

**ABALID:
Abbiegeassistent mit 3D-LIDAR-Sensorik**

**Abschlussbericht des Teilvorhabens
3D-Objekterkennung und semantische Analyse**

Laufzeit: 1.3.2013-31.5.2016

Johannes Gräter
Wei Tian
Martin Lauer
Christoph Stiller

Das diesem Bericht zugrunde liegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 16SV6029 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Inhaltsverzeichnis

1	Aufgabenstellung und Ablauf des Projektes	2
1.1	Aufgabenstellung	2
1.2	Voraussetzungen und Ablauf des Projektes	3
1.3	Stand der Forschung zu Beginn des Projektes	4
1.4	Projektkonsortium	5
2	Ergebnisse und Nutzen des Projektes	7
2.1	Technische Ergebnisse	7
2.1.1	Systemstruktur und Modulübersicht	7
2.1.2	Fahrradfahrerdetektion	7
2.1.3	Fahrradfahrerassoziation und -verfolgung	11
2.1.4	Eigenbewegungsschätzung mit Visueller Odometrie	13
2.1.5	Fahrradfahrer-Bewegungsschätzung	17
2.2	Ergebnisse aus Probandenstudie	20
2.3	Bewertung der Projektergebnisse	21
A	Veröffentlichungen	24
A.1	Publikationen, die innerhalb des Projektes entstanden	24
A.2	Literaturverzeichnis	25

Kapitel 1

Aufgabenstellung und Ablauf des Projektes

1.1 Aufgabenstellung

Fahrradfahrer sind im innerstädtischen Verkehr besonders gefährdet, da sie in Unfällen häufig die Leidtragenden sind. Die kritischsten Szenarien sind dabei Unfälle mit Güterkraftfahrzeugen (LKW), da diese zumeist mit schweren Verletzungen oder gar tödlich für den Radfahrer enden. Häufig geschehen derartige Unfälle im Kreuzungsbereich beim Rechtsabbiegen des LKWs. Die Ursache dafür ist in der erster Linie ein eingeschränktes Sichtfeld des LKW-Fahrers, sodass die Radfahrer schlicht übersehen werden. Konventionelle Maßnahmen wie zusätzliche Seitenspiegel am LKW oder Versetzen der Haltelinien in der Infrastruktur können diese Unfälle nicht vollständig vermeiden.

Um die Anzahl der Unfälle zwischen Fahrrad und LKW zu reduzieren und somit die Fahrradfahrer zu schützen, ist es deshalb wünschenswert, ein Assistenzsystem für LKW zu entwickeln, das den LKW-Fahrer dabei unterstützt die Umgebung des LKWs zu überwachen und auf sich anbahnende Unfälle zu reagieren. Die technischen Fortschritte im Bereich von Fahrerassistenzsysteme sowie der digitalen Umfelderkennung mit Hilfe von Kameras und anderen Sensoren bieten hierzu die technische Grundlage. Bei der Entwicklung solcher Systeme ist jedoch darauf zu achten, die technischen Anforderungen mit den Aspekten der Nützlichkeit und Akzeptanz für die LKW-Fahrer zu kombinieren.

Das Ziel des Projektes *ABALID: Abbiegeassistent mit 3D-LIDAR-Sensorik* war es, einen Prototypen für dieses Assistenzsystem zu entwickeln, um dadurch die LKW-Unfälle beim Rechtsabbiegen mit Fahrradfahrern auf innovative Weise zu vermeiden. Zum einen sollte das System mit Hilfe von einem Miniatur-3D-LIDAR-Sensor und einer CMOS-Kamera die Fahrradfahrer erfassen können. Zum anderen sollte eine Warn- und Anzeigeschnittstellen im LKW-Führerstand den LKW-Fahrer vor kritischen Situationen warnen, damit er die Fahrradfahrer in seinem Nahbereich nicht übersehen kann.

Das Gesamtprojekt umfasste sowohl die Entwicklung der Gerätehardware als auch der benötigten Algorithmen zur Informationsverarbeitung, ferner die Anforderungsanalyse sowie die experimentelle Evaluation des Systems in Form einer Probandenstudie. Das Teilprojekt

Fahrradfahrerdetektion und -bewegungsschätzung umfasste hierbei die Entwicklung der Verfahren zur Auswertung der Kamerabilder und LIDAR-Messungen. Dies beinhaltet alle Verfahren, die erforderlich sind, um aus den Bildfolgen der Kamera und den Messungen des LIDARs die benötigten Informationen zur Erfassung und Modellierung der Fahrzeugumgebung sowie der Fahrradfahrer zu extrahieren. Zusammenfassend sollten folgende Punkte umgesetzt werden:

- Die Bestimmung der Bewegung des LKWs.
- Die Erkennung der gefährdeten Fahrradfahrer.
- Die Schätzung der Bewegung der erkannten Fahrradfahrer.
- Die Fusion der Sensordaten von LIDAR und Kamera.

Alle Verfahren dazu müssen gleichzeitig robust, recheneffizient und zuverlässig sein, um im Fahrzeug mit begrenzter Rechenleistung, in Echtzeit und in variierenden Verkehrsszenarien einsetzbar zu sein. Die Schnittstelle der implementierten Software soll auch kompatibel zu weiteren Warnkomponenten sein. Die Erkennungsgenauigkeit muss hoch genug sein, um dem LKW-Fahrer einen Mehrwert zu bieten und ihn nicht durch viele fehlerhafte Informationen in die Irre zu leiten und sein Fahrverhalten zu stören.

1.2 Voraussetzungen und Ablauf des Projektes

Durch die vielen in den zurückliegenden Jahren am MRT durchgeführten Projekte im Bereich der sensorischen Umfeldwahrnehmung für autonome Fahrzeuge und Fahrerassistenzsystem besaß das MRT bereits zu Projektbeginn eine große Expertise in diesem Bereich, was dem Projekt ABALID durch vielfältige Synergien zugute kam. So konnte die neu entwickelte Software auf das bestehende Softwareframework aufbauen und die Vielzahl der vorhandenen Werkzeuge nutzen. Außerdem konnten bestehende algorithmische Lösungen als Startpunkt für die Algorithmenentwicklung genutzt und schrittweise an die Anforderungen des Projektes angepasst werden. Diese umfassen unter anderem einen Ansatz zur Visuellen Odometrie [Geiger 11], ein speziell für Fahrerassistenzfunktionen entwickeltes Merkmalspunktverfahren [Lategahn 14], sowie die am MRT entwickelte KITTI-Benchmark zur Evaluation der entwickelten Ansätze [Geiger 12].

Die Umsetzung des Teilprojektes Fahrradfahrerdetektion und -bewegungsschätzung war abhängig von der Verfügbarkeit der Sensorik, insbesondere des LIDAR-Sensors, der parallel vom Projektpartner Spies zu entwickeln war und somit erst zu einem späteren Zeitpunkt im Projektverlauf zur Verfügung stand. Um die Entwicklung nicht zu verzögern war es unabdingbar, aus eigener Vorentwicklung eine alternative Sensorplattform aufzubauen, die aus zwei Videokameras bestand und somit ein Stereokamerasystem ergaben. Mit Hilfe dieses Stereokamerasystems konnten die Sensormessungen des LIDAR-Sensors simuliert werden, wodurch die Entwicklung der Auswertelgorithmen vom Projektbeginn an möglich war. Durch Nutzung des am Institut für Mess- und Regelungstechnik (MRT) des KIT verfügbaren Versuchsfahrzeuges war es möglich, von Projektbeginn an Sensordaten aufzuzeichnen und damit eine große

Situationsvielfalt abzudecken. Zudem konnte auf bereits am MRT vorhandene Datensätze zurückgegriffen werden.

Durch diese Maßnahmen war es möglich, bereits nach 1,5 Jahren Entwicklungszeit einen ersten Prototypen – noch basierend auf einem Stereokamerasystem – zu demonstrieren. Nach gut 2 Jahren Entwicklungszeit war der LIDAR-Sensor sowohl Hardware- als auch Softwareseitig integriert und es konnten mit dem Versuchsfahrzeug des MRT auf einem Testgelände in Karlsruhe Testfahrten durchgeführt werden, um gemeinsam mit dem Projektpartner HFC typische Szenarien nachzustellen und Versuchsdaten zu erheben. Diese bildeten einerseits die Grundlage für die Entwicklung der Warnstrategie auf Seiten des Projektpartners HFC, als auch als Grundlage für die Weiterentwicklung der Algorithmen auf Seiten des KIT. Letzteres umfasste eine Optimierung der Algorithmen im Hinblick auf Genauigkeit und Robustheit, als auch im Hinblick auf Recheneffizienz als Voraussetzung für einen Einsatz in Echtzeit. Ferner wurde das Softwareframework verbessert, um eine höhere Robustheit und eine verbesserte Handhabbarkeit zu erzielen. Dadurch war es möglich, zu den Probandenstudien, die am Ende des dritten Projektjahres durchgeführt wurden, ein robustes und zuverlässiges Sensor- und Messsystem zur Verfügung zu stellen.

1.3 Stand der Forschung zu Beginn des Projektes

Die Entwicklung von Assistenzsystemen zur Detektion gefährdeter Verkehrsteilnehmer ist seit vielen Jahren ein aktives Forschungsfeld, das bis in die 1990er Jahre zurückgeht [Papageorgiou 98]. Ein erfolgreiches Beispiel ist das bildbasierte Detektionssystem für Fußgänger von Dollár et al. [Dollár 12], welches Bildverarbeitungsverfahren und Algorithmen des Machine-Learning nach dem aktuellen Stand der Technik nutzt und eine gute Leistung aufweist. Im Vergleich zur Fußgängerdetektion gibt es noch eine Lücke im Bereich der Fahrradfahrerdetektion. Die meisten auf dem Markt verfügbaren Systeme zur Detektion von Fahrradfahrern basieren auf Radar-, Infrarot- und Ultraschall-Sensorik [Dharmaraju 01]. Obwohl diese Systeme eine hohe Detektionsrate aufweisen können, ist es noch schwierig, Fahrradfahrer von anderen Objektklassen zu unterscheiden, wodurch viele Falschalarme nicht zu vermeiden sind. Außerdem folgt aus der kurzen Reichweite mancher Sensoren ein kurzes Zeitfenster, in dem der LKW-Fahrer reagieren und einen Unfall verhindern kann.

Anstatt als Komponente eines Fahrerassistenzsystems ist es auch möglich, das Detektionssystem in die Infrastruktur einzugliedern, wie z.B. induktive Sensoren unter der Straße einzusetzen [Krogmeier 08]. Allerdings sind die Instandhaltungskosten sehr hoch und die Fahrradfahrerdetektion nur an bestimmten Stellen möglich.

