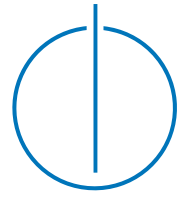




TECHNISCHE UNIVERSITÄT
MÜNCHEN
Institut für Informatik

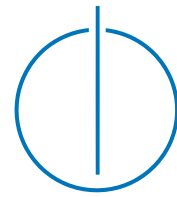


Maschinelle Erkennung von Verkehrsteilnehmern mittels heterogener Sensorik

Leonhard F. Walchshäusl



TECHNISCHE UNIVERSITÄT
MÜNCHEN
Institut für Informatik



Maschinelle Erkennung von Verkehrsteilnehmern mittels heterogener Sensorik

Leonhard F. Walchshäusl

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Bernd Radig

Prüfer der Dissertation:

1. Univ.-Prof. Gudrun J. Klinker, Ph.D.
2. Univ.-Prof. Dr. Darius Burschka

Die Dissertation wurde am 28.05.2008 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 15.10.2008 angenommen.

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

An erster Stelle steht meine Doktormutter Frau Prof. Gudrun Klinker, die mich während der letzten drei Jahre beständig betreut hat. Die ganze Zeit hindurch hatte sie stets ein offenes Ohr für meine arbeitsbezogenen Sorgen und Probleme. Für die zahllosen Anregungen und Impulse, sowie für ihren freundschaftlichen Umgang bin ich ihr sehr dankbar.

Des Weiteren danke ich Herrn Prof. Darius Burschka für die sofortige Bereitschaft das Zweitgutachten zu übernehmen, sowie für die wertvollen Ratschläge in der Endphase der Arbeit.

Ich danke der BMW Forschung und Technik GmbH und besonders Herrn Dr. Klaus Gresser, der mir die Durchführung dieser Arbeit ermöglichte. Vielen Dank für die zahlreichen unbürokratischen Hilfestellungen, sowie für den uneingeschränkten Rückhalt in der Anfangsphase der Dissertation. Hierfür bin ich gleichermaßen Herrn Dr. Dirk Wisselmann zu Dank verpflichtet.

Katrin Vogel, meiner Betreuerin bei BMW, möchte ich für die geförderte freie Arbeitsweise und die freundliche Unterstützung meinen Dank aussprechen. Dr. Marc Walessa danke ich für die fachlichen Anregungen und das wertvolle Feedback, sowie für die angenommenen Druckaufträge am Ende der Arbeit.

Meinem Tischkollegen und Studienfreund Rudi Lindl danke ich für die beispiellose Zusammenarbeit, die gemeinsam bewältigten Problemstellungen und das wechselseitige Aufrichten in schwierigen Phasen. Ferner danke ich allen Kollegen aus dem Sensorikteam – insbesondere den Doktorandenkollegen Antje Neve und Andreas Laika – für die konstruktiven fachlichen Diskussionen.

Weiterhin danke ich Vitali Lebsack, Florian Homm, Christina Gackstatter und Philipp Kemmeter, die sich in Form von Diplomarbeiten und Werkstudententätigkeiten aktiv an der Realisierung dieser Arbeit beteiligt haben.

Vielen Dank auch allen, die Anmerkungen und Verbesserungsvorschläge zur schriftlichen Ausarbeitung gegeben haben: Gudrun Klinker, Katrin Vogel, Marc Walessa und Miriam Häusler. Vivien und Robert Reed danke ich für das Korrekturlesen meiner englischsprachigen Veröffentlichungen.

Nicht zuletzt danke ich meinen Eltern, ohne deren vielseitige Unterstützung eine solche Arbeit schwerlich umsetzbar gewesen wäre. Meiner Verlobten Miriam Häusler danke ich für ihre Liebe und Fürsorge, sowie für das in Kauf nehmen der vielen Abende, die ich mit meinem viereckigen Freund verbracht habe.

München, im Mai 2008

Leonhard Walchshäusl

Zusammenfassung

In modernen Fahrzeugen halten verstärkt Fahrerassistenzsysteme zum Insassenschutz und zum Schutz anderer Verkehrsteilnehmer Einzug. Typische, bereits seit längerem verfügbare Systeme, wie Airbag und Antiblockiersystem, erreichen schon heute eine hohe Marktdurchdringung. Künftige aktive Sicherheitssysteme, die von der Kollisionswarnung bis hin zur autonomen Gefahrenbremsung reichen, sind Gegenstand aktueller Forschung. Insbesondere Systeme, die aktiv in das Fahrgeschehen eingreifen, erfordern eine hinreichend robuste und genaue maschinelle Wahrnehmung des Fahrumfeldes mittels bild- und/oder entfernungsgebender Sensorik.

Die vorliegende Arbeit untersucht, inwieweit durch eine verschränkte Betrachtung von heterogenen Sensordaten die Perzeptionszuverlässigkeit eines automotiven Sensorsystems gesteigert werden kann. Der Schwerpunkt liegt dabei auf Methoden zur Detektion und Verifikation von Fahrzeugen und Fußgängern. Dafür wird sowohl das Potential adäquater Merkmalsextraktionsverfahren als auch verschiedener Ensembletechniken zur Mustererkennung untersucht. Datenunvollständigheiten, die aufgrund inkongruenter Sensorerfassungsbereiche und unterschiedlicher Abstraten auftreten, werden gesondert diskutiert. Als Bestandteil einer einsatzfähigen Fahrumfelderfassung werden verschiedene Ansätze anhand von repräsentativen Testsequenzen, hinsichtlich Laufzeit und Erkennungsrate, gegenübergestellt und bewertet. Zur aussagekräftigen Durchführung der Evaluierung wurden speziell angepasste Bewertungstechniken entwickelt. Eine räumliche Datenvisualisierung vereinfachte die Systemumsetzung und ermöglicht eine medienwirksame Präsentation der Ergebnisse. Die Untersuchungen belegen einen Nutzen von multi- gegenüber monosensoriellen Systemen zur Erkennung von Verkehrsteilnehmern in künftigen Fahrerassistenzsystemen.

Abstract

Modern motor vehicles are increasingly being fitted with driver assistance systems, which are designed to protect vehicle occupants as well as other road users. A number of systems are well established and available, for example, airbags and anti-lock braking systems, both of which have a high market penetration. Future interactive safety systems range from collision warning to autonomous emergency braking, both of which are the subject of current research and development. These systems actively override the driver and require an extremely robust and accurate computer application which processes the perception awareness of the environment by means of imaging and/or ranging sensors.

This thesis examines the extent to which the performance of an automotive sensor system can be improved by combining heterogeneous sensor data. The main focus for improvement is on methods for the detection and verification of vehicles and pedestrians. Therefore, the potential of adequate feature extraction procedures and different ensemble methods for pattern recognition are examined. Missing features are considered separately. Various approaches are tested with respect to run-time and recognition rate. The testing data covers a number of representative driving scenarios. Special adapted assessment techniques have been developed, in order to provide a meaningful testing environment. A working integrated system is currently installed in an experimental car. A spatial visualization module simplified the system realisation and allows for presentations with public appeal. The results demonstrate a performance improvement of multisensory systems compared to single sensor systems which will provide greater safety to all road users.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	2
1.2	Randbedingungen und Zielsetzung	3
1.3	Ergebnisvorschau	4
1.4	Aufbau der Arbeit	5
2	Grundlagen	11
2.1	Fahrerassistenzsysteme	11
2.1.1	Überblick	12
2.1.2	Aktive Gefahrenbremsung	14
2.2	Automotive Sensorik	16
2.2.1	Funktionsweise und Verbreitung	16
2.2.2	Taxonomie und Vergleich	18
2.3	Sensorfusion	20
2.3.1	Vorteile	20
2.3.2	Fusionsmodelle	21
2.3.3	Fusionstaxonomien	22
2.3.4	Projektlandschaft im Kontext automotiver Fusionssysteme	24
2.4	Versuchsträger	25
2.4.1	Aufbau und Sensorsystem	25
2.4.2	Koordinatensysteme	26
2.5	Systemüberblick	30
3	Hypothesengenerierung	33
3.1	Entfernungsgebende Sensorik zur Hypothesenerzeugung	35
3.1.1	Segmentierungsverfahren	35
3.1.2	Einteilung der Segmente	40
3.1.3	Hypotheseninitialisierung	43
3.1.4	Bewertung	45
3.2	Bildgebende Sensorik zur Hypothesenerzeugung	46
3.2.1	„U“-Form für Fußgängerhypothesen im Wärmebild	46
3.2.2	Vertikale Kanten für Fahrzeughypothesen im Videobild	49

3.3	Verbunderzeugung und Bewertung	50
3.3.1	Zusammenführung von Fußgängerhypothesen	50
3.3.2	Hypothesenkombination und Bewertung der Aufmerksamkeitssteuerung	53
3.4	Zusammenfassung	58
4	Merkmalsgewinnung	63
4.1	Datensegmentierung	65
4.2	Gradientenhistogramme als Bildmerkmale	66
4.2.1	Parametervariation	68
4.2.2	Skalierungsinvarianz	70
4.2.3	Merkmalsrekonstruktion durch Achsenspiegelung	73
4.3	Reflektanzsignatur von Fahrzeugheckansichten	77
4.4	Bein-Pendelbewegung von Fußgängern	80
4.5	Merkmale aus der Kursverfolgung	82
4.6	Vergleich	84
4.7	Zusammenfassung	85
5	Hypothesenverifizierung	89
5.1	Monolithische Klassifikation	91
5.2	Ensembleverfahren	92
5.2.1	Ausgabenkalibrierung von SVM	94
5.2.2	Funktionale Kombinatoren	95
5.2.3	Metaklassifikation	99
5.3	Diskontinuierliches Merkmalsaufkommen	100
5.4	Bewertung und Vergleich	101
5.5	Zusammenfassung	103
6	Systembewertung	109
6.1	Referenzdatengewinnung	110
6.1.1	Semiautomatische Referenzdatenerzeugung	111
6.1.2	Algorithmusgestützte Referenzdatenerzeugung	114
6.1.3	Referenzsensor	117
6.1.4	Hybride Referenzdatenerzeugung	119
6.1.5	Zusammenfassung und Bewertung	120
6.2	Evaluierung	120
6.2.1	Metriken	121
6.2.2	Beschreibung der Testszenarien	123
6.2.3	Konfiguration des Testsystems	125
6.2.4	Ergebnisse	126
6.3	Zusammenfassung	134
7	Visualisierung von heterogenen Sensordaten	139
7.1	Grundlagen der Visualisierung	140
7.1.1	Daten und Informationsfluss	140
7.1.2	Designrichtlinien	140
7.2	Visualisierung des Informationsflusses	143

7.2.1	Sensorverbauorte und -erfassungsbereiche	144
7.2.2	Rohdaten und vorverarbeitete Sensordaten	144
7.2.3	Interpretierte Daten	148
7.2.4	Zeitlicher Verlauf der Messungen	149
7.2.5	Navigation und Interaktion	149
7.2.6	Werkzeuge der Implementierung	154
7.3	Präsentation und Kommunikation	155
7.3.1	Konventionelle Ausgabemedien	155
7.3.2	Erweiterte Realität	155
7.4	Fahrbahnerkennung mittels Flottenbewegung	166
7.5	Diskussion	170
7.6	Zusammenfassung	171
8	Schluss	173
8.1	Wissenschaftlicher Beitrag	173
8.2	Ausblick	175
8.3	Fazit	176
A	Datenkataloge	179
B	Bewertungskriterien für Klassifikatoren	183
B.1	Kontingenztafel	183
B.2	ROC- und DET-Kurve	184
B.3	Grundlagen der Support-Vektor-Maschine	186
C	Multispektrale Kamerakalibrierung	187
C.1	Projektives Kameramodell	187
C.2	Kalibrierungsverfahren	189
C.3	Bewertung	192
D	Exemplarische Systemleistung	195
Tabellenverzeichnis		201
Abbildungsverzeichnis		203
Eigene Veröffentlichungen und betreute Diplomarbeiten		207
Literaturverzeichnis		209

1

Einleitung

Menschen erfassen über ihre fünf verschiedenen Sinne (visuell, auditiv, olfaktorisch, gustatorisch und haptisch) heterogene Aspekte der Realität, die zu einem Gesamteindruck gebündelt werden. Durch die hohe Redundanz in der Wahrnehmung können Ausfälle einzelner Sinnesorgane kompensiert werden. Darüber hinaus ist das menschliche Gehirn in der Lage, mittels der Kombination von Wahrnehmungen, neue Informationen abzuleiten, wie beispielsweise das räumliche Sehvermögen. Ein anschaulicher Nachweis für die Fähigkeit des Denkapparats, Sinnesindrücke zu verknüpfen, kann durch die Beobachtung eines Ventriloquisten und seiner Bauchrednerpuppe erbracht werden. Während einer Darbietung scheint die Stimme des Bauchredners unmittelbar von der Puppe zu stammen. Die auditiven (Stimme des Bauchredners) und visuellen (Lippenbewegung der Bauchrednerpuppe) Modalitäten werden durch einen kognitiven Prozess zusammengeführt. Erst wenn die Entfernung zwischen Bauchredner und Puppe einen gewissen Distanzwert überschreitet, werden die Modalitäten getrennt wahrgenommen und unterschiedlichen Objekten zugeordnet [LK89, How82].

Auch auf einer höheren kognitiven Ebene verknüpfen Menschen vielfältige Informationen. So werden bei wichtigen Entscheidungen Vor- und Nachteile gegeneinander abgewägt. Steht beispielsweise ein medizinischer Eingriff bevor, wird oftmals nicht nur die Meinung eines einzelnen Arztes eingeholt. Vielmehr werden mehrere Experten konsultiert, in der Hoffnung dadurch ein aufschlussreiches Meinungsbild zu erhalten und infolgedessen eine bessere Entscheidung zu fällen.

Die angeführten Beispiele machen deutlich, dass eine Fusion von Informationen Vorteile bietet, und dass in der Natur Mechanismen existieren, um eine solche Zusammenführung zu bewältigen. Dies gibt der Wissenschaft den Ansporn, künstliche Systeme zu erforschen, die analog zu ihren natürlichen Vorbildern mehrere Informationsquellen kombinieren. In der multisensoriellen Mustererkennung beispielsweise werden Sensorausgaben verschränkt betrachtet, um im Datenkollektiv nach Regelmäßigkeiten zu suchen. Von den wissenschaftlichen Erkenntnissen aus diesem Forschungszweig können zahlreiche Anwendungen profitieren.

1.1 Motivation

Nach den Daten des Statistischen Bundesamtes, sinkt seit drei Dekaden, trotz signifikantem Wachstum des Straßenverkehrsaufkommens, die jährliche Zahl der Verkehrstoten. Dieser erfreuliche Trend ist auf eine restriktivere Gesetzgebung, eine verbesserte Verkehrsinfrastruktur und einen kontinuierlichen Zuwachs an Sicherheitsfunktionen in Kraftfahrzeugen zurückzuführen. Dennoch starben im Jahr 2006 immer noch mehr als 5000 Menschen in Deutschland und 40 000 Menschen in Europa im Straßenverkehr [Sta06]. Diese Zahlen machen deutlich, dass auch in Zukunft die Notwendigkeit besteht, neue Maßnahmen zum Schutz von Verkehrsteilnehmern zu ergreifen und voranzutreiben.

Die Automobilhersteller setzen schon seit Jahrzehnten Rückhaltesysteme ein, wie z. B. Airbags und Gurtstraffer, um die Fahrzeuginsassen zu schützen. Durch den enormen Fortschritt in der Rechenleistung, sowie einer beständigen Miniaturisierung von elektronischen Schaltungen, wurden in den vergangenen Jahren eine ganze Bandbreite von aktiven Sicherheitssystemen darstellbar. Fahrassistenzsysteme, wie Bremsen-Antiblockiersystem (ABS) und Elektronisches Stabilitätsprogramm (ESP), die versuchen kritische Fahrsituationen zu entschärfen, erreichen heute eine hohe Marktdurchdringung. Nachtsichtassistenten haben zum Ziel, den Fahrer bei schlechten Lichtverhältnissen rechtzeitig auf ungeschützte Verkehrsteilnehmer aufmerksam zu machen, und Bremsassistenten verstärken den Bremsdruck, wenn die Einleitung einer, durch den Fahrer initiierten, Gefahrenbremsung detektiert wurde.

Eine gänzlich neue Generation von automotiven Sicherheitssystemen bezieht relevante Daten des Fahrumfeldes mit ein. Diese Informationen werden eingeholt, indem das Fahrzeug mittels Sensoren die Fahrumgebung erfasst und auswertet (siehe auch Abbildung 1.1a). Die automotiv genutzte Sensorik beruht auf heterogenen Messprinzipien und kann in bild- (z. B. Video- und Wärmebildkameras) und entfernungsgebende Sensorsysteme (z. B. Radar, Laserscanner und Ultraschall) unterteilt werden. Dieses elektronische Auge ermöglicht dem „sehenden“ Fahrzeug selbsttätig einzuparken, dem Fahrer beim Spurwechsel zu assistieren, indem es den toten

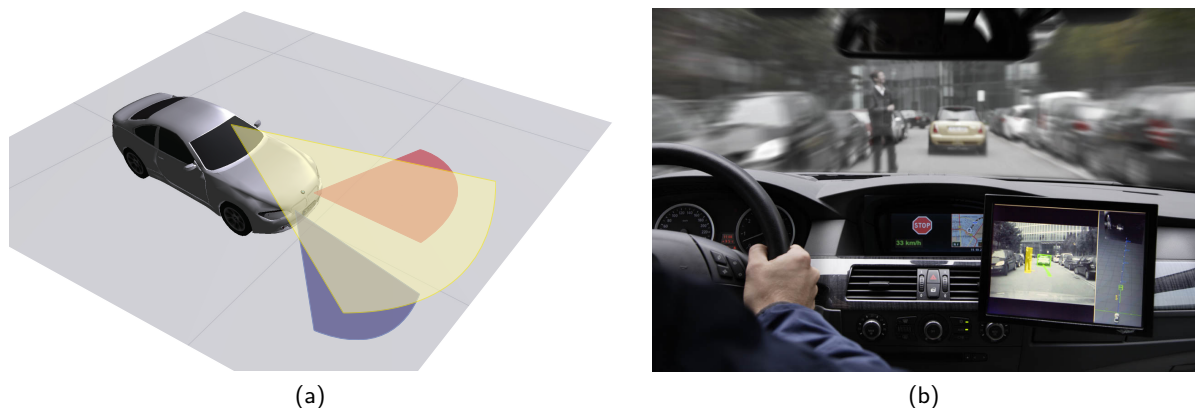


Abbildung 1.1: Illustration eines Versuchsfahrzeugs mit multisensorieller Fahrumfelderfassung und den entsprechenden Sensorerfassungsbereichen (a). Initiierte aktive Gefahrenbremsung (siehe Stop-Anzeige im Kombigerät) aufgrund einer vom Versuchsfahrzeug detektierten (siehe Ergebnisse der Umfelderfassung auf dem Monitorbild rechts unten in der Abbildung) Kollisiongefahr mit einem Fußgänger (b).

Winkel überwacht und ihn beim unbeabsichtigten Verlassen der Fahrspur zu alarmieren. Ein Beispiel für ein massiv in die Fahrzeugführung intervenierendes Fahrerassistenzsystem, das sich noch im Forschungsstadium befindet, ist die aktive Gefahrenbremsung. Erkennt das Fahrzeug, dass eine Kollision mit einem Verkehrsteilnehmer unmittelbar bevorsteht, betätigt es vollautomatisch die Bremsen, um dadurch den Unfall zu vermeiden oder zumindest die Aufprallenergie zu minimieren, insofern keine Reaktion vom Fahrer registriert wurde (siehe auch Abbildung 1.1b)

Derartige Assistenzfunktionen haben auf der einen Seite ein hohes Potential Menschenleben zu retten. Auf der anderen Seite ist bei Systemen, die stark in das Fahrgeschehen eingreifen, ein sehr hoher Zuverlässigkeitsgrad erforderlich. Fehlauflösungen sind nicht akzeptabel, da sie den Systemnutzen ins Gegenteil verkehren würden.

Das Fundament dieser „vorausschauenden“ Fahrerassistenzsysteme ist die Sensorik, sowie die maschinelle Verarbeitung der Messwerte. Da die Gesamtzuverlässigkeit direkt an die Zuverlässigkeit des Sensorsystems und der darauf operierenden Rechenverfahren gekoppelt ist, kann durch eine verschränkte Auswertung von redundanten und heterogenen Sensordaten die Systemzuverlässigkeit erhöht werden. Eine redundante Systemauslegung ist beispielsweise in der Lage einzelne Sensorausfälle zu kompensieren. Unter widrigen Witterungseinflüssen funktionieren viele Sensoren zudem nur eingeschränkt bzw. liefern ungenaue Daten aufgrund von Rauschen oder Messfehlern. Werden unterschiedliche (heterogene) Sensoren kombiniert, die auf verschiedenen Messprinzipien beruhen, können dadurch unzulängliche Messungen wechselseitig ausgeglichen werden. Nicht zuletzt finden isoliert genutzte Sensoren (z. B. Kamera für Spurerkennung und Radar für Einparkhilfe) ohnehin verstärkt Einzug ins Fahrzeug. Eine verschränkte Betrachtung der Sensordaten würde somit keine wesentlichen Mehrkosten verursachen. Allerdings würde ein erhöhter Bedarf an Rechenzeit für die Verarbeitung der größeren Datenmenge anfallen.

Gegenwärtig ist kein einzelner marktreifer Sensor verfügbar, der den hohen Anforderungen gerecht wird, die zukünftige automotiv Sicherheitssysteme beanspruchen. Einen möglichen Lösungsansatz bieten intelligente Systeme, die auf einem Sensorkollektiv operieren. Ein Ziel dieser Dissertation ist es, Synergieeffekte zwischen bild- und entfernungsgebenden Sensoren mit Hilfe geeigneter Algorithmen gewinnbringend nutzbar zu machen, um Fahrzeuge und Fußgänger im automotiven Kontext maschinell zu erkennen.

1.2 Randbedingungen und Zielsetzung

Die vorliegende Arbeit wurde bei der BMW Forschung und Technik GmbH im Rahmen des EU-Forschungsprojektes COMPOSE¹ erstellt. Erklärtes Projektziel war die Realisierung eines Versuchsträgerfahrzeugs zum aktiven Schutz von Verkehrsteilnehmern durch eine autonome Gefahrenbremsung. Dazu sollte neben einer Situationsanalyse, eine Fahrumfelderfassung umgesetzt werden, die mittels bild- und entfernungsgebenden Sensoren alle relevanten Verkehrsteilnehmer in Echtzeit lokalisiert. Die Konzeption und Umsetzung der Fahrumfelderfassung wurde durch zwei Dissertationen adressiert. Die hier vorliegende Dissertation thematisiert vorwiegend die Klassifizierung² von Fahrzeugen und Fußgängern, wohingegen die Dissertation von

¹ COMPOSE (collision mitigation and protection of road users) ist ein Teilprojekt des EU-Förderprojektes PReVENT.

² Vorgang oder Methode zur Einteilung von Objekten in Klassen oder Kategorien.

Rudi Lindl [Lin08] sich mit dem Tracking³ von Verkehrsteilnehmern beschäftigt. Die Systemkonzeption und -integration, die Datenvisualisierung, Teile der Datenvorverarbeitung sowie die Auswertung der Leistung des Gesamtsystems sind im Rahmen des Projektes in Zusammenarbeit abgestimmt und umgesetzt worden. Die grobe Aufteilung ist nochmals in Abbildung 1.2 illustriert.

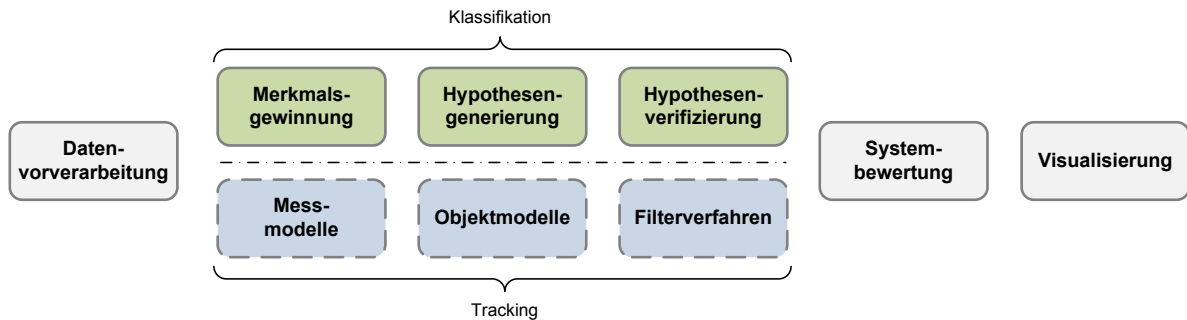


Abbildung 1.2: Die horizontale Struktur zeigt, mit Klassifikation (grün hinterlegte Felder) und Tracking (blau hinterlegte Felder), die wesentlichen Komponenten eines Umfelderkennungssystems. In der hier vorliegenden Dissertation wurden vorwiegend Aspekte der Klassifikation behandelt.

Ein wesentlicher Bestandteil dieser Arbeit war die Entwicklung eines robusten multisensoriellen Perzeptionssystems und dessen Integration in ein Versuchsträgerfahrzeug (siehe auch Abbildung 1.1b). Der Umsetzungsschwerpunkt lag dabei auf der Hypotheseninstantiierung und -klassifizierung. Des Weiteren sollten für eine medienwirksame Systempräsentation und einen unterstützten Entwicklungsprozess geeignete Visualisierungskonzepte umgesetzt werden. Eine weitere Zielsetzung dieser Dissertation war die Eruierung des multisensoriellen Potentials.

1.3 Ergebnisvorschau

Das Ergebnis dieser Arbeit ist ein funktionstüchtiges Fahrumfelderfassungssystem, das Fahrzeuge und Fußgänger in Echtzeit detektieren und lokalisieren kann. Dabei werden bis zu fünf heterogene Sensordatenströme, mittels adäquater Algorithmen, verschränkt ausgewertet. In Abbildung 1.3 sind, anhand von Schnappschüssen, einige Erkennungsergebnisse aus verschiedenen Versuchsfahrten aufgeführt. Sie zeigen das Videokamerabild des Versuchsträgers, das mit multisensoriell erkannten Objekten überlagert wurde.

In dieser Arbeit wird der Vorteil einer multisensoriellen Umgebungserfassung anhand von Beispielen und Auswertungen herausgestellt. Zudem werden neue Datenvorverarbeitungsverfahren und Detektionsmechanismen entwickelt und bewertet. Durch die massive Sensorintegration können hochwertige Leistungen für die simultane maschinelle Erkennung von Fahrzeugen und Fußgängern erzielt werden.

³ Der Begriff Tracking (engl. für das Verfolgen bzw. das Nachführen) bezeichnet Verfahren, die zum Verfolgen von Objekten und zur zeitlichen Zustandsschätzung dienen.

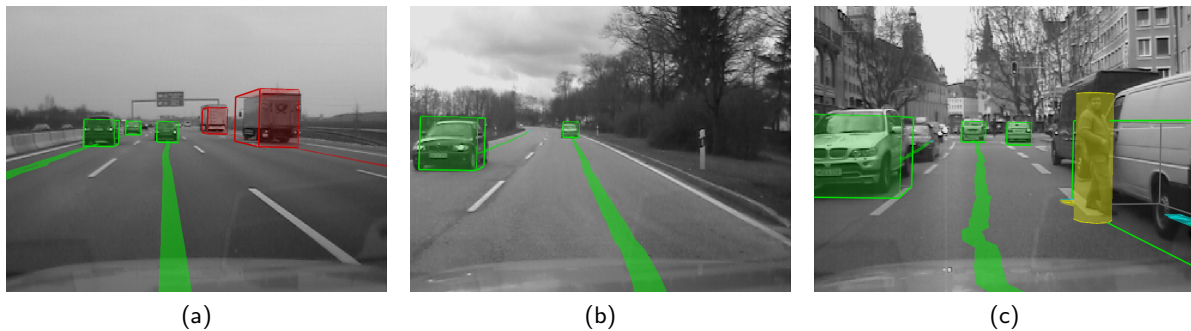


Abbildung 1.3: Ergebnisvorschau des entwickelten Umfelderkennungssystems. Fahrzeuge (grüne Quader), Fußgänger (gelber Zylinder) und Lastkraftwagen (rote Quader) werden mittels bild- und entfernungsgebender Sensorik auf der Autobahn (a), auf der Landstraße (b) und in der Stadt (c) maschinell erkannt und getrackt (Trajektorien).

1.4 Aufbau der Arbeit

Das nachfolgende Kapitel 2 führt Begrifflichkeiten ein, deren Verständnis für die Arbeit wichtig sind. Es werden aktuelle Fahrerassistenzsysteme vorgestellt, die Funktionsweise automotiver Sensorsysteme betrachtet und Sensorfusionskonzepte erläutert. Zudem wird die Sensorkonfiguration des in dieser Arbeit verwendeten Versuchsträgers beschrieben, sowie ein Überblick über das Gesamtsystem gegeben.

Im Anschluss folgen die drei Hauptkapitel Hypothesengenerierung (Kapitel 3), Merkmalsgewinnung (Kapitel 4) und Hypothesenverifizierung (Kapitel 5). Diese beinhalten die wesentlichen Bausteine zur multisensoriellen Erkennung von Verkehrsteilnehmern. Im ersten Hauptkapitel wird zunächst auf Datenvorverarbeitungstechniken eingegangen. Anschließend werden Verfahren zur mono- und multisensoriellen Erzeugung von Objekthypothesen vorgestellt. Das zweite Hauptkapitel widmet sich der Gewinnung von Merkmalen aus bild- und entfernungsgebenden Sensordaten. Es wird untersucht, welche Parametrisierung für Fahrzeuge im Wärme- und Videobild gewählt werden sollte, und wie sich Objektskalierungen auf die Erkennungsraten auswirken. Des Weiteren wird ein neues Verfahren zur Vergrößerung des Klassifikationsbereichs von bildgebenden Sensoren vorgestellt und bewertet. Das letzte Hauptkapitel adressiert verschiedene Klassifikatorfusionsverfahren und deren Eignung zur multisensoriellen Fahrzeug- und Fußgängerklassifikation. Datenunvollständigheiten, die aufgrund inkongruenter Sensorerfassungsbereiche und unterschiedlicher Abstraten auftreten können, werden gesondert diskutiert.

Die Leistungsfähigkeit der umgesetzten Fahrumfelderfassung wird in Kapitel 6 evaluiert. Dazu werden zunächst Techniken zur Referenzdatengewinnung vorgestellt. Im Anschluss daran werden die Erkennungs- und Fehlerraten des Gesamtsystems anhand von repräsentativen Fahrscenarien und unter verschiedenen Witterungseinflüssen ermittelt.

Visualisierungskonzepte für das im Rahmen dieser Arbeit umgesetzte Perzeptionssystem werden in Kapitel 7 vorgestellt und diskutiert. Darüber hinaus werden Darstellungslösungen für den Entwicklungsprozess und die Präsentation eines automotiven Perzeptionssystems thematisiert.

In Kapitel 8 werden die wichtigsten Ergebnisse zusammengefasst und ein Ausblick hinsichtlich möglicher Erweiterungen und Verbesserungen gegeben. Die Beschreibungen der verwendeten Testdatensätze und Bewertungskriterien finden sich in den Anhangskapiteln A und B. Im Anhang C wird auf Methoden zur multispektralen Kamerakalibrierung eingegangen. Einige exemplarische Szenendurchläufe, mit der in dieser Arbeit umgesetzten multisensoriellen Fahrumfelderfassung, werden in Anhang D illustriert.



„La condition humaine“ von René Magritte (1898 - 1967).

2

Grundlagen

In diesem Kapitel wird zunächst ein Überblick über aktuelle Fahrerassistenzsysteme gegeben und die Funktionsweise automotiver Sensorsysteme näher betrachtet. Anschließend werden wichtige theoretische und praktische Aspekte einer multisensoriellen Fahrumfelderfassung erläutert. Dazu zählen sowohl Sensorfusionskonzepte als auch formale Grundlagen. Abschließend werden die Sensorkonfiguration des in dieser Arbeit verwendeten Versuchsträgers, und die Komponenten des umgesetzten Fahrumfelderfassungssystems beschrieben.

2.1 Fahrerassistenzsysteme

Die Vorläufer der Fahrerassistenzsysteme, wie zum Beispiel der elektrische Anlasser, das Automatikgetriebe oder die Servolenkung, hatten zum Ziel, die Führungsaufgabe des Fahrers durch den Entfall von Maschinistentätigkeiten zu erleichtern [Fre04]. Seit Ende der 70er-Jahre wurden, angefangen vom Bremsen-Antiblockiersystem (ABS) bis hin zur dynamischen Fahrzeugstabilisierung (DSC), überwiegend Systeme eingeführt, die eine Steigerung der allgemeinen Sicherheit in schwierigen Fahrsituationen adressieren. Diese Fahrerassistenzsysteme beruhen im Allgemeinen auf einer Überwachung von fahrzeuginternen Zustandsgrößen wie Lenkwinkel, Querbeschleunigung oder Raddrehzahlen.

Bedingt durch den technischen Fortschritt in der Sensortechnologie werden heutzutage zunehmend Systeme entwickelt, die die Umgebung des Fahrzeugs bzw. das aktuelle Verkehrsgeschehen mit berücksichtigen. Hierdurch können zum einen bereits bestehende Systeme erweitert werden, und zum anderen „vorausschauende“ Anwendungen entwickelt werden, die vor gefährlichen Situationen warnen oder gegebenenfalls Gegenmaßnahmen ergreifen. Als erstes Beispiel solcher Systeme wurde 1995 „Adaptive Cruise Control“ (ACC) von Mitsubishi in Japan eingeführt [MS05]. Hierbei wird zusätzlich zur eigenen Geschwindigkeit, der Abstand zu einem vorausfahrenden Fahrzeug geregelt. Neben diesen aktiven Sicherheitssystemen wird fortlaufend auch die passive Sicherheit im Fahrzeug verbessert. Dazu zählen Maßnahmen des Fahrzeugaufbaus, wie zum Beispiel Sicherheitsgurte, Knautschzonen, Überrollbügel, Kindersitze und Airbags.

Im Folgenden wird zunächst ein Überblick aktueller Fahrerassistenzsysteme gegeben und mögliche Einteilungen vorgestellt. Als ein Beispiel für ein zukünftiges FAS wird abschließend die aktive Gefahrenbremsung für Fußgänger und Fahrzeuge herausgegriffen und ausführlicher betrachtet. Dieses fungierte als Zielapplikation, um die Leistungsfähigkeit der im Rahmen dieser Arbeit entwickelten Fahrumfelderfassung an einem realen Fahrerassistenzsystem erlebbar zu machen.

2.1.1 Überblick

Obwohl die im vorherigen Abschnitt aufgeführten Beispiele nur einen kleinen Ausschnitt aktueller Fahrerassistenzsysteme darstellen, sind sie dennoch ein Beleg dafür, wie vielfältig diese Systeme sein können. Eine Möglichkeit, Fahrerassistenzsysteme zu ordnen ist die Einteilung hinsichtlich der Art und Weise ihrer Unterstützung [Fre04]:

Systeme, die warnen oder informieren: Durch visuelle, haptische oder akustische Signale wird der Fahrer gewarnt oder informiert. So wird er zum Beispiel beim unbeabsichtigten Verlassen der eigenen Spur durch LDW bzw. LCW (lane departure warning bzw. lane change warning) alarmiert. Darüber hinaus werden dem Fahrer durch Navigationssysteme (NAVI) oder Kommunikation zwischen mehreren Fahrzeugen (Car2Car) Informationen über die aktuelle Verkehrslage weitergegeben. Bei Nachtfahrten unterstützen ihn Nachtsichtsysteme wie beispielsweise Wärmebildkameras, mit deren Hilfe wärmeemittierende Objekte wie Menschen oder Tiere visuell hervorgehoben werden können, schon bevor sie vom Scheinwerfer des Fahrzeugs erfasst werden (NiVi = night vision).

Systeme mit langsamer Reaktion, die führend unterstützen: Bei dieser Art von FAS wird unter wohldefinierten Randbedingungen moderat in die Fahrzeugführung eingegriffen. So regelt der Einparkassistent (PA = park assist) beispielsweise die Querführung und lenkt das Fahrzeug automatisch in eine Parklücke. Im höheren Geschwindigkeitsbereich übernehmen FAS wie das ACC teilautonom die Längsführung des Fahrzeugs um den Abstand zum vorausfahrenden Fahrzeug zu regulieren. Ein weiteres System, das sowohl in die Längs- als auch in die Querführung des Fahrzeugs eingreift, ist der Spurwechselassistent (LCA = lane change assistance). Er unterstützt den Fahrer bei einem Spurwechselmanöver. Für derartige Systeme gilt im Allgemeinen, dass sie in kritischen Situationen die Kontrolle wieder an den Fahrer übergeben.

Systeme mit schneller Reaktion, die stabilisieren bzw. intervenieren: In kritischen Fahrsituationen müssen Systeme wie das DSC oder ABS sehr schnell analysieren und gegebenenfalls geeignete Gegenmaßnahmen treffen, indem sie das Fahrzeug aktiv stabilisieren. Bremsassistenten (iBA = intelligent braking assistance) erkennen zum Beispiel anhand der Bewegungsgeschwindigkeit und -beschleunigung des Bremspedals die Intention des Fahrers für eine Vollbremsung und ergänzen gegebenenfalls deren unvollständige oder zu zaghafte Ausführung. Einen Schritt weiter gehen Systeme wie die aktive Gefahrenbremsung (AGB) oder die Kollisionsvermeidung (CA = collision avoidance), die auch ohne eine Initiierung des Fahrers in das Fahrgeschehen eingreifen, um Unfälle zu vermeiden oder zumindest deren Schwere zu verringern.

Eine weitere Einteilung von Fahrerassistenzsystemen kann anhand der jeweiligen Funktionsausprägung in Komfort- und Sicherheitssysteme vorgenommen werden [Fre04, Naa04]. Allerdings ist dabei eine eindeutige Zuordnung für manche Systeme wie beispielsweise das ACC nicht möglich, da sie sowohl den Komfort, als auch die Sicherheit erhöhen. In Abbildung 2.1

ist für die oben beschriebenen Fahrerassistenzsysteme eine Einteilung sowohl nach Komfort- und Sicherheitssystem als auch nach ihrer Unterstützungscharakteristik dargestellt.

