispiel menschlicher Bewegungsanalysen*,“ Dissertation, Technische Universität Dresden, 2012.
- [42] Z. Zhang, „*Flexible camera calibration by viewing a plane from unknown orientations*,“ *The Proceedings of the Seventh IEEE International Conference on Computer Vision*, S. 666–673, 1999.
- [43] D. Scaramuzza, „*A flexible technique for accurate omnidirectional camera calibration and structure from motion*,“ *IEEE International Conference on Computer Vision Systems*, 2006.
- [44] D. Scaramuzza, „*Omnidirectional vision: from calibration to robot motion estimation*,“ Dissertation, ETH Zürich, 2007.
- [45] M. Lindner und A. Kolb, „*Lateral and depth calibration of PMD-distance sensors*,“ *Advances in Visual Computing*, 2006.
- [46] M. Lindner, I. Schiller, A. Kolb, und R. Koch, „*Time-of-flight sensor calibration for accurate range sensing*,“ *Computer Vision and Image Understanding*, Vol. 114, Nr. 12, S. 1318–1328, 2010.
- [47] M. Lindner, A. Kolb, und T. Ringbeck, „*New insights into the calibration of ToF-sensors*,“ *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Juni 2008.
- [48] S. Oprisescu und D. Falie, „*Measurements with ToF cameras and their necessary corrections*,“ *International Symposium on Signals Circuits and Systems*, 2007.
- [49] S. Fuchs und G. Hirzinger, „*Extrinsic and depth calibration of ToF-cameras*,“ *IEEE Conference on Computer Vision and Pattern Recognition*, Juni 2008.
- [50] J. Radmer, P. Fusté, H. Schmidt, und J. Krüger, „*Incident light related distance error study and calibration of the PMD-range imaging camera*,“ *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.



- [51] J. Mure-Dubois und H. Hügli, „*Real-time scattering compensation for time-of-flight camera*,“ *Proceedings of the International Conference on Computer Vision Systems*, 2007.
- [52] D. D. Lichti, X. Qi, und T. Ahmed, „*Range camera self-calibration with scattering compensation*,“ *Journal of Photogrammetry and Remote Sensing*, 2012.
- [53] H. Rapp, M. Frank, F. A. Hamprecht, und B. Jähne, „*A theoretical and experimental investigation of the systematic errors and statistical uncertainties of time-of-flight cameras*,“ *International Journal of Intelligent Systems Technologies and Applications*, Vol. 5, Nr. 3, S. 402–413, 2008.
- [54] F. Mufti und R. Mahony, „*Statistical analysis of signal measurement in time-of-flight cameras*,“ *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 66, Nr. 5, S. 720–731, Sep. 2011.
- [55] S. Jamtsho und D. D. Lichti, „*Modelling scattering distortion in 3D range camera*,“ *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 38, Nr. 5, S. 299–304, 2010.
- [56] A. Lejeune, M. Piérard, und J. V. Van Droogenbroeck, „*A new jump edge detection method for 3D cameras*,“ *International Conference on 3D Imaging*, 2011.
- [57] A. A. Dorrington, J. P. Godbaz, M. J. Cree, A. D. Payne, und L. V. Streeter, „*Separating true range measurements from multi-path and scattering interference in commercial range cameras*,“ *Proc. SPIE 7864*, Jan. 2011.
- [58] M. Reynolds, J. Dobos, L. Peel, T. Weyrich, und G. J. Brostow, „*Capturing time-of-flight data with confidence*,“ *IEEE Conference on Computer Vision and Pattern Recognition*, S. 945–952, 2011.
- [59] L. Breiman, *Random forests*. Springer, 2001.