Um den im Bild erkannten Fahrradfahrer zur Kollisionswarnung nutzen zu können, muss dessen Bewegung in der Welt berechnet werden. Die Zustandsschätzung bewegter Objekte ist ein typisches Problem der maschinellen Wahrnehmung. Klassische Algorithmen wie das Kalman-Filter und dessen Erweiterungen [Kalman 60] sind etablierte Methoden um den nicht direkt beobachtbaren Systemzustand eines linearen oder linearisierbaren Systems zu schätzen. Ist das System nicht linearisierbar oder müssen multimodale Verteilungen geschätzt werden ist das Partikelfilter [Thrun 05] das Standardverfahren um die Verteilung des Zustandes durch viele einzelne Stichproben anzunähern. Ein großer Nachteil dieser Methoden ist die

notwendige Kenntnis der charakteristischen Parameter der verwendeten Verteilungen, wie im Kalman-Filter zum Beispiel die Kovarianzmatrizen des Zustandes und der Messungen. Diese sind in der Regel nicht bekannt und müssen daher händisch angepasst oder aufwendig gemessen werden.

Um die Fahrradfahrbewegung nicht nur lokal, relativ zum Sensor, sondern global in einem weltfesten Koordinatensystem beschreiben zu können, wird die Eigenbewegung des Fahrzeugs benötigt. Dies ist ein Spezialfall des *Simultaneous Localization and Mapping (SLAM)*-Problems in welchem der Fokus nur auf der Eigenbewegung beruht, was als *Visuelle Odometrie* bezeichnet wird. Ein sehr guter Überblick über verschiedene Methoden der Visuellen Odometrie sind in den Übersichtspapern von Scaramuzza und Fraundorfer gegeben ([Scaramuzza 11, Fraundorfer 12]). Zwar zeichnet sich klassischer, offline gerechneter SLAM bzw. eine offline gerechnete Kartierung durch sehr hohe Genauigkeit aus, jedoch ist der nötige Rechenaufwand zu hoch um auf einer kompakten mobilen Rechereinheit in Echtzeit eingesetzt werden zu können.

Auf der Hardware-Seite weist die Technik der Kamera-Sensorik einen großen Fortschritt im Bereich von Arbeitsfrequenz, Auflösung und Robustheit gegen variierende Umgebungsbedingungen auf. Zudem ist ein Aufschwung der Computer-Technik in den vergangenen Jahren deutlich zu sehen: die Baugröße und die Kosten der Rechner sinken, während deren Rechenleistung steigt. Obwohl die Bildauswertung und -verstehen rechnerisch aufwändig sind, kann ein solches Problem durch optimierte Implementierung der Algorithmen und Einsatz eines leistungsstarken Rechners aus der oberen Consumer-Preisklasse umgegangen werden. Mit Hilfe von solchen technischen Grundlagen wird gesichert, dass das Detektionssystem für ABALID-Projekt in Echtzeit realisierbar ist.

1.4 Projektkonsortium

Das Projektkonsortium bestand aus drei Partnern, die in Tabelle 1.1 aufgeführt sind. Das Teilvorhaben 3D-Objekterkennung und semantische Analyse war mit allen anderen Teilvorhaben eng verzahnt. Die wichtigsten Schnittstellen sind in der folgenden Aufzählung aufgeführt.

- mit dem Partner HFC wurde die gemeinsame Software des Systems entwickelt und die Warnstrategie mit der Umfeldwahrnehmung abgestimmt. Die Probandenstudien wurden gemeinsam durchgeführt.
- mit dem Partner Spies wurden die Anforderungen für die LIDAR-Sensorik erarbeitet und die Schnittstellen zwischen LIDAR-Sensorik und der ABALID-Software entwickelt.

Tabelle 1.1: Partner im Konsortium des Gesamtprojektes ABALID und ihre jeweiligen Aufgabenbereiche

Partner	Aufgaben
HFC Human Factors Consult GmbH, Berlin (HFC)	Gesamtprojektkoordination, Anforderungsanalyse, Probandenstudien, Warnstrategie, Schnittstelle zum Fahrer
Karlsruher Institut für Technologie, Institut für Mess- und Regelungstechnik (KIT-MRT)	3D-Objekterkennung und semantische Analyse
Spies GbR	Entwicklung der LIDAR-Technologie

Kapitel 2

Ergebnisse und Nutzen des Projektes

2.1 Technische Ergebnisse

In den folgenden Abschnitten sollen die technischen Ergebnisse des Projektes dargestellt werden. Da die entwickelte Software zur 3D-Objekterkennung und semantischen Analyse aus verschiedenen, miteinander interagierenden Modulen besteht, soll zunächst in Abschnitt 2.1.1 die Grundstruktur der Software beschrieben und ein Überblick über die einzelnen Module gegeben werden, während anschließend die algorithmischen Lösungen für die Teilprobleme detailliert dargestellt werden.

2.1.1 Systemstruktur und Modulübersicht

Die Systemfunktionalität wird mithilfe einfacher und seriennaher Hardware umgesetzt, deren Spezifikationen in Tabelle 2.1 gegeben sind.

Die Datenverarbeitung wird durch die in Abbildung 2.2 gezeigten Teilmodule umgesetzt. Aus den Kamerabildern wird der Fahrradfahrer im Bild detektiert und verfolgt. Parallel dazu wird die Fahrzeug-Eigenbewegung bestimmt. Diese Module sind in Abschnitten 2.1.2, 2.1.3 und in Abschnitt 2.1.4 beschrieben. Um die Bewegung des Fahrradfahrers zu schätzen sind Informationen über dessen dreidimensionale Position notwendig. Darum werden die Distanzmessungen aus dem, vom Partner Spies bereitgestellten, LIDAR in die Fahrradfahrerschätzung integriert, wie in Abschnitt 2.1.5 erläutert wird.

Das Ergebnis sind die geschätzten Trajektorien und Geschwindigkeiten aller erkannten Fahrradfahrer sowie des Egofahrzeugs. Diese werden an die Schnittstelle weitergegeben, welche die Daten an die Kollisionswarnung des Partners HFC weiterleitet.

2.1.2 Fahrradfahrerdetektion

Die Fahrradfahrerdetektion ist der erste Schritt im gesamten System (Abbildung 2.2) und dient als die Grundlage für die erfolgreiche Analyse eines wahrscheinlichen Unfallszenarios.



Abbildung 2.1: Hardware-Konfiguration im Abschlusstest. Beide Sensoren werden kompakt durch Saugnäpfe an der rechten Tür des LKWs angebracht (Bild links und Mitte). Die aufgenommenen Daten werden durch einen Rechner in der Fahrerkabine verarbeitet. Das Warnsignal wird durch einen Bildschirm dem LKW-Fahrer gezeigt (Bild rechts).

Tabelle 2.1: Technische Daten der beiden Sensoren.

Sensortyp	Arbeitsfrequenz	Auflösung
Monokulare Kamera	10 Hz	1500x960 Pixel
LIDAR	12,5 Hz	2°(horizontal), 0,8°(vertikal), max. 4 Scanebenen (alternierend)

Aus der Hardware-Perspektive muss die Frage zunächst beantwortet werden, welche Sensorik bei der Detektion angewendet werden soll. Im Vergleich der technischen Daten der beiden im Projekt verfügbaren Sensoren (Tabelle 2.1), ist nur die monokulare Kamera zur Fahrradfahrdetektion geeignet. Dafür gibt es mehrere Gründe. Die Kamera bietet eine wesentlich höhere horizontale und vertikale Auflösung, so dass die Fahrradfahrer in großem Detailreichtum dargestellt werden können. Dies ermöglicht erst die Unterscheidung von Fahrradfahrern von abgestellten Fahrrädern, Fußgängern oder anderen Objekten. Zudem arbeitet nur die Kamera mit einer Frequenz von 10 Hz, was die Echtzeit-Anforderung im Projekt erst erfüllt. Da die Kamera das beobachtete Szenario als Bildsignal erfasst, ist die Fahrradfahrdetektion in erster Linie bildbasiert. Der LIDAR wird lediglich zur Bestätigung erkannter Fahrradfahrer sowie zur Abstandsbestimmung genutzt.

Detektionsmethoden

Um die Zielobjekte, d.h. die Fahrradfahrer im Bild herauszufinden, wird hier die Methode mit dem Namen „Sliding Window Search“ angewendet, was auch dem Standard der Technik entspricht [Zhang 16a]. In diesem Verfahren wird ein rechteckiges Suchfenster mit derselben Größe wie das Zielobjekt auf das Bild gelegt. Innerhalb des Fensters werden Bildmerkmale berechnet, auf die ein auf Beispielfeldern trainierter Klassifikator angewendet wird. Um Objekte an jeder möglichen Position zu detektieren, wird das Suchfenster über das gesamte Bild geschoben und an jeder Stelle der Klassifikator angewendet. Das Verfahren wird in Abbildung 2.3 (a) veranschaulicht.

Eine weitere Schwierigkeit bei der Fahrradfahrdetektion besteht darin, dass das Aussehen eines Fahrradfahrers im Bild sehr stark durch seine Ansicht beeinflusst wird. Daher

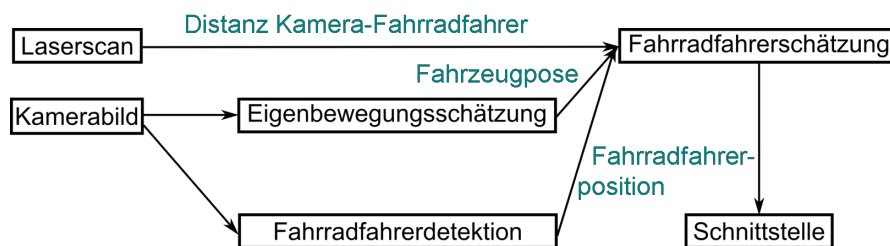


Abbildung 2.2: Software-Framework. Das aufgenommene Kamerabild wird zur Fahrradfahrerdetektion und der Bewegungsschätzung des LKWs angewendet. Die Fahrradfahrerschätzung basiert auf dem durch LIDAR gemessenen Abstand und die Fahrzeugpose, die aus der Eigenbewegungsschätzung folgt. Die ermittelte Fahrradfahrerposition wird an die Schnittstelle zur Warnstrategie weitergeleitet, um dort die Kritikalität der Situation zu bewerten und ggf. den LKW-Fahrer zu warnen.

ist ein einheitliches Suchfenster für alle Ansichten ungeeignet. Hier werden die Bilder der Fahrradfahrern je nach ihrer Ansicht in acht Gruppen unterteilt. Für jede Gruppe wird ein eigenes Suchfenster gewählt sowie ein eigener Klassifikator trainiert (Abbildung 2.3 (b)).