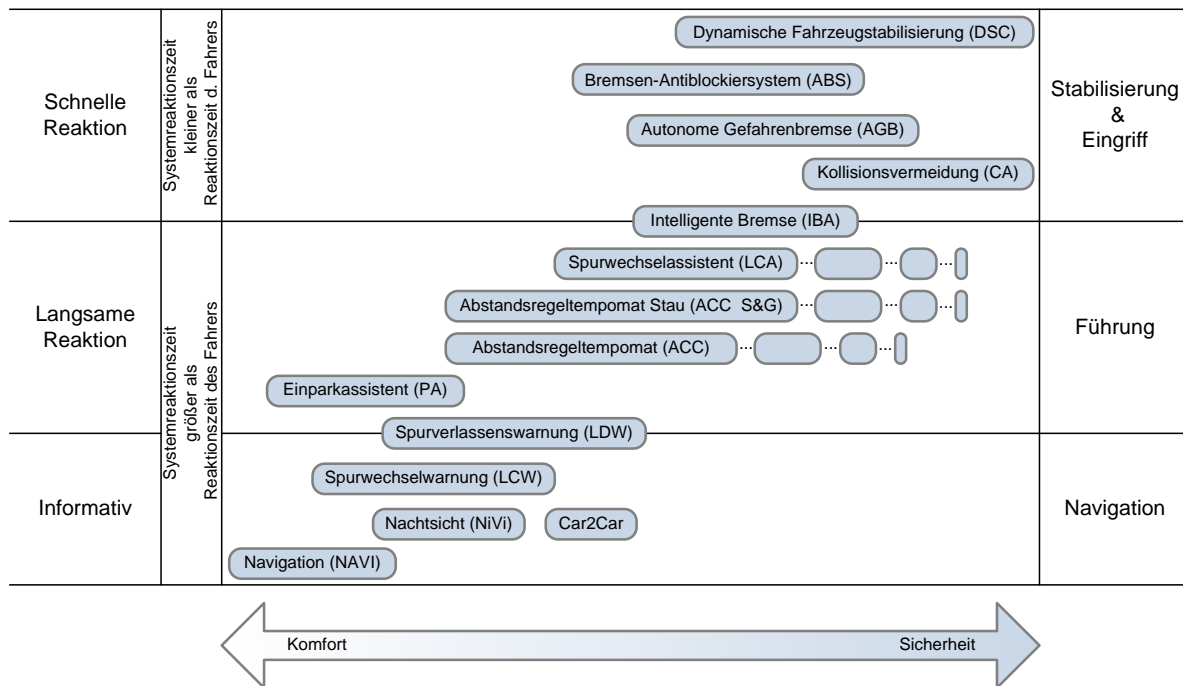


Abbildung 2.1: Einteilung von ausgewählten Fahrerassistenzsystemen nach Komfort und Sicherheit bzw. nach der Art ihrer Unterstützung [Fre04, Naa04]. Zukünftige Ausbaustufen einzelner Systeme sind durch leere Ovale gekennzeichnet. Das FAS „autonomes Fahren“ wurde aus Übersichtsgründen nicht eingezeichnet, da es den gesamten Diagrammbereich abdecken würde.

Für viele Fahrerassistenzsysteme ist der Zeitpunkt des Eingriffs charakteristisch. Deshalb kann hiermit eine weitere Einteilung von Fahrerassistenzsystemen erfolgen, wie sie in Abbildung 2.2 dargestellt ist [Fre04]: Die Mehrzahl an Systemen, wie z. B. die Navigation oder ACC, unterstützen den Fahrer zu einem Zeitpunkt, in dem sich das Fahrzeug in einer völlig unkritischen Fahrsituation befindet. Systeme wie beispielsweise ABS bzw. DSC intervenieren zu einem späteren Zeitpunkt, wenn sich das Fahrzeug bereits in einem instabilen Zustand (Blockieren der Räder bzw. Schleudern) befindet, und versuchen diesen zu stabilisieren. Ist ein Unfall hingegen nicht mehr zu vermeiden, kann die Energie des Zusammenstoßes durch Pre-Crash-Systeme wie die aktive Gefahrenbremse minimiert werden. Zum Unfallzeitpunkt werden die Insassen durch Airbags geschützt. Post-Crash-Systeme können nach einem Unfall in Anspruch genommen werden (z. B. Absetzen eines Notrufs).

Der perfekte Beifahrer ist eines der großen Ziele aktueller Forschungsbestreben. Nur wenn Notwendigkeit besteht, assistiert er dem Fahrer bezüglich des aktuellen Fahrgeschehens. Ähnlich anspruchsvoll ist die Vision vom autonomen Fahren, bei der die Fahraufgabe vollständig dem Fahrzeug überlassen wird, und der Fahrer somit selbst zum „Beifahrer“ wird. Zwar könnte hiermit menschliches Versagen als einer der wesentlichen Unfallfaktoren eliminiert werden – allerdings müsste dieser Vorteil durch ungeklärte Haftungsfragen und hohe Ansprüche in

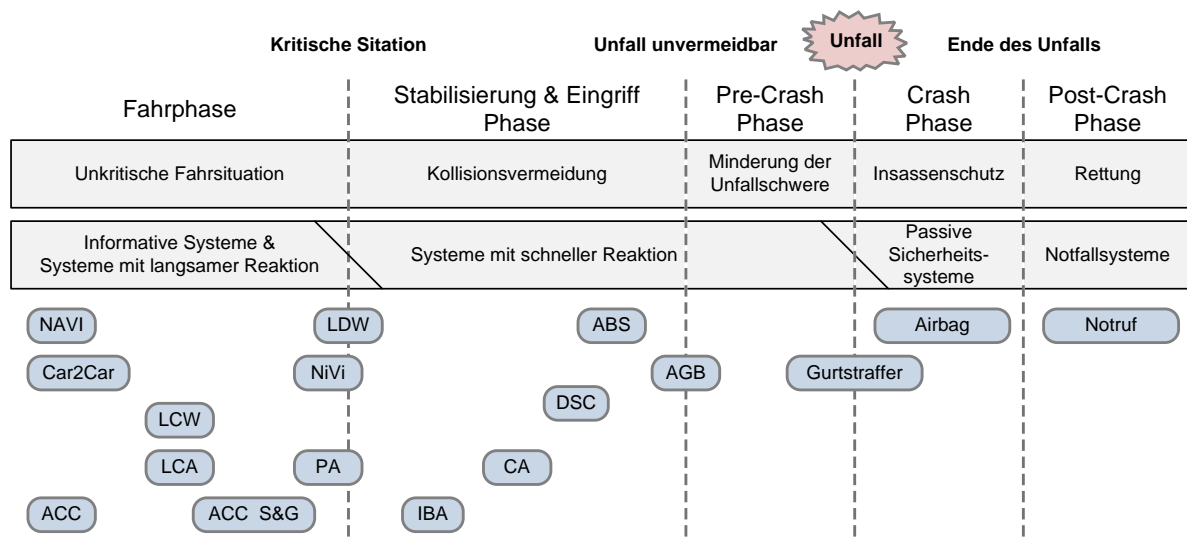


Abbildung 2.2: Einteilung von ausgewählten Fahrerassistenzsystemen nach dem Zeitpunkt ihres Eingriffs [Fre04].

Bezug auf Sensorverlässlichkeit und Systemleistung teuer erkauft werden. Zudem müsste das nichtdeterministische Verhalten anderer Verkehrsteilnehmer berücksichtigt werden. Unter wohldefinierten Randbedingungen und dem exzessiven Einsatz von Sensorik bzw. Rechenleistung ist autonomes Fahren im wissenschaftlichen Umfeld bereits Realität, wie die DARPA Urban Challenge [DAR07] eindrucksvoll belegt.

Bei aktiven Systemen, die autonom in das Fahrgeschehen eingreifen und somit die Verantwortung vom Fahrer auf das Fahrzeug übertragen, spielen rechtliche Aspekte eine große Rolle. Die zentrale zulassungsrechtliche Hürde liegt hier vor allem im Wiener Übereinkommen über den Straßenverkehr von 1968: „Jeder Fahrzeugführer muss unter allen Umständen sein Fahrzeug beherrschen“. Eine ausgiebige Diskussion dieser Fragestellung ist in [vdHvW01, SG06] zu finden. Nicht zuletzt aus psychologischen Gründen, wie zum Beispiel dem Verlust der Fahrfreude und technischen Berührungsängsten, wird in absehbarer Zukunft vollautonomes Fahren im realen Straßenverkehr eine Vision bleiben. An teilautonomen Systemen, wie zum Beispiel der „aktiven Gefahrenbremsung“, wird jedoch schon heute intensiv geforscht.

2.1.2 Aktive Gefahrenbremsung

Die deutsche Verkehrsstatistik von 2006 [Sta06] belegt, dass 60% aller tödlichen Unfälle auf einen Zusammenstoß mit einem fremden Fahrzeug zurückzuführen sind. An weiteren 14% der Unfälle sind Fußgänger beteiligt. In beiden Szenarien ist menschliches Versagen mit Abstand die häufigste Unfallursache. Fahrerassistenzsysteme wie der Bremsassistent oder die aktive Gefahrenbremsung besitzen großes Potential, die Zahl der Unfälle zu reduzieren bzw. die Unfallfolgen zu mindern, da sie schneller und effizienter als der Fahrer reagieren können. Um trotz dieses autonomen Eingriffs, die Souveränität des Fahrers zu bewahren, dürfen derartige Systeme erst dann agieren, wenn für den Fahrer keine Möglichkeit mehr existiert, den Unfall durch Ausweichen oder Bremsen selbst zu vermeiden. Obwohl sich unter diesen Randbedin-

gungen nicht mehr alle Unfälle verhindern lassen, ist es zumindest möglich, die Schwere der Verletzungen zu reduzieren, da ein Großteil der kinetischen Energie durch das Bremsmanöver absorbiert wird. An die AGB werden sehr hohe Anforderungen in Bezug auf die Häufigkeit von Fehlauflösungen gestellt, da schon eine einzige ungerechtfertigte autonome Vollbremsung sowohl die Akzeptanz für das System, als auch die allgemeine Verkehrssicherheit erheblich reduzieren würde.

Die Qualität der aktiven Gefahrenbremsung ist hauptsächlich von zwei Faktoren abhängig: Sie wird zum einen von der Zuverlässigkeit und Genauigkeit der zugrundeliegenden Fahrumfelderfassung beeinflusst. Geisterziele, Fehlklassifikationen oder falsche Zustandsschätzungen können zu einer Fehleinschätzung und somit zu einer ungerechtfertigten Initiierung des Bremsmanövers führen. Zum anderen wird die Qualität des Gesamtsystems, selbst unter der Annahme einer perfekten Fahrumfelderfassung, durch die Interpretation der aktuellen Situation und der daraus abgeleiteten Entscheidung bestimmt. So ist zum Beispiel das Verhalten von Fußgängern nur sehr schwer zu präzisieren.

Mertz [MDG⁺05] versucht diese Unsicherheit zu vermindern, indem für sich bewegende Objekte mehrere zufällige Trajektorien erzeugt werden, die in Abhängigkeit von Mittelwert und Varianz der zugehörigen Zustandsgrößen verteilt sind. Aus dem Verhältnis von den Trajektorien eines Objektes, die zu einer Kollision führen, zur Gesamtheit der Trajektorien eines Objektes, kann eine Kollisionswahrscheinlichkeit berechnet werden. Bei diesem Ansatz wird somit versucht, durch eine Monte-Carlo-Simulation die reale Kollisionswahrscheinlichkeit zu approximieren. Allerdings ist die Qualität und der Rechenaufwand direkt proportional zur Größe der Stichprobe.

In Abhängigkeit von der Geschwindigkeit des Eigenfahrzeugs bzw. des Fußgängers und der maximalen Beschleunigungen definiert Fürstenberg [FS05] einen sogenannten „Bereich ohne Entrinnen“¹. Befindet sich ein Fußgänger innerhalb dieses Bereiches, ist ein Unfall unvermeidbar.

Kämpchen [Käm07] berechnet alle physikalisch möglichen Trajektorien für das Eigen- und ein Fremdfahrzeug. In der Berechnung werden keine einschränkenden Annahmen über die Geschwindigkeit und Beschleunigung der beteiligten Objekte getroffen. Zusätzlich wird die Orientierung, Länge und Breite der Fahrzeuge berücksichtigt, was eine Anwendung dieses Verfahrens auch in Kreuzungsszenarien ermöglicht. Erst wenn alle Pfade zu einer Kollision führen, wird eine autonome Gefahrenbremsung initiiert.

In den meisten Verfahren werden Verkehrsteilnehmer jedoch isoliert von ihrer Umwelt betrachtet. Broadhurst et al. [BBK04] haben eine Simulationsumgebung entwickelt, die das gesamte Verkehrsgeschehen in komplexen Szenarien berücksichtigt und kritische Situationen erkennt. Neben Simulationen und regelbasierten Ansätzen, existieren in der Literatur Verfahren [JJ02, KJG04, LWP07], die eine probabilistische Situationsinterpretation verwenden. Unsicherheiten des Trackings und der Odometrie des Eigenfahrzeugs sowie deren zeitlicher Verlauf werden in einer gemeinsamen Wahrscheinlichkeitsdichtefunktion abgebildet, aus der die Wahrscheinlichkeit für einen Zusammenstoß abgeleitet werden kann.

Die aktive Gefahrenbremsung wird auch weiterhin Bestandteil von Forschungsaktivitäten bleiben, da sowohl die aktuellen Algorithmen der Situationsinterpretation, als auch die verfügbaren Sensoren nicht den hohen Qualitätsansprüchen genügen. Nicht nur für die AGB, sondern

¹ engl. „Region of no escape“

auch für viele andere Fahrerassistenzsysteme, ist die zugrundeliegende automotiv Sensorik ein entscheidender Faktor.

2.2 Automotive Sensorik

Sensoren² oder Messfühler sind technische Bauteile, die spezifische Eigenschaften ihrer Umgebung, aufgrund physikalischer oder chemischer Effekte (z. B. Wärmestrahlung, Temperatur, Feuchtigkeit, Druck, Schall, Helligkeit oder Beschleunigung) erfassen und in weiterverarbeitbare Größen (z. B. elektrische Spannung) umwandeln können. In modernen Fahrzeugen sind eine Vielzahl von Sensoren in Gebrauch. Diese müssen hohen Anforderungen genügen. Sie sollten unempfindlich gegenüber elektromagnetischen Interferenzen sein, in einem Temperaturbereich von -40°C bis zu $+125^{\circ}\text{C}$ funktionieren, geringe Fehlerraten aufweisen, vibrations- und schmutzunempfindlich sein und vieles mehr.

Fleming [FTSW01] kategorisiert automotiv Sensoren anhand ihres Einsatzgebietes im Fahrzeug und ihrer Kontrollfunktion. Er unterscheidet dabei die Bereiche Antriebsstrang, Fahrwerk und Fahrzeugaufbau. Sensoren des Antriebsstrangs sind für den Energieverbrauch, die Fahrbarkeit und die Fahrzeugleistung zuständig. Als Beispiele seien Druck- und Drehzahlsensoren für die Motorsteuerung genannt. Sensoren des Fahrwerks adressieren Fahreigenschaften und Sicherheitsbelange. Typische Vertreter sind Raddrehzahl- und Giergeschwindigkeitssensoren für die Fahrzeugstabilisierung und das Antiblockiersystem. Sensoren des Fahrzeugaufbaus decken zum einen Komfortfunktionen ab, wie beispielsweise Klimaregelung (Temperaturfühler, Feuchtigkeits- bzw. Regensensor) und automatisches Nachtfahrlicht (Helligkeitssensor). Zum anderen bilden sie die Basis für Sicherheitsfunktionen, wie z. B. Airbag (Piezosensor), Einparkhilfe (Ultraschall) und Spurwechselwarnung (Radar, Lidar). Wird in dieser Arbeit der Begriff (*automotiver*) *Sensor* verwendet, bezieht sich die Bezeichnung auf die Kategorie der sicherheitsrelevanten Sensoren des Fahrzeugaufbaus und zwar speziell auf automotiv Sensorik, die in der Lage ist, Informationen über das Fahrumfeld zu erlangen.

Im Folgenden wird zunächst die grobe Funktionsweise von der, für diese Arbeit relevanten, automotiven Sensorik erläutert. Der nächste Abschnitt kategorisiert die Sensoren anhand ihrer Erfassungseigenschaften und stellt einen Vergleich hinsichtlich der Sensorperformanz auf.

2.2.1 Funktionsweise und Verbreitung

Dieser Abschnitt erläutert die Funktionsweise der für diese Arbeit relevanten automotiven Sensoren. Details zum Funktionsprinzip finden sich in den angefügten Referenzen.

Videokamera: In den späten 60er Jahren wurden CCDs (charge coupled device) entwickelt, die heute günstig gefertigt werden können und als Videokamerasensor weit verbreitet sind. Grundsätzlich funktioniert eine einzelne Detektorzelle dieser elektronischen Bausteine ähnlich einer Photodiode. In jeder Zelle werden Photonen gesammelt, in elektrische Ladung umgewandelt und ausgelesen. Kamerasensoren auf CMOS (complementary metal oxide semiconductor) Basis ermöglichen ein schnelleres Auslesen der Bildinformationen was höhere Bildwiederholraten ermöglicht. Videokameras werden aktuell für Spur- und Verkehrszeichenerkennung im Automobil eingesetzt [Bor06].

² Sensor kommt von lateinisch *sentire* („fühlen“, „empfinden“).

Infrarotkamera: IR-Kameras arbeiten im nicht sichtbaren Lichtspektrum. Der Infrarotbereich liegt zwischen 780 nm und 14 μm , wobei aufgrund atmosphärischer Absorption nicht der gesamte Bereich für einen sinnvollen Sensoreinsatz zur Verfügung steht [RC02]. Zwei unterschiedliche Technologien sind derzeit als Nachtsichtsystem im automotiven Einsatz:

FIR (far infrared) Sensoren arbeiten im Spektralbereich von 8 μm bis 14 μm und erkennen mittels ungekühlter Mikrobolometer³-Matrizen Wärmestrahlung, die von jedem Körper mit einer Temperatur von mehr als 0 K ausgestrahlt wird. Aktuell werden Auflösungen von bis zu 320×240 Bildpunkten erreicht. Leistungsfähigere Systeme benötigen eine Kühlung und bleiben wegen ihrer hohen Kosten militärischen Anwendungen vorbehalten. Das erste automotive FIR-System wurde 1999 von Cadillac und Raytheon eingeführt [KKNK03].

NIR (near infrared) Systeme arbeiten im nahen Infrarotspektrum von 780 nm bis 1 μm . Sie müssen dabei die Szene aktiv im Infrarotbereich beleuchten, da die emittierte Energie der relevanten Objekte in diesem Spektrum nicht signifikant ist. Als Sensor fungieren herkömmliche CCDs auf Siliziumbasis. Vorteile von NIR-Systemen liegen in der für einen Menschen natürlicheren Darstellung von Objekten. Im Gegensatz zu FIR-Systemen werden auch Straßenmarkierungen und Verkehrsschilder abgebildet. Nachteile sind die geringere Reichweite und das gesondert erforderliche Beleuchtungsmodul.

RADAR: Das Akronym RADAR steht für „Radio Detection and Ranging“ und bezeichnet ein Verfahren zur Ortung von Objekten mittels elektromagnetischer Wellen im Gigahertzbereich (Funk). Dazu wird über eine Sendeeinheit ein Signal emittiert, welches besonders stark von metallischen Objekten zurückgestrahlt wird. Diese Reflexionen werden von einer Empfangseinheit detektiert. Über verschiedene Kriterien wie Laufzeit und Frequenzverschiebung des Signals können Informationen wie Entfernung, Winkel und radiale Geschwindigkeit über das erfasste Objekt gewonnen werden. Seit 1998 haben sich frequenzmodulierte Dauerstrichradare in Oberklassenfahrzeugen für adaptive Abstandhaltung etabliert. Nahbereichsradares mit 24 GHz haben in der Regel einen größeren Öffnungswinkel und werden für Spurwechselassistenten und Kollisionswarnungen eingesetzt [GJE⁺04]. Aktuelle automotive Radarentwicklungen zielen vermehrt in Richtung bildgebender Verfahren.

LIDAR: Das LIDAR (Light Detection and Ranging) Verfahren ist mit dem Funktionsprinzip des Radars verwandt. Statt elektromagnetischer Wellen im Gigahertzbereich werden Lichtimpulse emittiert und empfangen, die im ultravioletten, sichtbaren oder infraroten Spektralbereich liegen. Aufgrund der hohen Wellenfrequenz des Primärsignals sind die Auswirkungen des Dopplereffekts kaum messbar. Geschwindigkeiten können somit nicht direkt, sondern nur indirekt über zeitlich versetzte Messungen bestimmt werden. Mit Laserscannern, die auf Basis des Lidar-Prinzips die Umgebung abtasten, ist eine wesentlich genauere Winkelauflösung als mit Radar-Systemen möglich, da Lichtsignale stärker fokussiert werden können als Mikrowellen. Im Gegensatz zu Radar-Systemen fanden Lidar-Systeme bisher nur vereinzelt Einzug ins Automobil. Gründe dafür sind die niedrige Schlechtwetterperformanz und die begrenzten Verbauorte im Fahrzeug. Details zur Lidar-Technologie im Kontext von Fahrerassistenzsystemen findet sich in [RG05].

³ Mikrobolometer sind thermische Detektoren die Infrarotstrahlung durch eine wärmebedingte Änderung ihres elektrischen Widerstandes messen können.

2.2.2 Taxonomie und Vergleich

Eine mögliche Untergliederung von automotiven Sensorsystemen kann hinsichtlich ihrer Funktionsweise erfolgen. Aktive Sensoren emittieren ein elektromagnetisches Signal, dessen Reflexionen an Objekten von einer Empfangseinheit aufgenommen wird. Über Signallaufzeiten und -veränderungen können Rückschlüsse auf Aspekte der Realität geschlossen werden. Typische Vertreter sind Ultraschall, Radarsensoren und Laserscanner. Passive Sensoren verfügen lediglich über eine Empfangseinheit. Mittels dieser wird die Eigenstrahlung von Objekten wie beispielsweise sichtbares Licht oder Wärme detektiert. Passive Sensoren sind beispielsweise Video- und Wärmebildkameras.

Eine weitere Unterteilungsmöglichkeit der Perzeptionssensorik kann hinsichtlich des Messresultats erfolgen. Rein bildgebende Sensoren erzeugen ein Abbild der Realität in dem strukturelle Objektdetails sichtbar werden. Rein entfernungsgebende Sensoren geben Aufschluss über statische und dynamische Zustandsdaten der erfassten Objekte, wie beispielsweise Position, Orientierung, Geschwindigkeit und Entfernung. Aktuelle Sensoren sind selten reine entfernungs- bzw. bildgebende Sensoren. Es liegen vielmehr beide Messresultate in unterschiedlicher Qualität vor und die Kategorieübergänge sind fließend (siehe Abbildung 2.3).

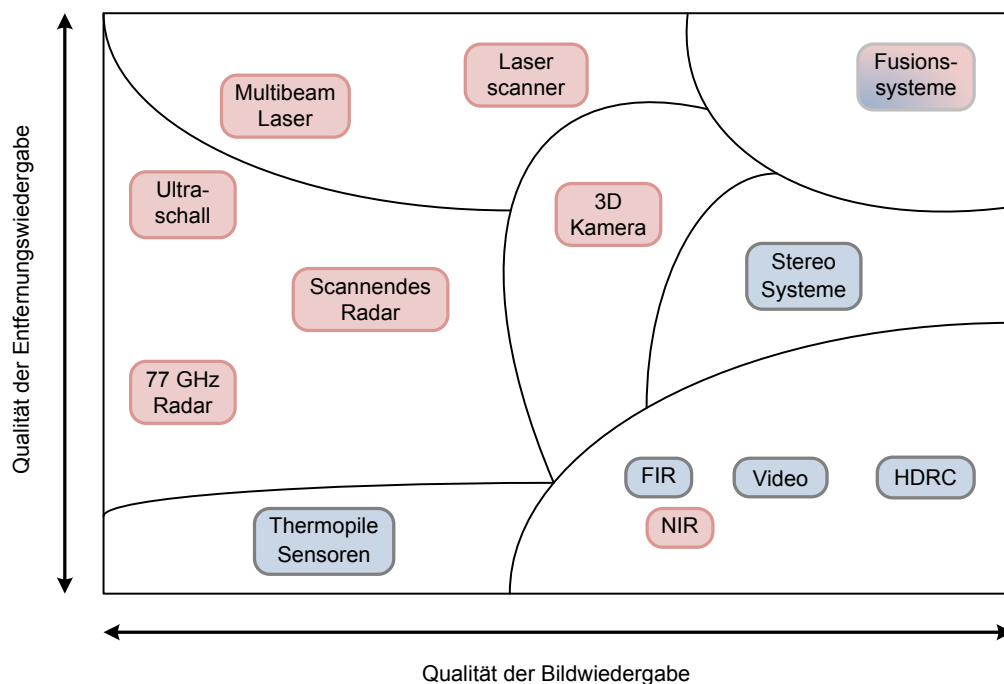


Abbildung 2.3: Kategorisierung automotiver Umfelderfassungssensorik hinsichtlich Bild- und Entfernungswiedergabe. Aktive Sensoren sind rot hinterlegt und passive Sensoren blau.

Videokameras oder HDRC (high dynamic range cameras) bilden beispielsweise vorwiegend die strukturellen Aspekte der Realität ab. Dennoch lassen sich Distanzen und somit Zustandsdaten mit begrenzter Genauigkeit schätzen, sofern Modellannahmen einfließen. Laserscanner, als vorwiegend entfernungsgebende Sensoren, liefern ebenfalls strukturelle Aspekte der Objekte,

wie beispielsweise Oberflächenbeschaffenheiten. Ein idealer Perzeptionssensor würde sowohl die Struktur als auch die statischen und dynamischen Zustandsdaten von Umfeldobjekten in sehr hoher Qualität wiedergeben. Zurzeit vermögen lediglich Fusionssysteme, die über geeignete Algorithmik unterschiedliche Sensordaten kombinieren, dem idealen Sensor einen Schritt weit näher zu kommen.

Die Qualität, mit welcher die Realität abgebildet wird, entscheidet nicht allein über die Performanz eines Perzeptionssensors. Ein wesentliches Kriterium perzeptionstauglicher Sensorik ist eine hohe zeitliche Erfassungsrate. Speziell für Sensoren, die in sicherheitsrelevanten Fahrerassistenzsystemen zum Einsatz kommen, sind Messfrequenzen von 10–100 Hz zwingend erforderlich. Weitere wichtige Gesichtspunkte im automotiven Kontext sind Sensorabmessungen, Kosten, Wettertauglichkeit und Reichweite (siehe Tabelle 2.1 für einen Sensorvergleich). Ein automotives Radar beispielsweise liefert im Vergleich zu einem Laserscanner eine begrenzte laterale Messgenauigkeit, verhält sich aber zugleich aufgrund des Messprinzips sehr robust bei Schlechtwetterbedingungen.

Tabelle 2.1: Vergleich automotiver Sensorsysteme zur Fahrumfelderfassung hinsichtlich ihrer Performanz ($\oplus\oplus$ sehr gut, \oplus gut, \circ befriedigend, \ominus ausreichend, $\ominus\ominus$ mangelhaft). Die betrachteten Stereosysteme können aus Videokameras/IR aufgebaut sein. Teile des Diagramms stammen aus [WGS⁺04]. Die ergänzten Bewertungen basieren auf der Beurteilung des Autors.

	Video	Nahes IR	Fern IR	Radar	Scan Radar	3D Kamera	Stereosysteme	Laserscanner	Ultraschall
max. Reichweite	100 m	80 m	>300 m	150 m	100 m	20 m	80 m	150 m	4 m
Longitudinale Genauigkeit	\ominus	\ominus	\ominus	\oplus	\oplus	\oplus	\oplus	$\oplus\oplus$	\oplus
Laterale Genauigkeit	$\oplus\oplus$	$\oplus\oplus$	$\oplus\oplus$	\ominus	\circ	\oplus	$\oplus\oplus$	\oplus	\circ
Objektbeschreibung	$\oplus\oplus$	$\oplus\oplus$	$\oplus\oplus$	\circ	\circ	\circ	$\oplus\oplus$	\oplus	\circ
Schlechtwetterverhalten	\ominus	\ominus	\ominus	$\oplus\oplus$	$\oplus\oplus$	\ominus	\ominus	\ominus	\oplus
Verhalten bei Dunkelheit	\ominus	\oplus	\oplus	\oplus	\oplus	\oplus	\ominus/\oplus	\oplus	\oplus
Verbaubarkeit	\oplus	\circ	\oplus	$\oplus\oplus$	\oplus	\circ	\ominus	\circ	$\oplus\oplus$
Kosten	$\oplus\oplus$	\circ	\ominus	\circ	\circ	\circ	$\circ/\ominus\ominus$	\ominus	$\oplus\oplus$
Fahrerassistenzsysteme	LDW	NiVi	NiVi	ACC, ACC S&G	ACC, ACC S&G	iBA	ACC, iBA	ACC, iBA	PA

2.3 Sensorfusion

Augenblicklich verfügbare automotiv Sensoren, wie sie im vorherigen Abschnitt 2.2 beschrieben wurden, können einzeln betrachtet den hohen Anforderungen, die neuartige Fahrerassistenzsysteme, wie beispielsweise die AGB, an die zugrundeliegende Sensorik in Bezug auf Zuverlässigkeit, Verfügbarkeit und Genauigkeit stellen, nur schwer genügen. Deshalb wird durch Theorien und Methoden der Sensorfusion versucht, Informationen aus mehreren Sensoren derart gewinnbringend zu kombinieren (fusionieren), dass das gemeinsam erzielte Ergebnis in gewisser Weise besser ist, als das Ergebnis der einzelnen Sensoren. Obwohl im Allgemeinen davon ausgegangen werden kann, dass zusätzliche Informationen einen positiven Einfluss auf das Ergebnis haben, sollte stets bedacht werden, dass sie unter Umständen keinen signifikanten oder sogar einen negativen Beitrag zum Gesamtergebnis leisten [Das00].

Im Folgenden wird zunächst auf Vorteile eingegangen, die von einem Multisensorsystem erwartet werden können. Die nachfolgenden Abschnitte beschreiben unterschiedliche Architekturen und Einteilungsschemata. Abschließend wird die Projektlandschaft zu automotiven Fusionssystemen beleuchtet, sowie ausgewählte Beispiele gegeben, in denen Sensorfusion in einem FAS zur Anwendung gebracht wurde.

2.3.1 Vorteile

In Bezug auf folgende Punkte können von einem Fusionssystem im Vergleich zu einem Einzelsensorsystem Vorteile erwartet werden:

Zuverlässigkeit: Multisensorsysteme zeichnen sich durch eine inhärente Redundanz aus, die die Zuverlässigkeit des Gesamtsystems erhöht. Existieren zum Beispiel redundante Informationen von mehreren Sensoren, können hardwarebedingte Sensorausfälle bis zu einem gewissen Umfang vollständig kompensiert oder zumindest ein eingeschränkter Betrieb des Systems gewährleistet werden. Darüber hinaus kann die Abhängigkeit des Gesamtsystems von externen Umwelteinflüssen wie Wetter oder Beleuchtung minimiert werden, wenn einzelne Sensoren nicht nur redundante Informationen liefern, sondern auch, bedingt durch verschiedene physikalische Messprinzipien oder Parametrisierungen, unterschiedlich auf diese externen Störeinflüsse reagieren.

Genauigkeit und Konfidenz: Durch redundante Messungen mehrerer Sensoren, können in der Kombination höhere Genauigkeiten für einzelne Messgrößen erzielt werden. Bild- und entfernungsgebende Sensorik ergänzen sich zum Beispiel sehr gut, da sie wechselseitig, in den Messgrößen Entfernung und laterale Ablage, geringe Varianzen aufweisen. Einzelne Sensoren können sich aber auch gegenseitig bestätigen, was zu einer höheren Konfidenz führt. Auf diese Weise lassen sich Geisterziele unterdrücken und Fehlklassifikationen verringern.

Erfassungsbereich und räumliche Auflösung: Die räumliche Abdeckung eines Multisensorsystems ist durch die Vereinigung der Erfassungsbereiche der beteiligten Sensoren bestimmt. Dadurch lassen sich, weitgehend unabhängig von den Einschränkungen der jeweiligen Sensorik, komplexe Erfassungsbereiche modellieren. Durch eine Überlagerung der Erfassungsbereiche von Sensoren mit redundantem Messprinzip, kann innerhalb dieses gemeinsamen Bereiches eine erhöhte räumliche Auflösung erreicht werden.

Zeitliche Auflösung: Neben der räumlichen Auflösung profitiert die zeitliche Auflösung des Gesamtsystems von zusätzlicher Sensorik. Hierbei werden mehrere Sensoren asynchron und zeitlich verschränkt zueinander betrieben. Im Idealfall, wenn alle Sensoren identische Taktraten aufweisen, steigt die zeitliche Auflösung linear mit der Anzahl der Sensoren. Allerdings ist zu beachten, dass Latenzzeiten, die durch das Messprinzip der jeweiligen Sensorik bedingt sind, hiervon nicht beeinflusst werden können. Im Allgemeinen wird jedoch die Systemantwortzeit eines Multisensorsystems verringert, da mehr Daten in der gleichen Zeit zur Verfügung stehen.

Messprinzip: Mittels geeigneter Algorithmen, lassen sich mit mehreren Sensoren völlig neue oder erweiterte Messprinzipien realisieren. Mit zwei Kameras und Methoden der Stereobildverarbeitung können zum Beispiel Entfernungen direkt gemessen werden. Dies ist mit einer Kamera allein nicht unmittelbar möglich.

Alle oben genannten Vorteile lassen sich nur selten in einem System gleichzeitig erreichen. Vielmehr werden im Allgemeinen, je nach gewünschter Anwendung und zur Verfügung stehender Sensorik, bestimmte Aspekte gezielt herausgegriffen, die durch den Einsatz mehrerer Sensoren optimiert werden.

2.3.2 Fusionsmodelle

Maschinelle Sensordatenfusion wurde erstmals in den 80er Jahren im militärischen Bereich angewendet [Var97]. Aus dieser Zeit stammen die ersten Fusionsmodelle, wie das JDL-Modell⁴ oder der Boyd-Steuerungszyklus [Boy87]. Unter einem Modell ist in diesem Zusammenhang eine Beschreibung von Prozessen zu verstehen, die für ein funktionelles Fusionssystem realisiert werden müssen. Aspekte der Softwareumsetzung oder Implementierungsdetails sind dabei nicht von Interesse.

Im erweiterten JDL-Modell [SBW99], das aktuell einen de facto Standard für Fusionssysteme im militärischen Bereich darstellt, wird die Informationsverarbeitung in fünf Ebenen unterteilt, wobei eine zunehmende Informationsverdichtung stattfindet [RPL07].

Vorverarbeitung (Ebene 0): Die Anfangsphase beinhaltet eine zeitliche und räumliche Registrierung der Daten sowie Vorverarbeitungsmaßnahmen auf Signal- oder Pixelebene.

Objekterkennung und Merkmalsextraktion (Ebene 1): Diese Ebene bezieht sich auf die Schätzung und Vorhersage von kontinuierlichen (Position und Kinematik) oder diskreten (Objektverhalten und -identität) Objektmerkmalen.

Situationsanalyse (Ebene 2): Alle detektierten Objekte werden in einen größeren Kontext gebracht, indem Objektbeziehungen analysiert werden. Dadurch wird eine Situationsinterpretation bezüglich des aktuellen Missionsstatus möglich.

Bedrohungsanalyse (Ebene 3): Situationsabhängig werden, im Sinne einer Risikominimierung, unterschiedliche Handlungsoptionen evaluiert.

Prozessoptimierung (Ebene 4): Diese Phase ist ein Metaprozess, der phasenübergreifend agiert und Laufzeitanforderungen überwacht sowie den Ressourcenbedarf (Sensoren und Rechenkomponenten) optimiert.

⁴ Das Modell wurde von der *Joint Directors of Laboratories (JDL) Data Fusion Working Group* entwickelt, die 1986 vom amerikanischen Militär zur Vereinheitlichung der Datenfusionsterminologie ins Leben gerufen wurde.

Das JDL-Modell beschreibt keinen Ablauf in einem Datenfusionssystem, sondern gibt lediglich die verschiedenen Hierarchieebenen wieder, in denen Informationen verarbeitet werden. Neben diesem Defizit wirken speziell die Ebenen 2 und 3 oftmals deplaziert im nicht militärischen Gebrauch. Ein etabliertes Fusionsmodell ist das Omnibus⁵ Modell [BO00], das Aspekte aus älteren Modellen aufgreift und zu einem neuen Modell kombiniert (siehe auch Abbildung 2.4). Die Prozesse „beobachten“, „ausrichten“, „entscheiden“ und „handeln“ werden als zyklisches Ablaufdiagramm – entsprechend dem Boyd-Steuerungszyklus [Boy87] – dargestellt und anhand des Wasserfallmodells [Bed92] detailliert. Die Prozessübergänge markieren verschiedene Fusionsstufen (Rohdatenfusion, Merkmalsfusion, Soft-decision Fusion, Hard-decision Fusion), die aus dem Dasarathy Modell [Das97] abgeleitet sind. Darüber hinaus wird eine Militärterminologie vermieden.

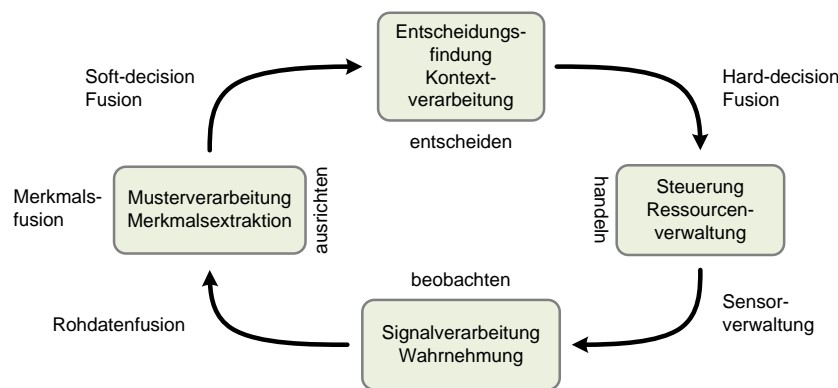


Abbildung 2.4: Omnibus Prozessmodell für Fusionssysteme. Zwischen den zyklisch ablaufenden Prozessen Signalverarbeitung, Musterverarbeitung, Entscheidungsfindung und Steuerung werden die entsprechenden Fusionsarten markiert [BO00].

Die vorgestellten Fusionsmodelle ermöglichen die Unterteilung eines komplexen Gesamtsystems in kleinere Teilsysteme. Analog dazu wird die Problemstellung in möglicherweise leichter zu bewältigende Teilprobleme zerlegt. Keines der Fusionsmodelle dient jedoch als optimale Vorlage für beliebige Aufgabenstellungen bezüglich der Sensorfusion. Vielmehr fungieren sie als Richtlinien, die stets an die speziellen Bedürfnisse angepasst werden müssen.