- [60] H. P. Moravec und A. Elfes, „*High resolution maps from wide angle sonar,*” *International Conference on Robotics and Automation*, S. 116–121, 1985.
- [61] K. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, und W. Burgard, „*OctoMap: A probabilistic, flexible, and compact 3D map representation for robotic systems,*” *Workshop on Best Practice in 3D Perception and Modeling for Mobile Manipulation*, 2010.
- [62] A. Elfes, „*Using occupancy grids for mobile robot perception and navigation,*” *Computer*, Vol. 22, Nr. 6, S. 46–57, Juni 1989.
- [63] S. Thrun, W. Burgard, und D. Fox, *Probabilistic robotics*. MIT press, 2005.
- [64] M. Bouzouraa, „*Belegungskartenbasierte Umfeldwahrnehmung in Kombination mit objektbasierten Ansätzen für Fahrerassistenzsysteme,*” Dissertation, Technische Universität München, 2011.
- [65] J. Bares, M. Hebert, T. Kanade, und E. Krotkov, „*Ambler: An autonomous rover for planetary exploration,*” *Computer*, Vol. 22, Nr. 6, S. 18–26, 1989.
- [66] M. Herbert und C. Caillas, „*Terrain mapping for a roving planetary explorer,*” *IEEE International Conference on Robotics and Automation*, S. 997–1002, 1989.
- [67] E. Fong und W. Adams, „*Representing a 3-D environment with a 1/2-D map structure,*” *IEEE International Conference on Intelligent Robotics and Systems*, Vol. 3, S. 2986–2991, 2003.
- [68] S. Kagami, K. Nishiwaki, J. J. Kuffner, K. Okada, M. Inaba, und H. Inoue, „*Vision-based 2.5 D terrain modeling for humanoid locomotion,*” *IEEE International Conference on Robotics and Automation*, S. 2141–2146, 2003.
- [69] J. Gutmann, M. Fukuchi, und M. Fujita, „*A floor and obstacle height map for 3D navigation of a humanoid robot,*” *IEEE International Conference on Robotics and Automation*, S. 1066–1071, 2005.



- [70] P. Michel, J. Chestnutt, S. Kagami, K. Nishiwaki, J. Kuffner, und T. Kanade, „*Online environment reconstruction for biped navigation*,“ *IEEE International Conference on Robotics and Automation*, S. 3089–3094, 2006.
- [71] P. Pfaff und W. Burgard, „*An efficient extension of elevation maps for outdoor terrain mapping*,“ *Field and Service Robotics*, 2006.
- [72] A. a. S. Souza und L. M. G. Goncalves, „*2.5-dimensional grid mapping from stereo vision for robotic navigation*,“ *2012 Brazilian Robotics Symposium and Latin American Robotics Symposium*, S. 39–44, 2012.
- [73] S. May, D. Droeschel, S. Fuchs, D. Holz, und A. Nüchter, „*Robust 3D-mapping with time-of-flight cameras*,“ *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, S. 1673–1678, 2009.
- [74] G. Arbeiter, J. Fischer, und A. Verl, „*3-D-environment reconstruction for mobile robots using fast-SLAM and feature extraction*,“ *6th German Conference on Robotics*, 2010.
- [75] M. Montemerlo, „*Fastslam: A factored solution to the simultaneous localization and mapping problem with unknown data association*,“ Dissertation, Carnegie Mellon University, 2003.
- [76] G. Arbeiter, R. Bormann, J. Fischer, M. Hägele, und A. Verl, „*Towards geometric mapping for semi-autonomous mobile robots*,“ *Spatial Cognition VIII*, 2012.
- [77] A. Geiger, J. Ziegler, und C. Stiller, „*StereoScan: Dense 3d reconstruction in real-time*,“ *2011 IEEE Intelligent Vehicles Symposium (IV)*, S. 963–968, Juni 2011.
- [78] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, und W. Stuetzle, „*Surface reconstruction from unorganized points*,“ *Proceedings of the 19th annual conference on Computer graphics and interactive techniques - SIGGRAPH '92*, S. 71–78, 1992.