Struktur des Fahrradfahrerklassifikators

Als Aufbau des Klassifikators wird eine Kaskade gewählt, deren Recheneffizienz und hohe Genauigkeit schon in zahlreiche Arbeiten gezeigt wurde [Dollár 12, Benenson 14]. Als Bildmerkmale werden hier HOG-Features (Histogram of Oriented Gradients) [Dalal 05] benutzt, da diese Merkmale die Kontur eines Objekts gut repräsentieren können. Wie in Abbildung 2.4 gezeigt, befindet sich auf jede Stufe der Kaskade ein einfacher Klassifikator, z.B. ein Entscheidungsbäume (DT) in den ersten Stufen der Kaskade sowie eine SVM (Support Vector Machine) am Ende. Zunächst dienen die DTs zum schnellen Herausfiltern der meisten negativen Samples (ohne Zielobjekt). Danach wird eine genauere Klassifikation für die wenigen übriggebliebenen Samples durch die SVM ausgeführt. Auf diese Weise werden sowohl eine hohe Rechengeschwindigkeit als auch eine hohe Genauigkeit für die Detektion erzielt. Zum Trainieren des Detektors wird der AdaBoost-Algorithmus angewendet [Tian 15a].

Ausnutzung der geometrischen Beschränkungen zur Fahrradfahrerdetektion

Obwohl eine schnelle Detektion durch die Kaskaden-Struktur ermöglicht wird, bleibt es dennoch schwierig, die Echtzeit-Anforderung zu erfüllen. Beispielsweise beträgt die Rechenzeit zur Detektion auf ein Bild mit einer Größe von 1500x960 Pixeln 150 ms, selbst bei Nutzung eines Mehrkernrechners (Tabelle 2.2). Bedenkt man allerdings, dass die Fahrradfahrer sich stets auf dem Boden befinden und die Kamera eine im wesentlichen konstante Einbauhöhe und Orientierung besitzt, so wird klar, dass Fahrradfahrer nur in bestimmten Bildbereichen zu sehen sein können. Somit ist es möglich, die Fahrradfahrerdetektion auf einen kleine Bildbereich zu beschränken. Zur Bestimmung des relevanten Bildbereiches wird die Methode von [Sudowe 11, Tian 15b] angewendet, die die geometrische Beziehung zwischen dem 3D Raum

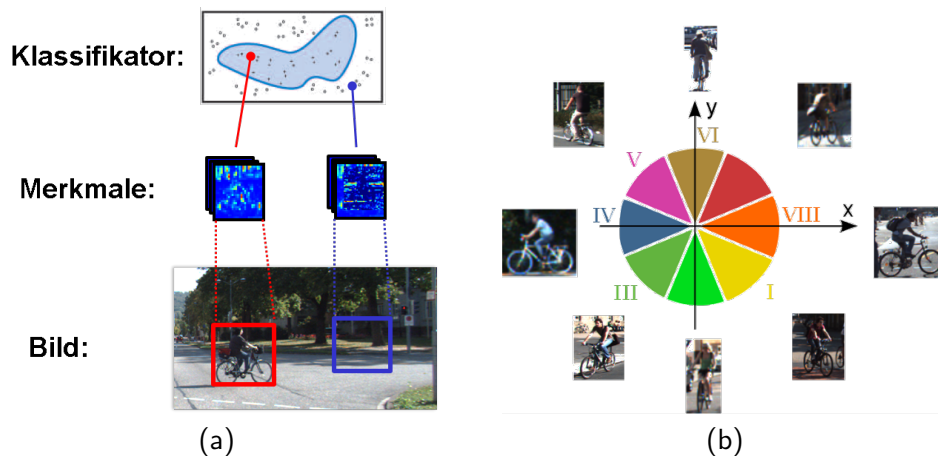


Abbildung 2.3: Bildbasiertes Detektionsverfahren „Sliding Window Search“ wird in (a) gezeigt. (b) zeigt die Aufteilung der Fahrradfahrersamples in acht Gruppen je nach ihrer Ansicht.

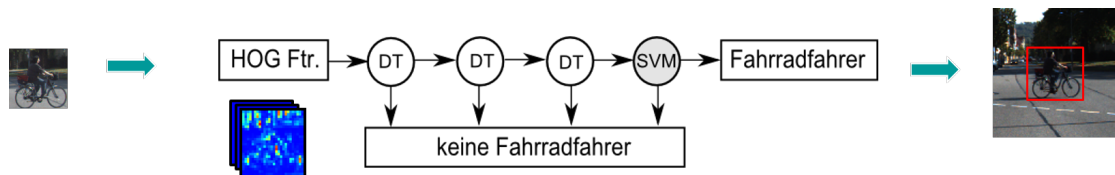


Abbildung 2.4: Fahrradfahrerdetektor mit Kaskaden-Struktur. Am Anfang der Kaskade sind Entscheidungsbäume angeordnet und am Ende eine Support Vector Machine. Als Eingaben dienen HOG-Merkmale der Bildsamples. Während die negativen Samples Stufe für Stufe herausgefiltert werden, bleiben die positiven Samples bis zum Ende der Kaskade erhalten. Der resultierenden Detektionen sind im Bild durch rote Rahmen gekennzeichnet.

und dem Kamerabild ausnutzt. Beispielsweise ist für einen Fahrradfahrer mit einer Höhe von 2 m bei einer Suchfensterhöhe von 96 Pixeln der relevante Bildbereich ein horizontaler Streifen (Abbildung 2.5 (a)). Für variierende Größen des Fahrradfahrers (1 m bis 2 m) ist der Streifen nur um ein paar Pixel breiter geworden (Abbildung 2.5 (b)). In diesem Fall beträgt der Rechenaufwand nur 50 ms pro Bild (Tabelle 2.2). Gleichzeitig können Falschdetektionen außerhalb des Bildbereichs vermieden werden. Die Methode ist auch dann anwendbar, wenn die Kamera während der Fahrt Nick- und Wankbewegungen durch die Bewegung des LKW-Führerhauses ausgesetzt ist [Tian 15b].

Validierung der Fahrradfahrerdetektion durch LIDAR-Sensorik

Aufgrund der geometrischen Beziehung zwischen dem 3D-Raum und dem Kamerabild ist es möglich, die aufgenommene Abstandsinformation vom LIDAR auszunutzen, um die bildbasierten Detektionsergebnisse zu validieren. Hier wird angenommen, dass die Größe der Fahrradfahrer in einem bestimmten Höhenbereich (1 m bis 2 m) liegt. Somit ist es möglich,

Tabelle 2.2: Rechenzeitvergleich für jeden Schritt der Detektion auf ein Bild mit einer Größe von 1500x960 Pixeln. Die Ausnutzung von Mehrkernrechnern wird durch den Eintrag „Parallelisierung“ gekennzeichnet.

	Merkmale	Detektion	Summe	Parallelisierung
Ohne Vorauswahl von Bildbereichen	80 ms	140 ms	220 ms	150 ms
Mit Vorauswahl von Bildbereichen	30 ms	50 ms	80 ms	50 ms

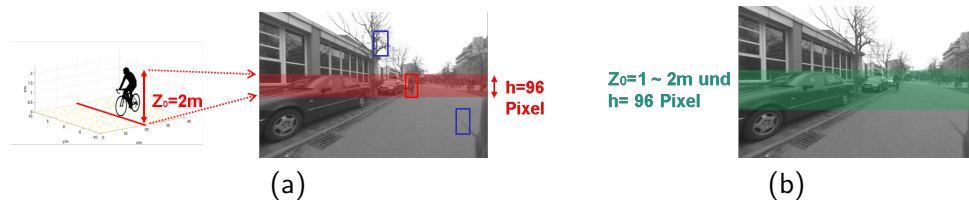


Abbildung 2.5: Geometrische Beziehung zwischen dem 3D-Raum und dem Kamerabild. Für einen Fahrradfahrer mit einer Höhe von 2 m und einer Suchfensterhöhe von 96 Pixeln ist der entsprechende Bildbereich, in dem der Fahrradfahrer auftreten kann, ein Streifen, der in (a) in rot gekennzeichnet ist. Die blauen Rechtecke sind möglichen Fehldetektionen, die durch Ausnutzen des Suchbereichs vermieden werden können. Der grüne Streifen in (b) repräsentiert den entsprechenden Suchbereich für Fahrradfahrer im Höhenbereich von 1 m bis 2 m.

aus der Bildposition eines Fahrradfahrers sowie dessen Höhe im Bild auf seine 3D-Position rückzuschließen. Zur Validierung werden zunächst die Positionen der Fahrradfahrer und die Laserstrahlen auf dieselbe horizontale Ebene, z.B. die Bodenebene, projiziert (Abbildung 2.6). Dann wird jeder Fahrradfahrer mit den Laserpunkten in seinem Nahbereich assoziiert. Die Annahme ist, dass der LIDAR so eingestellt wird, dass jedes Objekt in der Nähe des Fahrzeugs detektiert werden kann. Deswegen wird ein Fahrradfahrer ohne zugeordnete Laserpunkte als eine Fehldetektion betrachtet. Auf diese Weise wird die Detektionsgenauigkeit weiter erhöht.

2.1.3 Fahrradfahrerassoziation und -verfolgung

Um einen Unfall zwischen einem Fahrradfahrer und dem Fahrzeug vorhersagen zu können, muss nicht nur die Position des Fahrradfahrers sondern auch seine Trajektorie bekannt sein. Deswegen ist es notwendig, den erkannten Fahrradfahrer über die Bildfolgen hinweg zu verfolgen, mit neuen Detektionen zu assoziieren und dadurch zu aktualisieren. Zur Verfolgung werden drei Zustandsgrößen genutzt: die Position, die Geschwindigkeit und die Objektgröße im Bild, die die Dynamik des Fahrradfahrers im Bild gut repräsentieren können. Für jedes neue Bild werden die Zustände des erkannten Fahrradfahrers prädiziert und mit einer neuen Detektion assoziiert, falls die Überlappung dazwischen einen vordefinierten Schwellwert überschreitet (Abbildung 2.8). Dann werden die Zustände des Fahrradfahrers durch die assoziierte Detektion korrigiert und im nächsten Prädiktionsschritt weiter benutzt. Als Ansatz

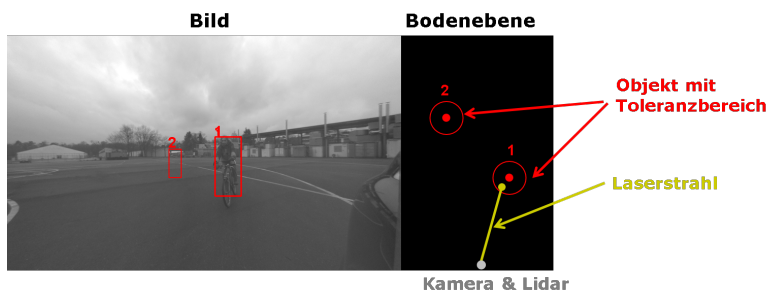


Abbildung 2.6: Detektor.