2.3.3 Fusionstaxonomien

Eine von Fusionsmodellen unabhängige Einteilung kann über den Zeitpunkt und die Art und Weise der Fusion gegeben werden. So kann eine Sensorfusion auf verschiedenen Abstraktionsebenen der Datenverarbeitungskette erfolgen. Unterschieden wird dabei zwischen einer Fusion auf Signal-, Merkmals- oder Objektebene [RPL07]. In diesem Zusammenhang wird in der Literatur häufig von früher oder später Fusion gesprochen, wobei sich „früh“ bzw. „spät“ auf den Fusionszeitpunkt in der Datenverarbeitungskette und somit auf die jeweilige Abstraktionsebene bezieht.

Fusion auf Signalebene (Rohdatenfusion): Die Fusion erfolgt zu einem frühen Zeitpunkt in der Datenverarbeitungskette, in dem Sensorrohdaten vorliegen. Als Rohdaten werden meist

⁵ omnibus = umfassend (AE)

Daten bezeichnet, die unmittelbar von einem Sensor stammen und noch nicht weiterverarbeitet wurden. Im Sensorsystem selbst kann jedoch schon eine Vorverarbeitung stattgefunden haben (z. B. Rauschunterdrückung bei einem Kamerasystem).

Fusion auf Merkmalsebene: Bei einer Merkmalsfusion werden unabhängig für jeden Sensor Attribute, wie beispielsweise Kanten im Bild, aus den Messdaten extrahiert. Anschließend werden die Merkmalsdaten der unterschiedlichen Sensoren kombiniert, um eine kollektive Objektbeschreibung abzuleiten.

Fusion auf Objektebene: Bei einer Fusion auf Objektebene wird die gesamte Datenverarbeitungskette für jeden Sensor eigenständig durchlaufen. Erst am Ende des Objektflusses werden die Objektdaten zu einer gemeinsamen Aussage kombiniert. Weitere oftmals synonym verwendete Bezeichnungen für diese Art der Fusion sind späte oder entscheidungsbasierte Fusion.

Bei der Frage nach dem optimalen Fusionszeitpunkt besteht ein Zielkonflikt hinsichtlich Effektivität und Effizienz. Auf späteren Abstraktionsebenen wird durch die Datenverarbeitung zwar die Datenmenge, aber zugleich auch der Informationsgehalt vermindert. Dadurch ist eine Fusion auf Objektebene in der Regel lauffzeiteffizienter als eine Fusion auf Signalebene, aber zugleich weniger effektiv. Aus informationstheoretischer Sicht sollte frühestmöglich fusioniert werden, wobei eine pauschale Aussage nicht getroffen werden kann, da der Fusionsnutzen stark von den beteiligten Sensoren und den eingesetzten Algorithmen abhängt.

Ein höherer Informationsgehalt ist durch die Kombination von redundanten oder komplementären Informationsquellen erzielbar. Sensorredundanz kann Unsicherheiten vermindern, Genauigkeiten erhöhen und die Ausfallsicherheit anheben. Sensorheterogenität erweitert den Merkmalsraum, indem durch verschiedenartige Messverfahren unterschiedliche Aspekte aus der Umgebung extrahiert werden [Var97]. Dieser beiden Prinzipien bedient sich Durrant Whyte [DW88] zur Aufgliederung der Fusionsarten hinsichtlich der Art und Weise, wie die Sensoren untereinander interagieren. Er unterscheidet komplementäre, konkurrierende und kooperative Fusionsansätze.

Komplementäre Fusion: Arbeiten zwei oder mehr Sensoren unabhängig voneinander und beobachten unterschiedliche räumliche, zeitliche oder physikalische Aspekte der Realität, kann eine komplementäre Fusion die einzelnen Sensordaten kombinieren. Von algorithmischer Seite ist diese Fusion einfach zu realisieren, da die Daten lediglich zeitsynchron vereinigt werden müssen. Ein Beispiel für eine komplementäre Fusion sind zwei Videokameras, die voneinander disjunkte Bereiche eines Raumes beobachten. Durch eine Vereinigung der beiden komplementären Sichtbereiche vergrößert sich der erfasste Gesamtbereich entsprechend.

Konkurrierende Fusion: Sensoren, die die gleichen Aspekte und Merkmale der Realität beobachten, können als konkurrierend aufgefasst werden. Durch diese redundante Perzeption können die beteiligten Sensoren kongruieren, aber auch im Widerspruch zueinander stehen. Geeignete Algorithmen werten die Einzelschätzungen aus und bilden daraus eine gemeinschaftliche Entscheidung. Beispielsysteme für eine konkurrierende Fusion finden sich in Flugzeugen (Fly-by-wire Systeme) bzw. auch in Fahrzeugen (Pedalwertgeber in der Motorsteuerung). Um die Fehlertoleranz zu erhöhen werden essentielle Sensoren und Kommunikationspfade mehrfach redundant ausgelegt. Durch einen verschränkten Datenabgleich können sowohl Sensorausfälle kompensiert als auch Übertragungsfehler erkannt und korrigiert werden.

Kooperative Fusion: Bei einer kooperativen Fusion werden die beteiligten Sensoren zu einem Kollektiv zusammengeschlossen. Geeignete Algorithmen extrahieren neue Informationen über die gemeinsame Beobachtung, die einem Einzelsensor im Kollektiv verborgen wären. Stereoskope Kamerasysteme beispielsweise erlauben über Epipolareometrie und Triangulation Entfernungsbestimmungen, die über ein monokulares System nicht direkt möglich sind. Robuste Algorithmen für eine kooperative Fusion zu finden ist schwierig, da kleine Einzelsensorunzuverlässigkeiten das Gesamtergebnis negativ beeinflussen können. Dadurch wird, im Gegensatz zur redundanten Perzeption einer konkurrierenden Fusion, die Robustheit und Genauigkeit des Multisensorsystems vermindert.

Bei einem realen Fusionssystem liegen in der Regel keine dieser Aufgliederungen in Reinform vor. Es existieren vielfältige Mischformen. In einem Fusionssystem zur Objekterkennung mit Stereobildverarbeitung beispielsweise, kann in überlappenden Bildbereichen kooperative (Entfernungsbestimmung über Triangulation) und in disjunkten Bildbereichen komplementäre Fusion (Erhöhung des Abdeckungsbereichs durch Vereinigung) stattfinden. Dies erfolgt auf Rohdaten- und Merkmalsebene. Ist an dem System zusätzlich entfernungsgebende Sensorik beteiligt, könnte überdies konkurrierend und entscheidungsbasiert fusioniert (redundante Fusion auf Objektebene) werden.

2.3.4 Projektlandschaft im Kontext automotiver Fusionssysteme

Von 1987 bis 1995 wurden in dem europäischen Forschungsprojekt PROMETHEUS⁶ die Grundlagen für aktuelle Systeme zur Fahrumfelderfassung im Fahrzeug gelegt. Basierend auf der Fusion von mehreren Kameras ist, anlässlich der Abschlusspräsentation von PROMETHEUS, 1994 ein Fahrzeug autonom mehrere hundert Kilometer im realen Straßenverkehr auf einer Autobahn in der Nähe von Paris gefahren [Dic02]. Seitdem beschäftigen sich nationale und internationale Nachfolgeprojekte in enger Zusammenarbeit zwischen Zulieferern, Automobilherstellern und Forschungsinstituten mit innovativer Sensorik, einer Erweiterung der Fahrumfelderfassung auf komplexe Szenarien und der Entwicklung neuartiger Fahrerassistenzsysteme. Hierbei wurden insbesondere von den Projekten PRéVENT [Eur07a], INVENT [Bun07], RADARNET [Eur07b] und CARSENSE [INR02] Aspekte der Sensorfusion näher betrachtet. Zum einen wurden unterschiedlichste Fusionsarchitekturen wie Frühe-, Späte oder Feedback-Fusion entwickelt und zum anderen verschiedene Sensorkonfigurationen bestehend aus Radar, Lidar und bildgebender Sensorik untersucht.

Als Ergebnis der genannten Forschungsaktivitäten befinden sich inzwischen erste Produkte auf dem Markt, die Methoden der Sensorfusion für Fahrerassistenzsysteme verwenden. Daimler beispielsweise ergänzt in der S-Klasse seit 2005 den Fernbereichsradar, der bisher für ACC verwendet wurde, um Radarsensorik für den Nahbereich. Durch die zusätzlich gewonnene Genauigkeit und den vergrößerten Erfassungsbereich wurden Einparkassistent, ACC S&G, Kollisionswarnung sowie ein Bremsassistent realisiert. Seit Herbst 2006 ist eine autonome Teilbremsung verfügbar, die ausgelöst wird, wenn der Fahrer die optische und akustische Kollisionswarnung ignoriert. In den aktuellen Lexus Modellen wird hingegen eine Kombination aus NIR-Stereobildverarbeitung und Radarsensorik für ACC S&G sowie für eine Kollisionswarnung verwendet. Durch die Fusion lässt sich ein größeres Spektrum an detektierbaren Objekten erreichen.

⁶ Akronym für „Program for European Traffic with Highest Efficiency and Unprecedented Safety“.

2.4 Versuchsträger

Im Rahmen des europäischen Forschungsprojektes COMPOSE⁷ zur Verkehrssicherheit, demonstrierte BMW, in Zusammenarbeit mit weiteren Projektpartnern, zukünftige Fahrerassistenzsysteme mit hohen Ansprüchen an die unterliegende Fahrumfelderfassung. Somit war neben der Entwicklung und dem Vergleich von Fusionskonzepten, die Systemintegration ins Fahrzeug ein Bestandteil dieser Arbeit. Dabei sind sowohl Echtzeitanforderungen als auch Randbedingungen, wie vordefinierte Koordinatensysteme und limitierte Ressourcen zu berücksichtigen.

Als Versuchsfahrzeug wurde eine BMW Limousine verwendet, um genügend Verbauplatz für die Sensorik und die Rechensysteme bereitzustellen. Alle für die Fahrumfelderfassung relevanten Sensoren wurden in die Frontpartie des Fahrzeugs integriert und deren Detektionsbereiche auf den Bereich vor dem Fahrzeug ausgerichtet. Das Rechensystem ist im Kofferraum verbaut und basiert auf handelsüblichen und fahrzeugtauglichen Komponenten. Durch die unterschiedlichen Öffnungswinkel und Verbauporte der Sensoren ergeben sich redundante und exklusive Abdeckungsbereiche. Eine Datenfusion ist nur in den redundanten Bereichen möglich. Durch die Ausrichtung der Sensoren werden vornehmlich Applikationen wie beispielsweise ACC und AGB adressiert.

2.4.1 Aufbau und Sensorsystem

Die im Versuchsfahrzeug verbauten Sensoren lassen sich in entfernungs- und bildgebende Sensoren gliedern. Unter die entfernungsgebenden Sensoren fallen ein Laserscanner und je zwei Nah- und Fernbereichsradare. Das bildgebende Sensorsystem setzt sich aus einer Videokamera mit Teleobjektiv, einer Videokamera mit Standardobjektiv und einer Wärmebildkamera zusammen. Die Verbauporte der Sensoren befinden sich hinter der Windschutzscheibe und im Bereich der Fahrzeug-Schürze (siehe Abbildung 2.5 für einen Verbaublick und Tabelle 2.2 für Sensordetails).

Tabelle 2.2: Details zur Sensorik des Versuchsfahrzeugs (horizontaler und vertikaler Öffnungswinkel, Reichweite für Fahrzeug/Fußgängererkennung, Winkelauflösung, Abtastfrequenz und Kurzbeschreibung).

	hor. \sphericalangle	vert. \sphericalangle	Reichweite	Auflösung	Frequenz	Beschreibung
FIR Kamera	$\pm 17.6^\circ$	$\pm 13.4^\circ$	80/60 m	320×240 (NTSC/4)	29.97 Hz	Mikrobolomertechnik, Shutterabgleich
Video (Tele)	$\pm 8.5^\circ$	$\pm 6.5^\circ$	100/80 m	320×240 (NTSC/4)	29.97 Hz	CMOS Farbbildkamera mit Teleobjektiv
Video (Normal)	$\pm 21.2^\circ$	$\pm 16.2^\circ$	60/30 m	320×240 (NTSC/4)	29.97 Hz	CMOS Farbbildkamera

⁷ COMPOSE (collision mitigation and protection of road users) ist ein Teilprojekt des EU-Förderprojektes PReVENT.

Tabelle 2.2: Details zur Sensorik des Versuchsfahrzeugs (horizontaler und vertikaler Öffnungswinkel, Reichweite für Fahrzeug/Fußgängererkennung, Winkelauflösung, Abtastfrequenz und Kurzbeschreibung).

	hor. \sphericalangle	vert. \sphericalangle	Reichweite	Auflösung	Frequenz	Beschreibung
Laserscanner	$\pm 61^\circ$	$\pm 3.6^\circ$	120/50 m	0.25° (lateral)	10 Hz	Vier vertikale Strahlebenen mit einem Versatz von 1.8° .
Fernbereichsradar	$\pm 3.5^\circ$	$\pm 2.5^\circ$	150 m	1° (lateral)	14 Hz	FMCW ⁸ -Verfahren (77 GHz)
Nahbereichsradar	$\pm 40^\circ$	$\pm 8.5^\circ$	25 m	ca. 2° (lateral)	14 Hz	Kurzpulsradar (24 GHz)

Während einer Versuchsfahrt kann zwischen zwei Systembetriebsmodi gewählt werden: Im Aufzeichnungsmodus werden alle Sensordaten, die über digitale (CAN und Ethernet) und analoge (Composite Video) Kommunikationspfade an einen zentralen Rechner übermittelt werden, auf einen Festspeicher geschrieben. Eine zeitliche Sensorsynchronisation erfolgt softwareseitig über die Ankunftszeiten der Daten. Der zweite Betriebsmodus ist der Perzeptionsmodus, in dem alle Sensordaten dem Fahrumfelderfassungssystem zugeführt werden. Dieses ermittelt eine Objektliste bestehend aus detektierten Fahrzeugen und Fußgängern und übersendet sie an die Applikationsschicht.

Abbildung 2.6 zeigt Momentaufnahmen der Sensoren von einer Versuchsfahrt. Um den maximalen Datendurchsatz des Festspeichers nicht zu überschreiten, werden die Bilddaten vor der Speicherung auf eine Viertel-NTSC Größe skaliert und von RGB-Farb- nach Grauwertbilder konvertiert. Durch diese Maßnahmen wird die Datenrate auf ca. 7 Megabyte pro Sekunde begrenzt.

2.4.2 Koordinatensysteme

Um eine Verknüpfung zwischen Daten zu ermöglichen, ist es besonders in Multisensorsystemen wichtig, die Messwerte in einer einheitlichen mathematischen Form zu repräsentieren. Die Messwerte entfernungsgebender Sensoren können als Messpunkte im dreidimensionalen Raum aufgefasst werden. Ein einzelner Messwert \mathbf{x}^S wird durch einen vierdimensionalen Vektor $(x \ y \ z \ 1)^T$ in homogenen Koordinaten dargestellt. Die Extradimension ist nützlich weil dadurch affine Transformationen wie Rotationen, Skalierungen und auch Translationen durch Multiplikationen mit 4×4 Matrizen repräsentiert werden können [HZ03]. Mittels dieser Transformationsmatrizen können Messwerte in andere Koordinatensysteme überführt werden. Ein Koordinatensystem fungiert als Basis im Raum – relativ zu dieser können homogene Messwerte dargestellt werden.

⁸ Frequency Modulated Continuous Wave

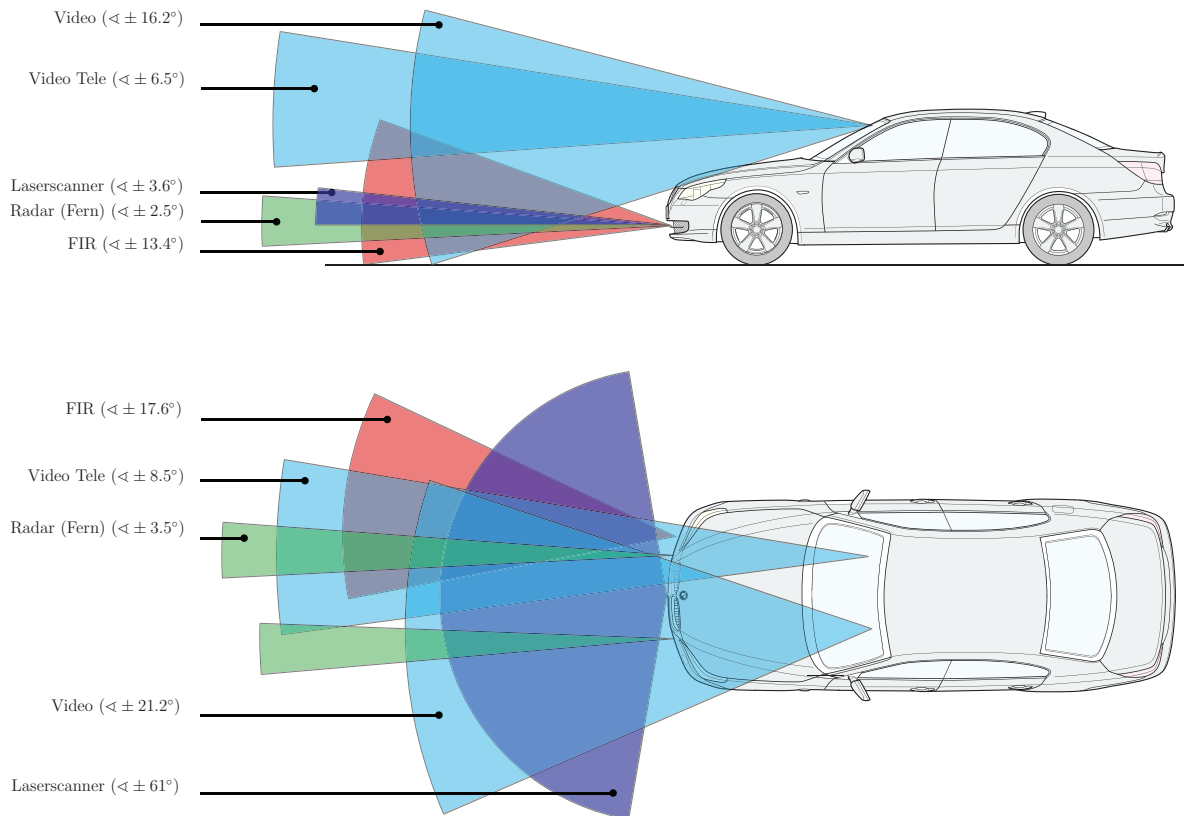


Abbildung 2.5: Sensorverbaupositionen und -sichtbereiche am Versuchsfahrzeug. Die vertikalen bzw. horizontalen Öffnungswinkel werden durch die obere bzw. untere Abbildung illustriert.

Abbildung 2.7 zeigt die relevanten rechtshändigen Koordinatensysteme einer multisensoriellen Fahrumfelderfassung. Unmittelbar nach einer Messung liegen die Daten zunächst im jeweiligen Sensorkoordinatensystem S_i vor. Die x -Achse weist entlang der zentralen Sichtachse des Sensors und der Ursprung liegt im Projektionszentrum. Sind mehrere Sensoren beteiligt, ist es wünschenswert, die verschiedenen Sensordaten in einem gemeinsamen Koordinatensystem darzustellen, um eine Datenfusion zu ermöglichen. Ein übergeordnetes Fahrzeugkoordinatensystem F kann dieses leisten. Die x -Achse des, durch die DIN Norm 70 000 [Nor94] beschriebenen fahrzeugfesten Koordinatensystems weist in Fahrzeuglängsrichtung, die y -Achse in Fahrzeugquerrichtung. Die z -Achse verläuft vertikal und ist orthogonal zur x - und y -Achse. Der Ursprung des Koordinatensystems ist die Projektion des vordersten Stoßfängerpunktes auf die Fahrbahnebene. Diese Wahl des fahrzeugfesten Koordinatensystemursprungs ermöglicht eine einfachere Vermessung und Kalibrierung der Sensorkoordinatensysteme. Die Koordinatensysteme H_k bzw. H_z sind lokale Koordinatensysteme von quaderförmigen bzw. zylindrischen Hypothesen, die datengetrieben initialisiert werden. Sowohl bei H_k als auch bei H_z liegt der Koordinatensystemursprung im, auf die Fahrbahnebene projizierten, Schwerpunkt der Hypothesen. Die x -Achse weist bei H_k entlang der Längsrichtung in Bewegungsrichtung und bei H_z entlang der Bewegungsrichtung.



Abbildung 2.6: Momentaufnahme der Sensordaten des Versuchsfahrzeugs während einer Aufnahme. Abbildung (a) und (b) zeigen die Bilder der beiden Videokameras mit Normal- bzw. mit Teleobjektiv. In Abbildung (c) wird das Bild der Wärmebildkamera dargestellt. Abbildung (d) zeigt die Daten des verbauten 4-Ebenen Laserscanners in einer virtuellen Umgebungsdarstellung.

Die Beziehungen zwischen verschiedenen Koordinatensystemen können als „Spatial Relationship Graph“ (SRG) [RH95, NWB04, PHBK06] beschrieben und visualisiert werden. Die Knoten repräsentieren dabei Koordinatensysteme und die Kanten starre oder dynamische räumliche Beziehungen zwischen den Systemen, die mittels Transformationsmatrizen \mathbf{T} darstellbar sind. Eine gute Einführung findet sich in [ABW01]. Ein typischer SRG einer Fahrumfelderfassung ist in Abbildung 2.8 dargestellt. Das Weltkoordinatensystem W ist das globale Koordinatensystem und beherbergt das eigene Fahrzeug F sowie das detektierte Fahrumfeld B . Relativ zum Fahrzeugkoordinatensystem sind die automotiven Sensoren S_1 bis S_n angebracht. Die räumlichen Verknüpfungen $F \leftrightarrow S_i$ zwischen Eigenfahrzeug und verbauten Sensoren sind starr und müssen über einen Kalibriervorgang bestimmt werden. Die Beziehung $W \leftrightarrow F$ zwischen Welt- und Fahrzeugkoordinatensystem ist dynamisch und ändert sich über die Zeit infolge der Eigenfahrzeugbewegung, die durch Verfahren der Odometrie ermittelt wird.

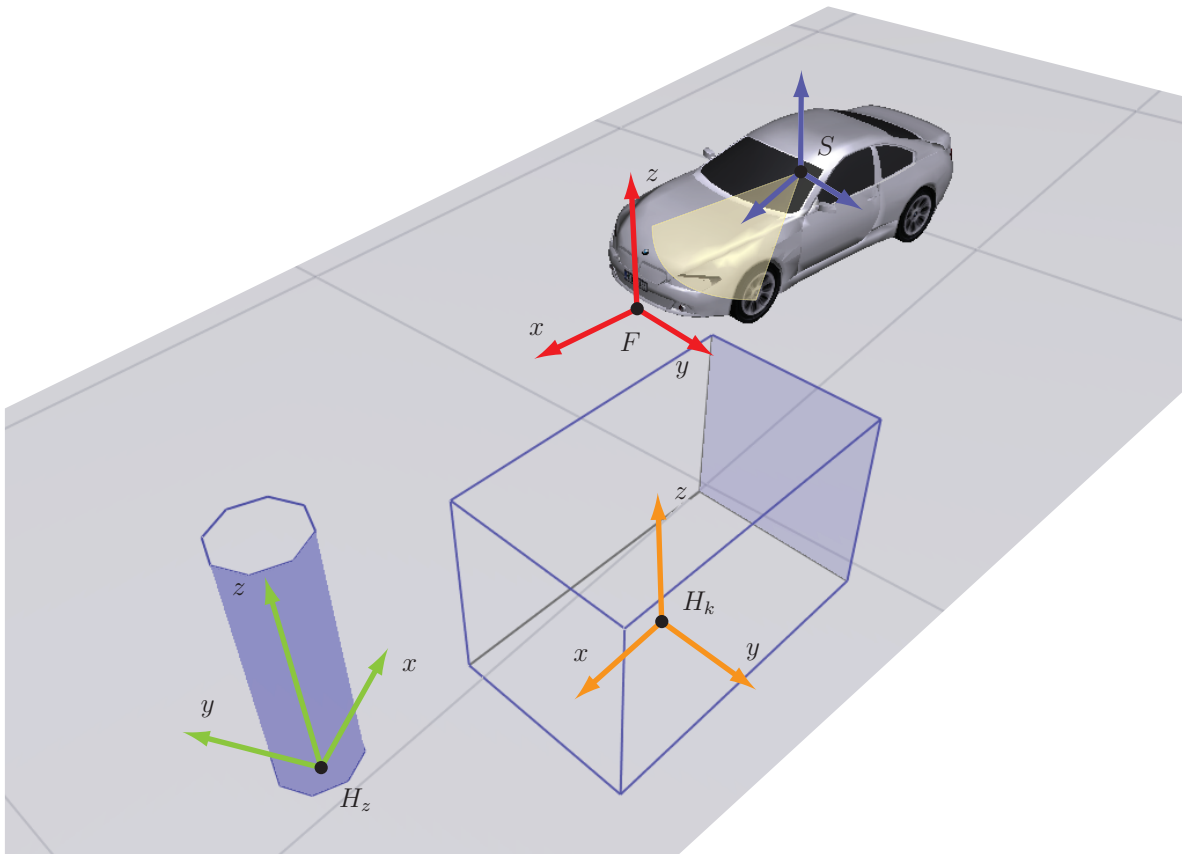


Abbildung 2.7: Die Abbildung zeigt den Ursprung und die Ausrichtung der verwendeten rechts-händigen Fahrzeug- (F), Sensor- (S), und Objektkoordinatensysteme (H_k bzw. H_z).

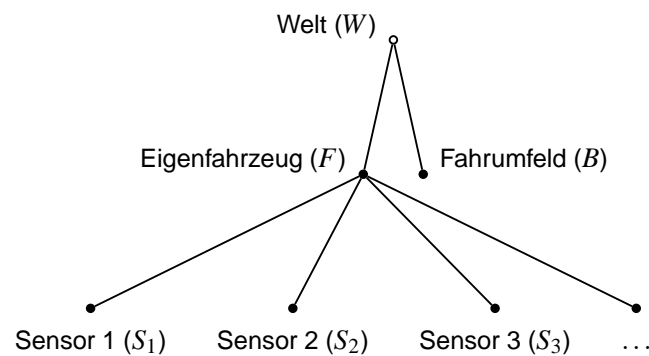


Abbildung 2.8: Dieser „Spatial Relationship Graph“ veranschaulicht die räumlichen Beziehungen von Sensoren, Eigenfahrzeug und Fahrumfeld.

Soll nun beispielsweise ein Messwert \mathbf{x}^S im Sensorkoordinatensystem S in einen Messwert \mathbf{x}^F in Fahrzeugkoordinaten F überführt werden, muss er lediglich mit der Transformationsmatrix

$\mathbf{T}_{S \rightarrow F}$ multipliziert werden. Über die Vorschrift $(\mathbf{T}_{A \rightarrow B})^{-1} = \mathbf{T}_{B \rightarrow A}$ können die Kanten des Szenengraphen in beiden Richtungen durchlaufen werden. Eine Umrechnung von einem Messwert \mathbf{x}^{S_1} , der im Sensorkoordinatensystem S_1 vorliegt, in ein anderes Sensorkoordinatensystem S_2 berechnet sich somit über $\mathbf{x}^{S_2} = \mathbf{T}_{F \rightarrow S_2} \mathbf{T}_{S_1 \rightarrow F} \mathbf{x}^{S_1} = (\mathbf{T}_{S_2 \rightarrow F})^{-1} \mathbf{T}_{S_1 \rightarrow F} \mathbf{x}^{S_1}$. Alternativ zu Transformationsmatrizen kann auch eine VQS-Darstellung mit Quaternionen [Vic01] verwendet werden, die speziell Rotationen kompakter kodieren kann.

2.5 Systemüberblick

Dieser Abschnitt gibt einen kurzen Überblick über die einzelnen Komponenten des Gesamtsystems und deren Zusammenspiel. Das Systemdesign und der zyklische Ablauf des Systems ist weitgehend an das Omnibus Fusionsmodell (siehe Abbildung 2.4 auf Seite 22) angelehnt. Das adaptierte Diagramm ist in Abbildung 2.9 dargestellt. Im Folgenden wird der Datenfluss innerhalb des Diagramms erläutert. Zudem werden die Komponenten Klassifikation, Hypothesenerzeugung und Tracking genauer beschrieben.

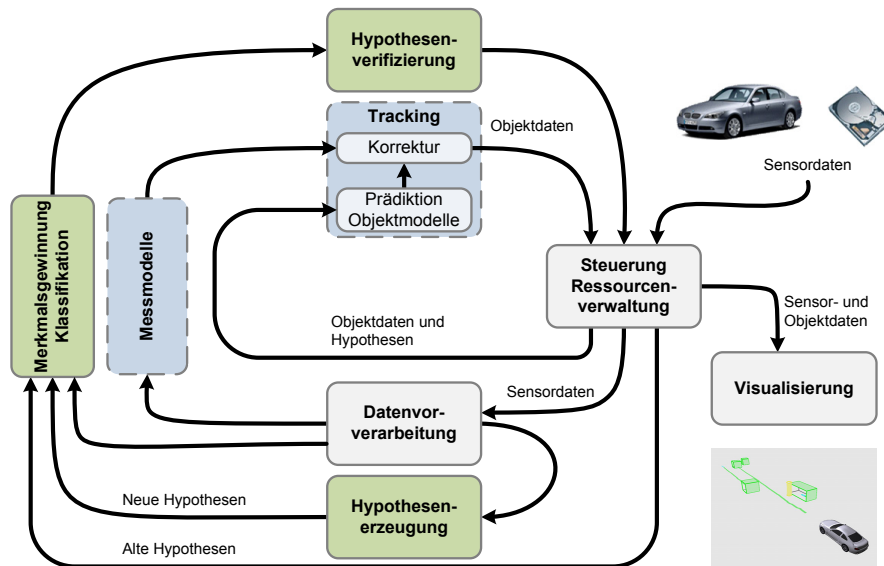


Abbildung 2.9: Klassifikations- (dunkel hinterlegt) und Trackingmodule (gestrichelt) als wesentliche Komponenten des Fahrumfelderfassungssystems.

Datenfluss: Der Datenfluss ist in die zwei Zyklen Klassifikation und Tracking unterteilt. Dabei werden Sensordaten, die entweder unmittelbar von der Sensorik stammen oder in aufgezeichneter Form vorliegen, in Objektmodelle umgewandelt und in die jeweiligen Objektzustände mit eingebracht. Werden neue Sensordaten über die Eingabesteuerung aufgenommen, werden diese zunächst vorverarbeitet und aggregiert. Anschließend werden basierend auf diesen vorverarbeiteten Daten neue Objektmodelle initialisiert. Über einen Klassifikationsprozess wird anhand von Merkmalen, die durch die Lage und Geometrie der Hypothesen gewonnen werden, eine Objekteinteilung vorgenommen. Zeitgleich werden durch das Tracking die Zustandsdaten der Hypothesen und der klassifizierten Objekte aktualisiert und zeitlich adaptiert. Eine Ressourcenverwaltung entfernt Objekte, die nicht länger durch Messwerte gestützt werden.

Klassifikation: Ein Klassifikator ist eine beliebige Funktion $D : \mathbb{R}^n \rightarrow \Omega$, die einem Merkmal $\mathbf{x} \in \mathbb{R}^n$ eine Klassenbezeichnung $\omega \in \Omega = \{\omega_1, \dots, \omega_c\}$ zuweist. Eine Klasse ω beinhaltet ähnliche Objekte und grenzt sich zu einer anderen Klasse ω' dadurch ab, dass deren Objekte unähnlich zu den eigenen Objekten sind. Objekte können anhand von charakteristischen Merkmalen beschrieben werden. In der statistischen Mustererkennung werden ausschließlich numerische Merkmale verwendet. Numerische Merkmale können zu einem n -dimensionalen Merkmalsvektor $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$ zusammengefasst werden. In Abbildung 2.9 sind die Klassifikationsmodule durch dunkel hinterlegte Rechtecke markiert.

Fahrzeuge und Fußgänger sind ein Beispiel für zwei Objektklassen. Gewicht und Abmessungen sind beispielsweise charakteristische Merkmale um Fußgänger von Fahrzeugen unterscheiden zu können. Weniger geeignete Merkmale wären die Farbe und das Baujahr. In dieser Arbeit werden ausschließlich binäre Klassenentscheidungen ($c = 2$) betrachtet. Durch die Vorklassifizierung im Hypothesenerzeugungsprozess, wird das Mehrklassenproblem (Fahrzeug, Fußgänger, Hintergrundobjekt) auf mehrere binäre Klassenprobleme abgebildet. Als Klassifikationsverfahren werden ausschließlich Support-Vektor-Maschinen (SVM) verwendet (siehe auch Abschnitt B.3 im Anhang für eine Einführung in SVM).

Für den Vergleich der Entscheidungsverfahren werden Bewertungskriterien wie Sensitivität bzw. Erkennungsrate und Fehlerkennungsrate sowie ROC- und DET-Kurven herangezogen. Eine detaillierte Beschreibung der Konzepte findet sich im Anhang B auf Seite 183.

Tracking: Der Begriff Tracking⁹ beschreibt den Prozess, mit dem Objekte über mehrere Zeitschritte hinweg verfolgt werden. Dabei wird versucht aus Messungen bzw. Beobachtungen des Objektes Zustandsgrößen wie zum Beispiel Position, Geschwindigkeit, Breite oder Orientierung abzuleiten bzw. zu schätzen. Ein weiterer wichtiger Aspekt des Trackings ist die Einbeziehung des zeitlichen Verlaufs bzw. der letzten Beobachtungen des zu verfolgenden Objektes. Dementsprechend wird ein Objekt im Kontext seiner Historie betrachtet. Auf diese Weise können zum einen Störeinflüsse wie zum Beispiel Messfehler kompensiert werden und zum anderen Zustandsgrößen wie die Beschleunigung abgeleitet werden, die sich einer direkten Beobachtung entziehen. In der Literatur wird der Prozess des Trackings auch oft als Filterung bezeichnet, da die Auswirkungen von Messfehlern abgeschwächt werden. In Abbildung 2.9 sind die Trackingmodule durch gestrichelte Rechtecke gekennzeichnet.

Hypothesengenerierung: Hypothesen sind mögliche Objekte des Fahrumfeldes. Sie werden in dieser Arbeit mittels quader- und zylinderförmiger Körper repräsentiert. Erstere approximieren Fahrzeuge und letztere werden für Fußgänger genutzt. Eine quaderförmige Hypothese $\mathcal{H}_k = \{\mathbf{x}, w, l, h, \phi\}$ ist über die Position \mathbf{x} im Weltkoordinatensystem, über die Orientierung ϕ , und über die Abmessungen Breite w , Länge l und Höhe h des Quaders eindeutig beschrieben. Analog dazu werden die Abmessungen einer zylindrischen Hypothese $\mathcal{H}_z = \{\mathbf{x}, r, h\}$ über Radius r und Höhe h festgelegt. Der Referenzpunkt der Objekte, der gleichzeitig auch der Ursprung der lokalen Koordinatensysteme H_k und H_z ist, liegt im Zentrum der Hypothesen auf der gedachten Straßenebene (vergleiche auch Abbildung 2.7 auf Seite 29).

Im nächsten Kapitel wird die Generierung von Hypothesen für Fahrzeuge und Fußgänger thematisiert.

⁹ engl. für das Verfolgen bzw. das Nachführen.

3

Hypothesengenerierung

Um relevante Verkehrsteilnehmer in den Sensordaten zu finden, ist eine erschöpfende Suche im Fahrumfeld eine pragmatische Lösung. Durch den hohen Aufwand ist diese Vorgehensweise mit heute erhältlichen Rechensystemen allerdings nur begrenzt in Echtzeit möglich. Mittels einer Erzeugung von punktuellen Objekthypothesen kann diese Suche vermieden werden, da der Bearbeitungsschwerpunkt hierbei lediglich auf messwertgestützte Vordergrundbereiche gelegt wird. Durch diesen Prozess der Aufmerksamkeitssteuerung kann sowohl der Verbrauch an Rechenzeit, als auch die Fehlerkennungsrate gesenkt werden.

Die Schätzung von Initialzuständen und die Vorklassifizierung von Objekten sind wesentliche Aspekte einer Hypothesengenerierung. Durch diese Schritte kann zum einen das Objekttracking erfolgreich angestoßen werden. Zum anderen erlaubt eine initiale Zustandsschätzung, in Verbindung mit einer Objektvorklassifizierung, eine gezielte und präzise Extraktion von multisensoriellen Merkmalen für die Hypothesenverifizierung. Im Bildverarbeitungsterminus entspricht eine Hypothesengenerierung einer Segmentierung des Sensorerfassungsbereichs in Vorder- und Hintergrundregionen bzw. in relevante und irrelevante Fahrumfeldobjekte.

Es werden die folgenden wesentlichen Anforderungen an ein Hypothesengenerierungsverfahren im automotiven Umfeld gestellt: Entscheidend ist eine umfassende und schnelle Objekterfassung, sowie eine gute initiale Zustandsschätzung der Hypothesen. Dies impliziert eine feine zeitliche und räumliche Granularität im Absuchen des Messwertraums nach Objekthypothesen. Da insbesondere bei Fahrumfelderfassungssystemen die Echtzeitfähigkeit einen hohen Stellenwert einnimmt, ist gleichzeitig eine hohe Verarbeitungsgeschwindigkeit erforderlich. Um diesen Zielkonflikt zu entschärfen, muss die Berechnungskomplexität der angewendeten Algorithmen begrenzt werden. Zudem sind adäquate Datenvorverarbeitungsschritte, wie Datenaggregationen und -filterungen nötig, um den Suchraum entsprechend zu verringern. Eine ideale Aufmerksamkeitssteuerung identifiziert somit bei minimalem Ressourcenverbrauch alle relevanten Objekte des Fahrumfeldes und extrahiert deren Initialzustände.

In der Literatur gibt es zahlreiche Verfahren zur Erzeugung von Fahrzeug- und Fußgängerhypothesen bzw. zur Aufmerksamkeitssteuerung. Ein guter Überblick findet sich in [SBM06, GT07]. In den folgenden Absätzen werden Literaturangaben zu Hypothesenerzeugungsverfahren für Fahrzeuge und Fußgänger mittels Videosensor, Wärmebildkamera, Stereosystem und Radar bzw. Laserscanner gegeben.

Die videobasierten Verfahren lassen sich in wissens- und bewegungsbasierte Vorgehensweisen untergliedern: Bei den wissensbasierten Verfahren zur Fahrzeugerkennung im Videobild werden Symmetrieeigenschaften [Kue91, ZBvS92], Farbe [BD98, CT88], Schatten [MC93, TvS98], Textur [KTVS98], Kantenverteilungen [MACH96, BHD00, SMBD02] und die Anordnung der Rücklichter [CP99] als charakteristische Merkmale herangezogen. Im Gegensatz dazu ist für Fußgänger im Videobild eine effiziente wissensbasierte Hypothesenerzeugung – aufgrund der hohen Gestaltvielfalt – problematisch. Demzufolge existieren hierfür nur vereinzelte Verfahren, die vornehmlich Symmetrieeigenschaften und Kantenverteilungen in der Beinregion zur Aufmerksamkeitssteuerung auswerten [HSS04, AS04]. Bewegungsbasierte Verfahren eignen sich zur Detektion von allen Objekten, die eine signifikante Bewegungsdifferenz zum Hintergrund aufweisen. Hierzu werden unterschiedliche Bewegungsbereiche im Bild identifiziert, indem der optische Fluss [HS81] anhand aufeinanderfolgender Bilder ermittelt wird. Diese Verfahren werden erfolgreich für die Erzeugung von Fahrzeug- und Fußgängerhypothesen angewendet [KER95, BPT97, GCTC98, HMYO03]. Allerdings können Objekte mit geringer Relativbewegung kaum wahrgenommen werden.