- [79] Z. C. Marton, R. B. Rusu, und M. Beetz, „*On fast surface reconstruction methods for large and noisy point clouds*,” *2009 IEEE International Conference on Robotics and Automation*, S. 3218–3223, 2009.
- [80] A. Swadzba, N. Beuter, J. Schmidt, und G. Sagerer, „*Tracking objects in 6D for reconstructing static scenes*,” *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, S. 1–7, Juni 2008.
- [81] C. Beder, B. Bartczak, und R. Koch, „*A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets*,” *IEEE Conference on Computer Vision and Pattern Recognition*, 2007.
- [82] Y. M. Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Mircusik, und S. Thrun, „*Multi-view image and ToF sensor fusion for dense 3D reconstruction*,” *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, S. 1542–1549, 2009.
- [83] M. Pollefeys, D. Nistér, J. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, M. P, C. Salmi, S. Sinha, B. Talton, L. Wang, H. Stewénus, R. Yang, G. Welch, und H. Towles, „*Detailed real-time urban 3d reconstruction from video*,” *International Journal of Computer Vision*, Vol. 78, S. 143–167, 2008.
- [84] M. Bajracharya, J. Ma, A. Howard, und L. Matthies, „*Real-time 3d stereo mapping in complex dynamic environments*,” *International Conference on Robotics and Automation - Semantic Mapping, Perception, and Exploration*, 2012.
- [85] W. Lorensen und H. Cline, „*Marching cubes: A high resolution 3D surface construction algorithm*,” *ACM Siggraph Computer Graphics*, 1987.
- [86] D. Meagher, „*Geometric modeling using octree encoding*,” *Computer graphics and image processing*, Vol. 19, Nr. 2, S. 129–147, 1982.
- [87] B. Curless und M. Levoy, „*A volumetric method for building complex models from range images*,” *Proceedings of the 23rd annual*



conference on Computer graphics and interactive techniques - SIGGRAPH '96, S. 303–312, 1996.

- [88] P. Neugebauer und K. Klein, „*Adaptive triangulation of objects reconstructed from multiple range images*,“ *IEEE Visualization'97*, 1997.
- [89] G. Graber, T. Pock, und H. Bischof, „*Online 3D reconstruction using convex optimization*,“ *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, S. 708–711, Nov. 2011.
- [90] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, und A. Fitzgibbon, „*KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera*,“ *Proceedings of the 24th annual ACM symposium on User interface software and technology*, S. 559–568, 2011.
- [91] H. Hemmat, E. Bondarev, G. Dubbelman, und P. H. N. de With, „*Evaluation of distance-aware KinFu algorithm for stereo outdoor data*,“ *Proceedings of the 9th International Conference on Computer Vision and Theory*, Vol. 2, S. 746–751, 2014.
- [92] E. Bondarev, F. Heredia, R. Favier, und P. H. de With, „*On photo-realistic 3D reconstruction of large-scale and arbitrary-shaped environments*,“ *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, S. 621–624, 2013.
- [93] M. Keller, D. Lefloch, M. Lambers, S. Izadi, T. Weyrich, und A. Kolb, „*Real-time 3D reconstruction in dynamic scenes using point-based fusion*,“ *3D Vision*, 2013.
- [94] M. Fischler und R. Bolles, „*Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*,“ *Communications of the ACM*, 1981.
- [95] C. Stewart, „*Bias in robust estimation caused by discontinuities and multiple structures*,“ *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, Nr. 8, S. 818–833, 1997.