Abbildung 2.7: Validierung der Detektionsergebnisse durch LIDAR-Daten. Die Positionen der erkannten Fahrradfahrer werden zunächst auf die Bodenebene projiziert. Dann wird jeder Fahrradfahrer mit den Laserpunkten in seinem Nahbereich assoziiert. Ein Fahrradfahrer ohne zugeordnete Laserpunkte wird als eine Fehldetektion betrachtet.



Abbildung 2.8: Prädiktion der Objektposition im neuen Bild. Die Assoziation mit der neuen Detektion erfolgt anhand der Überlappung zwischen prädizierter und detektierter Position.

für Prädiktion und Korrektur kommt das Kalman-Filter [Kalman 60] in Kombination mit dünnem Optischem Fluss [Beauchemin 95] bzw. Model-Matching zum Einsatz.

Fahrradfahrerverfolgung durch optischen Fluss

In dieser Methode werden zunächst Merkmalspunkte innerhalb des Bildbereichs ausgewählt, das den Fahrradfahrer zeigt. Im nächsten Bild werden die selben Punkte gesucht und ihre neue Positionen bestimmt (Abbildung 2.9 (a)). Anhand der Positionsänderungen solcher Punkte ist es möglich, die Bewegung des Objekts zwischen den Bildern zu schätzen. Da diese Methode nur ein paar Punkte auf dem Objekt beobachtet, ist der Rechenaufwand dafür sehr gering und beträgt weniger als 10 ms pro Bild.

Fahrradfahrerverfolgung durch Model-Matching

Hierbei wird für jedes erkannte Objekt im Bild ein eigenes Modell aufgebaut, das durch spezifische Merkmale (z.B. Gradienten) repräsentiert wird oder durch Lernverfahren aus den bisherigen Detektionen gelernt wird. Dieses Modell wird auf das nächste Bild angewendet

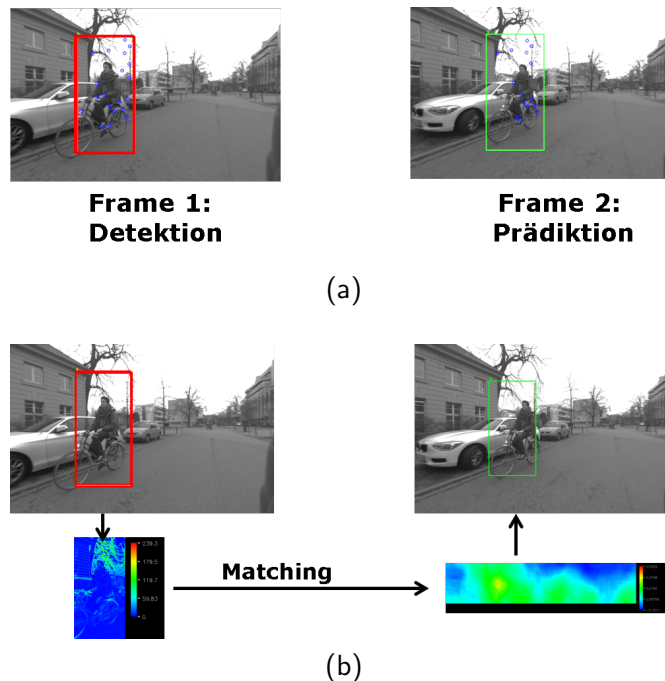


Abbildung 2.9: (a) zeigt das Objekttracking durch optischen Fluss. Ein paar Merkmalspunkte werden im Frame 1 auf dem detektierten Objekt ausgewählt und dann in Frame 2 gesucht. Die Objektposition im neuen Bild wird anhand der Positionsänderung der Merkmalspunkte geschätzt. In (b) wird für das erkannt Objekt ein Modell aufgebaut, das durch Gradienten-Merkmale repräsentiert ist. Im nächsten Bild wird das ähnlichste Objekt gesucht. Der Punkt mit dem höchsten Matching-Score entspricht der neuen Objektposition.

und durch ein Matching- oder Klassifikationsverfahren das ähnlichste Objekt gesucht. Der Punkt mit dem höchsten Matching-Score wird als die neue Position des Objekts angenommen (Abbildung 2.9 (b)). Ein Vorteil dieser Methode besteht darin, dass das Objekttracking sehr robust gegen starke Änderungen des Hintergrundes sowie teilweise Verdeckung des Objekts ist. Allerdings ist der Rechenaufwand (20 ms bis 30 ms pro Bild) deutlich größer als bei der Verwendung von Merkmalspunkten.

Im Software-System sind beide Verfahren als separate Module implementiert und können wahlweise eingesetzt werden.

2.1.4 Eigenbewegungsschätzung mit Visueller Odometrie

Flussschätzung

Die Basis für den Großteil der bildbasierten Bewegungsschätzungsalgorithmen beruht auf optischem Fluss. Da es sehr rechenaufwändig ist, diesen für jedes Pixel im Bild zu schätzen, wird nicht-dichter Fluss verwendet. Dabei wird wie folgt vorgegangen:

1. Es werden charakteristische Pixel, sogenannte *Keypoints* im Bild identifiziert, die über

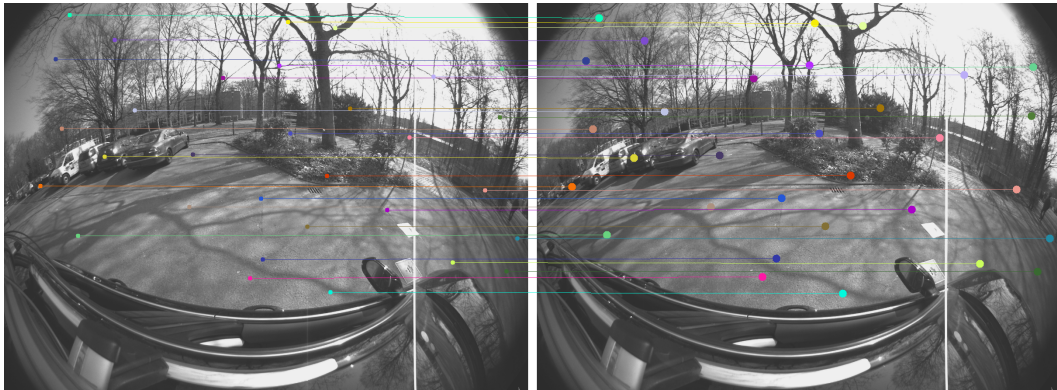


Abbildung 2.10: Schätzung des Flusses aus zwei aufeinanderfolgenden Bildern. Hier sind der Übersichtlichkeit wegen nur einige assoziierte Features gezeigt. Für die Bewegungsschätzung werden typischerweise 1000 dieser assoziierten Features benötigt.

eine Bildfolge gut wiederzufinden sind. Klassischerweise werden dafür Ecken, z.B. mit dem Harris-Eckendetektor [Harris 88] oder kreisförmige Strukturen, sogenannte *Blobs* genutzt.

2. Die Texturinformation der Region um das charakteristische Pixel wird mithilfe eines sogenannten *Deskriptors* komprimiert. Hierfür gibt es zahlreiche Beispiele wie SURF [Bay 06], BRIEF [Calonder 12] oder ORB [Rublee 11]. In dieser Arbeit werden die Deskriptoren von Lategahn et al. verwendet [Lategahn 14]. Keypoints und Deskriptoren werden zusammen als *Features* bezeichnet.
3. Schritt 1 und Schritt 2 werden auf einem zweiten Bild ausgeführt.
4. Die Deskriptoren beider Bilder werden auf Ähnlichkeit verglichen und eine optimale Zuordnung wird gefunden.

Somit kann die Bewegung der Keypoints zwischen den Bildern, also der optische Fluss, berechnet werden und in der folgenden Bewegungsschätzung verwendet werden, wie in Abbildung 2.10 gezeigt.

Bild-Zu-Bild-Schätzung

In einem ersten Schritt wird die Bewegung von einem Bild zum Nächsten berechnet. Dies geschieht mit dem von Hartley et al. [Hartley 03] entwickelten und von Geiger et al. [Geiger 11] implementierten 8-Punkt-Algorithmus. Hierbei wird der Epipolarfehler 2.1 minimiert.

$$(p_{i,0}, p_{i,1}, 1)^T F (p_{j,0}, p_{j,1}, 1) = 0 \quad (2.1)$$

wobei p_i, p_j die in den beiden Bildern beobachteten Features sind und $F \in \mathbb{R}^{3 \times 3}$ die Fundamentalmatrix [Hartley 03] darstellt. Durch geschicktes Umformen kann dieses Optimierungsproblem in ein lineares Gleichungssystem überführt und gelöst werden. Um Ausreißer auszuschließen wird der RANSAC Algorithmus [Fischler 81] angewandt.