Wärmebildsensorik spielt erst seit Kurzem eine größere Rolle zur Detektion von Objekten im automotiven Umfeld. Insbesondere Fußgängerhypothesen können aufgrund guter Bildkontrastverhältnisse robust erzeugt werden. Broggi [BFC⁺04] nutzt hierfür die spezifische Größe und das Seitenverhältnis von Fußgängern. Schwellwertbasierte Ansätze werden ferner in [THWN02, FYN⁺04, XLF05] betrachtet.

Bei den passiven entfernungsgebenden Sensoren werden Verfahren zur Stereobildverarbeitung mittels Videokameras [PP00, GGM⁺04, GT06, GM07] und seit Kurzem mittels Wärmebildkameras [LF04, BBLR05] genutzt, um Hypothesen zu erzeugen. Der Einsatz von Stereosystemen im Fahrzeug ist jedoch mit Schwierigkeiten verbunden, da abgesehen von der erschwerten Verbaubarkeit, thermische und mechanische Einflüsse eine automatische Kalibrierung des binokularen Systems notwendig machen.

Radar und Laserscanner spielen als aktive entfernungsgebende Sensoren eine große Rolle. Aufgrund der Messeigenschaften ist eine laterale Positionsschätzung mittels Radar jedoch ungenau. Für den Laserscanner werden im automotiven Umfeld vorwiegend einfache abstandsbaasierte Segmentierungen vorgenommen [EW00, FW01, FLD03], um Hypothesen zu initialisieren.

In der Literatur finden sich gegenwärtig kaum Ansätze zur kombinierten Hypothesenerzeugung für Fahrzeuge und Fußgänger in einem System. Zudem wird erst in den letzten Jahren eine multisensorielle Aufmerksamkeitssteuerung in Betracht gezogen [FYN⁺03, SCFW04]. Interessant ist in diesem Zusammenhang die Fragestellung, welche Auswirkungen eine multisensorielle Hypothesenerzeugung auf die Leistungsmerkmale eines Erkennungssystems hat.

In diesem Kapitel werden Hypothesenerzeugungsverfahren für Verkehrsteilnehmer mittels entfernungs- und bildgebenden Sensoren betrachtet. Dabei wird in Abschnitt 3.1.1 zunächst ein Rohpunkttaggregierungsverfahren für Laserscanner vorgestellt, auf dessen Basis Fahrzeug- und Fußgängerhypothesen simultan erzeugt werden können. Abschnitt 3.2 widmet sich der Hypothesenerzeugung im Wärme- und Videobild anhand von vertikalen Kanten. Im letzten Teil des Kapitels (Abschnitt 3.3) wird untersucht, über welches Potential eine kollektive Hypothesenerzeugung verfügt. Die räumliche Repräsentation und Notation der Fahrzeug- und Fußgängerhypothesen \mathcal{H}_k bzw. \mathcal{H}_z wurde im Grundlagenkapitel auf Seite 30 eingeführt.

3.1 Entfernungsgabende Sensorik zur Hypothesenerzeugung

Entfernungsgabende Sensoren zur Hypothesenerzeugung sind Radare und Laserscanner. Durch das unterschiedliche Messprinzip können verschiedene Hypothesenausprägungen erzielt werden. Die verwendeten Punktzielradare generieren Entfernungsmesswerte, die keine Information über die Objektausdehnung beinhalten. Dafür kann neben der Objektposition, die radiale Geschwindigkeit unmittelbar über den Dopplereffekt gemessen werden.

Laserscanner erzeugen durch die Umgebungsabtastung ein Entfernungsprofil (siehe Abbildung 3.1). Aus den Entfernungswerten können zusätzlich Objektgrößen und -orientierungen bestimmt werden. Objektgeschwindigkeiten sind nicht unmittelbar messbar. Sie können allerdings indirekt durch eine zeitliche Beobachtung abgeleitet werden.

In diesem Abschnitt werden zunächst zwei Segmentierungsverfahren für Laserscannerrohdaten beschrieben, die zur Erzeugung von Zusammenhangskomponenten genutzt werden können. Nach dieser Eingrenzung des Suchbereichs wird auf eine messwertgetriebene Fahrzeug- und Fußgängerhypothesenerzeugung eingegangen. Hypothesengenerierungsverfahren auf Basis eines Radarsensors werden nicht gesondert thematisiert, da die Vorgehensweisen im Wesentlichen den Vorgehensweisen zur Erzeugung von laserscannergestützten Hypothesen entsprechen. Aufgrund des Messprinzips und der internen Vorverarbeitung entfällt bei Punktzielradaren jedoch sowohl die Datensegmentierung, als auch die Orientierungsschätzung der Hypothesen.

3.1.1 Segmentierungsverfahren

Das Ziel einer Rohdatensegmentierung ist es, zum einen zusammenhängende Bereiche zu bestimmen, die einem Objekt aus der Realität entsprechen, und zum anderen für die nachfolgenden Verarbeitungsschritte das Datenaufkommen zu reduzieren. Viele Segmentierungsalgorithmen für Laserscannerrohpunkte stammen aus der Robotik. Dort ermitteln autonom agierende Roboter ihre Position, indem sie räumliche Umgebungsdaten mit elektronischen Karten abgleichen. In diesem Zusammenhang wird häufig zwischen Verfahren unterschieden, die entweder auf Kanten oder auf Regionen arbeiten [HGS07].

Laserscannerrohpunkte, die zu einem Objekt gehören, liegen im Allgemeinen in enger Nachbarschaft zueinander (siehe auch Abbildung 3.1). Mit dieser Annahme lässt sich die Frage der Objektzugehörigkeit auf ein einfaches Entfernungsmaß reduzieren. Dementsprechend werden Messwerte \mathbf{x}_i , die einen geringen euklidischen Abstand zu benachbarten Messpunkten \mathbf{x}_j aufweisen, in einer Zusammenhangskomponente \mathcal{R}_i gruppiert. Hierfür existieren einfache und effiziente Algorithmen, wie das „Region-Growing“, das seinen Ursprung in der Bildverarbeitung hat [ST88].

3.1.1.1 Region-Growing Algorithmus

Das Region-Growing Verfahren bildet Zusammenhangskomponenten, indem von einem Startpunkt aus schrittweise Nachbarkpunkte hinzugenommen werden, die ein gewisses Nachbarschaftskriterium erfüllen. Für die Segmentierung der räumlichen Rohdatenpunkte ist dies der euklidische Abstand. Der Startpunkt wird dabei solange auf neu hinzugenommene Punkte verlegt, bis kein weiterer Punkt das Nachbarschaftskriterium erfüllt. Damit ist eine Region abgeschlossen, und es wird ein neuer Punkt als Startpunkt gewählt, der noch keiner Regi-

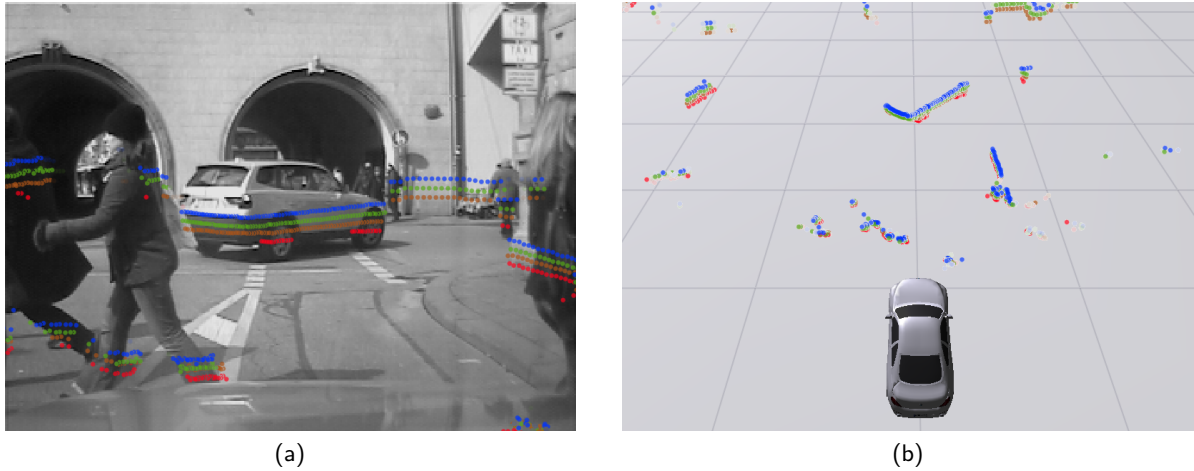


Abbildung 3.1: Beispielszene mit Laserscannerrohdaten. Objekte sind meist dadurch gekennzeichnet, dass sie durch eine kompakte Punktwolke repräsentiert werden (vergleiche das abbiegende Fahrzeug). Eine abstands-basierte Segmentierung ist jedoch nicht immer möglich (vergleiche querende Fußgänger).

on zugewiesen wurden. Eine detaillierte Beschreibung der Vorgehensweise für winkelsortierte Laserscannerrohdaten ist in [Tho05] zu finden.

Die Qualität der Segmentierung ist im Wesentlichen von der Wahl des Entfernungsschwellwertes τ abhängig. Wird dieser zu gering gewählt, so zerfallen Objekte in mehrere Zusammenhangskomponenten. Ist dieser hingegen zu groß, so verschmelzen benachbarte Objekte. Hierbei ist zu beachten, dass bedingt durch die limitierte Winkelauflösung β des Laserscanners, entfernte Objekte schlechter getrennt werden können. Aus diesem Grund sollte τ abhängig von der Entfernung r gewählt werden:

$$\tau = \max(\tilde{\tau}, r \tan(\beta)) \quad (3.1.1)$$

Ein minimaler Abstand $\tilde{\tau}$ verhindert, dass sehr nahe Objekte in Einzelobjekte zerfallen. Auf diese Weise ist sichergestellt, dass alle Regionen einen entfernungsunabhängigen Mindestabstand zueinander besitzen.

Nachdem alle Messwerte in zusammenhängende Bereiche zusammengefasst worden sind, werden für jede Region Merkmale wie der Schwerpunkt \mathbf{x}_c oder die Ausdehnung des kleinsten umschließenden Rechtecks bestimmt. Alle Rohdatenpunkte, sowie die daraus abgeleiteten Merkmale können als Grundlage für weitere Verarbeitungsschritte wie Klassifikation oder Tracking verwendet werden.

Bei modernen Laserscannern, die die Umgebung in mehreren Ebenen (vergleiche farblich gekennzeichnete Ebenen in Abbildung 3.1a) abtasten, kann das „Region-Growing“ entweder für alle Ebenen gemeinsam oder für jede Ebene einzeln durchgeführt werden. Ein Vorteil der ersten Variante ist, dass keine Mehrdeutigkeiten bzw. Überschneidungen entstehen, da der Mindestabstand τ zwischen allen Regionen gewahrt bleibt. Auf die Berechnung einzelner Merkmale kann die gemeinsame Betrachtung der Ebenen einen positiven oder negativen Einfluss haben. Sind zum Beispiel Messfehler (Ausreißer) über alle Ebenen gleichverteilt, so ist die Berechnung des Schwerpunktes robuster. Sind Messfehler auf einzelne Ebenen beschränkt, wie es häufig bei

LKW oder PKW zu beobachten ist (z. B. Messung der Fahrzeugbereifung und Teile des Unterbodens), so werden Merkmale wie der Schwerpunkt verfälscht. Wird das „Region-Growing“ für jede Ebene unabhängig durchgeführt, kann der Einfluss einzelner verrauschter Ebenen auf die Qualität der Merkmale minimiert werden. Allerdings entstehen Mehrdeutigkeiten und eine größere Anfälligkeit in Bezug auf globales Rauschen.

Um die Vorteile aus beiden Varianten zu kombinieren, wird ein zweistufiger Ansatz verwendet. In der ersten Stufe werden für jede Ebene Regionen gebildet. Anschließend werden diejenigen Regionen vereint (aggregiert), die sehr ähnliche Merkmale besitzen. Auf diese Weise können Mehrdeutigkeiten stark reduziert werden. Das Ähnlichkeitsmaß sollte so gewählt werden, dass Ebenen mit punktuellen Messfehlern nicht kombiniert werden. Hierfür kann zum Beispiel der euklidische Abstand zwischen den Schwerpunkten der einzelnen Regionen verwendet werden. Dieses Merkmal ist allerdings nicht invariant gegenüber Rotationen, was einen Ausgleich derartiger Fehler verhindert. Im Folgenden wird deshalb ein Verfahren vorgestellt, welches das Konzept der Regionen erweitert und die Berechnung weiterer Merkmale ermöglicht.

3.1.1.2 Inkrementelle Linienextraktion

Einer der schnellsten und genauesten Algorithmen zur Linienextraktion aus räumlichen Rohdaten ist der inkrementelle Line-Track Algorithmus [SKK⁺97, NMTS05]. Dieser stammt ursprünglich aus der Bildverarbeitung [FP03] und wurde auf die Aggregation von Laserscannerrohdaten übertragen. Analog zum „Region-Growing“-Algorithmus werden dabei Rohpunkte zusammengefasst, die einen geringen euklidischen Abstand zueinander aufweisen. Diese Regionen werden jetzt als Segmente $\mathcal{S}_i = [\mathbf{k}_1, \dots, \mathbf{k}_n]$ bezeichnet, da zusätzlich Ausgleichsgeraden in die Punktwolken eingepasst werden, die durch Kontrollpunkte \mathbf{k}_j repräsentiert werden. Dadurch können weitere Merkmale, wie Orientierung und Ausdehnung, abgeleitet werden. Die Eingabe für den Algorithmus sind winkelsortierte Rohdatenpunkte \mathbf{x}_i , die ebenenweise abgearbeitet werden.

Der detaillierte Ablauf der inkrementellen Linienextraktion (ILE) ist wie folgt: Zunächst wird durch zwei Ausgangspunkte \mathbf{x}_1 und \mathbf{x}_2 eine Gerade mit der Gleichung $y = a + bx$ gelegt. Anschließend werden so lange weitere Punkte \mathbf{x}_i zu den zwei Ausgangspunkten hinzugenommen und eine aktualisierte Ausgleichsgerade durch die anwachsende Punktmenge \mathcal{P} gelegt, bis ein Fehlerschwellwert σ_e überschritten wird. Als Fehler wird in der Regel der durchschnittliche Abstand $D = \frac{1}{n} \sum_{i=1}^n d_i$ der Rohpunkte $\mathbf{x} \in \mathcal{P}$ zur Geraden gewählt. Der orthogonale Abstand d_i eines Punktes $(x_i \ y_i)^T$ zur Geraden berechnet sich über $d_i = \frac{|y_i - (a + bx_i)|}{\sqrt{1+b^2}}$. Die Summe der Abstände $R = \sum_{i=1}^n d_i$ wird zudem verwendet, um die Parameter a und b der Ausgleichsgeradengleichung zu ermitteln [Wei03]. Dazu wird R^2 minimiert, indem über die partiellen Ableitungen

$$\begin{aligned} \frac{\partial R^2}{\partial a} &= \frac{2}{1+b^2} \sum_{i=1}^n [y_i - (a + bx_i)](-1) = 0 \\ \frac{\partial R^2}{\partial b} &= \frac{2}{1+b^2} \sum_{i=1}^n [y_i - (a + bx_i)](-x_i) + \sum_{i=1}^n \frac{[y_i - (a + bx_i)](-1)(2b)}{(1+b^2)^2} = 0 \end{aligned}$$

die Extremwerte der Funktion bestimmt werden. Nach Einsetzen und Umformen ergeben sich die Parameter der Geradengleichung zu $a = \bar{y} - b\bar{x}$ und $b = -B \pm \sqrt{B^2 + 1}$, mit $B = \frac{1}{2} \frac{(\sum_{i=1}^n y_i^2 - n\bar{y}^2) - (\sum_{i=1}^n x_i^2 - n\bar{x}^2)}{n\bar{x}\bar{y} - \sum_{i=1}^n x_i y_i}$.

Überschreitet der Fehler σ_e einen Schwellwert τ_σ , wird eine neue Ausgleichsgerade innerhalb des aktuellen Segments \mathcal{S} angelegt, die direkt am Endpunkt der alten Ausgleichsgerade anknüpft. Dazu wird \mathcal{P} neu initialisiert. Ein Segment besteht somit aus einem kreisfreien geschlossenen Streckenzug, der durch eine Menge von Kontrollpunkten $\mathcal{S} = [\mathbf{k}_1, \dots, \mathbf{k}_n]$, die die Schnittstellen der Einzelstrecken markieren, repräsentiert werden kann. Ein Kontrollpunkt wird berechnet, indem der Rohpunkt, bei dem der Fehler über dem Schwellwert τ_σ liegt, auf die aktuelle Gerade projiziert wird.

Das alte Segment \mathcal{S}_j wird abgeschlossen und ein neues Segment \mathcal{S}_{j+1} erzeugt, wenn der euklidische Abstand $\sigma_d = \|\mathbf{x}_{k+1} - \mathbf{x}_k\|_2$ zwischen dem aktuellen Rohdatenpunkt \mathbf{x}_k und dem neu hinzugenommenen Rohdatenpunkt \mathbf{x}_{k+1} einen Schwellwert τ_{dist} überschreitet. Der Algorithmus terminiert, wenn alle Rohpunkte aufgebraucht sind.

Damit der Algorithmus für den automotiven Einsatz verwendet werden kann, wurden zahlreiche Anpassungen vorgenommen. Ein ausschlaggebender Parameter für die Performanz ist der maximal zulässige Abstand τ_{dist} zwischen zwei Rohpunkten. Wird dieser Schwellwert zu klein gewählt, zerfallen die Segmente unter Umständen zu stark und die Ein-Segment-Ein-Objekt-Beziehung wird häufig verletzt. Eine zu große Wahl von τ_{dist} hat zur Folge, dass Segmente leichter zusammengeführt werden und dadurch Objekte, die in der realen Welt getrennt sind, als ein einzelnes Segment repräsentiert werden. Deshalb hat sich herausgestellt, dass τ_{dist} zunächst klein gewählt werden sollte, und anschließend in einem zweiten Schritt die Segmente mit den zusätzlichen Richtungsinformationen der Zusammenhangskomponenten verlängert, konkateniert und aggregiert werden sollten (siehe Abbildung 3.2).

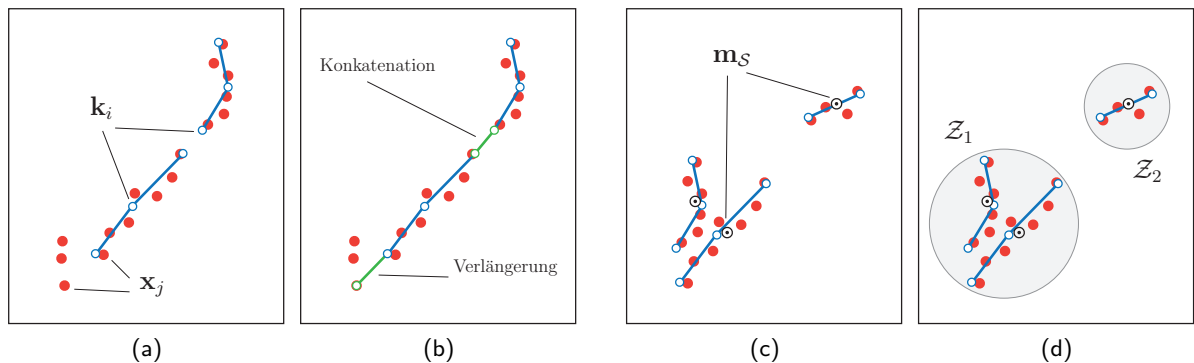


Abbildung 3.2: Verlängerung, Konkatenation (a,b) und Aggregation (c,d) für Laserscanner-Segmente. Mit \mathbf{k}_i werden die Kontrollpunkte und mit \mathbf{x}_i die zugrundeliegenden Rohpunkte des Segments notiert. Die Diagramme zeigen die Segmente aus einer Vogelperspektive.

Segmentverlängerung: Manche Rohpunkte können beim ersten Segmentierungsschritt keinem Segment zugeordnet werden, da der Abstand, den sie zu den benachbarten Rohpunkten aufweisen, größer als der Schwellwert τ_{dist} ist. Oftmals sind solche vereinzelt Rohpunkte Bestandteil eines angrenzenden Segments. Damit möglichst die Gesamtausdehnung aller Objekte erfasst wird, werden die Segmente nun an den beiden Endpunkten verlängert. Dazu wird für alle einzelnen Rohpunkte \mathbf{x} untersucht, ob sie einen Abstand τ_{verl} zur verlängerten Endstrecke der angrenzenden Segmente unterschreiten. Erfüllt ein Rohpunkt diese Bedingung, wird der Rohpunkt als neuer Kontrollpunkt an das Segment angehängt

$\mathcal{S} = [\mathbf{k}_1, \dots, \mathbf{k}_n] \rightarrow \mathcal{S}' = [\mathbf{k}_1, \dots, \mathbf{k}_n, \mathbf{x}]$, wodurch sich das Segment entsprechend vergrößert.

Segmentkonkatenation: Bei diesem Schritt werden kollineare Segmente zusammengefügt. Dadurch können kurze Unterbrechungen überbrückt werden, die durch Verdeckungen oder Messfehler verursacht werden. Zunächst werden dazu die Segmente nach dem Schwerpunkt $\mathbf{m} = \frac{1}{n} \sum_1^n \mathbf{k}_i$ in Polarkoordinaten winkelsortiert. Anschließend werden zwei benachbarte Segmente \mathcal{S}_j und \mathcal{S}_{j+1} auf Kollinearität überprüft. Dazu muss der Winkel zwischen den beiden Endvektoren $\mathbf{v}_j = \mathbf{k}_n^{(j)} - \mathbf{k}_{n-1}^{(j)}$ bzw. $\mathbf{v}_{j+1} = \mathbf{k}_2^{(j+1)} - \mathbf{k}_1^{(j+1)}$ der Segmente einen Schwellwert τ_α unterschreiten:

$$\arccos \frac{\mathbf{v}_j \cdot \mathbf{v}_{j+1}}{\|\mathbf{v}_j\| \|\mathbf{v}_{j+1}\|} \leq \tau_\alpha \quad (3.1.2)$$

Gleichzeitig darf der Abstand $\|\mathbf{k}_n^{(j)} - \mathbf{k}_1^{(j+1)}\|$ zwischen den beiden Segmentendpunkten einen Schwellwert τ'_{dist} nicht überschreiten. Es gilt jedoch $\tau'_{\text{dist}} > \tau_{\text{dist}}$, da durch die Kollinearitätsbedingung eine zusätzliche Einschränkung gegeben ist. Sind die Bedingungen erfüllt, wird ein neues konkateniertes Segment $\mathcal{S}' = [\mathbf{k}_1^{(j)}, \dots, \mathbf{k}_n^{(j)}, \mathbf{k}_1^{(j+1)}, \dots, \mathbf{k}_n^{(j+1)}]$ erstellt. Die Ausgangssegmente \mathcal{S}_j und \mathcal{S}_{j+1} werden entfernt.

Segmentaggregation: Moderne Laserscanner tasten die Umgebung ebenenweise ab (siehe auch Abschnitt 2.4). Werden die Rohpunkte in den Ebenen gemeinsam zu Segmenten zusammengefasst, führt dies oftmals zu fehlerhaften Segmentverläufen, da eine einzelne verrauschte Ebene das gesamte Segment beeinflusst. Es hat sich deshalb herausgestellt, dass es von Vorteil ist, jede Ebene getrennt zu segmentieren und erst im Anschluss überlappende Segmente zu aggregieren. Zunächst werden dafür alle Segmente \mathcal{S}_j anhand des jeweiligen Segmentschwerpunktes $\mathbf{m}_{\mathcal{S}}$ winkelsortiert. Segmente \mathcal{S}_j deren Schwerpunkte einen geringen euklidischen Abstand zueinander aufweisen, werden zu Segmentmengen $\mathcal{Z}_i = \{\mathcal{S}_1, \dots, \mathcal{S}_n\}$ zusammengefasst. Segmente \mathcal{S} , die nicht aggregiert werden können, werden in eine einelementige Segmentmenge $\mathcal{Z}_i = \{\mathcal{S}\}$ eingefügt. Auch nach der Aggregation können die Segmente noch unabhängig voneinander betrachtet werden.

3.1.1.3 Zusammenfassung

Die Erweiterungen des vorgestellten Segmentierungsalgorithmus haben viele Vorteile im Vergleich zum einfachen „Region-Growing“-Vorgehen. In Abbildung 3.3 ist das Segmentierungsergebnis einer typischen Straßenszene mit sowohl dem typischen „Region-Growing“- (die Regionen werden durch rechteckige Bereiche visualisiert) als auch der erweiterten Segmentierung (Linienzüge) zu sehen. Die Entfernungsschwellwerte τ wurden bei beiden Algorithmusvarianten auf einen identischen Wert von 0.7 m festgelegt.

Die Ergebnisse sind wie folgt: Fahrzeuge werden durch die Segmentverlängerung vollständiger segmentiert. Dieser Vorteil wird insbesondere bei Strukturen erzielt, die nahezu in Messrichtung verlaufen und dadurch hohe Messwertabstände verursachen (vergleiche Lieferwagen bei Position (1)). Durch die Segmentkonkatenation werden Unterbrechungen von kollinearen Segmenten korrigiert. Diese Unterbrechungen bzw. Rohdatenverluste werden neben Teilverdeckungen durch z. B. Bäume oder Ampeln, ebenso durch flache Auftreffwinkel auf stark reflektierende Materialien verursacht (vergleiche abbiegendes Fahrzeug bei Markierung (2)). Dennoch können Objekte, die eine geringe räumliche Nähe zueinander aufweisen, wie Fußgänger voneinander getrennt werden, solange keine Kollinearitäten zwischen den Segmenten auftreten (vergleiche die drei Fußgänger bei Markierung (3)).

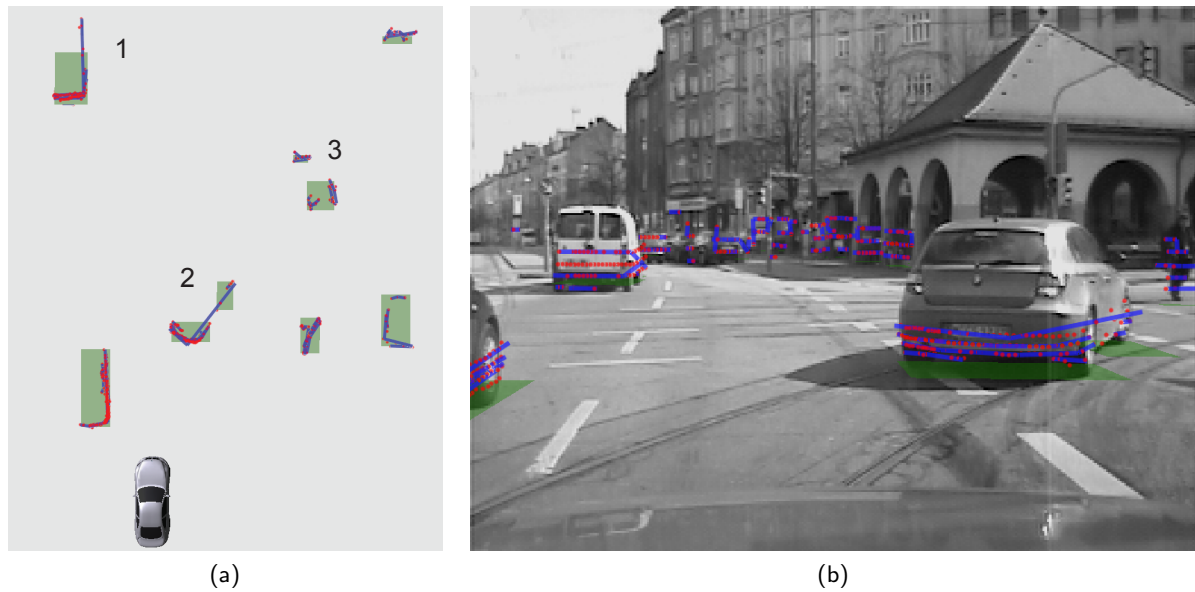


Abbildung 3.3: Exemplarischer Vergleich zwischen dem „Region-Growing Algorithmus“ und der erweiterten inkrementellen Linienextraktion für Laserscannerrohdaten (rote Punkte). Die ermittelten Regionen durch den „Region-Growing“-Algorithmus sind durch grüne rechteckige Bereiche, und die Resultatsegmente der inkrementellen Linienextraktion durch blauen Linienzüge in einer Vogelperspektiven- (a) und einer Videoansicht (b) der Szene markiert.

3.1.2 Einteilung der Segmente

Nach der Laserscannervorverarbeitung stehen Segmente \mathcal{S} zur Verfügung, die aus Rohpunkten aggregiert wurden. Diese Segmente sind aus Linienzügen aufgebaut und entsprechen im Idealfall einem einzelnen Objekt aus dem Fahrumfeld. Anhand der Segmentlänge und der Segmentform kann somit eine Aussage bezüglich der möglichen Objektklasse sowie deren Position und Ausrichtung getroffen werden. Fußgänger sind beispielsweise vorwiegend durch kurze Segmente charakterisiert, wohingegen Fahrzeuge anhand längerer Segmente erkennbar sind. In den nächsten beiden Abschnitten wird auf die Berechnungsroutinen zur Ableitung von Segmentverläufen und -längen eingegangen.

3.1.2.1 Segmentverläufe

Die Segmente, die durch die erweiterte inkrementelle Linienextraktion gewonnen werden, weisen unterschiedliche Verläufe auf. Typische Erscheinungsbilder von Segmenten für Fahrzeuge sind „I“- und „L“-förmige Linienzüge (siehe auch Abbildung 3.4a und 3.4b). Diese Formen entstehen, da die meisten Fahrzeuge näherungsweise quaderförmig sind, und abhängig vom Erfassungswinkel entweder eine oder zwei Seiten dieses Quaders für den Sensor sichtbar sind. Fußgänger nehmen aufgrund der Beinbewegung vielfältige Segmentverläufe an (siehe Abbildung 3.4c). Hintergrundobjekte (Ampeln, Häuserfronten, Randbebauung) können beliebige Segmentformen aufweisen (siehe Abbildung 3.4d). Vergleiche auch Abbildung 3.3 für exemplarische Segmentverläufe.

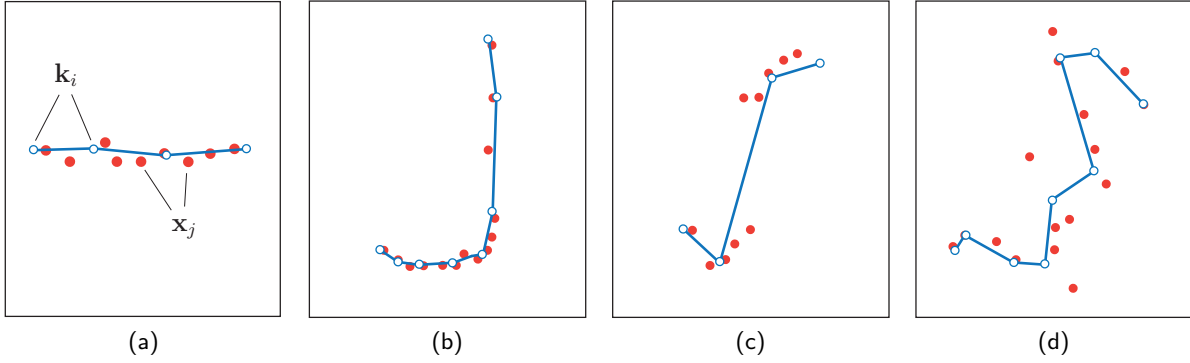


Abbildung 3.4: Laserscanner-Segmente $\mathcal{S} = \{\mathbf{k}_1, \dots, \mathbf{k}_n\}$ aus einer Vogelperspektive von einem Fahrzeugheck („I“-Verlauf, Abbildung (a)), einer Fahrzeugschrägansicht („L“-Verlauf, Abbildung (b)), einem Fußgänger (kein dedizierter Verlauf, Abbildung (c)) und einem Hintergrundobjekt (kein dedizierter Verlauf, Abbildung (d)). Mit \mathbf{k}_i werden die Kontrollpunkte und mit \mathbf{x}_i die zugrundeliegenden Rohpunkte des Segments notiert. Die Darstellungen sind nicht maßstabstreu.

Um die Segmentformen zu bestimmen, werden folgende Rechenvorschriften durchgeführt. Zunächst wird durch alle zum Segment $\mathcal{S} = \{\mathbf{k}_1, \dots, \mathbf{k}_n\}$ gehörenden Rohpunkte $\mathbf{x} \in \mathcal{S}_r$ eine Ausgleichsgerade $y = ax + b$ gelegt. Als Regressionsverfahren wird die gleiche Vorgehensweise wie in Abschnitt 3.1.1.2 gewählt. Unterschreitet der durchschnittliche Abstand $d_i = \frac{|y_i - (a + bx_i)|}{\sqrt{1 + b^2}}$ der Rohpunkte zur Ausgleichsgeraden einen Schwellwert τ_e , dann wird ein „I“-Verlauf angenommen. Wird der Schwellwert überschritten, wird überprüft, ob das Segment anstatt dessen einen „L“-Verlauf aufweist.

Dazu wird zunächst ein „L“-Verlauf hypothetisiert und ein Rohpunkt $\mathbf{x}_c \in \mathcal{S}_r$ gesucht, der dem Eckpunkt des „L“-Verlaufs am Nächsten liegt. Der Punkt \mathbf{x}_c kann ermittelt werden, indem der Punkt gesucht wird, der von Start- und Endpunkt \mathbf{k}_1 bzw. \mathbf{k}_n des Segments den größten Abstand aufweist:

$$\mathbf{x}_c = \arg \max_{\mathbf{x} \in \mathcal{S}_r} (\|\mathbf{x} - \mathbf{k}_1\| + \|\mathbf{x} - \mathbf{k}_n\|) \quad (3.1.3)$$

In einer anderen Variante zur Bestimmung des Eckpunkts, wird der Rohpunkt als Eckpunkt gewählt, der den geringsten Abstand zum Sensor aufweist. Die Rechenvorschrift (3.1.3) hat sich jedoch als robuster gegenüber Ausreißern erwiesen.

Ist der hypothetisierte Eckpunkt nun bekannt, wird die Menge \mathcal{S}_r , der zum Segment gehörigen Rohpunkte anhand des Eckpunkts in zwei Teilmengen $\mathcal{P}_1 = \{\mathbf{x} \in \mathcal{S}_r \mid \mathbf{x}^\theta \leq \mathbf{x}_c^\theta\}$ und $\mathcal{P}_2 = \mathcal{P} \setminus \mathcal{P}_1 \cup \{\mathbf{x}_c\}$ unterteilt. \mathbf{x}^θ bzw. \mathbf{x}_c^θ bezeichnen dabei die Winkel der Rohpunkte in Polarkoordinaten im Sensorkoordinatensystem. Durch die Rohpunkte $\mathbf{x} \in \mathcal{P}_1$ bzw. $\mathbf{x} \in \mathcal{P}_2$ der beiden Teilmengen wird je eine Ausgleichsgerade g_1 bzw. g_2 gelegt. Liegen die Regressionsgeradenfehler unter einem Schwellwert τ_f und liegt der spitze Schnittwinkel der beiden Ausgleichsgeraden in einem Intervall $[90^\circ - \alpha, 90^\circ]$, wird das Segment als ein Segment mit „L“-Verlauf ausgewiesen. Die empirisch ermittelten Schwellwerte sind $\tau_e = 0.28$, $\tau_f = 0.38$ und $\alpha = 15^\circ$.

3.1.2.2 Segmentabmessungen

Die absolute Länge der einzelnen Segmente \mathcal{S}_i wird mit $\delta_{\mathcal{S}}$ bezeichnet und berechnet sich über die Summe aller Teilsegmentlängen.

$$\delta_{\mathcal{S}} = \sum_{i=1}^{n-1} \|\mathbf{k}_i - \mathbf{k}_{i+1}\| \quad (3.1.4)$$

Fehler in der Vorverarbeitung bzw. Messrauschen kann zu alternierenden Teilsegmenten führen. Folglich werden durch die Längenbestimmung nach Gleichung (3.1.4) oftmals zu hohe Segmentlängen berechnet. Um dadurch verursachte Fehlinterpretationen zu vermeiden, kann die Längenbestimmung eines Segments alternativ durch die Länge

$$\tilde{\delta}_{\mathcal{S}} = \|\mathbf{k}_1 - \mathbf{k}_n\| \quad (3.1.5)$$

der Sekantenstrecke approximiert werden. Durch diese Gleichung wird die korrekte Segmentlänge zwar unterschätzt – gleichzeitig ist dieses Vorgehen jedoch robuster gegenüber fehlerbehafteten Segmentverläufen und wird deshalb zur Erzeugung der Hypothesen herangezogen.

Wurde ein Segment als „L“-Verlauf erkannt, kann neben der absoluten Segmentlänge zusätzlich eine Breiten und Tiefenschätzung vorgenommen werden. Dazu werden die Rohpunkte $\mathbf{x} \in \mathcal{S}_r$ orthogonal auf die beiden Regressionsgeraden g_1 bzw. g_2 des „L“-Verlaufs projiziert (siehe auch Abbildung 3.7c). Über die maximale Strecke $\delta_{\mathcal{S}}^1 = \|\arg \max_{\mathbf{x} \in \mathcal{S}_r} ((\mathbf{x} \rightarrow g_1)_x) - \arg \min_{\mathbf{x} \in \mathcal{S}_r} ((\mathbf{x} \rightarrow g_1)_x)\|$ zwischen den beiden Extrempunkten, kann die Ausdehnung $\delta_{\mathcal{S}}^1 \times \delta_{\mathcal{S}}^2$ eines Segments bestimmt werden ($\delta_{\mathcal{S}}^2$ wird analog zu $\delta_{\mathcal{S}}^1$ berechnet, indem die Rohpunkte orthogonal auf g_2 projiziert werden).

3.1.2.3 Verteilung der Segmentlängen

Um zu untersuchen, ob spezifische Fahrzeug- bzw. Fußgängerhypothesen anhand der jeweiligen Segmentlängen $\tilde{\delta}_{\mathcal{S}}$ erzeugt werden können, wurde die Verteilung der Segmentlängen $\tilde{\delta}_{\mathcal{S}}$ für Fahrzeugheckansichten, Fußgänger und Hintergrundobjekte anhand von 20 000 Testdatensätzen untersucht. Die Ergebnisse der Auswertung sind in Abbildung 3.5 dargestellt.