- [96] Y. Kanazawa und H. Kawakami, „*Detection of planar regions with uncalibrated stereo using distributions of feature points.*” *BMVC*, 2004.
- [97] M. Zuliani, „*The multiransac algorithm and its application to detect planar homographies.*” *IEEE International Conference on Image Processing*, 2005.
- [98] O. Gallo, R. Manduchi, und A. Rafii, „*CC-RANSAC: Fitting planes in the presence of multiple surfaces in range data.*” *Pattern Recognition Letters*, Vol. 32, Nr. 3, S. 403–410, 2011.
- [99] D. Nistér, „*Preemptive RANSAC for live structure and motion estimation.*” *Machine Vision and Applications*, S. 1–29, 2005.
- [100] R. Toldo und A. Fusiello, „*Robust multiple structures estimation with j -linkage.*” *Computer Vision-ECCV 2008*, 2008.
- [101] A. Sabov und E. Uhlmann, *Virtuelle Rekonstruktion von Innenräumen basierend auf Messdaten von Tiefenkameras*. Fraunhofer Verlag, 2010.
- [102] T. G. Stahs, „*Ein aktives 3D-Robotersystem auf der Grundlage eines verallgemeinerten Ansatzes zur Erstellung modellbasierter Objekterkennungsverfahren.*” Dissertation, Technische Universität Braunschweig, 1995.
- [103] D. Scaramuzza und F. Fraundorfer, „*Visual odometry: Part I: The first 30 years and fundamentals.*” *IEEE Robotics & Automation Magazine*, Vol. 18, Nr. 4, S. 80–92, 2011.
- [104] B. Lucas und T. Kanade, „*An iterative image registration technique with an application to stereo vision.*” *IJCAI*, Vol. 130, S. 121–130, 1981.
- [105] B. F. Fraundorfer, D. Scaramuzza, und F. Fraundorfer, „*Visual Odometry: Part II: Matching, Robustness, Optimization, and Applications.*” *Robotics & Automation Magazine, IEEE*, Vol. 19, Nr. 2, S. 78–90, 2012.



- [106] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, und P. Sayd, „*Real time localization and 3d reconstruction*,“ *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, S. 363–370, 2006.
- [107] G. Klein und D. Murray, „*Parallel tracking and mapping for small AR workspaces*,“ *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, S. 1–10, 2007.
- [108] K. Konolige, M. Agrawal, und J. Solà, „*Large-scale visual odometry for rough terrain*,“ in *Robotics Research*, Serie Springer Tracts in Advanced Robotics. Springer, 2011, Vol. 66, S. 201–212.
- [109] A. S. Huang, A. Bachrach, P. Henry, M. Krainin, D. Fox, und N. Roy, „*Visual odometry and mapping for autonomous flight using an RGB-D camera*,“ *International Symposium on Robotics Research*, 2011.
- [110] A. Prusak, O. Melnychuk, H. Roth, I. Schiller, und R. Koch, „*Pose estimation and map building with a pmd-camera for robot navigation*,“ *Dynamic 3D Imaging*, 2007.
- [111] D. Droschel, S. May, D. Holz, P. Plöger, und S. Behnke, „*Robust ego-motion estimation with ToF cameras*.“ *Proceedings of the European Conference on Mobile Robots (ECMR)*, S. 187–192, 2009.
- [112] S. May, D. Droschel, D. Holz, und S. Fuchs, „*3D pose estimation and mapping with time-of-flight cameras*,“ *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2008.
- [113] A. Mian, M. Bennamoun, und R. Owens, „*A novel representation and feature matching algorithm for automatic pairwise registration of range images*,“ *International Journal of Computer Vision*, Vol. 66, Nr. 1, S. 19–40, 2006.
- [114] P. Besl und N. McKay, „*A Method for registration of 3-D shapes*,“ *Robotics-DL tentative*, 1992.
- [115] H. J. Hemmat, E. Bondarev, G. Dubbelman, und P. H. N. de With, „*Improved ICP-based pose estimation by distance-aware 3D mapping*,“ *Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, S. 360–367, 2014.