Bündelausgleich

Der von Scaramuzza beschriebene Bündelausgleich [Scaramuzza 11] ist ein Standardverfahren für die visuelle Kartierung. Diese ist eine mögliche Art das SLAM-Problem zu lösen. Hierbei werden gleichzeitig die 3D-Struktur der Umgebung rekonstruiert und die Kameraposen berechnet. Benötigt werden hierbei sogenannte *Featuretracks*, welche erzeugt werden indem die in Abschnitt 2.1.4 beschriebene Flusschätzung solange pro Keypoint ausgeführt wird, bis dieser nicht mehr wiederzufinden ist. Typischerweise erhält man somit Featuretracks von bis zu 20 Bildern Länge. Das Grundprinzip des Bündelausgleichs besteht darin, dass die in den Featuretracks enthaltene 3D-Umgebungsinformation abhängig von den Kameraposen ist, jedoch die Positionen der Keypoints in den Bildern wiederum in Abhängigkeit zu den Kameraposen steht. Diese wechselseitige Abhängigkeit kann durch gemeinsame Optimierung gelöst werden. Zuerst wird ein Startwert für Posen und die 3D-Position der Keypoints, hier *Landmarken* genannt, ermittelt. Für die Posen $p_j \in P$, mit Gesamtheit aller Posen P , wird die Bild-Zu-Bild-Methodik aus Abschnitt 2.1.4 verwendet und mit deren Hilfe die Landmarken $x_i \in X$, mit Gesamtheit aller Landmarken X , trianguliert. Dazu wird das Optimierungsproblem 2.2 gelöst, in welchem der sogenannte *Rückprojektionsfehler* minimiert wird.

$$\hat{P}, \hat{X} = \operatorname{argmin}_{P, X} \sum_j \sum_i \|A^{-1}(p_j \cdot \mathbf{x}_i) - \mathbf{p}_{i,j}\|_2^2 \quad (2.2)$$

hierbei stellen P und X die geschätzten Werte für die Gesamtheit aller Posen und die Gesamtheit aller Landmarken dar. Die Matrix $A \in \mathbb{R}^{3 \times 3}$ ist die intrinsische Matrix, welche aus der Kamerakalibrierung bekannt ist. In der gesamten Arbeit werden ausschließlich Lochkameramodelle behandelt. Die Laufindices i, j laufen respektive über alle Landmarken und alle Bilder. Idealerweise wird also das globale Optimum für alle Posen und alle Landmarken mithilfe aller gemessenen Featuretracks gefunden. Diese Algorithmik auf einem Consumer-Rechner echtzeitfähig umzusetzen ist bei der momentanen Elektronikleistung nicht möglich. Um den Rechenaufwand zu reduzieren wird darum nur innerhalb eines Fensters von beispielsweise 12 Bildern optimiert. Zusätzlich werden die Startwerte für die Posen nur berechnet falls noch nicht genug Messungen für die Optimierung vorhanden sind. Andernfalls wird eine vorherige Schätzung als Startwert verwendet. Durch geschicktes abwechselndes Festhalten von Posen und Landmarken muss die rechenintensive Optimierung innerhalb des gesamten Fensters nur ca. alle 5 Bilder erfolgen. Obwohl diese letzte Operation mit ca. 130 ms nicht echtzeitfähig ist, kann so insgesamt eine durchschnittliche Rechenzeit von 90 ms erreicht werden, womit alle Kamerabilder verarbeitet werden können. Um das System als Ganzes echtzeitfähig zu machen werden die Flusschätzung und der Bündelausgleich in einer Pipeline implementiert. In Abbildung 2.11 sind Beispiele aus einer Evaluation auf dem KITTI-Benchmark [Geiger 12] zu sehen.

Skalenschätzung

Ein grundlegendes Problem bei Visueller Odometrie mit nur einer Kamera ist mangelnde Beobachtbarkeit eines Parameters: der Skale. Anschaulich bedeutet das, dass die Kamera

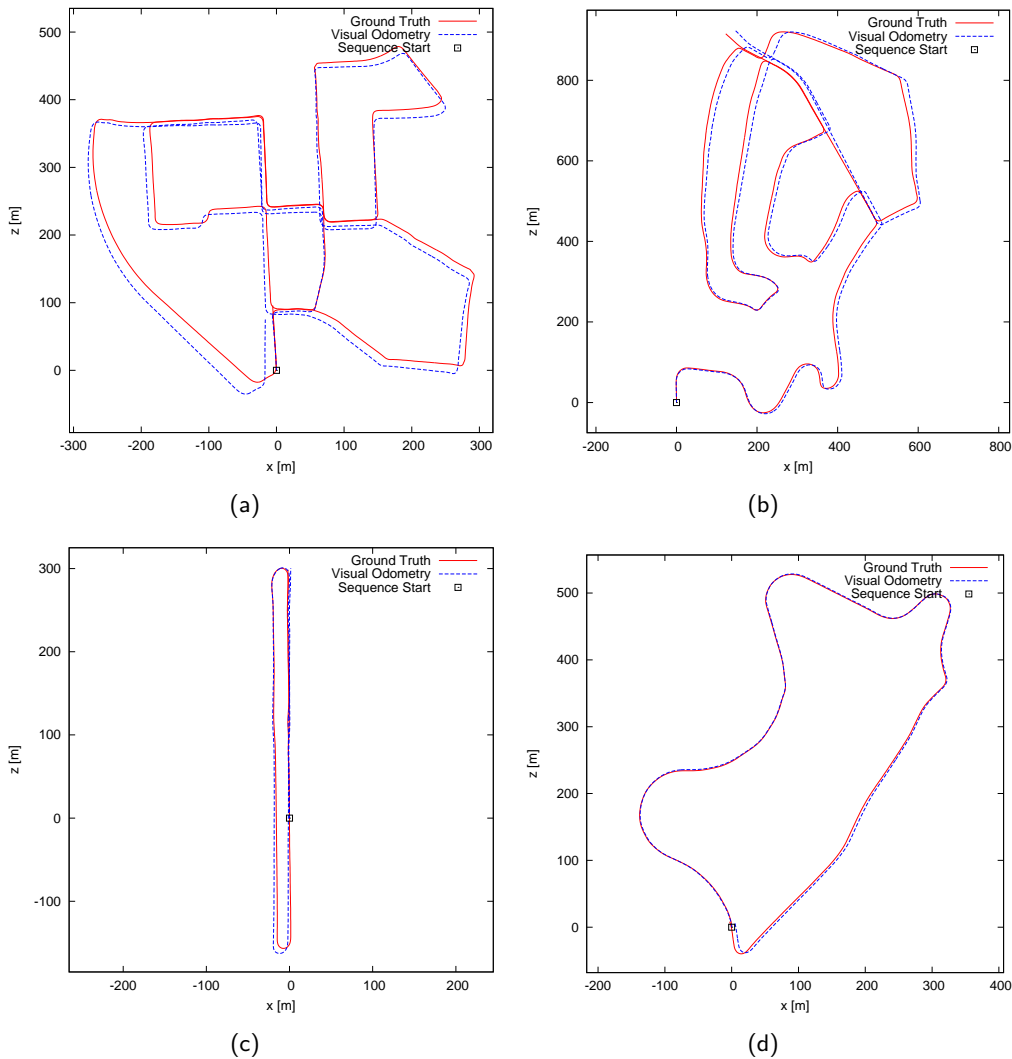


Abbildung 2.11: Beispiele aus dem KITTI-Benchmark, welche die hohe Genauigkeit der monoskopischen Visuellen Odometrie darstellen. Dargestellt sind in der Draufsicht Fahrzeugtrajektorien über mehrere Kilometer hinweg. In rot ist der mit GPS gemessene, tatsächliche Verlauf der Fahrt eingezeichnet, in blau der mit visueller Odometrie geschätzte Verlauf.

nicht metrisch die zurückgelegte Strecke messen kann, sondern nur die Richtung. Diese Information geht bei der Aufnahme der 3D-Umgebungsinformationen auf dem 2D-Kamerachip verloren. Um diese Information abzuschätzen muss das in der Kamera Gesehene mit einem metrischen Längenmaß verglichen werden. Hierzu treffen wir die Annahme, dass die Kamera sich in konstanter Höhe über dem Boden befindet. Die, durch die Federung des LKWs erzeugten, Höhenänderungen sind im Vergleich zur Montagehöhe tolerierbar. Anschließend werden die Features wie in Abschnitt 2.1.4 zu 3D-Punkten rekonstruiert, allerdings noch ohne die Skale. Nun kann die Grundebene rekonstruiert werden. Dies geschieht mit einem robusten

Least-Squares-Schätzer, dem Trimmed-Least-Squares-Verfahren. Die Grundebenenschätzung wurde zusätzlich durch Fluchtpunkte gestützt und die Ergebnisse veröffentlicht [Gräter 15b].

Mögliche Erweiterungen

Die Bewegungsschätzung zeigt sehr gute Ergebnisse, die die Anforderungen an dieses System erfüllen. Allerdings ist das Fahren bei Nacht bisher noch völlig unbetrachtet. Hier könnten andere Flusssschätzer, welche auch bei geringer Beleuchtung funktionieren, den innerstädtischen Einsatz in der Nacht ermöglichen. Vielversprechend ist auch der Einsatz erst kürzlich erschienener Kameras mit einem deutlich höheren dynamischen Umfang als die hier genutzten.

2.1.5 Fahrradfahrer-Bewegungsschätzung

Die Bewegungsschätzung der Fahrradfahrer basiert auf einer klassischen Regression. Eine Herausforderung ist hierbei jedoch, dass zwei verschiedene Arten von Messungen vorliegen: Die Sichtstrahlen zum Fahrradfahrer aus der bildbasierten Detektion des Fahrradfahrers, siehe Abschnitt 2.1.2 und Distanzmessungen mithilfe des Laserscanners. Aus der Kamera ist die Distanz der Fahrradfahrers nicht messbar, wohingegen die Laserscannermessung nicht den genauen Sichtwinkel des Fahrradfahrers messen kann. Um dieses Problem zu lösen wird in diesem Projekt eine Regression mit latenten Variablen eingesetzt, wobei die Distanz des Fahrradfahrers zur Kamera die latente Variable ist. Durch das hohe Rauschen in der Laserscanner-Tiefenmessung müssen des Weiteren starke Modellannahmen getroffen werden. In diesem Fall wird der Fahrradfahrer durch eine Punktmasse mit linearer Bewegung approximiert, welche sich mit konstanter Geschwindigkeit im 3D-Raum bewegt. Diese Annahme trifft nicht auf die reale Fahrradfahrerbewegung zu. Daher wird diese entschärft, indem die Regression nicht über alle Messungen erfolgt, sondern nur in einem Fenster von n Bildern. So können auch die Beschleunigung und sogar Kurvenfahrten modelliert werden.