Objekte wie Ampeln und Straßenschilder weisen eine durchschnittliche Breite von $\tilde{\delta}_{\mathcal{S}} = 0.25$ m auf. Der leichte Anstieg bei 2.0 m deutet aber darauf hin, dass es durchaus breitere Hintergrundobjekte gibt. Sie sind an dieser Stelle allerdings unterrepräsentiert, da in den Testdaten vornehmlich Bäume und Ampeln als Hintergrundobjekte markiert wurden. Die Verteilung für Fußgänger weist bei 0.55 m ein deutliches Maximum auf. Fußgängerquerschnitte über 0.7 m können darauf zurückgeführt werden, dass Fußgänger durch das Tragen von Koffern, durch das Auftreten in kleinen Gruppen, aber auch durch die Schrittgröße breitere Laserscannersegmentlängen verursachen.

Bei Fahrzeugheckansichten liegt das Maximum der Verteilung bei ca. 1.5 m. Ein zweites Maximum befindet sich bei 2.7 m und spiegelt die durchschnittliche Breite von Lastkraftwagen wider, die bei der Datenerhebung mitberücksichtigt wurden. Bei Fahrzeugen werden stark von der Durchschnittslänge abweichende Segmentlängen vorwiegend durch Fehlsegmentierungen verursacht.

Es ist deutlich zu erkennen, dass anhand der Segmentlänge $\tilde{\delta}_{\mathcal{S}}$ die Fahrzeugklasse sehr gut von der Fußgängerklasse getrennt werden kann. Eine Separierung von der Hintergrundklasse

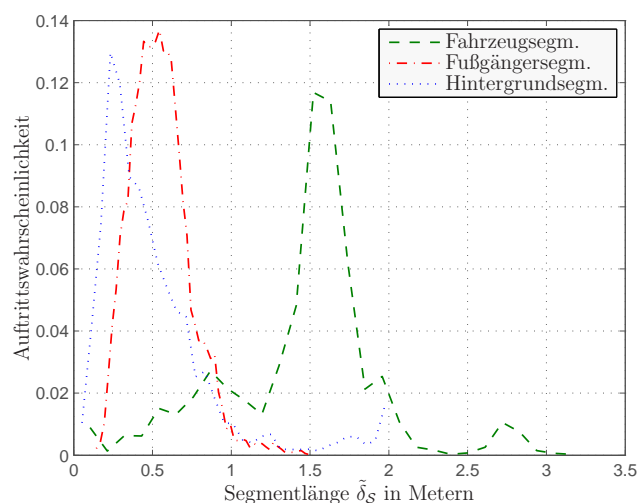


Abbildung 3.5: Verteilung der Laserscanner-Segmentlängen $\tilde{\delta}_S$ von Fahrzeugheckansichten, Fußgängern, und Hintergrundobjekten. Die Fahrzeug und Fußgängerklasse kann anhand von $\tilde{\delta}_S$ sehr gut separiert werden.

ist nur bedingt möglich. Es können durch diese Vorklassifizierung jedoch objektspezifische Hypothesen \mathcal{H}_k und \mathcal{H}_z erzeugt werden.

Im nächsten Abschnitt wird der Entscheidungsprozess zur laserscannergestützten Hypothesenerzeugung zusammengefasst. Zudem wird auf die konkrete Initialisierung der Hypothesen eingegangen.

3.1.3 Hypotheseninitialisierung

Ist die Segmenterzeugung abgeschlossen und wurden die Segmentformen und -längen berechnet, können im Anschluss daran die quaderförmigen Hypothesen \mathcal{H}_k für Fahrzeuge bzw. die zylinderförmigen Hypothesen \mathcal{H}_z für Fußgänger auf Basis der Segmente initialisiert werden. Ein detailliertes Ablaufdiagramm, das den regelbasierten Mechanismus zur Erzeugung der laserscannergestützten Fahrzeug- und Fußgängerhypothesen zeigt, ist in Abbildung 3.6 zu sehen.

3.1.3.1 Quaderhypothesen

Eine Quaderhypothese wird generiert, wenn das Segment einen „I“- oder „L“-Verlauf aufweist und die Segmentausdehnungen $\tilde{\delta}_S$ bzw. $\tilde{\delta}_S^1 \times \tilde{\delta}_S^2$ in einem fahrzeugtypischen Größenintervall liegen (siehe Abbildung 3.6). Sind diese Vorbedingungen erfüllt, werden zunächst die entsprechenden Ausgleichsgeraden ($g = [\mathbf{a}, \mathbf{b}]$ für „I“-Verlauf und $g_1 = [\mathbf{c}, \mathbf{d}]$ bzw. $g_2 = [\mathbf{d}, \mathbf{e}]$ für „L“-Verlauf) durch die Rohpunkte gelegt (siehe auch Abschnitt 3.1.2.1).

- Bei einem „I“-Verlauf wird der Quader bündig an der Regressionsgeraden ausgerichtet. Dabei liegt das Zentrum des Quaders, welches innerhalb des realen Objekts liegen sollte, stets im Sichtschatten des Sensors (siehe Abbildung 3.7b). Der Orientierungsvektor $\mathbf{q}_I = (b_2 - a_2 \quad b_1 - a_1)^T$ des Quaders ist orthogonal zu g , und die Breite b sowie die Länge

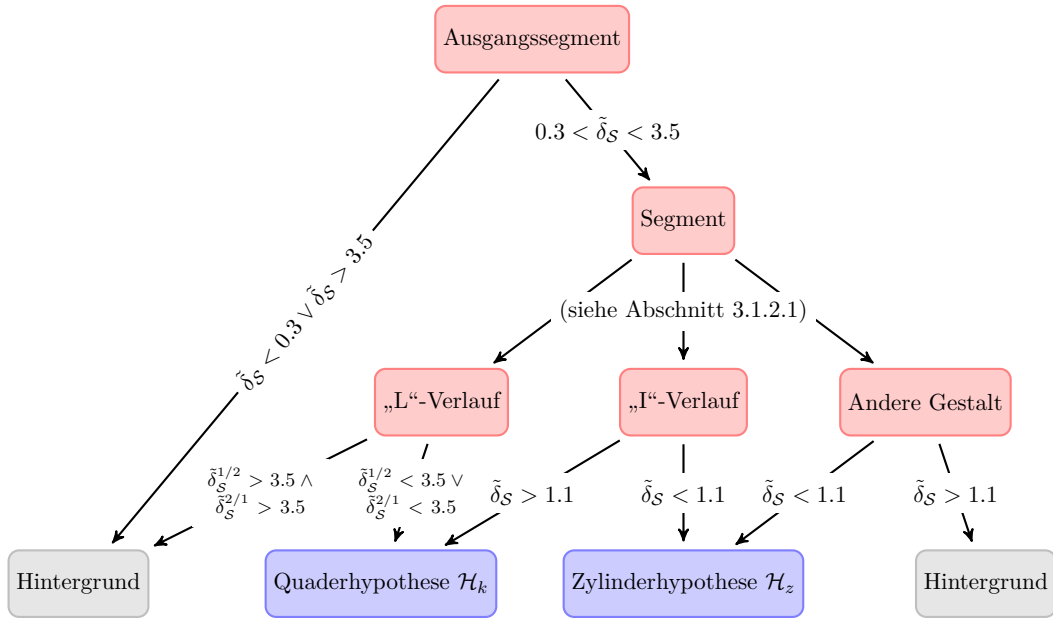


Abbildung 3.6: Laserscanner-Segmente zur Generierung von Hypothesen. Die angegebenen Ausdehnungsschwellwerte sind in Metern und wurden für einen Laserscanner mit einer Winkelauflösung von 0.25° empirisch ermittelt.

l des Quaders werden durch die Länge des Vektors $\mathbf{b} - \mathbf{a}$ vorgegeben. Somit ist $b = l = \|\mathbf{b} - \mathbf{a}\|$. Die Höhe kann allerdings nicht unmittelbar gemessen werden und wird somit auf eine Standardhöhe von $h = 1.5$ m gesetzt. Das Hypothesenzentrum \mathbf{m}_k , das auch zeitgleich den Ursprung für das lokale Objektkoordinatensystem darstellt, berechnet sich über $\mathbf{m}_k = \frac{\mathbf{a} + \mathbf{b}}{2} + \frac{\mathbf{q}_L \|\mathbf{b} - \mathbf{a}\|}{2\|\mathbf{q}_L\|}$.

- Bei einem „L“-Verlauf des Segments wird der Quader mit dem Eckpunkt auf den Schnittpunkt \mathbf{d} der Geraden g_1 und g_2 eingepasst. Die Breite $h = \|\mathbf{d} - \mathbf{c}\|$ bzw. Länge $l = \|\mathbf{e} - \mathbf{d}\|$ ergibt sich aus den Längen der beiden Regressionsgeraden (siehe Abbildung 3.7c). Der Mittelpunkt \mathbf{m}_k berechnet sich über $\mathbf{m}_k = \frac{\mathbf{c} + \mathbf{e}}{2}$. Als Orientierung \mathbf{q}_L wird die um 45° rotierte Richtung $\mathbf{w} = \mathbf{d} + \frac{\mathbf{e}\|\mathbf{c} - \mathbf{c}\|}{2\|\mathbf{e}\|\|\mathbf{c}\|}$ der Winkelhalbierenden des „L“-Segments herangezogen. Diese aus den beiden Seitengeraden g_1 bzw. g_2 gemittelte Orientierungsschätzung liefert bessere Genauigkeiten als eine Ableitung der Orientierung von einer einzelnen Seitengerade, da die Fahrzeugkontur oftmals keine exakte Quaderform ausweist.

Zusammenfassend wird eine quaderförmige Hypothese abhängig vom Segmentverlauf wie folgt erzeugt:

$$\mathcal{H}_k = \begin{cases} \left[\mathbf{m}_k = \frac{\mathbf{a} + \mathbf{b}}{2} + \frac{\mathbf{q}_L \|\mathbf{b} - \mathbf{a}\|}{2\|\mathbf{q}_L\|}, b = \|\mathbf{b} - \mathbf{a}\|, l = \|\mathbf{b} - \mathbf{a}\|, h = 1.5, \mathbf{q}_L \right] & \text{„I“-Verlauf} \\ \left[\mathbf{m}_k = \frac{\mathbf{c} + \mathbf{e}}{2}, b = \|\mathbf{d} - \mathbf{c}\|, l = \|\mathbf{e} - \mathbf{d}\|, h = 1.5, \mathbf{q}_L \right] & \text{„L“-Verlauf} \end{cases}$$

3.1.3.2 Zylinderhypothesen

Zylindrische Hypothesen werden auf Laserscannersegmenten erzeugt, die eine beliebige Gestalt aufweisen und deren Segmentlängen $\tilde{\delta}_S$ im Intervall $[0.3, 1.1]$ liegen. Der Mittelpunkt des neu-

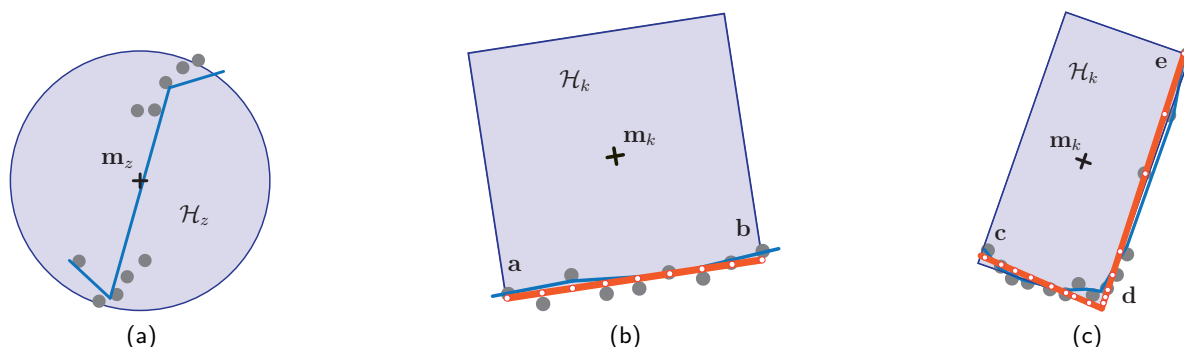


Abbildung 3.7: Hypothesenerzeugung anhand von Laserscanner-Segmenten. Abbildung (a) zeigt die Initialisierung einer zylindrischen Fußgängerhypothese \mathcal{H}_z . In den Rohdaten sind die beiden Beine zu erkennen. Eine quaderförmige Fahrzeughypothese \mathcal{H}_k wird anhand eines „I“-Segments (b) und eines „L“-Segments (c) generiert. Die Sensorverbauposition liegt bei allen Skizzen unterhalb der Objekte. Die Darstellungen zeigen eine Vogelperspektivenansicht und sind nicht maßstabsgetreu.

en Hypothesenobjekts ergibt sich aus dem Schwerpunkt $\mathbf{m}_z = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$, der aus den zum aggregierten Segment gehörenden Rohdaten $\mathbf{x}_i \in \mathcal{Z}_r$ abgeleitet wird (siehe Abbildung 3.7a). Da die Höhe und vertikale Position der Fußgängerhypothese nicht unmittelbar gemessen werden kann, wird von einer initialen Größe $h = 1.8 \text{ m}$ ausgegangen. Zudem wird angenommen, dass der Fußpunkt der Hypothese auf der Grundebene liegt (vergleiche auch Abbildung 2.7 auf Seite 29). Das Hypothesentupel $\mathcal{H}_z = [\mathbf{m}_z, \sigma_S, h]$ ist über die Position, den Durchmesser $\sigma_S = \tilde{\delta}_z$ und die Höhe h eindeutig definiert.

Um das Hypothesenaufkommen zu reduzieren, wird die zylindrische Hypothese nur initialisiert, wenn in der aggregierten Segmentmenge \mathcal{Z} eine adäquate Menge von Segmentebenen enthalten sind. Das bedeutet, dass die Kardinalität $|\mathcal{Z}|$ abhängig von der Distanz der Hypothese und vertikalem Öffnungswinkel/Verbauort des Laserscanners auf ihren Wert überprüft wird. Dadurch wird sichergestellt, dass es sich um ein Objekt handelt, welches eine ausreichende vertikale Ausdehnung aufweist. Allerdings wird der Schwellwert so gering gewählt, dass keine (kleinen) Fußgänger fälschlicherweise ignoriert werden.

3.1.4 Bewertung

Durch einen entfernungsgebenden Sensor mit einer hohen Winkelauflösung, wie einem Laserscanner mit mehreren Scanebenen, können präzise Objekthypothesen erzeugt werden. Neben den Objektklassen kann zudem die Position von Fahrzeugen und Fußgängern initialisiert werden. Bei der Erzeugung von Fahrzeughypothesen kann darüber hinaus eine Orientierungsschätzung vorgenommen werden. Einen klaren Vorteil bietet das Abtasten der Umgebung in mehreren Ebenen. Beim Erfassen eines Fahrzeugs beispielsweise, werden durch einzelne Scanebenen oftmals nicht ausschließlich die Fahrzeugkontur, sondern zusätzlich der Unterboden oder ein Teil der Räder abgetastet. Durch das Messen in mehreren Ebenen ist somit eine gewisse Redundanz vorhanden.

Eine ungenaue Segmentierung bzw. Hypothesenerzeugung tritt auf, wenn Objekte einen zu geringen Abstand zueinander aufweisen. Durch die vorgestellte inkrementelle Liniensextraktion

wird diese Problematik erfolgreich adressiert. Unterschreitet der Objektabstand jedoch die Winkelauflösung des Sensors, ist keine Trennung mehr möglich.

In [KBD05] und [Lin08] werden iterative Verfahren beschrieben, die auf Basis der „L“- bzw. „I“-Segmentgestalt eine genauere Positions- und Orientierungsschätzung von Fahrzeugen erlauben. Diese Ansätze sind jedoch um ein Vielfaches rechenaufwendiger und sollten demnach ausschließlich zum Tracken von bereits klassifizierten Objekten angewendet werden, da die Anzahl der klassifizierten Objekte erheblich unter der Anzahl der Hypothesen liegt und insbesondere bei klassifizierten Objekten eine genauere Zustandsschätzung von Interesse ist. Zur Vorklassifizierung und initialen Hypothesenerzeugung sind die in diesem Abschnitt geschilderten Vorgehensweisen ausreichend robust und erheblich ressourcensparender.

Im nächsten Abschnitt wird auf die Hypothesenerzeugung mittels bildgebender Sensoren eingegangen.

3.2 Bildgebende Sensorik zur Hypothesenerzeugung

Sowohl das Video- als auch das Wärmebild kann als Informationsquelle zur Erzeugung von Hypothesen genutzt werden. Geeignete Merkmale müssen zum einen schnell zu berechnen und zum anderen klassendurchgängig verfügbar sein. Vertikale Kantenpaare sind ein starkes Indiz für Fahrzeuge und Fußgänger, da diese meist an den Objektkonturen vorhanden sind. Im folgenden Abschnitt werden „U“-förmige Kantenbegrenzungen im Kopfbereich genutzt, um Fußgängerhypothesen im Wärmebild zu erzeugen. Vertikale Kanten werden zur Initialisierung von Fahrzeughypothesen im Videobild herangezogen.

3.2.1 „U“-Form für Fußgängerhypothesen im Wärmebild

Durch das Messprinzip einer Wärmebildkamera können Umgebungsbereiche voneinander differenziert werden, die eine unterschiedliche Temperatur aufweisen. Insbesondere Objekte, wie Fahrzeuge und Fußgänger, die sich in ihrer Eigentemperatur signifikant von der Umgebungstemperatur abheben, werden zuverlässig detektiert. Bei genauerer Betrachtung eines Fahrzeughecks sind die Rücklichter, die Reifen und die Abgasanlage markante warme Bereiche. Die Temperaturverteilung von Fußgängern ist abhängig von deren Kleidung. In der Regel sind vornehmlich die schwachisolierten Körperpartien, wie Kopf und Hände besonders markant. Insbesondere der Gesichtsbereich ist unabhängig von den Witterungsverhältnissen meist unverdeckt und somit ein kennzeichnendes Merkmal für einen Fußgänger im Wärmebild (siehe auch Fußgängerabbildungen A.2 auf Seite 181).

3.2.1.1 Umsetzung

Um diese Kopfbereiche im Wärmebild zu detektieren, wurde folgendes Verfahren implementiert. Zunächst wird, ausgehend vom Wärmebild $f(x, y)$, das zugehörige Gradientenbild f_x mit dem zentrierten Filterkern $[1, 0, -1]$ zur Gewinnung von vertikalen Kanten berechnet. Eine zweifach durchgeführte Schwellwertoperation, mit den Schwellwerten τ_s bzw. $-\tau_s$ auf dem Gradientenbild, diskretisiert die Gradientenmaxima und erzeugt zwei diskrete Bilder f_x^- und f_x^+ . Die Maximabereiche dieser diskretisierten Kantenbilder können als Zusammenhangskomponenten interpretiert werden (siehe Abbildung 3.8a). Mittels einer Tiefensuche auf diesen

Bereichen werden die Regionen linearisiert und als Kantenstruktur \mathcal{K} repräsentiert (siehe Abbildung 3.8b). Die Tiefensuche ermittelt dabei die vertikale und horizontale Ausdehnung δ_h und δ_w der Bereiche in Bildpunkten sowie den Referenzpunkt $\mathbf{v} = (v_x, v_y)^T$ der Kante in Bildkoordinaten. Eine Kantenstruktur $\kappa \in \mathcal{K}$ ist somit durch das Tupel $[\mathbf{v}, \delta_w, \delta_h, \sigma]$ eindeutig definiert. Der Tuppeltrag $\sigma \in \{-1, 1\}$ bezeichnet das Gradientenvorzeichen der Kante (dunkel-hell bzw. hell-dunkel Kante) und kodiert somit deren Extraktionsort (f_x^- bzw. f_x^+). Zu kurze Kanten ($\delta_h < \tau_l$) sowie Kanten, deren Hauptverlaufsrichtung nicht vorwiegend vertikal ist ($\frac{\delta_w}{\delta_h} > 1$), werden aus der Menge \mathcal{K} entfernt. In einem nächsten Schritt werden nun Kantenpaare $[k_i, k_j]$ mit $k_i, k_j \in \mathcal{K}$ gesucht, für die folgende Bedingungen gelten:

- Die Referenzpunkte der beiden Kanten müssen im Bildbereich auf einer ähnlichen Höhe liegen, damit eine „U“-Form des Kopfbereichs gewährleistet werden kann. Es gilt somit $|((v_y)_i - (v_y)_j)| < \tau_h$.
- Da die warme Kopfpattie auf helle Intensitätswerte abgebildet wird, muss die linke Kante einen positiven und die rechte Kante einen negativen Gradientenwert aufweisen. Es muss somit $\sigma_i = -1 \wedge \sigma_j = 1$ gelten.
- Der horizontale Abstand $d_h = ((v_x)_j - (v_x)_i)$ der Kanten ist durch das Intervall $[\tau_m, \tau_n]$ begrenzt. Diese Minimal- bzw. Maximalabstände definieren sich über die beiden Schwellwerte τ_m und τ_n und sind abhängig vom Öffnungswinkel der Kamera, von der minimalen und maximalen Detektionsentfernung und vom durchschnittlichen Kopfdurchmesser eines Menschen.

Einzelkanten ohne Partner werden verworfen. Die Laufzeit der Suche ist quadratisch, da jede Kante mit jeder anderen Kante verglichen werden muss.

Wurden nun die linken und rechten Kantenpartner ermittelt, wird abschließend sichergestellt, dass die Kopfhypothese durch eine horizontale Kante abgeschlossen wird. Dazu wird zusätzlich zum Gradientenbild f_x das Gradientenbild f_y mittels des zentrierten Filterkerns $[1, 0, -1]^T$ berechnet, um gleichermaßen horizontale Kantenverläufe zu extrahieren. Es wird durch die Schwellwertoperation das diskrete Kantenbild f_y^- ermittelt, da hell-dunkel-Kanten von Interesse sind (siehe horizontale Kanten in Abbildung 3.8b). Abschließend wird für jedes Kantenpaar $k_i, k_j \in \mathcal{K}$ überprüft, ob einer der beiden Referenzpunkte auf einer horizontalen Kante liegt. Es muss also $f_y^-(\mathbf{v}_i) = 1 \vee f_y^-(\mathbf{v}_j) = 1$ gelten. Kantenpaare die diese Bedingung nicht erfüllen, werden ausgesondert. Die restlichen Kantenpaare fungieren als Kopfkandidaten und somit als Fußgängerhypothesen (siehe Abbildung 3.8c). Die optimale Parametrisierung der notwendigen Schwellwerte τ_s, τ_l, τ_m und τ_n ist kameraabhängig. Für das verwendete Testsystem (siehe auch Abschnitt 2.4.1) wurden die Werte empirisch ermittelt und auf $\tau_s = 50$, $\tau_l = 5$, $\tau_m = 30$ und $\tau_n = 60$ gesetzt.

Die Kantenberechnung wird aus Performanzgründen auf den Bereich oberhalb des Bildhorizonts begrenzt, da lediglich dort Kopfhypothesen vermutet werden. Dieser Bereich ist abhängig von der Kameraausrichtung. Zusätzlich wird ein kleiner Sicherheitsbereich hinzugenommen, um Nickbewegungen des Eigenfahrzeugs, unterschiedliche Fußgängergrößen und Straßenunebenheiten auszugleichen.

Die erzeugten Hypothesen werden über eine Größenannahme der Fußgänger von 1.8 m nach Weltkoordinaten transformiert. Durch diese Unsicherheit kann die Entfernung lediglich sehr ungenau geschätzt werden. Die Ergebnishypothese eines Kantenpaares (k_1, k_2) ergibt sich schließlich zu

$$\mathcal{H}_z = \left[\vartheta \left(\frac{\mathbf{v}_1 + \mathbf{v}_2}{2} \right), 0.5, h \right],$$

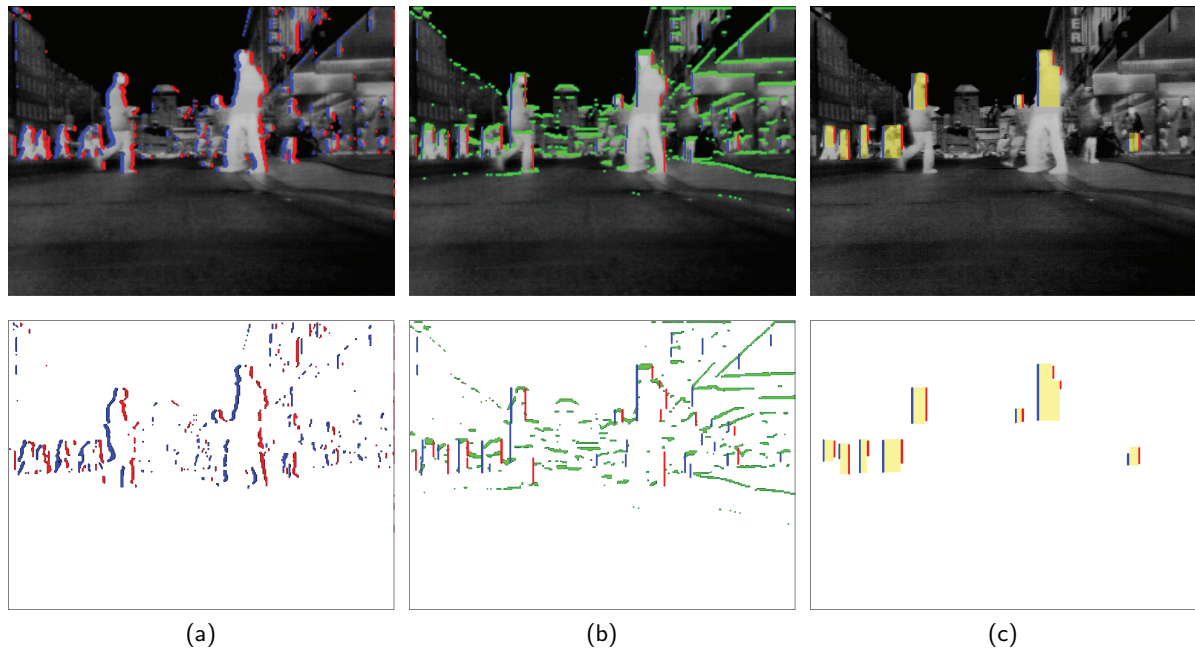


Abbildung 3.8: Segmentierungsprozess von „U“-Formen zur Fußgängerhypothesierung im Wärmebild. Zunächst werden vertikale Bildkanten extrahiert (a). Diese werden linearisiert (b). Über horizontale Kanten (b) und Partnerbildung können „U“-förmige Bereiche als Kopfkandidaten extrahiert werden (c).

wobei $\vartheta(\cdot)$ die Rückprojektionsfunktion eines Bildpunktes nach Weltkoordinaten unter der Flachen-Welt-Annahme ist.

3.2.1.2 Bewertung

Im Nahbereich wird vorwiegend der Kopfbereich durch die Kantenpaare detektiert, da durch den Schulterbereich die vertikalen Kanten oft unterbrochen werden (siehe auch rechten Fußgänger in Abbildung 3.8c). Im Fernbereich verschmilzt die Fußgängerkontur hingegen oftmals zu einer Gesamtsilhouette, und die vertikalen Kantenpaare reichen vom Kopf- bis zum Fußbereich (siehe Fußgänger im Fernbereich in Abbildung 3.8c). Bei letzterem Fall ist somit der Fußpunkt des Fußgängers bekannt, wodurch über die „Flache-Welt-Annahme“ auf Entfernungswerte geschlossen werden kann. Wird im anderen Fall der Referenzpunkt nur durch den Kopf vorgegeben, können lediglich sehr ungenaue Distanzinformationen abgeleitet werden, da die Größe des Fußgängers in der Regel nicht bekannt ist.

Bei dem vorgestellten Hypothesenerzeugungsverfahren treten Fehler erster und zweiter Art auf. Zum einen werden Fußgänger nicht gefunden, da sie Bestandteil einer Gruppe sind (siehe Abbildungen 3.8c und 3.9a), die Kontur mit anderen ähnlichwarmen Hintergrundobjekten verschmilzt (siehe rechte Fußgänger in Abbildung 3.8c), oder sich die Umgebungstemperatur der Körperwärme annähert (siehe Abbildung 3.9c). Zum anderen werden Hintergrundobjekte fälschlicherweise der Fußgängerklasse zugeordnet, da sie eine ähnliche „U“-Gestalt im Wärmebild aufweisen. Dazu zählen beispielsweise Verkehrsschilder, Fahrzeugbestandteile und Gebäude (siehe Abbildungen 3.9b und 3.9c). Erst durch die zusätzliche Erzeugung von Hypothesen

durch entfernungsgebende Sensorik (siehe Abschnitt 3.1), sowie die Verifizierung der Hypothesen (siehe Kapitel 5), können beide Fehlerraten gesenkt werden.

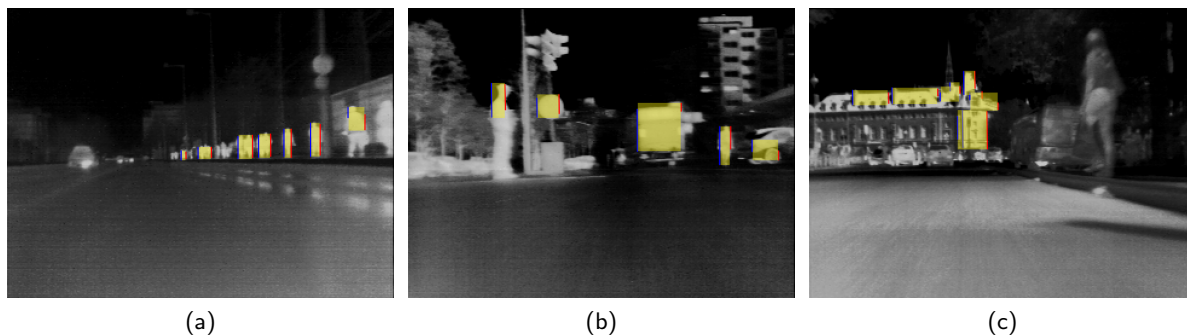


Abbildung 3.9: Beispielaufnahmen für die Fußgängersegmentierung anhand von Wärmebildkanten bei einer Außentemperatur von 0 °C (a), 12 °C (b) und 25 °C (c).

Das Verfahren ist auf eine Fußgängerhypothesenerzeugung im Videobild nicht unmittelbar übertragbar. Zum einen ist der Rechenzeitbedarf höher, da Videobilder meist stärker texturiert sind als Wärmebilder und es dadurch ein größeres Kantenaufkommen gibt. Zum anderen ist die Intensitätssignatur von Fußgängern im Videobild, aufgrund von vielfältiger Bekleidung, sehr heterogen und wesentlich schwächer ausgeprägt als im Wärmebild. Fahrzeughypothesen können jedoch im Videobild durch vertikale Kanten erzeugt werden.

3.2.2 Vertikale Kanten für Fahrzeughypothesen im Videobild

Charakteristische Kanten treten bei Fahrzeugheckansichten im Videobild zwischen den dunklen Reifen des Fahrzeugs und der hellen Straße auf. An diesen Stellen finden sich im Wärmebild gleichermaßen vertikale Kanten, da sich die durch die Fahrt aufgewärmten Reifen von der meist kühleren Straße abheben. Die Extraktion dieser Kantenpaare kann analog zu den Schritten aus Abschnitt 3.2.1 erfolgen. Es müssen lediglich die Schwellwerte und Abstandsparameter entsprechend adaptiert werden. Zum einen sind die Abstandsschwellwerte nun entsprechend größer zu wählen. Zum anderen müssen die Gradientenrichtungen der Kanten im Videobild umgedreht werden, da dort ein dunkles Objekt (schwarze Reifen) vor einem hellen Objekt (Straße) detektiert wird. Zudem liegt der Referenzpunkt \mathbf{p} der Kanten nun am Fußpunkt der Kante. Horizontale Kanten werden nicht zusätzlich extrahiert. Die Höhe des Fahrzeugs wird nicht direkt gemessen, sondern über das durchschnittliche Breiten-Höhenverhältnis von Fahrzeughecks bestimmt.

Eine praktische Durchführung der Hypothesengenerierung von Fahrzeughecks war aufgrund des hohen Kantenaufkommens und der daraus resultierenden Verletzung der Echtzeitfähigkeit nicht unmittelbar möglich. Das Verfahren konnte jedoch in Verbindung mit einer initialen Segmentierung durch die Laserscannerdaten zum Einsatz kommen. In Abbildung 3.11 auf Seite 55 sind Ergebnisse der kantenbasierten Fahrzeughypothesenerzeugung im Videobild mit einer lasergestützten Aufmerksamkeitssteuerung zu sehen.

3.3 Verbunderzeugung und Bewertung

In den vorherigen Abschnitten wurden Verfahren zur Hypothesenerzeugung vorgestellt, die ausschließlich auf den Messwerten eines einzelnen Sensors operieren. In diesem Abschnitt werden die Hypothesen auf Basis mehrerer Sensoren erzeugt. Zudem werden Leistungsunterschiede experimentell ermittelt.

Eine einfache und effektive Möglichkeit, Hypothesen sensorübergreifend zu erzeugen, ist die Zusammenführung von monosensoriell generierten Hypothesen. Der interessanten Fragestellung, welche Auswirkungen ein dadurch verursachtes höheres Hypothesenaufkommen auf die Hypothesenabdeckung und die Erkennungsraten hat, wird in Abschnitt 3.3.1 nachgegangen. Einen Schritt weiter geht eine Kombination von Hypothesen, die in Abschnitt 3.3.2 vorgestellt wird. Dazu muss zunächst ermittelt werden, durch welche Hypothesen das gleiche Objekt im Fahrumfeld repräsentiert wird. Anschließend kann der Objektzustand durch eine sensorübergreifende Hypothesenmenge beschrieben werden.

Für die folgenden Verfahren ist es notwendig, dass die Orientierungen und Verbauorte der verwendeten Sensoren zueinander bekannt sind. Diese Werte können durch eine Sensorkalibrierung ermittelt werden (siehe Anhang C für Verfahren zur multispektralen Kamerakalibrierung bzw. [Lin08] für weitere Methodiken zur Kalibrierung).

3.3.1 Zusammenführung von Fußgängerhypothesen

Werden Hypothesen nicht ausschließlich von einem einzelnen Sensor, sondern von mehreren Sensoren $\{S_1, \dots, S_k\}$ zeitgleich erzeugt, entsteht ein Hypothesenkollektiv

$$\mathcal{K} = \{\mathcal{H}_1^{S_1}, \dots, \mathcal{H}_m^{S_1}, \dots, \mathcal{H}_1^{S_k}, \dots, \mathcal{H}_n^{S_k}\}.$$

Durch eine einheitliche räumliche Repräsentation der Hypothesen \mathcal{H} (siehe auch Abschnitt 3.2.1) abstrahiert \mathcal{K} von den sensorspezifischen Eigenschaften. Die Güte der räumlichen Darstellung wird durch das Sensorprinzip bedingt. Für bildbasierte Hypothesen beispielsweise ist eine Entfernungsbestimmung oftmals ungenau.

3.3.1.1 Untersuchung

Um abschätzen zu können, welche Auswirkungen eine multisensorielle Hypothesenerzeugung auf die Performanz der Fahrumfelderfassung hat, wird im Folgenden die Leistungsfähigkeit einer kombinierten Erzeugung anhand eines Beispielszenarios überprüft. Dazu werden durch einen Laserscanner, und zeitgleich durch das Wärmebild, Fußgängerhypothesen nach den monosensoriellen Erzeugungsverfahren aus Abschnitt 3.1 bzw. 3.2.1 generiert. Da lediglich die Leistungsfähigkeit der Hypothesenerzeugungsroutinen und deren Auswirkung von Interesse sind, wird zu Evaluierungszwecken in jedem Szenenzeitschritt eine Neuinitialisierung der Zustandswerte vorgenommen und keinerlei Historie betrachtet.

In Abbildung 3.10 sind die Trajektorienverläufe von Fußgängern und Radfahrer aus einer typischen Stadtszene zu sehen. Das Diagramm zeigt die Referenztrajektorienverläufe der Verkehrsteilnehmer, sowie algorithmisch generierte Hypotheseninitialisierungen, die anhand der Verfahren aus Abschnitt 3.1 bzw. 3.2.1 erzeugt wurden. Dabei wurden alle zeitdiskreten Ergebnisse aufintegriert und in gemeinsame Weltkoordinaten eingetragen. Die Referenztrajektorien

wurden semiautomatisch erstellt (siehe auch Abschnitt 6.1.1), wobei Objektreferenzen nur im gemeinsamen Sichtbereich von Wärmebildkamera und Laserscanner attribuiert wurden. Durch die unterschiedlichen Markierungen der Hypothesen werden die jeweiligen erzeugenden Sensoren gekennzeichnet, wobei Hypothesen, die keinen Verkehrsteilnehmer darstellen gesondert ausgewiesen sind. Fehlerkennungen werden somit durch f_p , und korrekte Detektionen durch r_p markiert.