- [116] S. Winkelbach, S. Molkenstruck, und F. M. Wahl, „*Low-cost laser range scanner and fast surface registration approach*,” *Pattern Recognition*, S. 718–728, 2006.
- [117] R. C. Smith und P. Cheeseman, „*On the representation and estimation of spatial uncertainty*,” *The International Journal of Robotics Research*, S. 56–68, 1986.
- [118] S. Thrun, „*A probabilistic on-line mapping algorithm for teams of mobile robots*,” *The International Journal of Robotics Research*, Vol. 20, Nr. 5, S. 335–363, 2001.
- [119] T.-D. Vu, „*Vehicle perception : Localization, mapping with detection, classification and tracking of moving objects*,” Dissertation, Institut National Polytechnique de Grenoble, 2009.
- [120] R. Iser, „*Where am I? Contributions to the localization problem of mobile robots*,” Dissertation, Technische Universität Braunschweig, 2012.
- [121] A. Eliazar und R. Parr, „*DP-SLAM: Fast, robust simultaneous localization and mapping without predetermined landmarks*,” *Proceedings of the 18th International Joint Conference on Artificial intelligence*, S. 1135–1142, 2003.
- [122] N. S. Boroujeni, S. A. Etemad, und A. Whitehead, „*Fast obstacle detection using targeted optical flow*,” *Proceedings of the 19th IEEE International Conference on Image Processing (ICIP'12)*, S. 65–68, Sep. 2012.
- [123] H.-H. Bock, „*Clustering methods: a history of K-Means algorithms*,” in *Selected contributions in data analysis and classification*. Springer, 2007.
- [124] M. Shih, Y. Chang, B.-C. Fu, und C.-C. Huang, „*Motion-based background modeling for moving object detection on moving platforms*,” *Proceedings of the 16th International Conference on Computer Communications and Networks*, S. 1178–1182, 2007.



- [125] Talukder A. und L. Matthies, „*Real-time detection of moving objects from moving vehicles using dense stereo and optical flow*,“ *IEEE Conference on Intelligent Robots and Systems*, Vol. 4, 2004.
- [126] A. Talukder, S. Goldberg, L. Matthies, und A. Ansar, „*Real-time detection of moving objects in a dynamic scene from moving robotic vehicles*,“ *IEEE Conference on Intelligent Robots and Systems*, 2003.
- [127] S. Vacek, T. Schamm, J. Schröder, J. M. Zöllner, und R. Dillmann, „*Collision avoidance for cognitive automobiles using a 3D PMD camera*,“ in *Intelligent Autonomous Vehicles*, Vol. 6, Nr. 1, 2007, S. 379–384.
- [128] T. Schamm, J. M. Zollner, S. Vacek, und J. Schroder, „*Obstacle detection with a Photonic Mixing Device-camera in autonomous vehicles*,“ *International Journal of Intelligent Systems Technologies and Applications*, Vol. 5, Nr. 3, S. 315–324, 2008.
- [129] B. Fardi, J. Dousa, G. Wanielik, B. Elias, und A. Barke, „*Obstacle detection and pedestrian recognition using a 3D PMD camera*,“ in *IEEE Intelligent Vehicles Symposium*, 2006.
- [130] Osram, „*OSLON Black Series (850 nm) - Datasheet*,“ Datenblatt, 2012.
- [131] J. L. Bentley, „*K-d trees for semidynamic point sets*,“ in *Proceedings of the sixth annual symposium on Computational geometry*. ACM, 1990, S. 187–197.
- [132] A. Kadambi, R. Whyte, A. Bhandari, L. Streeter, C. Barsi, A. Dorrington, und R. Raskar, „*Coded time of flight cameras: sparse deconvolution to address multipath interference and recover time profiles*,“ *ACM Transactions on Graphics (TOG)*, 2013.
- [133] T. Huang, G. Yang, und G. Tang, „*A fast two-dimensional median filtering algorithm*,“ *IEEE Transactions on Acoustics, Speech and Signal Processing*, 1979.