Algorithmik

Im Folgenden wird die Algorithmik näher beschrieben: Um die Position $\mathbf{x} = (x, y, z)^T$ und die Geschwindigkeit $\mathbf{v} = (v_x, v_y, v_z)^T$ zu Zeitpunkt t_i zu schätzen, wird der quadratische Fehler aus Gleichung 2.3 minimiert.

$$\operatorname{argmin}_{\mathbf{x}, \mathbf{v}} \sum_{i=1}^n \|\mathbf{x} + t_i \mathbf{v} - \mathbf{p}_i\|_2^2 \quad (2.3)$$

Wobei $\mathbf{p}_i \in \mathbb{R}^3$ die gemessene Fahrradfahrerposition in Weltkoordinaten zum Zeitpunkt t_i beschreibt. Durch Nullsetzen der ersten Ableitung von \mathbf{x} und \mathbf{v} erhält man das lineare Least-Squares-Problem aus Gleichung 2.4.

$$\begin{bmatrix} I \cdot n & I \sum_i t_i \\ I \sum_i t_i & I \sum_i t_i^2 \end{bmatrix} \cdot \begin{pmatrix} \mathbf{x} \\ \mathbf{v} \end{pmatrix} = \begin{pmatrix} \sum_i \mathbf{p}_i \\ \sum_i t_i \mathbf{p}_i \end{pmatrix}, i = 1 \dots n \quad (2.4)$$

I bezeichnet hierbei die Einheitsmatrix aus \mathbb{R}^3 . Die Fahrradfahrerposition \mathbf{p}_i wird vom Fahrzeug aus Position \mathbf{P}_i gemessen, wobei durch den Laserscanner die Tiefe r_i und durch die Kamera die Richtung \mathbf{w}_i bekannt ist. Daraus folgt:

$$\mathbf{p}_i = \mathbf{P}_i + r_i \mathbf{w}_i, \quad \|\mathbf{w}_i\|_2 = 1 \quad (2.5)$$

Dadurch, dass Kamera und Laserscanner nicht synchronisiert sind, kann die Distanz r_i nicht zu jedem Zeitpunkt bestimmt werden. Um diese latente Variable nicht explizit schätzen zu müssen wird ein impliziter Schätzwert für r_i genutzt. Darum wird die lokale Fahrradfahrerposition $\mathbf{x} + t_i \mathbf{v} - \mathbf{P}_i$ zu jedem Zeitpunkt t_i auf die aus der Kamera bekannten Richtung \mathbf{w}_i projiziert.

Der Tiefen-Schätzwert $r_{i,proj}$ für nicht beobachtete r_i kann daher durch Gleichung 2.6 ausgedrückt werden.

$$r_{i,proj} = \mathbf{w}_i^T \mathbf{x}_i \quad (2.6)$$

Für diese Methodik können also drei verschiedene Typen von Messungen festgestellt werden:

1. Eine LIDAR-Messung kann zu einer Kameramessung assoziiert werden. Damit ist w_i mithilfe der Kamera beobachtet und r_i durch den LIDAR bekannt.
2. Es kann keine LIDAR-Messung zur Kameramessung assoziiert werden. Hier nutzen wir Gleichung 2.6 um r_i mit $r_{i,proj}$ abzuschätzen, w_i erhalten wir weiterhin von der Kamera.
3. Ein Punktecluster nahe einer vorherigen Schätzung kann aus dem LIDAR extrahiert werden. In diesem Fall können sowohl r_i , als auch w_i aus dem Mittelstrahl des Clusters extrahiert werden. Weitere Informationen zur Cluster-Extraktion finden sich in Abschnitt 2.1.5.

Bemerkenswert ist hierbei, dass wir die Fahrradfahrerbewegung mit einer linearen Methode schätzen können. Die Rechenzeit ist daher unter $1ms$ mit einer Fenstergröße von $n = 12$ Bildern auf einem handelsüblichen Laptop. Im Allgemeinen hat sich $n = 12$ als ein guter Kompromiss zwischen Modellierung nicht-linearer Bewegung und Ausgleich des Messrauschens erwiesen.

Clustering und Assoziation von LIDAR- und Kameramessungen

Um die vom Laserscanner gemessenen Punktwolken nutzen zu können, identifizieren wir charakteristische Muster in der Wolke, welche zum Fahrradfahrer gehören könnten. Zuerst wird daher die Punktwolke in Cluster segmentiert, welche sich durch un stetige Tiefe von einander abheben. Implementiert ist dies durch die *Linkage*-Metrik von Frank Moosmann [Moosmann 13]. Zu jeder Detektion aus der Kamera möchten wir das zugehörige Punktwolken-Cluster assoziieren.

Dazu nutzen wir die folgenden Kriterien:

1. Der zu erwartende Abstand des Fahrradfahrers zur Kamera.

2. Der räumliche Abstand des Sichtstrahls des Fahrradfahrers zum Cluster.

Für das erste Kriterium nutzen wir vorherige Schätzungen um die Position des Fahrradfahrers vorherzusagen und zusätzlich dessen detektierte Höhe und Breite im Bild. Nimmt man eine durchschnittliche Höhe für den Fahrradfahrer als gegeben an, kann man mithilfe der Kameraparameter dessen zu erwartende Höhe im Bild schätzen. Die Tiefe z ergibt sich also zu $z = f \frac{\Delta h}{\Delta v}$ mit f der Brennweite der Kamera, Δh der Höhe des Fahrradfahrers in Metern und Δv der im Bild gemessenen Höhe in Pixeln. Mithilfe einer Taylorreihenentwicklung erster Ordnung um Δh und Δv , können wir den Fehler in z fortpflanzen und erhalten

$$\sigma_z = \frac{f}{\Delta v} \sqrt{\left(\frac{\Delta h}{\Delta v}\right)^2 \cdot \sigma_{\Delta v}^2 + \sigma_{\Delta h}^2} \quad (2.7)$$

Um diese Information zur Assoziation nutzen zu können müssen sie auf das Intervall $[0, 1]$ skaliert werden. Dazu nutzen wir die Normalverteilungsfunktion $\mathbb{N}(z, \sigma_z)$.

Als zweites Kriterium zur Zuordnung der Cluster, wird der räumliche Abstand des Cluster-Mittelpunkts \mathbf{m} zum durch die Fahrradfahrerdetektion bekannten Sichtstrahl \mathbf{r} genutzt. Der Sichtstrahl \mathbf{r} folgt direkt aus der Anwendung des inversen Kameramodells auf den Mittelpunkt der Detektion aus der Bildebene.

Nun kann der minimale Abstand d von \mathbf{r} zu \mathbf{p} mithilfe von Gleichung 2.8 ermittelt werden.

$$d = (\mathbf{r}^T \mathbf{p}) \mathbf{r} - \mathbf{p} \quad (2.8)$$

Um beide Kriterien vergleichen zu können, skalieren wir d wiederum auf das Intervall $[0, 1]$ und kombinieren beide Kriterien mithilfe einer *Fuzzy-Logik*. Dies ist mithilfe einer parametrisierten Sigmoidfunktion $\text{sigm}(x, o, g) = \frac{1}{1 + g \cdot \exp(x - o)}$ umgesetzt, wobei o den offset in x und g den exponentiellen Verstärkungsfaktor darstellen und angepasst werden müssen. Die hier verwendete Sigmoidfunktion ist in Abbildung 2.12 zu sehen.

Somit werden Abstände bis zu ca. $2m$ toleriert, bevor der Wert der Sigmoidfunktion stark absinkt.

Für jedes potentielle Cluster-Bild-Paar wird anschließend der Mittelwert der beiden Kriterien gebildet und die am besten bewerteten Paarungen für die Bewegungsschätzung verwendet.

Kamera-Laser-Kalibrierung

Um zu wissen welches Pixel im Bild dem gemessenen Abstand des Lasers entspricht muss die Translation und Rotation, die das Laserkoordinatensystem in das Kamerakoordinatensystem überführt, bekannt sein. Die geringe Auflösung sowie das große Tiefenrauschen des Laserscanners stellten uns hierbei vor eine Herausforderung, da hierfür keine Kalibrierungsalgorithmen verfügbar sind. Darum wurde im Rahmen dieses Projektes eine neuartige Laserscanner zu Kamerakalibrierung entwickelt, die die genaue Bestimmung dieser Transformation ermöglicht. Kernidee ist hierbei, dass die Wellenlänge des emittierten Laserlichts zwar nicht für das menschliche Auge, aber für den Lichtsensor der Kamera sichtbar ist. Somit können die Lichtpunkte mit deren Hilfe die Laufzeitmessung für den LIDAR geschieht in der Kamera

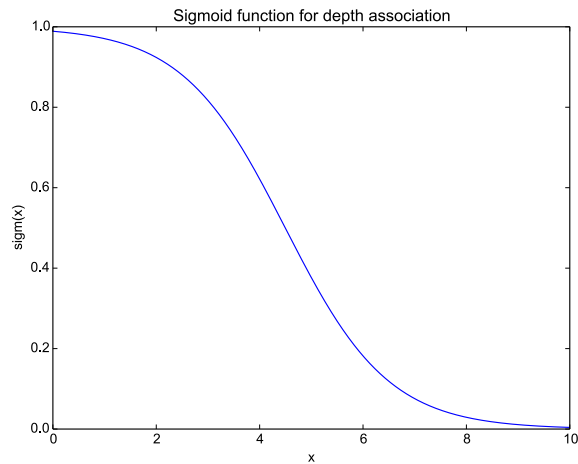


Abbildung 2.12: Sigmoidfunktion für offset $o = 1.5$ and Verstärkungsfaktor $g = 0.05$.

beobachtet werden, wie in Abbildung 2.13 zu sehen. Nun nutzt man mehrere solcher Aufnahmen und wendet darauf eine Art Bündelausgleich wie in Abschnitt 2.1.4 beschrieben an, wobei jedoch die 3D-Position der Landmarken durch die Lasermessung bekannt ist. Somit kann die Pose zwischen Kamera und Laserscanner trotz geringer Auflösung des Laserscanners und dessen hohe Ungenauigkeit ermittelt werden. Weitere Einzelheiten zu dieser Methode sind in [Gräter 16] zu finden.

2.2 Ergebnisse aus Probandenstudie

Zu Projektende wurde eine praktische Studie mit 6 Probanden und einem LKW vom Projektpartner Meyer&Meyer auf eine echte Straße durchgeführt. Während diesen Tests konnte einerseits das generelle Funktionsprinzip des Systems vor allem bezüglich der Warnsignale für Fahrradfahrer beim Rechtsabbiegen des LKWs evaluiert werden, andererseits konnte die Robustheit der Fahrradfahrerererkennung und Bewegungsschätzung in realitätsnahen Verkehrsszenarien getestet werden. Grundlage für die Studie waren die kompakt gebauten Sensoren (Abbildung 2.1), die den Fahrradfahrer und seinen Abstand zum Sensor jeweils durch Kamerabild und Laserstrahlen aufnehmen. Dazu stand auch ein Rechner zur Verfügung, der die Bild- und Laserdaten auswertet. Zusätzlich wurden die Warnsignale durch die beim Partner HFC entwickelten Schnittstelle auf einem Bildschirm angezeigt, damit der LKW-Fahrer den Fahrradfahrer in gefährliche Szenarien rechtzeitig erkennen kann.