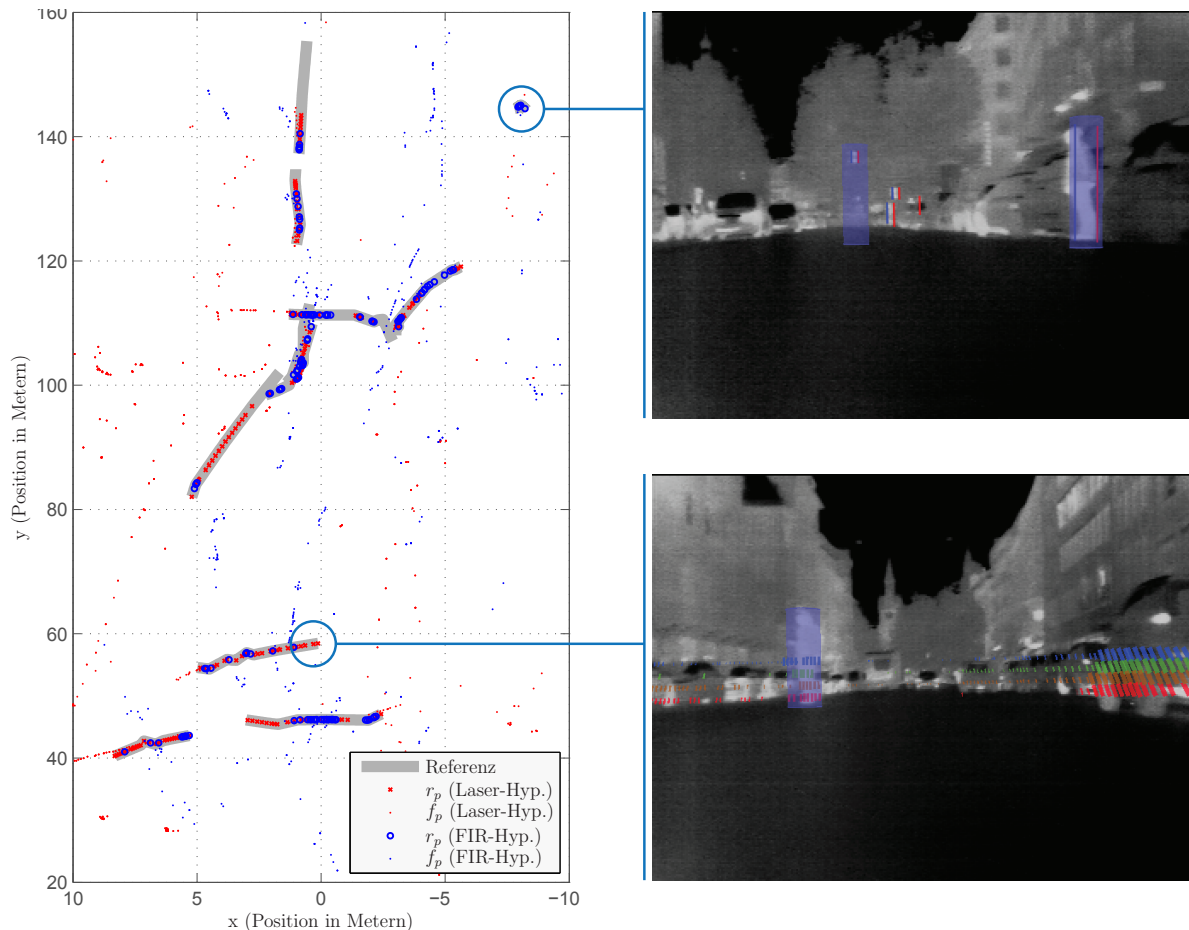


Abbildung 3.10: Verbunderzeugung von Fußgängerhypothesen mittels Laserscanner und Wärmebildkamera im städtischen Umfeld. In der Vogelperspektivendarstellung auf der linken Seite sind Referenzpositionen und Hypothesenerzeugungskoordinaten für Fußgänger eingetragen. Die Wärmebilder auf der rechten Seite greifen einzelne Szenenzeitpunkte heraus.

3.3.1.2 Auswertung

Bei einer genauen Betrachtung des Diagramms in Abbildung 3.10 werden verschiedene Aspekte deutlich: Nicht alle Referenztrajektorien werden gleichermaßen durch Hypothesenmessungen abgedeckt. Der Fußgänger im oberen Bereich des Diagramms wird ausschließlich anhand von Wärmebildhypothesen wahrgenommen. Durch seine räumliche Nähe zu den am Straßenrand

parkenden Fahrzeugen kann er durch die Laserscannersegmentierung nicht separiert werden. Im Gegensatz dazu wird der Fußgänger im unteren Bereich des Diagramms vorwiegend durch Laserscannerhypothesen detektiert, da im Wärmebild die Kontrastverhältnisse zum Hintergrund zu schwach sind und damit keine signifikanten Kantenpaare detektiert werden können. Durch diese kombinierte Aufmerksamkeitssteuerung, kann somit der Wahrnehmungsraum sinnvoll ergänzt werden. Für eine zahlengestützte Auswertung werden zunächst einige Evaluierungskriterien eingeführt:

- Das *Hypothesenaufkommen* bezeichnet die durchschnittlichen Hypotheseninitialisierungen pro Sekunde. Für diese Auswertung wird keinerlei Tracking berücksichtigt. Das bedeutet, dass Hypothesen nach einem Rechenzyklus sofort wieder entfernt werden.
- Mit der *Rechenzeit* wird die durchschnittliche Berechnungszeit auf einem aktuellen Fahrzeugversuchsrechner¹ angegeben.
- Die *Initialisierungszeit für die Hypothesenerzeugung* bezeichnet die durchschnittliche Zeit in Millisekunden, die zwischen dem Erzeugen eines Referenzobjekts und dem erstmaligen Aufsetzen einer, das Referenzobjekt abdeckenden, Hypothese vergeht. Die *Initialisierungszeit für die Hypothesenerkennung* ist die durchschnittliche Zeit, die benötigt wird um das Objekt mittels eines wärmebildbasierten Klassifikators² zu erkennen.
- Die *Abdeckungsrate* für Hypothesen bzw. für Erkennung bezeichnet die durchschnittliche Anzahl der Hypothesen/Objektinitialisierungen pro Referenzobjekt pro Sekunde.

Die Leistungsdaten der monosensoriellen und kollektiven Hypothesenerzeugung für Fußgänger mittels Laserscanner und Wärmebildkamera sind in Tabelle 3.1 aufgeführt. Die angegebenen Zahlenresultate sind Durchschnittswerte, die durch die Evaluierung der obigen Stadtszene mit 12 Fußgängern (Dauer ca. eine Minute) ermittelt wurden. Der Laserscanner wurde mit einer Frequenz von 10 Hz betrieben, und die Wärmebildkamera mit einer Frequenz von 30 Hz. Sensorlatenzzeiten wurden bei der nachfolgenden Bewertung nicht berücksichtigt.

Tabelle 3.1: Bewertung der Hypothesenerzeugung für Fußgänger.

Bezeichnung	Sensorkonfiguration		
	Laserscanner	Wärmebild	Kombiniert
Hypothesenaufkommen	24.9 s^{-1}	39.3 s^{-1}	62.8 s^{-1}
Rechenzeit	7.7 ms	4.8 ms	12.1 ms
Initialisierungszeit für Hypothesenerzeugung	670 ms	584 ms	229 ms
Initialisierungszeit für die Objekterkennung	1.71 s	1.42 s	1.14 s
Abdeckungsrate Hypothesen	4.96 s^{-1}	3.73 s^{-1}	8.63 s^{-1}
Abdeckungsrate Erkennung	0.79 s^{-1}	0.77 s^{-1}	1.45 s^{-1}
Fehlerkennungsrate der Erkennung	0.17 s^{-1}	0.79 s^{-1}	0.79 s^{-1}

¹ Als Fahrzeugversuchsrechner kam ein Pentium 4, 1.8 GHz, 1Gb Speicher zum Einsatz.

² Als Klassifikator wurde eine SVM verwendet, die auf einem Merkmalsvektor aus Gradientenhistogrammen operiert (siehe auch Abschnitt 4.2).

Bei der kombinierten Hypothesenerzeugung akkumuliert sich das Hypothesenaufkommen auf ca. 60 Hypotheseninitialisierungen in der Sekunde. Analog dazu verhält sich der Anstieg der benötigten Rechenzeit für die Vorverarbeitungsberechnungen. Durch das erhöhte Hypothesenaufkommen sinkt allerdings gleichzeitig die durchschnittliche Zeit, die für die Initialisierung einer Hypothese benötigt wird von 670 ms, bzw. 584 ms auf 229 ms. Analog dazu verdoppelt sich die durchschnittliche Abdeckungsrate der Referenzobjekte. Zudem steigt die Erkennungsrate von 0.77 Hypothesen pro Sekunde bei gleichbleibender Fehlerkennungsrate auf 1.45 Hypothesen pro Sekunde an. Die Fehlerkennungsrate wird nicht tangiert. Die Auswertung verdeutlicht, dass eine kollektive Hypothesenerzeugung wesentliche Vorteile birgt. Das erhöhte Leistungspotential kommt insbesondere bei einem partiellen oder totalen Sensorausfall (siehe beispielsweise Abbildung 3.9c auf Seite 49) zum Tragen, da die Hypothesenerzeugung in diesem Fall durch den funktionierenden Sensor übernommen werden kann. Allerdings ist für einen anschließenden Trackingeinsatz die Distanzschätzung der bildgebenden Sensorik prinzipbedingt begrenzt.

3.3.2 Hypothesenkombination und Bewertung der Aufmerksamkeitssteuerung

Bei dem im vorherigen Abschnitt eingeführten Hypothesenkollektiv werden die Hypothesen unabhängig voneinander erstellt. Das hat zum Nachteil, dass ein Objekt möglicherweise durch mehrere monosensorielle Hypothesen beschrieben wird, wodurch zum einen mehr Rechenzeit beansprucht wird, und zum anderen ein nachgeschaltetes Trackingverfahren etwaige Mehrdeutigkeiten auflösen muss. Des Weiteren werden Informationsquellen, die durch eine verschränkte Hypothesenerzeugung genutzt werden können, nicht mit einbezogen.

In diesem Abschnitt wird das Potential unterschiedlicher Kombinationsmethoden von entfernungs- und bildbasierten Hypothesen untersucht. Dabei steuert der entfernungsgebende Sensor die Aufmerksamkeit, in welchem Bildbereich die Hypothese zu finden ist. Zunächst wird auf die verschiedenen Vorgehensweisen der Kombination eingegangen. Diese werden mit Verfahren ohne Aufmerksamkeitssteuerung verglichen. Am Ende des Abschnitts werden die Methodiken anhand einer Fahrzeugerkennung bewertet und gegenübergestellt. In Tabelle 3.2 findet sich vorab eine Übersicht der untersuchten Verfahren, die im Anschluss detailliert beschrieben werden.

Tabelle 3.2: Übersicht über die unterschiedlichen Hypothesenerzeugungsverfahren im Bild. Die durchschnittliche Bildhypothesenanzahl bezieht sich auf ein Bild und eine Hypothese.

Verfahren	Abkürzung	Durchschnittliches Bildhypothesenauf- kommen	t_{Zyk} (s)
Erschöpfende Suche	HypoScan	ca. 28000	77 s
Geometriegestützte Suche	HypoGeoScan	ca. 5000	15 s
Geometrie- und Lagegestützte Suche (Schrittweite 1)	HypoGeoPosScan ₁	ca. 500	1.5 s
Geometrie- und Lagegestützte Suche (Schrittweite 3)	HypoGeoPosScan ₃	ca. 200	0.6 s

Tabelle 3.2: Übersicht über die unterschiedlichen Hypothesenerzeugungsverfahren im Bild. Die durchschnittliche Bildhypothesenanzahl bezieht sich auf ein Bild und eine Hypothese.

Verfahren	Abkürzung	Durchschnittliches Bildhypothesenauf- kommen	t_{Zyk} (s)
Direkte Projektion	GeoLage	1	4 ms
Kantenbasierte Suche	GeoLageKanten	1	10 ms

3.3.2.1 Bildhypothesenerzeugung ohne sensorgestütztes Vorwissen

Bei einer Hypothesenerzeugung im Bild, die gänzlich ohne Vorwissen durchgeführt wird, muss der gesamte Suchbereich durch Bildhypothesen abgedeckt werden. Lediglich a priori Annahmen zur Objektgröße und zum Objektaufenthaltort können zur Reduzierung des Suchraums herangezogen werden. Zudem kann eine feste Schrittweite verwendet werden, um das Hypothesenaufkommen etwas zu begrenzen. Grundsätzlich ist das Hypothesenaufkommen und somit die Rechenzeit durch diese Vorgehensweise jedoch sehr hoch. Um eine Auswertung in angemessener Zeit durchführen zu können, wurden die möglichen Größenwerte auf zehn gestaffelte Bildhypothesenausdehnungen begrenzt. Durch die folgenden Methodiken kann der Suchbereich weiter reduziert werden. Diese Vorgehensweise wird mit *HypoScan* abgekürzt.

3.3.2.2 Geometrie- und Lagegestützte Bildhypothesenerzeugung

Anhand der Geometrie- und Entfernungsinformationen der Fahrzeughypothesen, die mittels eines winkelauflösenden entfernungsgebenden Sensors gewonnen werden, ist die Pixelausdehnung der Hypothese im Bild bekannt. Somit kann im Videobild ein dichtes Netz an Bildhypothesen erzeugt werden, deren Ausdehnung durch die entfernungsbasierte Hypothese determiniert ist. Diese Vorgehensweise wird mit *HypoGeoScan* abgekürzt. Weitere Suchbereichseinschränkungen sind möglich, wenn zudem die Lage der Hypothese herangezogen wird.

Sind die maximal möglichen Nickbewegungen des Eigenfahrzeugs bekannt, sowie die zu erwartenden Kalibrierungsungenauigkeiten, kann für jede Entfernungshypothese H_i^L ein entsprechendes rechteckiges Suchfenster \mathcal{F}_i im Bild generiert werden. Bildhypothesen H_i^V werden anschließend nur in den jeweiligen Suchfenstern erzeugt. Dazu wird ein dichtes Netz mit einer vertikalen und horizontalen Pixelschrittweite δ generiert. Für ein Suchfenster mit einer Ausdehnung von $b \times h$ Bildpunkten ergeben sich bei einer Bildhypothesengröße von $b^V \times h^L$ Bildpunkten $(b - b^V + 1)(h - h^V + 1)\delta^{-2}$ mögliche Bildhypothesen. Diese Vorgehensweise wird mit *HypoGeoPosScan $_{\delta}$* abgekürzt.

Ein besonderer Fall liegt vor, wenn die Größe des Suchfensters der Größe der Hypothesenprojektion ($b = b^V \wedge h = h^V$) entspricht. Das Ergebnis ist ein einzelner Bildausschnitt, der das zugehörige Objektabbild enthält (siehe auch Abbildung 4.1 auf Seite 65). Durch ungenaue Kalibrierung oder nicht messbare Nickbewegungen des Versuchsfahrzeugs, kann jedoch

eine durchgängige Überlappung nicht gewährleistet werden (vergleiche auch Abbildung 3.11). Diese Vorgehensweise wird mit *GeoLage* abgekürzt.

3.3.2.3 Kantengestützte Bildhypothesenerzeugung

Eine zusätzliche Verfeinerung der Bildhypothesenposition kann objektspezifisch durchgeführt werden. Am Beispiel von Fahrzeugheckansichten wird dieses Vorgehen aufgezeigt. Es ist vor Allem dann besonders anspruchsvoll zugehörige Kantenpaare in einem Videobild zu finden, wenn die Szenerie stark texturiert ist. Dies führt zu einem hohen Kantenaufkommen, wodurch es vermehrt zu Fehlzuordnungen kommt. Kann der Suchbereich jedoch durch einen weiteren Sensor eingeschränkt werden, wird zum einen die Anzahl der Kantenkandidaten verringert, und zum anderen können zusätzliche Informationen die Korrespondenzfindung von Kanten vereinfachen. Ist beispielsweise die Breite und Distanz eines Fahrzeugs durch den entfernungsgebenden Sensor bekannt, ist eine erheblich restriktivere Wahl der Abstands- und Größenschwellwerte möglich.

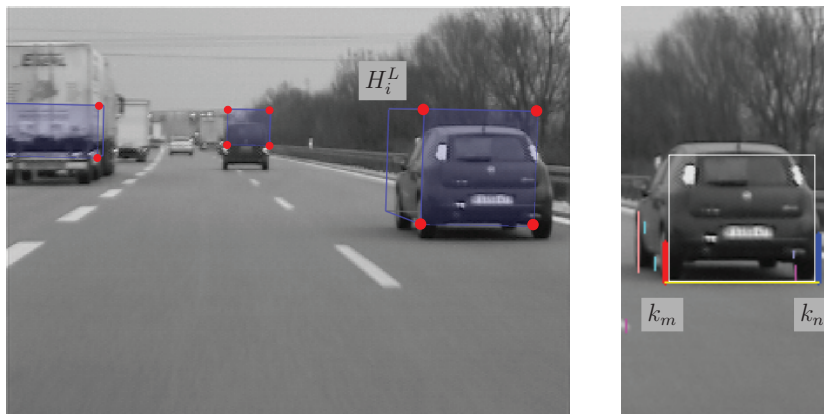


Abbildung 3.11: Korrektur der Bildprojektion von entfernungs-sensorbasierten Hypothesen durch kantenbasierte Bildhypothesen. Durch Nickbewegung bzw. ungenauer Kalibrierung abweichende Hypothesen können durch vertikale Kantenpaare in der Position korrigiert werden.

Die Bildhypothese wird anhand vertikaler Kanten gestützt (siehe auch Abschnitt 3.2.1). Die Lage- und Geometrieinformationen der Laserscannerhypothese kann zur Begrenzung des Suchraums verwendet werden. Geeignete Kantenpaare werden anhand einer Kostenfunktion bestimmt. In die Kostenfunktion fließen die Differenzen von laserscanner- und kantenbasierter Fahrzeugbreite (b_s und b_i) sowie laserscanner- und kantenbestimmten Fußpunkt ($y_s^{+/-}$ und $y_i^{+/-}$) mit ein. Diese Kostenfunktion

$$C(k_i, k_j) = |b_s - b_i| + |y_s^+ - y_i^+| + |y_s^- - y_i^-| \quad (3.3.1)$$

wird für alle möglichen Kantenpaar (k_i, k_j) für $i \neq j$ berechnet. Das Kantenpaar (k_m, k_n) , welches bezüglich der Funktion $C(\cdot)$ kostenminimal ist, wird für die Bildhypothesenerzeugung herangezogen (siehe weißes Rechteck in Abbildung 3.11). Diese Vorgehensweise wird mit *GeoLageKanten* abgekürzt.

Abbildung 3.11 zeigt quaderförmige Hypothesen, die durch einen Laserscanner erzeugt wurden. Aufgrund einer Nickbewegung des Eigenfahrzeugs – verursacht durch einen starken Abbremsvorgang – korrespondieren die Kameraprojektionen der Hypothesen nicht mit den Aufenthaltsorten der Fahrzeuge. Dadurch werden extrahierte Bildmerkmale verfälscht. Am Beispiel der markierten Hypothese H_i^L wird eine Korrektur der Bildprojektion durch Kombination mit einer kantengestützten Bildhypothese durchgeführt.

3.3.2.4 Bewertung und Vergleich

Damit die oben vorgestellten Möglichkeiten zur Bildhypothesenerzeugung hinsichtlich ihrer Leistungsfähigkeit verglichen werden können, wurde die, mit Referenzdaten versehene, Autobahnszene S_4 herangezogen (siehe auch Abschnitt 6.2.2 auf Seite 123). Für alle erzeugten Bildhypothesen wurden Bildmerkmale³ im Videobild extrahiert und mittels eines trainierten Fahrzeugklassifikators bewertet. Dadurch kann die Leistungsfähigkeit der Hypothesengenerierung indirekt an den Auswirkungen auf die Erkennungsleistung überprüft werden, da die Klassifikatoren auf einen präzisen Bildbereich angewiesen sind. Die Ergebnisse der Evaluierung sind in Abbildung 3.12a zu sehen. Weitere exemplarische Resultate werden in den Abbildungen 3.12b und 3.13 dargestellt.

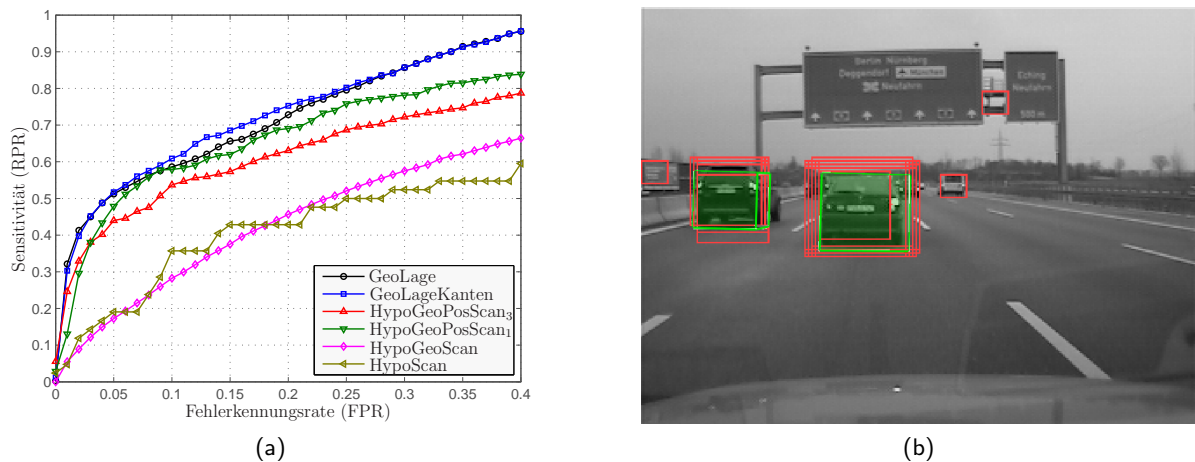


Abbildung 3.12: Bewertung der verschiedenen Aufmerksamkeitssteuerungsverfahren zur Fahrzeugdetektion anhand der Erkennungsraten (a). Als Referenzszene wurde ein Autobahnszenario herangezogen. Abbildung (b) zeigt einen Szenenschnappschuss und das Ergebnis von zwei Verfahren. Durch rote Rahmen wird das Resultat von HypoGeoScan und durch die grünen Objekte das Resultat von GeoLageKanten visualisiert.

Die besten Ergebnisse liefern die Verfahren mit Aufmerksamkeitssteuerung GeoLage und GeoLageKanten. Gleichzeitig sind diese Verfahren auch am leistungseffizientesten, da lediglich eine Hypothese pro Objekt ausgewertet muss. Da bei der verwendeten Szene lediglich geringe Nickbewegungseinflüsse auftraten, ist der Vorsprung von GeoLageKanten gegenüber GeoLage gering. Auf unebenen Straßen, z.B. in Stadtszenarien, vergrößert sich der Leistungszu-

³ Es wurden die videobasierten Gradientenmerkmale, die in Abschnitt 4.2 eingeführt werden verwendet.

wachs jedoch erheblich. Eine niedrigere Erkennungsleistung erzielen die scannenden Verfahren HypoGeoPosScan₃ bzw. HypoGeoPosScan₁. Durch eine Verkleinerung der Fensterabdeckungsschrittweite auf einen Bildpunkt (HypoGeoPosScan₁) kann, auf Kosten der Laufzeit, ein leichter Leistungszuwachs erzielt werden. Erheblich niedrigere Erkennungsleistungen werden bei den Verfahren HypoGeoScan bzw. HypoScan erzielt, das gesamte Bild mit Hypothesen abdecken.

Die durchschnittlich benötigte Rechenzeit t_{Zyk} , zur Ausführung eines Bewertungszyklus, ist direkt korreliert mit der Anzahl der erzeugten Bildhypothesen (siehe Tabelle 3.2). Für die kantenbasierte Suche und die direkte Projektion ist t_{Zyk} am niedrigsten, da lediglich eine einzelne Hypothese ausgewertet werden muss. Der geringfügig höhere Rechenzeitanspruch für die kantenbasierte Suche ist auf die zusätzlichen Bildverarbeitungsschritte zurückzuführen. Die Suchverfahren HypoGeoPosScan₃ bzw. HypoGeoPosScan₁ sind bei kleinem Suchbereich bedingt echtzeitfähig. HypoGeoScan bzw. HypoScan können ohne Modifikationen nicht für ein System mit Echtzeitanforderungen eingesetzt werden.

Durch die Aufmerksamkeitssteuerung kann die Erkennungsleistung bei geringerer Algorithmuskomplexität erhöht werden. Gleichzeitig wird durch die effiziente Suchraumbegrenzung viel Rechenzeit gespart. Eine redundante Systemauslegung ist allerdings nur bedingt möglich. Durch das enge Zusammenspiel der Sensoren kann ein Sensorausfall, im Gegensatz zum Hypothesenkollektiv, nur durch eine gleichzeitige Anpassung des Verarbeitungsalgorithmus kompensiert werden.

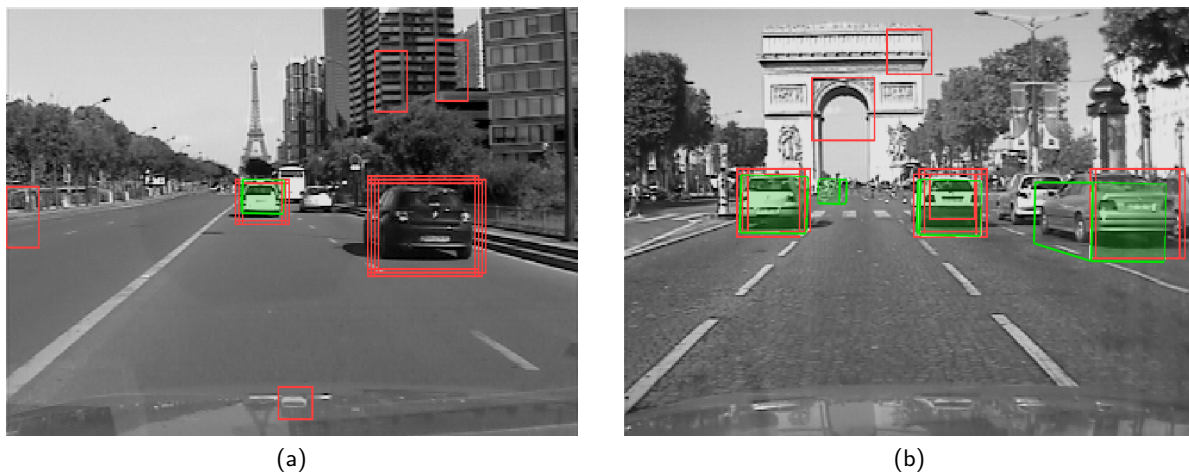


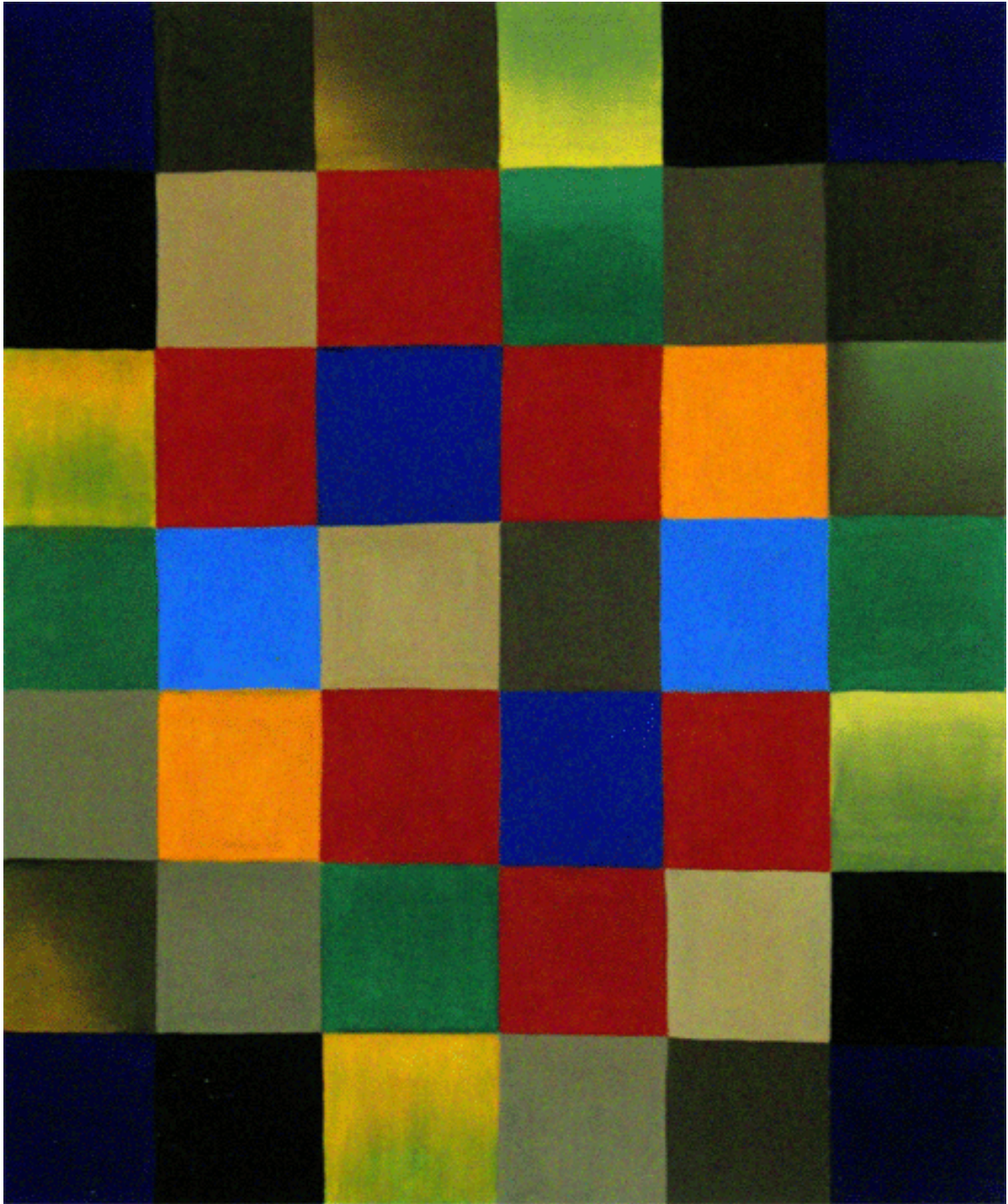
Abbildung 3.13: Bildbasierte Klassifikation mit kantenbasierter laserscannergestützter Suche (grüne Quader) und mit den Verfahren HypoGeoScan (a) und HypoScan (b) (rote Rahmen). Durch die scannenden Verfahren können Objekte erkannt werden, die durch den Laserscanner nicht abgedeckt werden (a). Zusätzlich sind jedoch Fehlerkennungen möglich, die anhand des entfernungsgebenden Sensors vermieden werden könnten (z. B. Fahrzeugspiegelung in der Motorhaube).

3.4 Zusammenfassung

In diesem Kapitel wurden Hypothesenerzeugungsverfahren für Fahrzeuge und Fußgänger mittels bild- und entfernungsgebenden Sensoren betrachtet. Es wurde zunächst eine inkrementelle Linienextraktion für Laserscannerrohdaten vorgestellt, auf deren Basis sowohl Fahrzeug- als auch Fußgängerhypothesen erzeugt werden können. Im Anschluss daran wurde eine Hypothesenerzeugung auf Basis von Kanten im Wärmebild thematisiert. Im letzten Teil des Kapitels wurde untersucht, welches Potential eine sensorübergreifende Hypothesenerzeugung aufweist.

Es konnte gezeigt werden, dass durch den Laserscanner eine hohe Hypothesenabdeckungsrate sowohl für Fahrzeuge als auch für Fußgänger erzielt werden kann. Probleme treten bei partiellen Verdeckungen oder bei zu geringen Objektzwischenabständen auf. Eine aggregierte Aufmerksamkeitssteuerung mit Laserscanner und Wärmebild kann diese Probleme adressieren und zu einer ganzheitlicheren Wahrnehmung führen. Eine Zusammenführung von Hypothesen führt gleichermaßen zu einer erheblich gesteigerten Erkennungsleistung bei geringerer algorithmischer Komplexität. Allerdings müssen dabei Sensorausfälle gesondert berücksichtigt werden.

Im nächsten Kapitel werden geeignete Merkmale identifiziert und bewertet, um die erzeugten Hypothesen mittels Klassifikatoren verifizieren zu können.



„Neue Harmonie“ von Paul Klee (1879 - 1940).

4

Merkmalsgewinnung

Im Kontext der Mustererkennung entspricht die Merkmalsextraktion einer Dimensionsreduktion von Daten, wobei repräsentative Informationen erhalten bleiben. Da das Datenaufkommen aufgrund von Redundanz (viele Daten, wenig Informationsgehalt) oftmals sehr hoch ist, würden Algorithmen, die unmittelbar auf diesen Daten operieren, ein hohes Ressourcenaufkommen hinsichtlich Speicherverbrauch und Rechenzeit aufweisen. Zudem würde bei Mustererkennungsalgorithmen das Problem der Überanpassung¹ begünstigt werden [DHS00]. Ein Merkmalsextraktionsverfahren bildet somit redundante Daten auf dimensionsreduzierte Merkmalsvektoren ab, die vorzugsweise ausschließlich relevante Informationsanteile enthalten. Die resultierenden Merkmalsvektoren können anschließend als Eingabe für ein Lernverfahren verwendet, oder mit einer bekannten Vorlage verglichen werden, um eine Unterscheidung zwischen Vorder- und Hintergrundklasse zu treffen.

Es existieren eine Reihe von automatischen Dimensionsreduktionsalgorithmen, wie beispielsweise Principal Component Analysis, Isomap und Kernel PCA. Merkmalsextraktionsverfahren, die applikationsabhängig von einem menschlichen Experten erstellt werden, sind jedoch oftmals effektiver. In der Literatur werden für Fußgänger- und Fahrzeugerkennungssysteme eine Vielzahl von unterschiedlichen Merkmalen herangezogen. Aufgrund der größeren Verbreitung von bildgebender Sensorik, werden diese vorwiegend aus Bildern gewonnen. Ein guter Überblick über charakteristische Merkmale zur Fußgängererkennung findet sich sowohl in [Gav01] als auch in [BBC⁺02]. Eine Zusammenfassung aktueller Vorgehensweisen wird in [GT07] beschrieben. Darüber hinaus werden in [SBM06] geeignete Merkmale zur Fahrzeugerkennung diskutiert.

Merkmale, die als Vorlage fungieren, sind beispielsweise probabilistische Templates, die eine Wahrscheinlichkeitsverteilung der Intensitätswerte widerspiegeln. Diese werden im Erkennungsprozess über eine Korrelationsfunktion zur Fußgängererkennung im Wärmebild herangezogen [ND02]. Die charakteristische Wärmeverteilung von Fußgängern kann zudem mit künstlich generierten Modellen nachgebildet werden [BFC⁺04, BBLR05]. Ähnliche Eigenschaften werden über Helligkeitshistogramme, Kontrast- und Grauwertverteilungen [FYN⁺04] kodiert. In

¹ Das Problem der Überanpassung (overfitting) kann auftreten, wenn ein statistisches Trainingsmodell auf zufälligen Musteranordnungen in den redundanten Daten passt, und dadurch eine Generalisierung des Modells zur Erkennung von Mustern außerhalb der Trainingsdaten beeinträchtigt wird.

[SGRB05] werden einzelne Körperteile als Knoten eines Graphs extrahiert und mit einem fußgängertypischen Mustergraphen verglichen. Im Bereich der Fahrzeugerkennung zählen die „U“-Form des Fahrzeughecks [HKT⁺00], die Anordnung der Rückleuchten [CP99] und das Nummernschild [PP95] zu klassenkennzeichnenden Merkmalen.

Zu den Merkmalen, die über Lernverfahren als Trainingsdatensatz verwendet werden, zählen sogenannte „Haar“-Merkmale. Seit ihrer Veröffentlichung und Anwendung von Viola und Jones [VJ01] wurden sie im sichtbaren Videospektrum für zahlreiche Objektklassen, wie Fußgänger [PP00, VJS05, XCQW06, CLS⁺07] und Fahrzeuge [SK00] herangezogen. Sie kodieren lokale Kontrastverhältnisse an diskreten Kantenrichtungen. Gaborfilter, die gleichermaßen Kanten und Linien kompakt kodieren können, wurden für die Fußgängererkennung [HHK02, CZQ05] und die Fahrzeugerkennung im Videobild [CZS06, SBM02, SBM05] untersucht. Einen Schritt weiter gehen Gradientenhistogramme, die neben der Kantenintensität zusätzlich deren Größen- und Orientungsverteilung in einem lokalen Bildbereich kodieren. Sie stellen aktuell das „state-of-the-art“-Merkmal für Personenerkennung im sichtbaren Videobild dar [DT05, JZ07]. Aber auch im Wärmebild werden diese Merkmale erfolgreich eingesetzt [SRBB06]. Aus Laserscannerrohdaten werden gegenwärtig vor allem Objektausdehnungen als Merkmale herangezogen [FW01, FDW02].

In diesem Kapitel werden alle Merkmale vorgestellt, die für die hypothesengetriebene Fußgänger- und Fahrzeugerkennung dieser Arbeit Verwendung finden. Als Bildmerkmale werden für beide Objektklassen und bildgebenden Sensoren normierte Gradientenhistogramme genutzt. Es wird untersucht, welche Parametrisierung für Fahrzeuge im Wärme- und Videobild gewählt werden sollte, und wie sich Objektskalierungen auf die Erkennungsraten auswirken. Des Weiteren wird ein neues Verfahren für eine Sichtbereichsvergrößerung des bildgebenden Sensors durch eine Achsenspiegelung vorgestellt und bewertet. Aus den Laserscannerrohdaten wird für Fußgänger die Bein-Pendelbewegung verwendet. Für Fahrzeugheckansichten wird untersucht, inwiefern die typische Reflektanzsignatur als Merkmal verwendet werden kann. Zusätzlich wird das Dynamikverhalten von Fahrzeugen als Kennzeichen herangezogen. Eine Gegenüberstellung der verwendeten Merkmale mit den entsprechenden Objektklassen ist in Tabelle 4.1 zu sehen. Die Datenkataloge, die in diesem Kapitel zur Merkmalsbewertung verwendeten werden, sind im Anhang A auf Seite 179 zu finden.

Tabelle 4.1: Verwendete Merkmale bezüglich der Objektklassen und Sensoren.

Objektklassen	Merkmale			
	Gradienten- histogramme	Reflektanz- signatur	Pendelbewegung	Dynamikwerte
PKW	Video und FIR (Abschnitt 4.2)	Laserscanner (Abschnitt 4.3)		Radar und Laserscanner (Abschnitt 4.5)
Fußgänger	Video und FIR (Abschnitt 4.2)		Laserscanner (Abschnitt 4.4)	

4.1 Datensegmentierung

Ausgangsbasis für einen effizienten Merkmalsgewinnungsprozess sind Objekthypothesen, mittels derer eine Aufmerksamkeitssteuerung erzielt werden kann (siehe auch Kapitel 3 zur Hypothesengenerierung). Lediglich an diesen Stellen werden Merkmale aus den Sensordaten extrahiert. Dadurch wird das Merkmalsaufkommen begrenzt, wodurch Rechenzeit gespart und die Fehlerkennungsrate (FPR)² gesenkt werden kann. In Abbildung 4.1 wird der Segmentierungsprozess auf den verfügbaren Sensordaten illustriert (siehe auch Informationen zum Versuchsträger in Abschnitt 2.4).