- [134] L. W. Abele und F. M. Wahl, „*A digital procedure for boundary detection and elimination of background in cytologic images,*” *Medinfo*, Vol. 77, S. 1035–1038, 1977.
- [135] F. M. Wahl, *Digitale Bildsignalverarbeitung*, Serie Nachrichtentechnik. Springer Berlin Heidelberg, 1984, Vol. 13.
- [136] C. Tomasi und R. Manduchi, „*Bilateral filtering for gray and color images,*” *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, S. 839–846, 1998.
- [137] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009.
- [138] W. Karel, S. Ghuffar, und N. Pfeifer, „*Modelling and compensating internal light scattering in time of flight range cameras,*” *The Photogrammetric Record*, Vol. 27, Nr. 138, S. 155–174, 2012.
- [139] J. E. Bresenham, „*Algorithm for computer control of a digital plotter,*” *IBM Systems journal*, Vol. 4, Nr. 1, S. 25–30, 1965.
- [140] Y. Dalbah, D. Koltermann, und F. M. Wahl, „*Robust real time extraction of plane segments from time-of-flight camera images,*” *International Conference on Digital Image Processing (ICDIP)*, 2014.
- [141] I. Jolliffe, *Principal component analysis*. Springer, 2005.
- [142] M. Sonka, V. Hlavac, und R. Boyle, *Image processing, analysis, and machine vision*. Thomson Learning, 1999.
- [143] M. Baer, M. E. Bouzouraa, C. Demiral, und U. Hofmann, „*EgoMaster: A central ego motion estimation for driver assist systems,*” *IEEE International Conference on Control and Automation*, S. 1708–1715, 2009.
- [144] Y. Dalbah, M. K. Assoy, M. Roehder, D. Rosebrock, und F. M. Wahl, „*Visuelle Odometrie mit einer hochauflösenden Time-of-Flight Kamera,*” *Forum der Bildverarbeitung*, 2012.



- [145] J. Shi und C. Tomasi, „*Good features to track*,“ *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, S. 593–600, 1994.
- [146] E. Rosten und T. Drummond, „*Machine learning for high-speed corner detection*,“ *Computer Vision-ECCV 2006*, S. 1–14, 2006.
- [147] D. G. D. Lowe, „*Distinctive image features from scale-invariant keypoints*,“ *International journal of computer vision*, Vol. 60, Nr. 2, S. 1–28, 2004.
- [148] H. Bay, T. Tuytelaars, und L. V. Gool, „*Surf: Speeded up robust features*,“ *Computer Vision-ECCV*, 2006.
- [149] A. Alahi, R. Ortiz, und P. Vandergheynst, „*Freak: Fast retina keypoint*,“ *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- [150] B. Horn, „*Closed-form solution of absolute orientation using unit quaternions*,“ *JOSA A*, 1987.
- [151] S. May, D. Droeschel, D. Holz, S. Birlinghoven, S. Augustin, S. Fuchs, und N. Andreas, „*Three-dimensional mapping with time-of-flight cameras*,“ *Journal of Field Robotics (JFR), Special Issue on Three-dimensional Mapping*, Vol. 26, S. 934–965, 2009.
- [152] S. J. Julier und J. K. Uhlmann, „*Unscented filtering and nonlinear estimation*,“ *Proceedings of the IEEE*, Vol. 92, Nr. 3, 2004.
- [153] U. Kramer, *Kraftfahrzeugführung*. Carl Hanser Verlag, 2008.
- [154] S. Blackrnan und A. House, „*Design and analysis of modern tracking systems*,“ *Boston, MA: Artech House*, 1999.
- [155] Y. Dalbah, S. Rohr, und F. M. Wahl, „*Detection of dynamic objects for environment mapping by time-of-flight cameras*,“ *IEEE International Conference on Image Processing*, 2014.
- [156] G. Welch und G. Bishop, „*An introduction to the Kalman filter*,“ *Annual Conference on Computer Graphics & Interactive Techniques*, 1995.



- [157] A. J. Davison, „*Real-time simultaneous localisation and mapping with a single camera,*” *IEEE International Conference on Computer Vision*, 2003.
- [158] Zoller + Fröhlich GmbH, „*Web Page Zoller + Fröhlich,*” 2014. [Online]. Verfügbar: <http://www.zf-laser.com/>
- [159] Nikon, „*iGPS/iSpace modular positioning, tracking and measurement system for factory-wide deployment,*” 2014. [Online]. Verfügbar: http://www.nikon.com/products/instruments/lineup/industrial/metrology_assisted_production/large_volume/igps/
- [160] C. Depenthal, „*iGPS used as kinematic measuring system,*” *FIG Congress*, 2010.
- [161] W. Kim, W. Yibing, I. Ovsianikov, S. Lee, Y. Park, C. Chung, und E. Fossum, „*A 1.5 Mpixel RGBZ CMOS image sensor for simultaneous color and range image capture,*” *IEEE International Solid-State Circuits Conference*, S. 392–394, 2012.