Diese insgesamt achtstündige Studie war in zwei Phasen gegliedert. Sie fand im Innovationspark Wuhlheide in Berlin statt - einem realistischen innerstädtischen Szenario mit anspruchsvollen, engen Kurven und variablem Abstand der Fahrradwege zum Fahrzeug.

In der ersten Phase wurde die Robustheit und die Detektionsgenauigkeit des Systems getestet. Der Fahrradfahrer fuhr mit unterschiedlichen Geschwindigkeiten und Abständen an der rechten Seite des LKWs. Es wurden diverse Szenarien mit Überholen, Ein- und Ausscheren

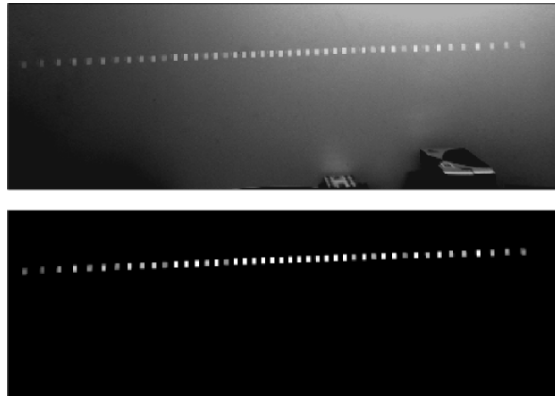


Abbildung 2.13: Oben sieht man die Szene mit Licht, der Laserscanner ist auf eine Wand gerichtet, hier sind die LIDAR-Messpunkte schon schwach sichtbar. Im unteren Bild wird das Umgebungslicht ausgeschaltet, sodass Bedingungen wie in einer Dunkelkammer herrschen. Somit werden die Punkte an der Wand, an welchen der LIDAR misst, deutlicher sichtbar.

und Verdeckung des Fahrradfahrers behandelt. Somit konnten reale Verkehrsszenarien so gut wie möglich im Test simuliert und ausgewertet werden.

Die zweite Phase bezog die Wahrnehmung des LKW-Fahrer mit ein, wobei ähnliche Testszenarien wie in Phase 1 wiederholt wurden. Mithilfe des Feedbacks des LKW-Fahrers zum Nutzen der Warnung konnte das System anwenderfokussiert evaluiert werden.

Während der gesamten Dauer der Testfahrten arbeiteten die Verfahren zur Erkennung von Fahrradfahrern sowie zur Bestimmung von deren Bewegung sowie der Eigenbewegung des LKW zuverlässig und robust. Lediglich bei schwierigen Lichtverhältnissen sowie bei umfangreicher Verdeckung der Fahrradfahrer konnten diese im Kamerabild nicht detektiert werden. Dennoch wurde das System von den beteiligten LKW-Fahrern als ausgesprochen nützlich und hilfreich empfunden. Eine genauere Auswertung der Probandenstudie ist im Bericht des Projektpartners HFC zu finden.

2.3 Bewertung der Projektergebnisse

Im Rahmen des Projektes ABALID ist ein System entstanden, das einem LKW-Fahrer gezielt Informationen über kritische Situationen im Stadtverkehr zur Verfügung stellt und damit einen Beitrag dazu leistet, Unfälle zwischen LKW und Fahrradfahrern zu verhindern. Im Rahmen des Teilprojektes *3D-Objekterkennung und semantische Analyse* hat hierzu die Algorithmen zur echtzeitfähigen Auswertung der kombinierten Kamera/LIDAR-Sensorik entwickelt und in ein für mobile Anwendungen geeignetes Softwareframework integriert. Beim Entwurf des Systems wurden neben den technischen Randbedingungen stets auch die Anforderungen aus Nutzersicht berücksichtigt, um nicht nur eine technisch ausgereifte, sondern auch eine anwenderfreundliche, nützliche Lösung zu entwickeln.

Eine Besonderheit des Projektes stellte das unkonventionelle Sensorsystem aus einer monoskopischen Kamera und einem hierzu unsynchronisierten LIDAR-System mit niedriger



(a)



(b)

Abbildung 2.14: Durchführung der Tests im realen Szenario: der LKW versuchte bei der Kreuzung rechts abzubiegen mit einem Fahrradfahrer an der rechten Seite vorbeizufahren. (a) zeigt ein schwierigeres Szenario als (b), da der Fahrradfahrer zur Zeit durch die Bäume zwischen dem Radweg und der Fahrbahn teilweise verdeckt ist.

Auflösung dar. Die Wahl dieses Systems war in einer kostengünstigen, seriennahen Konzeption des Systems begründet, stellte andererseits aber auch erhebliche Anforderungen an die zu entwickelnde Sensordatenauswertung. Aufgrund der wenigen Arbeiten, die sich bisher wissenschaftlich mit solchen Sensorplattformen befassen, waren umfangreiche und grundlegende Arbeiten erforderlich. Bedeutende Neuentwicklungen umfassten

- eine Kreuzkalibrierung zwischen LIDAR und Kamera
- eine monoskopische Eigenbewegungsschätzung mit hoher Genauigkeit
- einen robusten, bildbasierten Fahrradfahrerdetektor und -tracker
- eine robuste Bewegungsschätzung für Fahrradfahrer basierend auf der Kombination von Video- und LIDAR-Sensorik

Alle Verfahren mussten in Echtzeit auf einer mobilen Rechnerplattform mit begrenzter Rechenkapazität arbeiten können, so dass neben der erreichbaren Genauigkeit auch die erforderliche Rechenleistung zu berücksichtigen waren. In mehreren Fahrzeugtests wurde die Software auf einem handelsüblichen Laptop aus dem Jahr 2013 ausgeführt und damit gezeigt, dass dieses Ziel erreicht wurde. Angesichts der kontinuierlich steigenden Leistungsfähigkeit der Rechner (Moorsches Gesetz) kann geschlussfolgert werden, dass bei heutigen Rechnern sogar noch Leistungsreserven verfügbar sind.

Trotz der Echtzeitbedingungen haben alle Algorithmen auch im Hinblick auf die Genauigkeit ihrer Ergebnisse gute bis sehr gute Ergebnisse geliefert. Die monoskopische Eigenbewegungsschätzung hat eine sehr geringe Drift und ist daher auch über mehrere Hundert Meter Wegstrecke zuverlässig. Die Fahrradfahrererkennung, ergänzt um den Fahrradfahrertracker, ist in der Lage, Fahrradfahrer zuverlässig über lange Zeiträume hinweg zu verfolgen. Allerdings stellen größere Verdeckungen des Fahrradfahrers, z.B. durch parkende Fahrzeuge, Gebüsch,

Masten, Bänke oder Mülltonnen eine Schwierigkeit dar, die noch nicht vollständig beherrscht wird. Daher stellen Fahrradfahrer auf baulich getrennten, von Parkstreifen oder Baumreihen getrennten Fahrradwegen, für das im Rahmen des Projektes ABALID entwickelte System noch immer eine Schwierigkeit dar.

Zwar wurden die genannten Verfahren in erster Linie für die im Projekt ABALID konzipierte Fahrerassistanzanwendung entwickelt, allerdings sind die Verfahren auch für eine Vielzahl anderer Anwendungen im Bereich Fahrerassistenz und autonome Fahrzeuge, Robotik und Videoauswertung von großem Nutzen und versprechen daher Verbesserungen auch in weiteren Bereichen über das Projekt ABALID hinaus. Die entwickelte Methodik ist grundsätzlich auch für die Erkennung von Fußgängern und anderen Verkehrsteilnehmern geeignet. In Rahmen von fünf wissenschaftlichen Konferenzbeiträgen wurden die Arbeiten veröffentlicht. Ferner wurden eine Masterarbeiten und ein studentisches Forschungsprojekte [Zou 15, Vuyets 15] zu einzelnen Fragestellungen erstellt.

Neben der technischen Reife des Systems wurde mit den abschließenden Probandentests auch die Nützlichkeit des Systems für LKW-Fahrer gezeigt. Der Testfahrer empfand das System als sehr hilfreich und nützlich und er würde ein derartiges System – sobald es serienreif ist – gerne weitenutzen.

Allerdings haben sich im Rahmen des Projektes auch einige Schwächen herausgestellt, die mit dem erarbeiteten System noch nicht vollständig gelöst werden konnten. Neben der bereits erwähnten Detektion teilweise verdeckter Fahrradfahrer stellen schwierige Lichtverhältnisse (Gegenlicht, Dunkelheit) das größte Problem dar. Da das erarbeitete System sich wegen der geringen Auflösung des LIDARs größtenteils auf die Kamerabilder stützt, führt eine Störung der Kamerabilder zwangsläufig zu Fehlern in der Wahrnehmung. Dies könnte in einem weiterentwickelten System durch einen höher aufgelösten Sensor (LIDAR und/oder RADAR) kompensiert werden. Überdies hinaus bieten auch neuere Entwicklungen im Bereich derameratechnik bzw. der Kamerabildauswertung Potentiale für Verbesserungen auf der Basis der Video-Sensorik.

Das nun bestehende System stellt sowohl im Hinblick auf die Hardware wie auch auf die Software einen Prototypen für ein neuartiges Abbiegeassistenzsystem für LKW dar, das mit hoher Genauigkeit die Fahrradfahrer in der Fahrzeugumgebung detektieren sowie ihre Bewegungen schätzen kann, damit LKW-Fahrradfahrer-Unfälle vermieden werden können. Viele der realisierten Verfahren zur Objekterkennung und Bewegungsschätzung sind geeignet, auch in einem späteren Produkt genutzt zu werden.

Anhang A

Veröffentlichungen

A.1 Publikationen, die innerhalb des Projektes entstanden

- [Gräter 15a] J. Gräter & M. Lauer. Monoscopic automotive ego-motion estimation and cyclist tracking. In: Doctoral Consortium at International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2015.
- [Gräter 15b] J. Gräter, T. Schwarze & M. Lauer. Robust scale estimation for monocular visual odometry using structure from motion and vanishing points. In: IEEE Intelligent Vehicles Symposium, S. 475–480, 2015.
- [Gräter 16] J. Gräter, T. Strauß & M. Lauer. Photometric laser scanner to camera calibration for low resolution sensors. In: IEEE Intelligent Transportation Systems Conference, 2016.
- [Tian 15a] W. Tian & M. Lauer. Fast and robust cyclist detection for monocular camera systems. In: Doctoral Consortium at International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 2015.
- [Tian 15b] W. Tian & M. Lauer. Fast cyclist detection by cascaded detector and geometric constraint. In: IEEE Conference on Intelligent Transportation Systems, S. 1286–1291, 2015.
- [Vuyets 15] T. Vuyets. Evaluation of different approaches for bicyclist recognition. Forschungsprojekt, Karlsruhe School of Optics and Photonics, 2015.
- [Zou 15] T. Zou. Decision tree based deformable part models. Masterarbeit, Karlsruher Institut für Technologie, 2015.