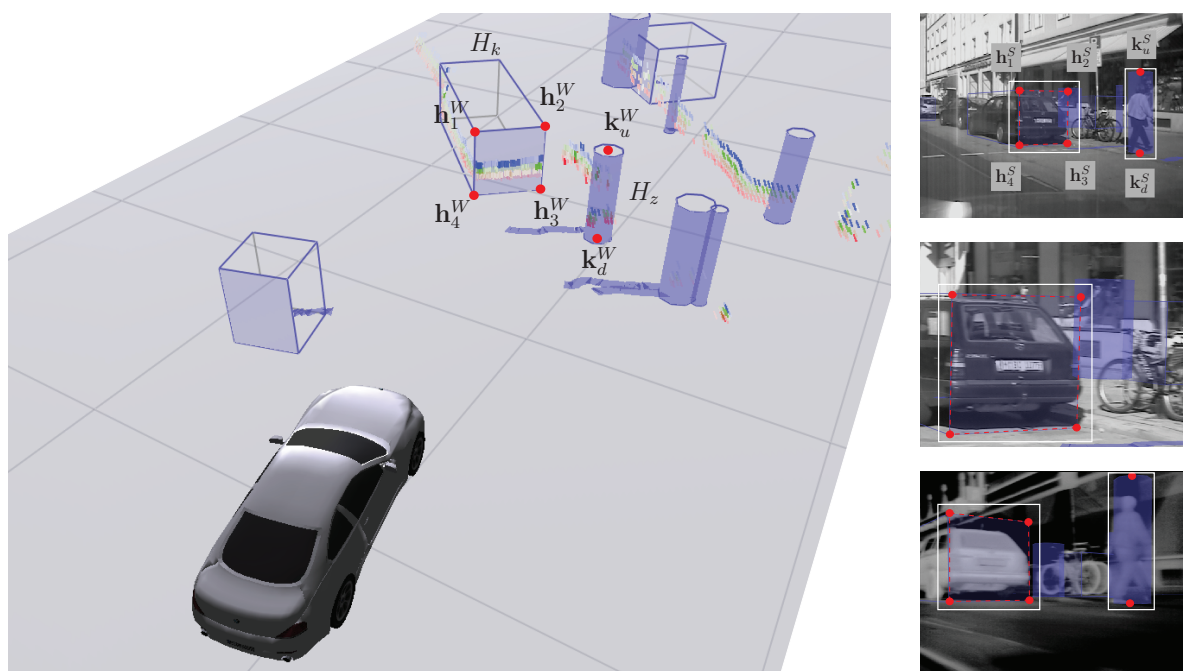


Abbildung 4.1: Illustration der Datensegmentierung zur Gewinnung von Merkmalen. Laserscannerrohdaten (Isometrische Ansicht auf linker Seite) werden den quader- und zylinderförmigen Hypothesen zugeordnet. Alle Hypothesen werden in die bildgebende Sensorik projiziert und auf rechteckige Bereiche abgebildet (Bilder auf der rechten Seite von oben nach unten: Videokamerabild mit Normalobjektiv, Videokamerabild mit Teleobjektiv und Wärmebild).

Auf der linken Seite der Illustration werden das Eigenfahrzeug, die Laserscannerrohdaten und mehrere quaderförmige und zylindrische Hypothesen mit ihren Bewegungstrajektorien dargestellt. Eine Fahrzeughypothese H_k und eine Fußgängerhypothese H_z wird exemplarisch für beliebige Hypothesen herausgegriffen und in die Bilder aller verfügbaren bildgebenden Sensoren S_i projiziert (rechte Seite der Illustration). Zur Bestimmung der resultierenden Bildbereiche werden für H_k zunächst die Eckvektoren \mathbf{h}_1^W , \mathbf{h}_2^W , \mathbf{h}_3^W und \mathbf{h}_4^W der sensorzugewandten, kurzen Rechtecksseite \mathcal{H}_k^W von Weltkoordinaten in zweidimensionale Bildkoordinaten \mathbf{h}_1^S , \mathbf{h}_2^S , \mathbf{h}_3^S und

² Die Fehlerkennungsrate oder Falsch-Positiv-Rate ist das Verhältnis von falsch eingeordneten Hintergrundobjekten zu der Gesamtanzahl von Hintergrundobjekten definiert (siehe auch Tabelle B.2 im Anhang).

\mathbf{h}_4^S überführt. Aufgrund der perspektivischen Verzerrung wird durch diese Rechtecksprojektion ein konkaves Viereck \mathcal{H}^S erzeugt. Um den Bereich, zur Vereinfachung der Merkmalsextraktion, erneut in ein Rechteck zu transformieren, wird von diesem konkaven Viereck das kleinste umfassende Rechteck

$$\tilde{\mathcal{H}}_k^S = \left\{ \tilde{\mathbf{h}}_1^S = (\min[h_{1x}, h_{4x}] \quad \min[h_{1y}, h_{2y}])^T, \tilde{\mathbf{h}}_2^S = (\max[h_{2x}, h_{4x}] \quad \min[h_{1y}, h_{2y}])^T, \right. \\ \left. \tilde{\mathbf{h}}_3^S = (\max[h_{2x}, h_{3x}] \quad \max[h_{3y}, h_{4y}])^T, \tilde{\mathbf{h}}_4^S = (\min[h_{1x}, h_{4x}] \quad \max[h_{3y}, h_{4y}])^T \right\}$$

gebildet.

Die zylindrischen Fußgängerhypothesen H_z werden in Weltkoordinaten durch die Mittelpunkte der Zylinderdeck- bzw. -grundplatte \mathbf{k}_d^W bzw. \mathbf{k}_u^W und den Radius r^W beschrieben. Nach der Projektion der Mittelpunkte in Sensorkoordinaten kann das Rechteck

$$\tilde{\mathcal{H}}_z^S = \left\{ \tilde{\mathbf{k}}_1^S = (k_{ux} - r^S \quad k_{uy})^T, \tilde{\mathbf{k}}_2^S = (k_{ux} + r^S \quad k_{uy})^T, \right. \\ \left. \tilde{\mathbf{k}}_3^S = (k_{dx} + r^S \quad k_{dy})^T, \tilde{\mathbf{k}}_4^S = (k_{dx} - r^S \quad k_{dy})^T \right\}$$

mit

$$r^S = \frac{\arctan\left(\frac{r_w}{|\mathbf{k}_d^W - \mathbf{p}^W|_2}\right) w_S}{\phi_S}$$

in Bildkoordinaten bestimmt werden. \mathbf{p}^W bezeichnet hierbei den Sensorverbauort in Weltkoordinaten, ϕ_S den horizontalen Öffnungswinkel und w_S die horizontale Pixelauflösung des bildgebenden Sensors.

In einem letzten Schritt werden die resultierenden Rechtecke um objektspezifische Skalierungsfaktoren m_k bzw. m_z prozentual vergrößert (weiße Rechtecke in Abbildung 4.1), da der Umgebungskontext der Objekte eine Rolle für die Erkennungsraten spielt. Aufgrund von ungenauen Sensorkalibrierungen oder Nickbewegungen des Eigenfahrzeugs sind zusätzliche Detailkorrekturen der Bildextraktionsbereiche notwendig. Diese Verfahren wurden in Abschnitt 3.3.2 auf Seite 53 behandelt.

Die Extraktion von Laserscannerrohdaten wird hypothesen- und merkmalspezifisch durchgeführt. Die grundsätzlich Zuordnung der Rohdaten erfolgt durch modifizierte Nearest Neighbour Verfahren. Eine genaue Segmentierungsbeschreibung findet sich in den Abschnitten 4.3 und 4.4. Im nächsten Abschnitt werden zunächst die Bildmerkmale thematisiert.

4.2 Gradientenhistogramme als Bildmerkmale

Ein Histogramm H beschreibt die statistische Verteilung der Häufigkeit einer Variablen X auf η Intervalle (Klassen) h_1, \dots, h_η . Diese Häufigkeit kann entweder absolut oder relativ ausgedrückt werden. Ein Beispiel in der klassischen Bildanalyse ist die Erstellung eines Histogramms über die Häufigkeit der vorkommenden Farbwerte in einem Bild $f(x, y)$. Bei Gradientenhistogrammen³, die auf McConnell [McC86] zurückgehen, erfolgt die Klasseneinteilung

³ Englische Bezeichnungen für Gradientenhistogramm sind „edge histogram“ und „histogram of oriented gradients (HOG)“.

nicht nach dem Intensitätswert eines Bildpunktes, sondern nach der jeweiligen Kantenrichtung $f_{\theta^+}(x, y) = \text{atan2}(f_y(x, y), f_x(x, y))^4$ bzw. $f_{\theta^-}(x, y) = \arctan\left(\frac{f_y(x, y)}{f_x(x, y)}\right)$ der Gradientenbilder f_x und f_y , welche mittels eines Kantenoperators (z. B. Sobel, Prewitt) erzeugt werden. Die möglichen Richtungen von $-\pi$ bis π werden hierfür in η Klassen unterteilt. Entsprechend der Kantenrichtung wird nun für jeden betrachteten Bildpunkt die zugehörige Gradientenmagnitude $f_{\nabla}(x, y) = \sqrt{f_x(x, y)^2 + f_y(x, y)^2}$ auf die entsprechende Histogrammklasse h_k aufaddiert:

$$\forall x, y \quad h_k \rightarrow h_k + f_{\nabla}(x, y)$$

Wie sich die Histogrammklasse h_k berechnet ist davon abhängig, ob neben der Gradientenrichtung zusätzlich die Gradientenorientierung⁵ mit berücksichtigt werden soll (siehe auch Abbildung 4.2).

$$k = \begin{cases} \left\lfloor \frac{1}{2\pi} (f_{\theta^+}(x, y) + \pi) \cdot \eta \right\rfloor + 1 & \text{Kantenorientierung wird berücksichtigt.} \\ \left\lfloor \frac{1}{\pi} (f_{\theta^-}(x, y) + \pi) \cdot \eta \right\rfloor + 1 & \text{Kantenorientierung wird nicht berücksichtigt.} \end{cases}$$

Welche Kantenberechnung gewählt werden sollte, ist abhängig von der zu trainierenden Objektklasse. Bei gleichbleibenden Kontrastverhältnissen zwischen Objekt und Objekthintergrund (z. B. schwarze Autoreifen vor heller Straße) enthalten die Kantenorientierungen wertvolle Informationen. Wechseln jedoch die Hintergrund-Vordergrund Kontrastverhältnisse wie z. B. bei unterschiedlich gekleideten Fußgängern vor beliebigen Hintergründen ist es im Allgemeinen von Vorteil, wenn die Orientierung der Kanten ignoriert wird.

Wird ein Gradientenhistogramm H über einen kleinen Bildausschnitt berechnet, werden dadurch die dominanten lokalen Kantenrichtungen kodiert – der genaue Kantenverlauf wird ignoriert. Rauschen und unerhebliche Details werden somit weitgehend gefiltert. Um die Gradientenverläufe eines ganzen Objektes zu erfassen, berechnet Lowe [Low99, Low04] eine Schar von normierten Gradientenhistogrammen an markanten Bildpositionen (Keypoints), die in einem Vorverarbeitungsschritt ermittelt werden. Die von Lowe eingeführte euklidische Normierung der Histogramme kompensiert Beleuchtungsänderungen:

$$\forall h_i \in H \quad \tilde{h}_i \rightarrow \frac{h_i}{\sqrt{\sum_{j=1}^{\eta} (h_j)^2}} \quad (4.2.1)$$

Die Performanz des Erkennungsalgorithmus von Lowe ist unmittelbar korreliert mit der Güte der Keypoint-Suchroutine. Dalal und Triggs [DT05] überziehen in ihrem alternativen Ansatz

⁴ atan2 liefert den Winkel des Vektors $(x \ y)^T$ im korrekten Quadranten zurück und ist definiert als

$$\text{atan2}(y, x) := \begin{cases} \arctan \frac{y}{x} & \text{für } x > 0 \\ \arctan \frac{y}{x} + \pi & \text{für } x < 0, y \geq 0 \\ \arctan \frac{y}{x} - \pi & \text{für } x < 0, y < 0 \\ +\pi/2 & \text{für } x = 0, y > 0 \\ -\pi/2 & \text{für } x = 0, y < 0 \\ 0 & \text{für } x = 0, y = 0 \end{cases}$$

⁵ Wird die Orientierung einer Kante mit berücksichtigt, so gibt es für eine Kantenrichtung zwei mögliche Orientierungen. Eine Kante mit vertikaler Richtung beispielsweise kann sowohl durch einen hell-dunkel als auch einen dunkel-hell Übergang verursacht werden und somit zwei mögliche Orientierungen aufweisen.

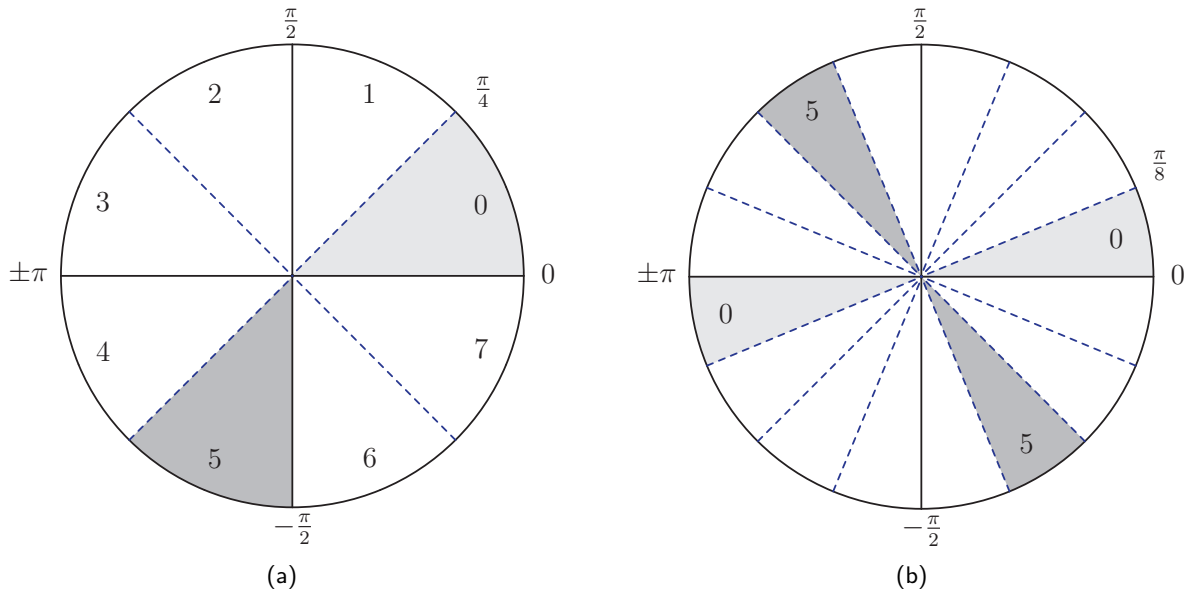


Abbildung 4.2: Beispielhafte Unterteilung der Kantenrichtungen von $-\pi$ bis π in $\eta = 8$ gleich große Intervalle (Klassen). Abhängig von der Objektklasse und den Abbildungseigenschaften des bildgebenden Sensors sollten die Orientierungen der Kanten unterschieden (a) oder ignoriert (b) werden.

das Bild mit einem dichten Netz von rechteckigen Zellen, in denen normierte Gradientenhistogramme H_i bestimmt werden. Dadurch ist dieses Verfahren unabhängig von einer Keypointsuche und resultiert speziell für Fußgänger in wesentlich höheren Erkennungsraten. Komplexe globale Objektstrukturen werden somit durch einen Verbund von lokalen Gradientenhistogrammen abgebildet. Ein Bild mit einer Größe von $w \times h$ Bildpunkten, das in Histogramme H_i mit η Gradientenintervalle und einer rechteckigen Zellengröße von $\kappa \times \kappa$ Bildpunkten aufgeteilt wird, resultiert in einen $\frac{\eta \cdot w \cdot h}{\kappa^2}$ -dimensionalen Merkmalsvektor

$$\mathbf{x}_H = \left(\underbrace{(h_1 \quad \dots \quad h_\eta)^T}_{H_1} \quad H_2 \quad \dots \quad H_{\frac{w \cdot h}{\kappa^2}} \right)^T. \quad (4.2.2)$$

Die ideale Parameterwahl für normierte Gradientenhistogramme ist sensor- und objektklassen-spezifisch. Im folgenden Abschnitt werden die relevanten Größen experimentell ermittelt.

4.2.1 Parametervariation

Viele Stellgrößen zur Erzeugung der Gradientenhistogramme üben einen mehr oder weniger großen Einfluss auf die Erkennungsraten aus. Dalal und Triggs wiesen experimentell nach [DT05], dass die ungeglättete Erzeugung der Gradientenbilder mittels einer diskreten Faltung, durch den zentrierten Filterkern $[1, 0, -1]$ bzw. $[1, 0, -1]^T$, die besten Resultate liefert. Des Weiteren ist es vorteilhaft, wenn das Objekt bildzentriert vorliegt und zusätzlich von einem schmalen Hintergrundrand umgeben ist. Dieser Hintergrundanteil fungiert als Objektkontext und ist ausschlaggebend für eine robuste Erkennung. Der Performanzunterschied zwischen quadratischen, rechteckigen und kreisförmigen Histogrammzellen ist marginal. Wegen der einfacheren

Handhabung und effizienteren Berechnung werden im Folgenden ausschließlich quadratische Zellen verwendet. Die wesentlichen Parameter eines Gradientenhistogramms $\kappa \times \kappa \times \eta^{-/+}$ sind dessen Zellenausdehnung κ in Bildpunkten, die Anzahl der Gradientenklassen η und die Berücksichtigung bzw. Ignorierung der Kantenorientierung ($^-/+$). Für Fußgänger im Videobild sind nach [DT05] $8 \times 8 \times 9^+$ Gradientenhistogramme für 64×128 große Trainingsbilder am geeignetsten. Für Fahrzeugheckansichten sowie für Fußgängeransichten im Wärmebild werden die optimalen Parameterkonfigurationen im nächsten Abschnitt ermittelt.

Wird die Orientierung der Gradienten bei Fahrzeugen ignoriert, sinkt die Erkennungsrate im Wärmebild um 15 % (10^{-4} FPR) und um 27 % (10^{-5} FPR) im Videobild. Offensichtlich liefern charakteristische hell-dunkel Kontrastverhältnisse sowohl bei Fahrzeugheckansichten (vgl. auch obere Bildreihe in Abbildung 4.6) im Videobild (z. B. weißes Nummernschild vs. dunklerer Rückfront und schwarzer Reifen vs. heller Straße) als auch im Wärmebild (z. B. heißer Reifen vs. kalter Straße und warmer Rückleuchten vs. kühlerer Umgebung) informationstragende Kantenorientierungen.

Wie in Abbildung 4.3 zu sehen ist, bewirkt ein feines Netz von Gradientenhistogrammen höhere Erkennungsraten als ein grobes Netz. Die optimale Zellenabmessungen $\kappa \times \kappa$ zur Detektion von Fahrzeugansichten liegt bei 5×5 . Im Wärmebild liefern feinere Partitionierungen, wie 2×2 große Zellen vergleichbare Resultate. Durch die damit verbundene Vervierfachung der Merkmalsvektorstärke sind jedoch gleichfalls 5×5 große Zellen vorzuziehen. Der ideale Bereich für die Anzahl der Orientierungsklassen η liegt sowohl im Wärmebild als auch im Videobild zwischen 8 und 16 Unterteilungen. Bei kleineren Werten bricht die Erkennungsrate stark ein. Größere Werte bewirken keine zusätzliche Verbesserung der Erkennungsrate. Für Fahrzeugheckansichten mit einer Abmessung von 60×60 Bildpunkten, sind mit den Konfigurationen $5 \times 5 \times 8^-$ (Video) bzw. $5 \times 5 \times 16^-$ (FIR) die höchsten Erkennungsraten zu erzielen.

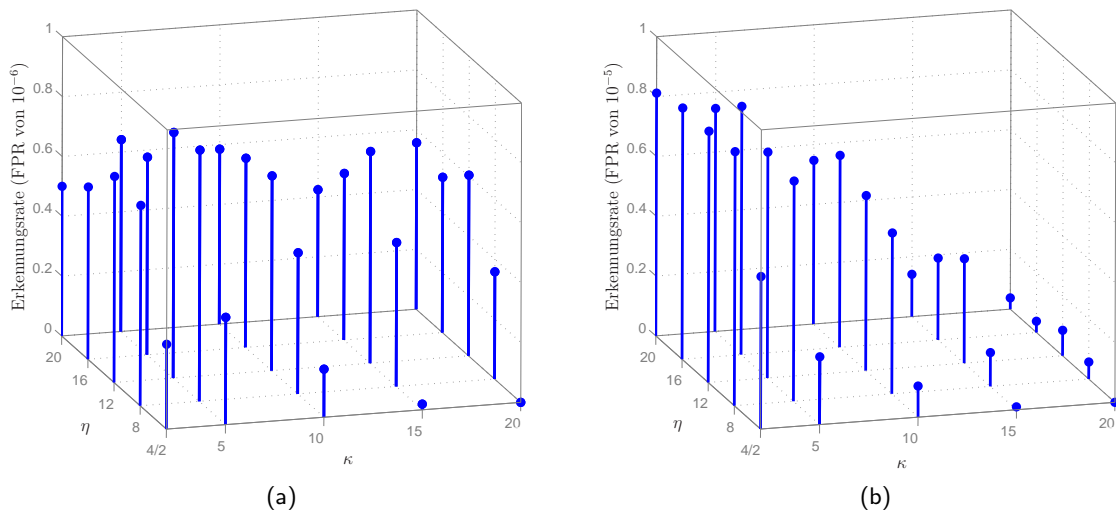


Abbildung 4.3: Erkennungsraten von PKWs im Videobild (a) und Wärmebild (b) über Gradientenhistogramme $\kappa \times \kappa \times \eta$ abhängig von den Zellengrößen κ in Bildpunkten und der Gradientenklassenanzahl η . Die Konfigurationen $5 \times 5 \times 8^-$ (Video) bzw. $5 \times 5 \times 16^-$ (FIR) liefern die höchsten Erkennungsraten.

Anders als bei Fußgängeransichten im Videobild, profitiert die Erkennung von Fußgängeransichten im Wärmebild von der Berücksichtigung der Kantenorientierung. Dies gilt jedoch nur für Umgebungstemperaturen bis ca. 20°C, da sich bei höheren Temperaturen der Kontrast zwischen Hintergrund und wärmeemittierenden Objekten verschlechtert. Bei signifikant höheren Umgebungstemperaturen invertieren sich die Kontrastverhältnisse und damit gleichermaßen die Kantenorientierungen an der Fußgängerkontur. Enthält der Trainingsdatensatz Heißlandaufnahmen sind somit ungerichtete Kanten vorzuziehen (7% Erkennungsratenverbesserung bei einer FPR von 10^{-4}). Die folgenden Betrachtungen konzentrieren sich auf Ansichten, die bei einer Umgebungstemperatur unter 20°C aufgenommen wurden.

Um die optimalen Parameter für die Gradientenhistogramme zu ermitteln, wurden von den infrarotbasierten Fußgängeransichten (Bildabmessungen liegen bei 64×128 Bildpunkten) aus dem Datenkatalog 4 (siehe auch Anhang A auf Seite 179) die Erkennungsraten mit unterschiedlichen Parameterkonfigurationen ermittelt. Die Gradientenhistogramme wurden in quadratischen Zellen mit einer Breite/Höhe von 4, 8, 16 und 32 Bildpunkten und mit Gradientenintervalle im Bereich $[4, 8, 12, \dots, 32]$ berechnet. Bei Zellengröße von 8×8 bzw. 16×16 Bildpunkten und einem Gradientenintervall zwischen 16 und 32 werden die höchsten Erkennungsraten erzielt (zwischen 63% und 72% bei einer FPR von 10^{-4}). Kleinere oder größere Zellengrößen führen zu einem starken Abfall der Erkennungsraten. Höhere Gradientenintervalle als 16 bewirken keine Erhöhung der Erkennungsraten. Die optimale Gradientenhistogrammkonfiguration, bei gleichzeitig minimaler Merkmalsvektorenlänge, liegt bei $8 \times 8 \times 12^-$. Damit wird eine Erkennungsrate von 72% erzielt.

4.2.2 Skalierungsinvarianz

Das vorgestellte Verfahren zur Berechnung der Gradientenhistogramme ist an konstante Bildabmessungen $w \times h$ gebunden. Erhöht bzw. verringert sich die Distanz von Objekten zur Kamera, so verkleinert bzw. vergrößert sich aufgrund der perspektivischen Projektion deren Abbildungsausdehnung im Bildbereich. Wird die Gradientenberechnung auf einem verkleinerten Bildbereich $w' \times h'$ durchgeführt, müssen entweder die Zellengröße κ , oder die Anzahl der Zellen pro Bild vermindert werden, damit ausschließlich das Objekt von Gradientenhistogrammen überdeckt wird. Analog dazu müssen bei einem vergrößerten Bildbereich die Zellengrößen angehoben werden, oder aber die Anzahl der Zellen pro Bild erhöht werden, um das gesamte Objekt abzudecken. Beide Vorgehensweisen führen zu einer Änderung der resultierenden Merkmalsvektorenlänge, wodurch Klassifikationsverfahren mit statischen Merkmalsvektorkardinalitäten (z. B. SVM, Neuronale Netze) nicht länger verwendbar sind.

4.2.2.1 Umsetzung

Um eine konstante Länge des Merkmalsvektors zu erhalten, sollten demnach die Abmessungen des Bildausschnitts, in welchem die Gradientenhistogramme berechnet werden, konstant bleiben. Dies kann erzielt werden, indem das Eingabebild, vor der Merkmalsextraktion mit den Faktoren $\frac{w}{w'}$ bzw. $\frac{h}{h'}$ in der Breite bzw. Höhe skaliert wird. Das resultierende Bild hat somit stets die Abmessungen $w \times h$, wodurch gleichermaßen die Kardinalität $|\mathbf{x}_H|$ des Merkmalsvektors mit $\frac{\eta \cdot w \cdot h}{\kappa^2}$ konstant bleibt.

In der Regel werden die einzelnen Pixel des Ausgangsbildes durch einen Skalierungsvorgang nicht auf ganzzahlige Koordinaten abgebildet. Es stehen jedoch verschiedene Interpolations-

verfahren zur Auswahl, um die Bildpunkte des Resultatbildes wieder auf ganzzahlige Koordinatenwerte zu projizieren.

Nearest Neighbour: Die „Nearest Neighbour“ Interpolation ist in der Berechnung am effizientesten unter den Interpolationsverfahren. Der Intensitätswert des zu interpolierenden Bildpunktes wird von dem am nächsten (euklidischer Abstand) liegenden Bildpunkt übernommen. Im Grunde genommen werden dadurch Bildpunkte bei einer Bildverkleinerung entfernt und bei einer Bildvergrößerung ausgedehnt.

Bilineare Interpolation: Bei einer bilinearen Interpolation werden die vier nächsten Nachbarn des zu interpolierenden Bildpunktes berücksichtigt. Der Intensitätswert des Interpolationspunktes bildet sich aus dem gewichteten Mittelwert der vier Nachbarintensitätswerte.

Bikubische Interpolation: Den höchsten Berechnungsaufwand, aber auch die beste Qualität bietet die bikubische Interpolation. Es werden die Intensitätswerte eines 4×4 Umfeldes des zu interpolierenden Punktes genutzt. In den Intensitätswert, des zu interpolierenden Bildpunktes, fließen diese, zusätzlich entfernungsabhängig gewichtete, 16 Umgebungswerte ein.

4.2.2.2 Bewertung

Es stellt sich nun die Frage welche Auswirkungen eine Bildskalierung in Verbindung mit den jeweiligen Interpolationsverfahren auf die Erkennungsraten hat. Um dieser Frage nachzugehen wurden Testbilder der Datenkataloge 1 (Fahrzeuge im Videobild), 2 (Fahrzeuge im Wärmebild), 3 (Fußgänger im Videobild) und 4 (Fußgänger im Wärmebild) mittels einer bikubischen Skalierung, auf vorgegebene Bildgrößen $w' \times h'$ skaliert. Dadurch können beliebige Objekt-Kamera Entfernungsverhältnisse simuliert werden. Da vereinfachend von einem idealen Lochkamera-modell ausgegangen wird, werden die ursprünglichen Seitenverhältnisse $\frac{w}{h} = \frac{w'}{h'}$ beibehalten. Im Anschluss daran wurden die Bildbereiche, durch obige Interpolationsverfahren, auf die Bewertungsgröße $w \times h$ zurückskaliert (60×60 für Fahrzeuge und 64×128 für Fußgänger) und hinsichtlich der Erkennungsgüte überprüft (siehe Abbildung 4.4 für Fahrzeuge).

Grundsätzlich erfolgt keine Änderung der Erkennungsrate bei einer Verkleinerung ($w' > w \wedge h' > h$) der Testbereiche auf die initiale Trainingsgröße von $w \times h$ Bildpunkten, da durch den Skalierungsprozess keine relevante Information verloren geht. Wird jedoch ein kleinerer Bereich vergrößert brechen die Erkennungsraten, unabhängig von der Objekt-/Sensorklasse, ab einer Halbierung der Breiten- und Höhenabmessungen ein. Ein verzögerter Abfall der Erkennungsraten kann mit einer verkleinerten initialen Trainingsgröße erzielt werden. Gleichzeitig wird dadurch die maximale Erkennungsrate herabgesenkt (vgl. 20×20 Ergebnisse in Abbildung 4.4a).

Zur Bildbereichsvergrößerung erzielen die bilineare und bikubische Interpolation, im Vergleich der verschiedenen Interpolationsverfahren, nahezu identische Ergebnisse. Bei Bildbereichsverkleinerungen ist die Wahl des Interpolationsverfahrens unerheblich. Unter Miteinbeziehung der Rechenzeiten ist somit für Bildvergrößerungen eine bilineare Interpolation, und für Bildverkleinerungen die Nearest Neighbour Interpolation vorzuziehen.

Um einen Anhaltspunkt für entfernungsabhängige Erkennungsleistungen zu bekommen, können die skalierten Bildbereiche für ein typisches Sensorsystem (siehe auch Abschnitt 2.4.1) in Bezug zu den jeweiligen Entfernungswerten gesetzt werden. Anhand der durchschnittlichen Fahrzeug- bzw. Fußgängerbreite w_r in Metern, der Breite w_p des Fahrzeugs/Fußgängers in

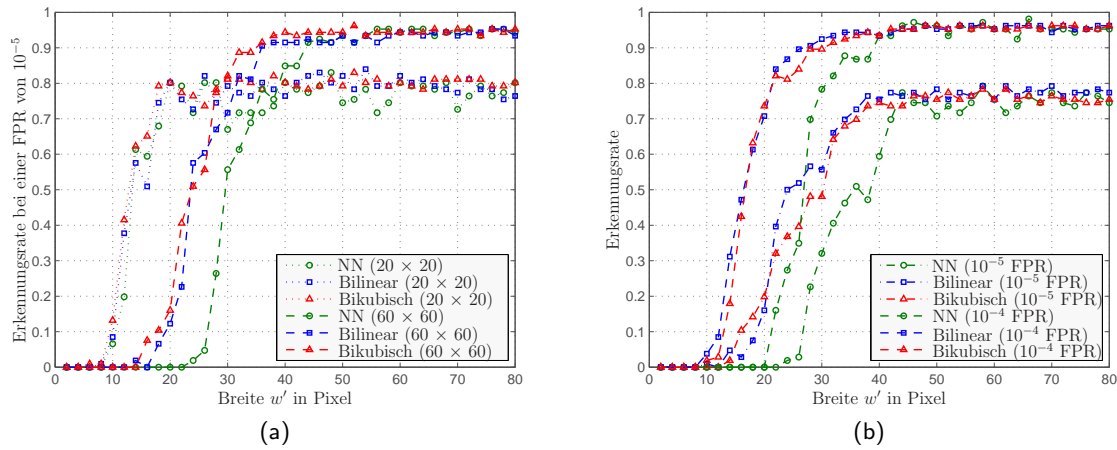


Abbildung 4.4: Vergleich der Erkennungsraten für PKW-Heckansichten im Videobild (a) und Wärmebild (b) bei verschiedenen Skalierungen $w' \times h'$ (entspricht einer Entfernungssimulation von der Kamera) mit Nearest Neighbour, bilinearen und bikubischen Interpolationsverfahren. Die Abmessungen der Trainingsdaten betrug, wenn nicht anders gekennzeichnet, 60×60 Bildpunkte.

Bildpunkten, des Öffnungswinkels ϕ_S und der horizontalen Pixelauflösung λ_S des bildgebenden Sensors sind approximierte Objektentfernungen

$$d = \frac{w_r}{\tan\left(\frac{\phi_S \frac{w_p}{2}}{\lambda_S}\right)}$$

ableitbar. Die Tabelle 4.2 setzt die Fahrzeug- und Fußgängerbreiten in Bezug zu den dazu entsprechenden Entfernungswerten, abhängig vom verwendeten bildgebenden Sensor. Zusätzlich sind die entsprechenden Erkennungsraten angegeben, die mittels der Trainingskataloge 1-4 (siehe auch Anhang A auf Seite 179) ermittelt wurden.

Tabelle 4.2: Entfernungsabhängige Erkennungsraten mit Gradientenhistogrammen für Fahrzeugheckansichten und Fußgänger im Wärme- und Videobild. Siehe Abschnitt 2.4.1 auf Seite 25 zu Sensordetails.

Breite in Bildpunkten	FIR Entfernung	Erkennungsraten (10^{-4} FPR)	Video	Video Tele	Erkennungsraten
Fahrzeugheckansichten (10^{-4} FPR)					
60	13 m	0.95	10.8 m	27 m	1.0
50	15.6 m	0.96	13 m	32.4 m	1.0
40	19.5 m	0.93	16.2 m	40.4 m	1.0

Tabelle 4.2: Entfernungsabhängige Erkennungsraten mit Gradientenhistogrammen für Fahrzeugheckansichten und Fußgänger im Wärme- und Videobild. Siehe Abschnitt 2.4.1 auf Seite 25 zu Sensordetails.

Breite in Bildpunkten	FIR Entfernung	Erkennungsrate (10^{-4} FPR)	Video	Video Tele	Erkennungsrate
30	26 m	0.93	21.6 m	53.9 m	0.99
25	31.2 m	0.82	25.9 m	64.7 m	0.82
20	39.1 m	0.71	32.4 m	80.9 m	0.59
15	39.1 m	0.32	52.1 m	107.8 m	0.17
10	78.1 m	0.04	64.9 m	161.8 m	0.0
Fußgänger (10^{-3} FPR)					
40	6.5 m	0.99	5.4 m	13.5 m	0.79
30	8.7 m	0.99	7.2 m	18.0 m	0.80
20	13.0 m	0.97	10.8 m	27.0 m	0.77
15	17.4 m	0.94	14.4 m	36.0 m	0.63
10	26.0 m	0.88	21.6 m	54.0 m	0.48
8	32.6 m	0.87	27.0 m	67.4 m	0.46
6	43.4 m	0.65	36.0 m	89.9 m	0.16
4	65.1 m	0.73	54.1 m	134.8 m	0.17

Mit den verbauten Videokameras ist bei einer FPR von 10^{-4} für Fahrzeuge eine Klassifikationsreichweite von ca. 26 m mit dem Normalobjektiv und von ca. 60 m mit dem Teleobjektiv möglich. Die Wärmebildkamera kann Fahrzeuge bis zu einer Entfernung von ca. 40 m robust klassifizieren. Für Fußgänger bricht die Erkennungsrate im Videobild ab einer Entfernung von ca. 20 m (54 m mit Teleobjektiv) und im Wärmebild ab einer Entfernung von 30 m ein.

4.2.3 Merkmalsrekonstruktion durch Achsenspiegelung

Dadurch dass die Histogrammberechnung über den gesamten zu verifizierenden Bildausschnitt durchgeführt werden muss, beschränkt sich die Merkmalsextraktion auf Objekte, die im bildgebenden Sensor vollständig sichtbar sind. Folglich verringern sich die horizontalen und vertikalen Öffnungswinkel des effektiven Sichtbereichs des Videosensors um die doppelten Bildausschnittsabmessungen. Im konkreten Fall der Fahrzeugheckerkennung müsste somit das gesamte Fahrzeugheck im Videosensor sichtbar sein, um eine Merkmalsextraktion auf Histogrammbasis ausführen zu können. Dadurch würden nützliche Informationen verworfen werden, die zu einer Erkennung von teilweise sichtbaren Fahrzeugen (z. B. Einscherern) beitragen könnten.

Eine Möglichkeit, diese Information dennoch zu nutzen, ist das Training einer auf Teilan-sichten basierenden Klassifikatorenschar. Für jeden möglichen Verdeckungsgrad an den Sen-

sorgrenzen bzw. bei Objekt-Objektverdeckungen muss ein passender Komponentenklassifikator trainiert werden, der im Erkennungsfall selektiert wird und eine entsprechende Bewertung abgeben kann. Ein Nachteil dieser Vorgehensweise ist der erhöhte Verwaltungs- und Trainingsaufwand, da für jedes Objekt eine Vielzahl an Komponentenklassifikatoren vorgehalten werden muss.

Ein wesentlich einfacheres Verfahren, das jedoch ausschließlich bei der Erkennung von symmetrischen Objekten angewendet werden kann, ist die Ergänzung von nicht sichtbaren Informationen mittels Spiegelung. Der verdeckte bzw. außerhalb des Sensorsichtbereichs liegende Teilabschnitt ist vollständig rekonstruierbar, sofern der Anteil der Verdeckung und die Lage der Symmetrieachse bekannt sind, und von der Verdeckung nicht einander entsprechende redundante Spiegelbereiche betroffen sind. Insbesondere digitalisierte Fahrzeughecks eignen sich für eine künstliche spiegelungsbasierte Rekonstruktion, da sie in der Regel eine konstruktionsbedingte vertikale Achsensymmetrie aufweisen. Die Lage der Symmetrieachse wird in den folgenden Ausführungen durch die Hypothesenerzeugung anhand des Laserscanners ermittelt (siehe auch Abschnitt 3.1.3 auf Seite 3.1.3).

4.2.3.1 Umsetzung

Zunächst wird für eine quaderförmige Hypothese H_k ermittelt, ob eine Spiegelung durchgeführt werden muss. Das ist der Fall wenn sich die rektifizierte Rückfront $\tilde{\mathcal{H}}_k^S$ in Sensorkoordinaten, mit den Sensorsichtgrenzen schneidet (siehe auch Abschnitt 4.1 und Abbildung 4.5a). Da nur vertikale Spiegelungen von Interesse sind, müssen somit entweder die Eckpunkte $\tilde{\mathbf{h}}_1^S$ und $\tilde{\mathbf{h}}_4^S$ oder die Eckpunkte $\tilde{\mathbf{h}}_2^S$ und $\tilde{\mathbf{h}}_3^S$ außerhalb der Bildebene liegen. Zur Vereinfachung der Schreibweise werden im Folgenden die Eckpunkte ohne Koordinatensystemsymbol und Akzentuierung notiert.