A.2 Literaturverzeichnis

- [Bay 06] H. Bay, T. Tuytelaars & L. Van Gool. SURF: Speeded up robust features. In: European Conference on Computer Vision, S. 404–417, 2006.
- [Beauchemin 95] S. S. Beauchemin & J. L. Barron. The computation of optical flow. ACM Computing Surveys, Band 27, Nr. 3, S. 433–466, 1995.
- [Benenson 14] R. Benenson, M. Omran, J. Hosang & B. Schiele. Ten years of pedestrian detection, what have we learned? IEEE European Conference on Computer Vision, S. 613–627, 2014.
- [Calonder 12] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha & P. Fua. BRIEF: Computing a local binary descriptor very fast. IEEE Transactions on Pattern Analysis and Machine Intelligence, Band 34, Nr. 7, S. 1281–1298, 2012.
- [Dalal 05] N. Dalal & B. Triggs. Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition, S. 886–893, 2005.
- [Dharmaraju 01] R. Dharmaraju, D. A. Noyce & J. D. Lehman. An evaluation of technologies for automated detection and classification of pedestrians and bicyclists. In: 71st ITE Annual Meeting, 2001.
- [Dollár 12] P. Dollár, C. Wojek, B. Schiele & P. Perona. Pedestrian detection: An evaluation of the state of the art. IEEE Transactions on Pattern Analysis and Machine Intelligence, Band 34, S. 743–761, 2012.
- [Fischler 81] M. A. Fischler & R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, Band 24, Nr. 6, S. 381–395, 1981.
- [Fraundorfer 12] F. Fraundorfer & D. Scaramuzza. Visual odometry: Part II: Matching, robustness, optimization, and applications. IEEE Robotics & Automation Magazine, Band 19, Nr. 2, S. 78–90, 2012.
- [Geiger 11] A. Geiger, J. Ziegler & C. Stiller. Stereoscan: Dense 3d reconstruction in real-time. In: IEEE Intelligent Vehicles Symposium, S. 963–968. IEEE, 2011.
- [Geiger 12] A. Geiger, P. Lenz & R. Urtasun. Are we ready for autonomous driving? the KITTI vision benchmark suite. In: IEEE Conference on Computer Vision and Pattern Recognition, S. 3354–3361, 2012.
- [Harris 88] C. Harris & M. Stephens. A combined corner and edge detector. In: Alvey Vision Conference, S. 147–151, 1988.

- [Hartley 03] R. Hartley & A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [Kalman 60] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, Band 82, Nr. 1, S. 35–45, 1960.
- [Krogmeier 08] J. V. Krogmeier & D. M. Bullock. Statewide wireless communications project. volume 2: Inductive loop detection of bicycles and inductive loop signature processing for travel time estimation. Forschungsbericht, Purdue University, 2008.
- [Lategahn 14] H. Lategahn, J. Beck & C. Stiller. Dird is an illumination robust descriptor. In: *IEEE Intelligent Vehicles Symposium*, S. 756–761. IEEE, 2014.
- [Moosmann 13] F. Moosmann. *Interlacing Self-Localization, Moving Object Tracking and Mapping for 3d Range Sensors*. Dissertation, Karlsruher Institut für Technologie, 2013.
- [Papageorgiou 98] C. Papageorgiou, T. Evgeniou & T. Poggio. A trainable pedestrian detection system. In: *IEEE Intelligent Vehicles Symposium*, S. 241–246, 1998.
- [Rublee 11] E. Rublee, V. Rabaud, K. Konolige & G. Bradski. ORB: An efficient alternative to SIFT or SURF. In: *International Conference on Computer Vision*, S. 2564–2571. IEEE, 2011.
- [Scaramuzza 11] D. Scaramuzza & F. Fraundorfer. Visual odometry. *IEEE Robotics & Automation Magazine*, Band 18, Nr. 4, S. 80–92, 2011.
- [Sudowe 11] P. Sudowe & B. Leibe. Efficient use of geometric constraints for sliding-window object detection in video. In: *International Conference on Computer Vision Systems*, S. 11–20, 2011.
- [Thrun 05] S. Thrun, W. Burgard & D. Fox. *Probabilistic Robotics*. MIT Press, 2005.
- [Zhang 16a] S. Zhang, R. Benenson, M. Omran, J. Hosang & B. Schiele. How far are we from solving pedestrian detection? *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

Berichtsblatt

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Schlussbericht
3. Titel ABALID: Abbiegeassistentz mit 3D-LIDAR Sensorik Abschlussbericht des Teilprojektes 3D-Objekterkennung und semantische Analyse	
4. Autor(en) [Name(n), Vorname(n)] Gräter, Johannes Tian, Wei Lauer, Martin Stiller, Christoph	5. Abschlussdatum des Vorhabens 31.5.2016
	6. Veröffentlichungsdatum
	7. Form der Publikation Broschüre
8. Durchführende Institution(en) (Name, Adresse) Karlsruher Institut für Technologie Institut für Mess- und Regelungstechnik Engler-Bunte-Ring 21 76131 Karlsruhe	9. Ber. Nr. Durchführende Institution -
	10. Förderkennzeichen 16SV6029
	11. Seitenzahl 27
12. Fördernde Institution (Name, Adresse) Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 30
	14. Tabellen 3
	15. Abbildungen 13
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum) Teilergebnisse wurden bereits auf verschiedenen Fachkonferenzen vorgestellt (siehe Anhang A1 im Schlussbericht)	
18. Kurzfassung Das Teilprojekt 3D-Objekterkennung und semantische Analyse beschäftigte sich im Rahmen des Projektes ABALID mit der Entwicklung von Sensordatenauswerteverfahren für eine kombinierte Sensorplattform bestehend aus einer monoskopischen Kamera und einem LIDAR-Sensor. Ziel war die Erfassung der Eigenbewegung der Sensorplattform, die Erkennung und Verfolgung von Fahrradfahrern im Erfassungsbereich der Sensorik sowie die Schätzung der Bewegung der detektierten Fahrradfahrer. Die Sensorplattform stellt einen Bausteins eines im Projekt ABALID entwickelten Toter-Winkel-Warn-Systems für LKWs dar, das Unfälle zwischen rechtsabbiegenden LKWs und Fahrradfahrern verhindern soll. Zur Realisierung der Umfeldwahrnehmung wurden bildbasierte Fahrradklassifikatoren mit bildbasierten Trackingverfahren kombiniert. Der Abstand der Fahrradfahrer wurde aus den LIDAR-Messungen übernommen, wofür eine LIDAR-zu-Kamera-Kalibrierung entwickelt wurde. Die Eigenbewegungsschätzung wurde mit Hilfe von visueller Odometrie und Bündelausgleich realisiert, wobei zur Skalenschätzung eine dünne 3D-Rekonstruktion realisiert wurde. Im Ergebnis wurde eine robuste Sensorauswerteeinheit realisiert, die die Eigenbewegungsschätzung mit hoher Genauigkeit und die Fahrradfahrerdetektion mit großer Zuverlässigkeit gewährleistet. Alle Verfahren wurden im Hinblick auf Effizienz entwickelt, um im mobilen Einsatz mit begrenzter Rechenleistung eine echtzeitfähige Verarbeitung zu gewährleisten. Schwächen zeigte das System vor allem bei teilweiser Verdeckung der Fahrradfahrer durch andere Objekte sowie bei ungünstigen Beleuchtungsverhältnissen. Der abschließende Probandentest bestätigte einerseits die technische Reife der Lösung als auch die Nützlichkeit zur Unterstützung der LKW-Fahrer im Stadtverkehr.	
19. Schlagwörter Fahrerassistenz, Umfelderkennung, monoskopische Kamera, LIDAR, Fahrradfahrer-Detektion, Eigenbewegungsschätzung	
20. Verlag	21. Preis

Document Control Sheet

1. ISBN or ISSN	2. type of document (e.g. report, publication) final report
3. title ABALID: Turn assist with 3D-LIDAR sensor Final report of subproject 3D obstacle detection and semantical analysis	
4. author(s) (family name, first name(s)) Gräter, Johannes Tian, Wei Lauer, Martin Stiller, Christoph	5. end of project 31.5.2016
	6. publication date
	7. form of publication brochure
8. performing organization(s) (name, address) Karlsruher Institute of Technology Institute of Measurement and Control systems Engler-Bunte-Ring 21 76131 Karlsruhe	9. originator's report no. -
	10. reference no. 16SV6029
	11. no. of pages 27
12. sponsoring agency (name, address) Federal Ministry of Education and Research 53170 Bonn	13. no. of references 30
	14. no. of tables 3
	15. no. of figures 13
16. supplementary notes	
17. presented at (title, place, date) Partial results have been published at different symposiums (refer to Appendix A1 of final report)	
18. abstract The subproject <i>3D obstacle detection and semantical analysis</i> as part of the project ABALID addressed the development of algorithms for analyzing measurements from a combined camera/lidar sensor setup. The aim of the project was to determine the ego motion of the sensor setup, to detect and track bicycle riders, and to estimate their movement. The sensor setup is one piece within the project ABALID that aimed at developing a death angle warn assist for trucks that is designed to reduce the number of accidents between trucks and bicycles at crossroads. The design of the perception methods was mainly based on a camera based bicycle rider detector and an image based bicycle tracker. The distance of the bicycles from the sensor setup was evaluation based on the lidar measurements which required an accurate cross calibration between lidar sensor and camera which was also developed in the project. The ego motion estimation was based on visual odometry and bundle adjustment. The scale was estimated from a sparse 3d reconstruction of the scene. In total, a robust system for environment perception was implemented which allows to determine the ego motion with high precision and to detect and track bicycles reliably. All approached were designed to operate in real time to allow their usage in mobile applications. Shortcomings of the developed approaches have been encountered in situations with partly occluded bicyclists and in cases of adverse lighting conditions. The final test of the system on a truck has shown the technical reliability of the system as well as its usefulness to support truck drivers in urban traffic.	
19. keywords driver assistance, environment perception, monocular camera, lidar, bicycle detection, ego motion estimation	
20. publisher	21. price