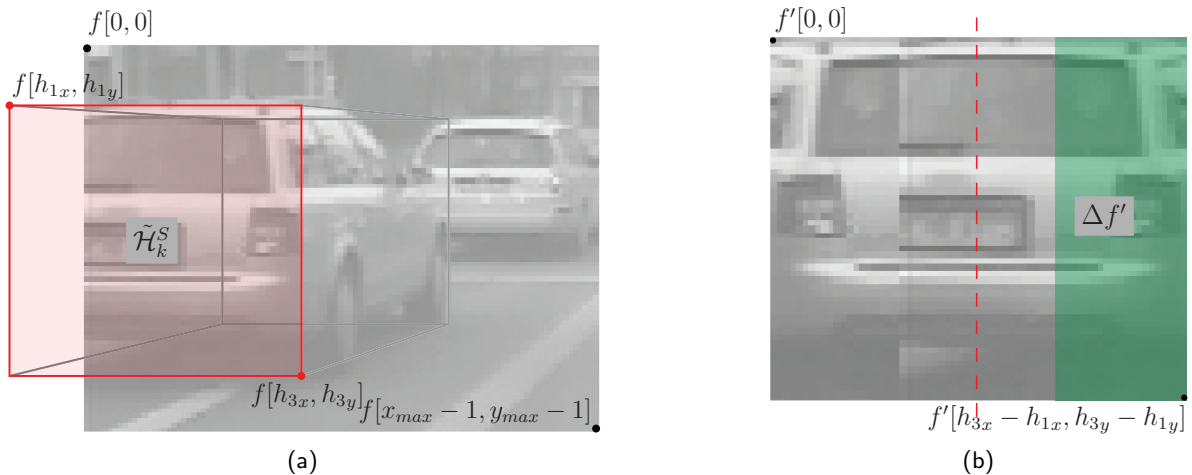


Abbildung 4.5: Bildmerkmalsergänzung durch Achsenspiegelung. Abbildung (a) zeigt die Projektion einer quaderförmigen Fahrzeughypothese in ein Kamerabild, wobei das Fahrzeugheck nicht vollständig sichtbar ist. Abbildung (b) zeigt die mittels Symmetrieannahmen vervollständigte Heckansicht durch eine Achsenspiegelung des hinterlegten Bereichs $\Delta f'$ an der geschätzten Symmetrieachse (gestrichelte Linie).

Gilt $(h_{1x} < 0) \wedge (h_{3x} > 0) \wedge (h_{1y} \geq 0) \wedge (h_{3y} < y_{max})$ so liegt ein linksseitiger Verdeckungsfall vor. Bei $(h_{1x} < x_{max}) \wedge (h_{3x} > x_{max}) \wedge (h_{1y} \geq 0) \wedge (h_{3y} < y_{max})$ befindet sich der rechte Ausschnitt der Rückfront $\tilde{\mathcal{H}}_k^S$ außerhalb des Sensorsichtbereichs (Pixelkoordinatenprojektionen jenseits des Bildbereichs werden auf Zahlen < 0 bzw. $> x_{max}, y_{max}$ abgebildet). In den nachfolgenden Ausführungen wird ohne Beschränkung der Allgemeinheit auf den linksseitigen Verdeckungsfall eingegangen. Rekonstruktionen von rechtsseitigen Verdeckungen können analog zum linksseitigen Fall berechnet werden. Bevor die Spiegelung durchgeführt wird, ist zunächst der Verdeckungsgrad

$$\tau = \begin{cases} \frac{-h_{1x}}{h_{3x}} & h_{1x} < 0 \\ 0 & \text{sonst} \end{cases}$$

von Interesse. Für $0 < \tau \leq 0.5$ wird die Symmetrieachse nicht verdeckt und eine vollständige Reproduktion $f'[x', y']$ des Bildausschnitts $[h_{1x}, h_{1y}], [0, h_{3y}]$ ist möglich (siehe auch Abbildung 4.5b):

$$\forall (0 < i < h_{3x} - h_{1x}) \wedge (0 < j < h_{3y} - h_{1y}) \quad f'[i, j] = \begin{cases} f[h_{1x} + i, h_{1y} + j] & i > |x_1| \\ f[h_{3x} - i, h_{1y} + j] & \text{sonst} \end{cases}$$

Im Fall $\tau > 0.5$ ist keine vollständige Rekonstruktion mehr möglich, da Objektbereiche auf beiden Seiten der vertikalen Symmetrieachse verdeckt werden. Es hat sich jedoch gezeigt, dass eine Spiegelung des verbliebenen Ausschnitts zu Erkennungszwecken oftmals genügt. Die Schnittachse fungiert dann als Symmetrieachse:

$$\forall (0 < i < 2 \cdot x_2) \wedge (0 < j < y_2 - y_1) \quad f'[i, j] = \begin{cases} f[-h_{3x} + i, h_{1y} + j] & i > x_2 \\ f[h_{3x} - i, h_{1y} + j] & \text{sonst} \end{cases}$$

4.2.3.2 Bewertung

Es ist zu erwarten, dass der beschriebene Spiegelungsprozess Auswirkungen auf die Erkennungsraten hat. Zum einen kann die Rekonstruktion von schiefperspektivischen Ansichten vertikale Phantomkanten an der Verknüpfungsgeraden verursachen (siehe Abbildungen 4.6a und 4.6d). Zum anderen nimmt der Ähnlichkeitsgrad von Fahrzeug- und Hintergrundklasse zu, da asymmetrische Hintergrundobjekte gleichermaßen von der Spiegelung betroffen sind.

Um quantitativ zu überprüfen inwiefern sich diese Effekte bemerkbar machen, wurde ein Erkennungsratenvergleich von ungespiegelten ($\tau = 0$) und gespiegelten ($\tau \geq 0$) Heckansichten vorgenommen. Zum Training wurden die Hälfte des Datensatzes von Katalog 1 und 2 (siehe auch Anhang A auf Seite 179) mit je 750 Heckansichten und 3000 Hintergrundbildern sowohl im Video- als auch Wärmebild verwendet. Auf Basis von diesen Bilddaten wurden für die Video- und Wärmebilddaten je eine SVM erzeugt. Zum Testen der SVM Erkennungsraten wurden eine vom Trainingssatz disjunkte Menge von je 750 Heckansichten und 3000 Hintergrundansichten herangezogen. Dieses Testset wurde gezielt verändert, indem Spiegelungen der Heckansichten mit zufällig gewähltem Verdeckungsgrad τ in den Intervallen $]0, 0.5]$, $]0.5, 0.75]$ und $]0.75, 1.0[$ durchgeführt wurden (siehe Abbildung 4.6).

Steigt der Verdeckungsgrad τ , sinkt gleichzeitig der ursprüngliche Informationsgehalt. Als Konsequenz nehmen die Erkennungsraten ab (siehe Abbildung 4.7). Im Videobild ist für $\tau < 0.5$ die Abnahme der Erkennungsraten im Vergleich zum ungespiegelten Referenzkatalog

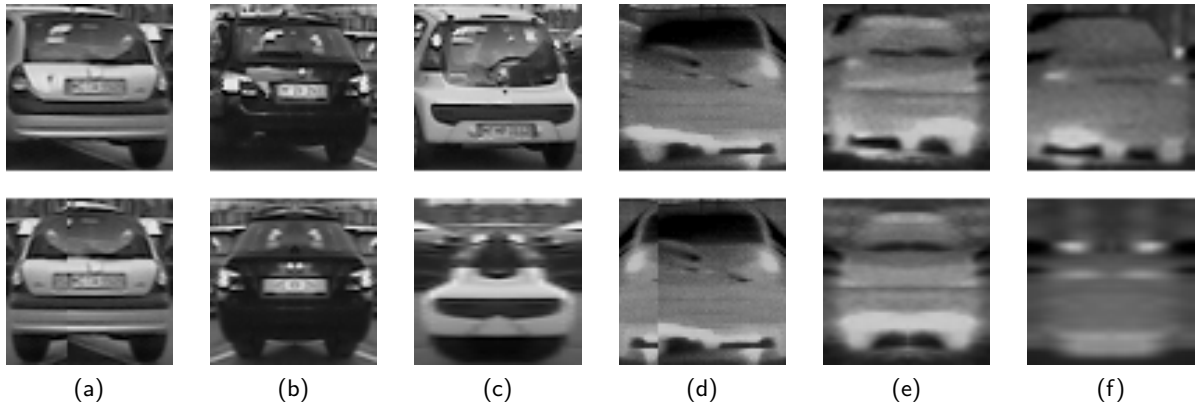


Abbildung 4.6: Beispielsichten von gespiegelten Fahrzeughecks. Die obere Reihe zeigt die Originalansichten von Fahrzeughecks und die untere Reihe die Spiegelungen mit $0 < \tau \leq 0.5$ für (a) bzw. (d), $0.5 < \tau \leq 0.75$ für (b) und (e) und $0.75 < \tau < 1.0$ für (c) und (f). Abbildung (a), (b) und (c) wurden mit einer CCD-Videokamera aufgenommen. Die Abbildungen (d), (e) und (f) entstammen aus einem Wärmebild.

($\tau = 0$) gering (bei einer Fehlerkennungsrate von 10^{-4} von 99.1 % um ca. 5 % auf 95.2 %). Im Wärmebild sind die Auswirkungen stärker (ca. 12 % von 90.7 % auf 78.3 %), da der Symmetriegrad der Heckansichten im Vergleich zum Videobild nicht so ausgeprägt ist (vgl. beispielsweise die asymmetrisch angebrachte Auspuffanlage in Abbildung 4.6d und 4.6f).

Ist keine vollständige Rekonstruktion des Fahrzeughecks mehr möglich ($0.5 < \tau < 0.75$), sinkt die Erkennungsrate im Videobild um weitere 25 %, auf einen Wert von 73.6 %. Im Wärmebild ist lediglich ein geringer Abfall der Erkennungsraten zu beobachten. Grund dafür könnte sein, dass der Informationsgehalt im äußeren Bereich des Fahrzeughecks am höchsten ist (z. B. warme Reifen und Rücklichter) und durch eine Spiegelung mit $0.5 < \tau < 0.75$ nicht tangiert wird (siehe auch Abbildung 4.6e). Ein signifikanter Einbruch der Erkennungsraten auf 19 % im Videobild und auf 17 % im Wärmebild zeigt sich erst, wenn weniger als ein Viertel ($\tau > 0.75$) des Fahrzeughecks sichtbar ist.

Unter der Voraussetzung, dass symmetrische Objekte wie Fahrzeughecks detektiert werden sollen, und die Abmessungen des Objektes durch andere Sensorik bekannt sind, kann durch die vorgestellte Achsenspiegelung fehlende Information ergänzt werden. Bis zu einem Verdeckungsgrad von 75 % sind nur geringe Einbrüche der Erkennungsraten zu beobachten. Vor allem im Nahbereich kann dadurch der effektive Klassifikationsbereich des Kamerasystems erheblich vergrößert werden. Bei einer angenommenen Fahrzeugbreite von 1.5 m vergrößert sich der effektive horizontale Sichtbereich des bildgebenden Sensors um $\pm 15^\circ$. Prinzipiell ist das Verfahren gleichermaßen für alle symmetrischen Objektklassen anwendbar, wie beispielsweise Lastkraftwagen. Bei rigidem Objekten, wie z. B. Fußgängern, ist eine Durchführung jedoch nicht sinnvoll. Voraussetzung ist stets ein weiterer Sensor/Verfahren, wodurch die Lage der Symmetrieachse bestimmt werden kann.

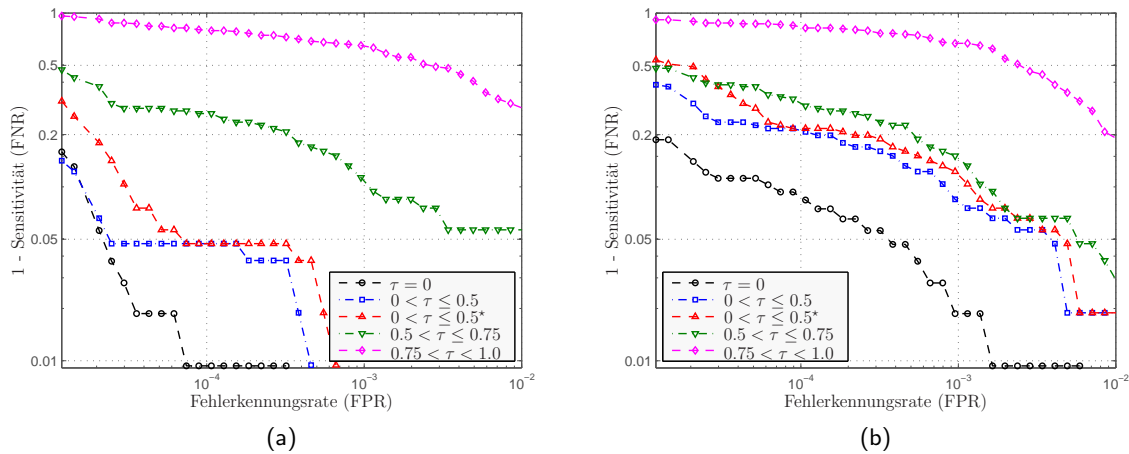


Abbildung 4.7: DET-Kurven zum Vergleich der Erkennungsraten von ungepiegelten ($\tau = 0$) mit gespiegelten ($\tau > 0$) Fahrzeughecks im Video- (a) und im Wärmebild (b). Bei den DET-Kurven $0 < \tau \leq 0.5^*$ wurden, zusätzlich zu den Heckansichten, alle 20 000 Hintergrundansichten im entsprechenden Intervall gespiegelt. Dadurch steigt der Ähnlichkeitsgrad zwischen Hintergrund- und symmetrischer Vordergrundklasse, und die Fehlerkennungsrate steigt. Die quantitativen Auswirkungen sind jedoch gering (0.5 % Erkennungsratenabfall im Videobild und 1 % im Wärmebild) und zu vernachlässigen.

4.3 Reflektanzsignatur von Fahrzeugheckansichten

Ein Laserscanner liefert in der Regel nicht nur Entfernungsinformationen, sondern, über die Messung der Rückstreuintensitäten, ebenso die Reflektanzverteilung von erfassten Objekten. Die Stärke der Rückstreuung ist, neben der Oberflächenbeschaffenheit (die definiert welche Anteile des eintreffenden Lichts absorbiert, reflektiert und diffus gestreut werden) des jeweiligen Objekts, im Wesentlichen von der Sendeleistung des Lasers, der Objektentfernung und dem Einfallswinkel abhängig. Sind die letzten drei Parameter bekannt, kann über die Laser-Radar-Gleichung [Sko90] auf die Objektreflektanz geschlossen werden.

Um Verkehrsteilnehmer, Verkehrsschilder und Leitpfosten bei Dunkelheit besser kenntlich zu machen, werden diese Objekte mit Reflektorfolie oder Rückstrahlern versehen. Diese sogenannten Retroreflektoren weisen besondere Eigenschaften hinsichtlich ihres Reflexionsverhaltens auf. Durch die Anordnung vieler kleiner dreiflächiger Winkelreflektoren, totalreflektierender Prismen oder winziger Glaskugeln wird ein großer Anteil des eintreffenden Lichts in die ursprüngliche Ausgangsrichtung zurückgeworfen. Retroreflektoren im Straßenverkehr reflektieren das Licht nicht exakt mit dem Einfallswinkel. Vielmehr wird durch gezielte Fertigungsvarianzen eine angemessene Streuwirkung erzielt, um dadurch den Winkel zwischen dem zurückgeworfenen Licht der Fahrzeugscheinwerfer und den Augen des Fahrers zu überbrücken. In der Näherung kann ein Retroreflektor trotz dieser minimalen Streueigenschaften als optischer Spiegel angesehen werden. Im Vergleich zu anderen Materialien liefern somit Rückstrahler und Reflektorfolien unabhängig vom Einfallswinkel die größte Rückstreuleistung [FD04].

Diese Eigenschaft kann insbesondere bei Fahrzeugheckansichten als ein Erkennungsmerkmal genutzt werden, da sowohl die Abmessungen als auch die Anbauorte von einem retroreflektieren-

den amtlichen Kennzeichen und den Rückstrahlern gesetzlich vorgeschrieben sind. So müssen nach § 53 Absatz (4) der Straßenverkehrszulassungsordnung (StVZO) [KW94] „Kraftfahrzeuge an der Rückseite mit zwei roten Rückstrahlern ausgerüstet sein. Die Seitenlänge solcher Rückstrahler muss mindestens 150 mm betragen. Der äußerste Punkt der leuchtenden Fläche der Rückstrahler darf nicht mehr als 400 mm vom äußersten Punkt des Fahrzeugumrisses, und ihr höchster Punkt, der leuchtenden Fläche, nicht mehr als 900 mm von der Fahrbahn entfernt sein“. Nach der Fahrzeug-Zulassungsverordnung (FZV), die als Bundesrechtsverordnung zulassungstechnische Teile der StVZO ablöst, und die Zulassung von Fahrzeugen zum öffentlichen Straßenverkehr normiert, „haben einzeilige Kennzeichen eine maximale Abmessung von 520 mm in der Breite und 110 mm in der Höhe“. Abgesehen von diversen Ausnahmeregelungen bewirken diese Vorschriften ein einheitliches Erscheinungsbild von Personenkraftwagen, in Bezug auf amtliches Kennzeichen und Rückstrahler, in Deutschland. Aber auch im Ausland befinden sich die Rückstrahler meist an der äußersten Fahrzeugbegrenzung, und gewöhnlich wird das Nummernschild vertikal zentriert angebracht.

Umsetzung

Ziel der nachfolgenden Merkmalsextraktion ist es, die relative Lage und Ausdehnung der schwach bzw. stark reflektierenden Bereiche eines Fahrzeughecks, entfernungs- und rotationsinvariant zu bestimmen. Ausgehend von einer quaderförmigen Hypothese H_k wird die, dem Sensor zugewandte Seite, in äquidistante horizontale Intervalle unterteilt. In jedem dieser Intervalle, die auch als Klasse eines Histogramms aufgefasst werden können, wird die maximale Rückstreuintensität ermittelt (siehe Abbildung 4.8a). Aufgrund der Retroreflektoranordnungen auf dem Fahrzeugheck ergibt sich eine charakteristische Signatur von Intervallen mit höherer bzw. niedrigerer Reflexionsintensität. Diese spezifische Intervallverteilung, die Hintergrundobjekte in der Regel nicht aufweisen, kann als ein kennzeichnendes Merkmal genutzt werden (siehe auch Abbildung 4.9a).

Die Berechnungen zur Erzeugung des Merkmalsvektors werden im Koordinatensystem der Fahrzeughypothese vorgenommen (siehe auch Abbildung 4.1). Zunächst werden somit die Messdaten \mathbf{x}_i^S des Laserscanners transformiert ($\mathbf{x}_i^{H_k} \leftarrow \mathbf{T}_{S \rightarrow H_k} \mathbf{x}_i^S$) und anschließend orthogonal auf die Rückseite der Hypothese projiziert: $(x_1 \ x_2 \ x_3 \ 1)^T \rightarrow \tilde{x}_i = x_2$. Messwerte $\mathbf{x}_i^{H_k}$, die eine zu große Distanz zur Hypothesenrückseite $\mathcal{H}_k^{H_k}$ aufweisen werden ignoriert (Nearest Neighbour). Nach der Unterteilung der Rückseite in ρ gleichgroße vertikale Intervalle $K_j = \left[(j-1) \frac{|h_{2y}^{H_k} - h_{1y}^{H_k}|}{\rho}, j \frac{|h_{2y}^{H_k} - h_{1y}^{H_k}|}{\rho} \right]$, für $j \in \{1, \dots, \rho\}$, werden die Rückstreuintensitäten $I(\mathbf{x}_i^{H_k})$ der Messdaten den Intervallen zugeordnet:

$$\forall i \quad K(\tilde{x}_i) \leftarrow \max \left(K(\tilde{x}_i), I(\mathbf{x}_i^{H_k}) \right)$$

Die Anzahl der Intervalle ρ ist ein ausschlaggebender Parameter. Wird ρ zu hoch gewählt, bleiben einzelne Intervalle unbesetzt und verfälschen die extrahierte Signatur. Bei zu niedrigem ρ sinkt die Auflösung des Merkmalsvektors. Aus diesem Grund wird ρ in der Art und Weise gewählt, dass auf jedes Intervall zumindest ein Messwert bzw. Intensitätswert fällt. Die Wahl von ρ ist somit abhängig von der Winkelauflösung γ_{S_l} des Laserscanners und von dem Winkel α der vom Sensor überstrichenen Fahrzeugrückseite. Dieser Winkel $\alpha = \arccos \left(\left\langle \mathbf{h}_1^{S_l}, \mathbf{h}_2^{S_l} \right\rangle \left(|\mathbf{h}_2^{S_l}| \cdot |\mathbf{h}_1^{S_l}| \right)^{-1} \right)$ liegt zwischen den beiden Eckpunkten $\mathbf{h}_1^{S_l}$ und $\mathbf{h}_2^{S_l}$ der

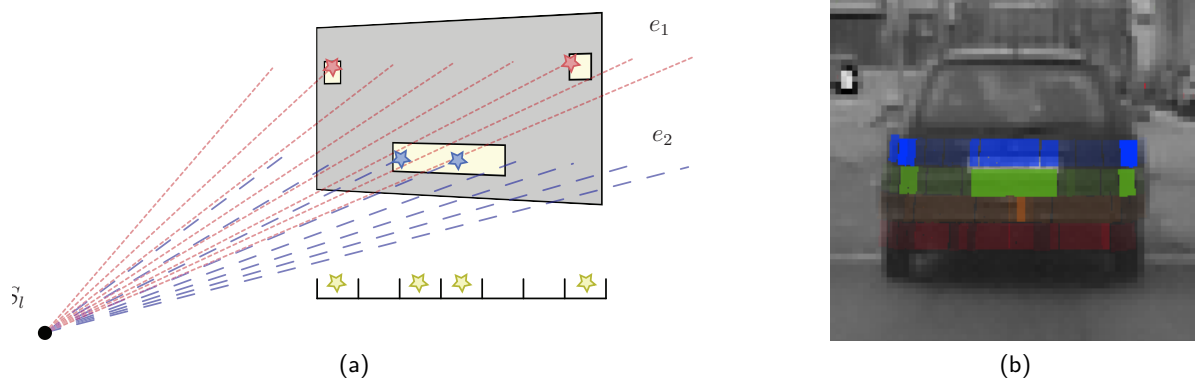


Abbildung 4.8: Abbildung (a) zeigt die Skizze eines Fahrzeughecks mit Nummernschild und zwei Rückstrahlern. Die Retroreflektoren werden durch einen Laserscanner S_l mit zwei Strahlebenen e_1 und e_2 detektiert (Sternsymbole). Die Rückstreuintensitäten werden anhand des horizontalen Öffnungswinkels in ein Intervall einsortiert. In Abbildung (b) ist das Kamerabild eines Fahrzeughecks, das mit den Daten eines vierzeiligen Laserscanners überlagert wurde, zu sehen. Die Stärke der Reflexionen ist durch den Transparenzgrad des jeweiligen Farbwertes visualisiert.

Hypothesenrückseite. Damit Unter- bzw. Überabtastungen vermieden werden, sollte $\rho < \frac{\alpha}{\gamma_{S_l}}$ gelten. Abschließend werden die Intervalle K_i durch lineare Interpolation auf eine einheitliche Merkmalsvektorenlänge $|\mathbf{x}_R| = 10$ skaliert.

Bewertung

Durch die Reflektanzsignatur können Fahrzeugheckansichten von Hintergrundobjekten unterschieden werden. Die erzielten Erkennungsraten liegen bei ca. 80 % (FPR von 10^{-2}), sofern eine SVM mit nichtlinearem Kernel verwendet wird (vgl. Abbildung 4.9b). Zum Training der SVM wurde Katalog 5 (siehe auch Anhang A auf Seite 179) herangezogen. Erkennungsprobleme ergeben sich zum einen bei mangelhafter Merkmalsgewinnung, die durch eine ungenaue Segmentierung oder falsche Trackingdaten verursacht wird. Zum anderen treten Fehlerkennungen gehäuft an retroreflektierenden Verkehrsschildern auf, wenn aufgrund von Verschmutzungen die Reflektorverteilung von Fahrzeugheckansichten nachgebildet wird.

Abhängig vom Verbauort des Laserscanners am Versuchsfahrzeug kann die Reflektanzsignatur von Fahrzeughecks nur in einem gewissen Entfernungsbereich extrahiert werden. Ist das Fahrzeug zu weit entfernt, verringert sich die horizontale Abtastung des Fahrzeugs. Ist der Abstand zum Fahrzeug zu klein, werden die retroreflektierenden Bereiche womöglich nicht überstrichen. Bei dem verwendeten Versuchsträger (siehe Abschnitt 2.4) liegt die Verbauposition des Laserscanners 20 cm über dem Boden und der vertikale Öffnungswinkel bei 3.2° . Dadurch ergibt sich, bei einer durchschnittlichen Höhe von Nummernschild und Rückleuchten von 0.7 m, ein Extraktionsintervall zwischen 10 und 60 m.

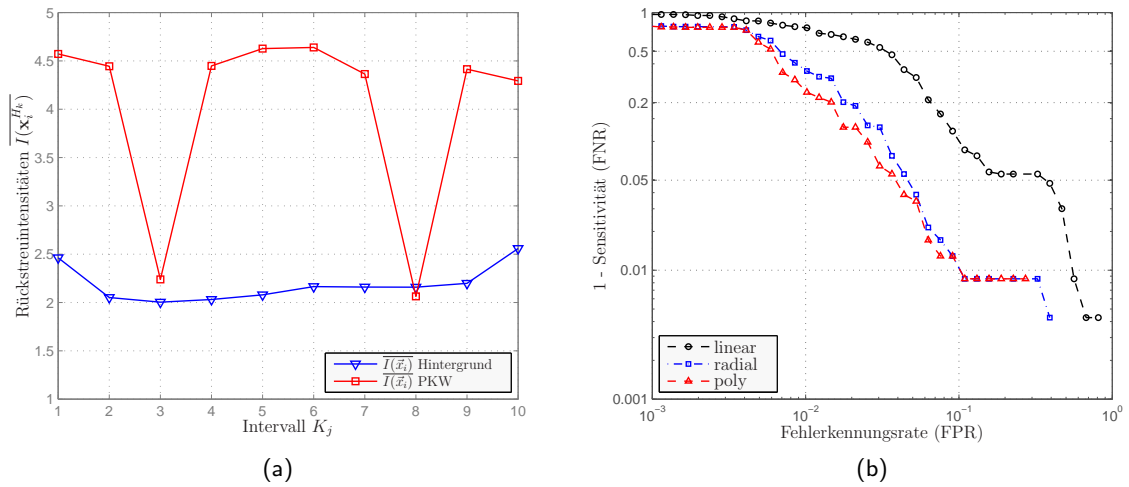


Abbildung 4.9: Durch einen Laserscanner ermittelte Durchschnittsreflexionsstärken $\overline{I(\mathbf{x}_i^H)}$ an 10 horizontalen Intervallen K_j von 600 Fahrzeugheckansichten und 700 Hintergrundobjekten. Die Retroreflektoren (Nummernschild und Rückleuchten) sind bei Fahrzeugen deutlich sichtbar (a). Abbildung (b) zeigt die Erkennungsraten von Fahrzeugheckansichten anhand der Reflektanzsignatur. Die SVM wurden mit linearem, radialem und polynomiellen Kernel trainiert.

4.4 Bein-Pendelbewegung von Fußgängern

Gehende oder laufende Fußgänger weisen regelmäßige Beinbewegungen auf. Diese sogenannte Pendelbewegung kann mit Sensoren erfasst und als charakteristisches Merkmal für einen Fußgänger verwendet werden. Nach einer Ganganalyse [Whi96] von Männern und Frauen zwischen 13 und 80 Jahren macht eine Person, in normaler Gangart, zwischen 81 und 150 Schritte in der Minute. Dies resultiert in einer Schrittfrequenz f_S , die zwischen 1.35 Hz und 2.5 Hz liegt. In der Literatur finden sich zahlreiche Ansätze, diese kennzeichnende Pendelbewegung im Kamerabild, über oszillierende Bereiche [LB98, LP98, FSWG06] oder Modellannahmen [Gav99, RZW⁺05, EBMM03], zu extrahieren. Speziell für nichtstationäre Kamerasysteme und somit vor wechselndem Hintergrund sind diese Herangehensweisen oftmals schwierig umzusetzen.

Wesentlich effektiver lässt sich die Pendelbewegung aus Laserscannerdaten gewinnen (siehe Abbildung 4.10). Einige Ansätze [FLD03, TK04, SZN⁺06] basieren auf der Segmentierung der beiden Beine eines Fußgängers. Eine Separierung ist jedoch aufgrund der Beinstellung und der begrenzten Winkelauflösung des Laserscanners nicht durchgängig gewährleistet. Alternativ dazu kann die Ausdehnungsänderung des gesamten Fußgängers längs seiner Bewegungsrichtung bestimmt werden [Tho05], da die Beine bei Gehbewegungen über den Rumpf hinausragen.

Umsetzung

Die Berechnungen dazu werden in Weltkoordinaten und im \mathbb{R}^2 in der xy -Ebene (Straßenebene) durchgeführt. Nach der Aggregation der Laserscannerrohdaten $\mathbf{x}_{i,j}^W$ in Zusammenhangsbereiche s_j^W (siehe auch Abschnitt 3.1.1.2 auf Seite 37), werden die homogenen Vektoren $\mathbf{x}_{i,j}^W$

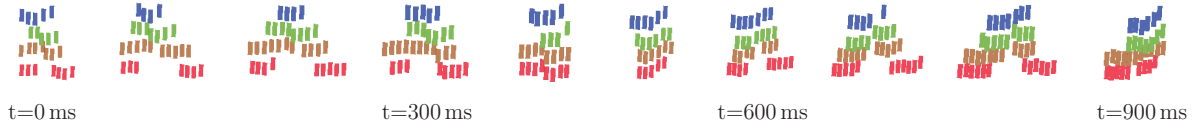


Abbildung 4.10: Über die Zeit t gewonnene Laserscannerrohdaten eines querenden Fußgängers während einer Annäherungsfahrt. Die Schrittbewegungen sind deutlich zu erkennen.

zunächst auf die xy -Ebene projiziert ($(x_1 \ x_2 \ x_3 \ 1)_{i,j}^T \rightarrow (x_1 \ x_2)_{i,j}^T$). Über eine zeitliche Beobachtung (Tracking) der Zusammenhangskomponenten kann der Bewegungsvektor $\mathbf{d}_j \in \mathbb{R}^2$ abgeleitet werden. Die Folgebetrachtungen werden nur durchgeführt, sofern eine signifikante Geschwindigkeitskomponente ($|\mathbf{d}_j| > 1.5 \text{ km/h}$) messbar ist. Unter der Annahme, dass das Maxima der Beinbewegung eines Fußgängers längs seiner Bewegungsrichtung \mathbf{d}_j erfolgt, ist über die Orthogonalprojektion der Laserscannerrohdaten auf die Gerade $\lambda \mathbf{d}_j$, die jeweilige Schrittweite $\Delta f_{t,j}$ zum Zeitpunkt t für die Zusammenhangskomponente s_j^W bestimmbar:

$$\forall s_j \quad \Delta f_{t,j} = \max \left(\frac{\langle \mathbf{d}_j, \mathbf{x}_{1,j} \rangle}{|\mathbf{d}_j|}, \dots, \frac{\langle \mathbf{d}_j, \mathbf{x}_{i,j} \rangle}{|\mathbf{d}_j|} \right) - \min \left(\frac{\langle \mathbf{d}_j, \mathbf{x}_{1,j} \rangle}{|\mathbf{d}_j|}, \dots, \frac{\langle \mathbf{d}_j, \mathbf{x}_{i,j} \rangle}{|\mathbf{d}_j|} \right)$$

Werden die mittelwertbereinigten Schrittweiten $[\Delta f_1, \dots, \Delta f_{2n}]$ über ein Zeitfenster $2n$ beobachtet, kann mittels einer diskreten Fouriertransformation das Frequenzspektrum von s_j ermittelt werden.

$$k_l = \sum_{k=0}^{2n-1} (\Delta f_{k,j} - \overline{\Delta f}) e^{-\frac{2\pi i}{2n} jk} \quad j = 0, \dots, 2n - 1.$$

Um das Ursprungssignal verlustfrei rekonstruieren zu können, muss nach dem Nyquist-Shannon-Abtasttheorem $f_{\text{abtast}} > 2 f_{\text{max}}$ gelten. Eine Laserscannerabtastrate von $f_{\text{abtast}} = 10 \text{ Hz}$ ist somit ausreichend zur Detektion der Pendelbewegung, da Frequenzbereiche bis $f_{\text{max}} = 5 \text{ Hz}$ rekonstruiert werden sollen. Eine Unterscheidung zwischen Fußgängern und Hintergrundobjekten kann schließlich über eine Maximabestimmung im Frequenzbereich getroffen werden (siehe Abbildung 4.11).

Bewertung

Die Bestimmung der Erkennungsraten wurde auf dem Datenkatalog 6 (siehe auch Anhang A auf Seite 179) durchgeführt. Dieser wurde aus 80 Annäherungsfahrten auf einzelne querende oder in Längsrichtung gehende/laufende Fußgänger erzeugt. Zudem sind fußgängerfreie Stadt- und Überlandfahrten bei Nacht im Katalog enthalten. Während dieser Fahrten werden 21 800 Fußgängerhypothesen H_z aufgezeichnet. Davon kann aber lediglich eine Teilmenge von 72 Hypothesen zur Bestimmung der Fehlerkennungsrate herangezogen werden, da die Pendelbewegung nur bestimmbar ist, wenn die Hypothese eine Minimalgeschwindigkeit \mathbf{d}_{min} von 1.5 km/h aufweist und eine Beobachtungsdauer von mindestens 1.6 s ($\Delta t = 16$) eingehalten wird. Basierend auf diesem Testkatalog wurde eine Erkennungsrate von 98% und eine Fehlerkennungsrate von 20% ermittelt, wobei die Fehlerkennungsrate bezüglich der 72 Hypothesen bestimmt wurde.

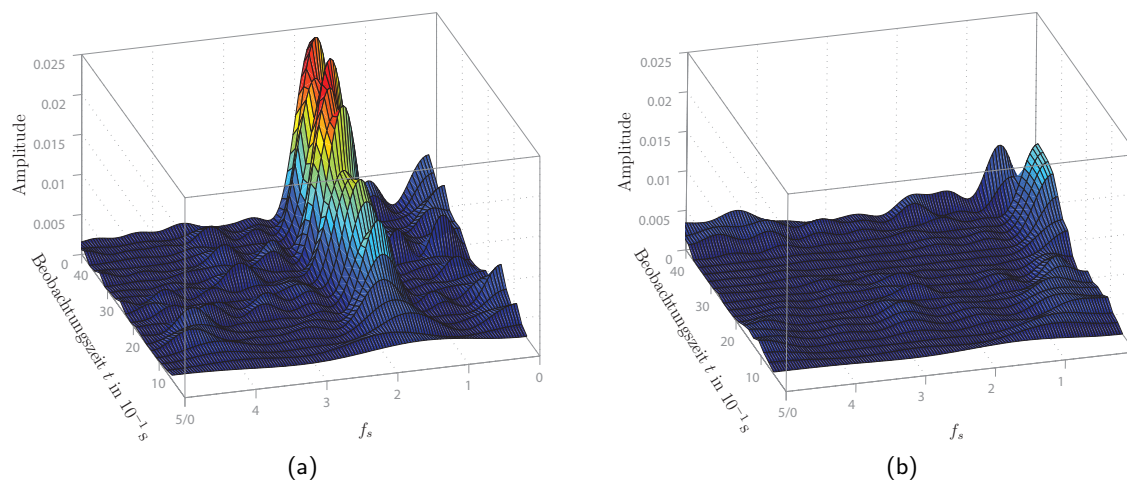


Abbildung 4.11: Frequenzanalyse der Ausdehnungsänderungen von Laserscannerrohdaten bei einem Fußgängerobjekt (a) und einem Hintergrundobjekt (b). In Abbildung (a) ist mit steigender Beobachtungszeit eine deutliche Pendelbewegung bei ca. 1.7 Hz zu erkennen. Das Frequenzspektrum des Hintergrundobjektes weist abgesehen von einem Maximum bei der Frequenz 0 (Gleichanteil) nur Rauschen auf.

Für eine schnelle zeitliche Reaktion auf Fußgänger eignet sich dieses Merkmal nur bedingt, da die minimale Beobachtungszeit zur Bestimmung des Frequenzspektrums eingehalten werden sollte. Darüber hinaus muss der Fußgänger in Bewegung sein, damit eine Pendelbewegung ableitbar ist. Fußgänger, die sich langsam oder schnell fortbewegen, werden durch das Verfahren robust erkannt. Gleichzeitig ist jedoch die Fehlerkennungsrate hoch.

4.5 Merkmale aus der Kursverfolgung

Nachdem virtuelle Objekte durch die Hypothesengenerierung erzeugt wurden, versuchen Trackingverfahren die neu initialisierten Hypothesen H_k bzw. H_z in nachfolgenden Zeitschritten wiederzufinden. Für jedes Objekt wird eine Spur⁶ angelegt, indem neben dem aktuellen Objektzustand zusätzlich zeitliche und spezifische Zustandsdaten verwaltet werden. Basierend auf der Objektdynamik wird ein zukünftiger Objektzustand berechnet, um darauf aufbauend eine Messwertprädiktion vorzunehmen. Um neue Messwerte der richtigen Vorhersage zuzuordnen, werden über die Differenzen von Position, Ausdehnung und Orientierung Ähnlichkeitsmaße definiert. Liegen diese Werte im Rahmen der zugrundeliegenden Objektdynamik, wird der Messwert der jeweiligen Spur zugeordnet und der Objektzustand messwertgetrieben gefiltert und aktualisiert. Kann keine Messwertzuordnung erfolgen, vermindert sich ein spurspezifisches Zuordnungsmass χ . Unterschreitet dieses einen Schwellwert γ , wird die entsprechende Spur entfernt. Die für das Tracking erforderlichen Filterverfahren, wie Kalman- und Partikelfilter, werden in der Dissertation von Rudi Lindl [Lin08] detailliert betrachtet.

⁶ engl. Track

Die Objektgeschwindigkeit und das -alter stellen aussagekräftige Merkmale für einen Klassifikator dar. Beide können durch Trackingverfahren abgeleitet werden. Bei Objekten mit hohem Altersindex t (Zeitspanne die seit der Trackinstantiierung vergangen ist) und einer signifikanten Geschwindigkeit v handelt es sich mit hoher Wahrscheinlichkeit um Fahrzeuge. Für stationäre Objekte, sowie für Objekte mit einem geringen Altersindex, kann keine gesicherte Aussage über die Klassenzugehörigkeit getroffen werden. Die Abbildungen 4.12a und 4.12b quantifizieren diese Beobachtungen. Die beiden Streudiagramme zeigen die Abhängigkeiten der Merkmale Geschwindigkeit und Altersindex von Fahrzeugen und Nichtfahrzeugen aus mehreren Aufnahmesequenzen. In diesen sind sowohl Standsequenzen als auch Stadt- und Autobahnfahrten enthalten. Negative Geschwindigkeitswerte sind das Resultat von entgegenkommenden Fahrzeugen. Es ist deutlich zu erkennen, dass Fahrzeuge den gesamten zweidimensionalen Wertebereich ausfüllen, wohingegen Nichtfahrzeuge mit zunehmendem Altersindex geringere Geschwindigkeitswerte aufweisen.

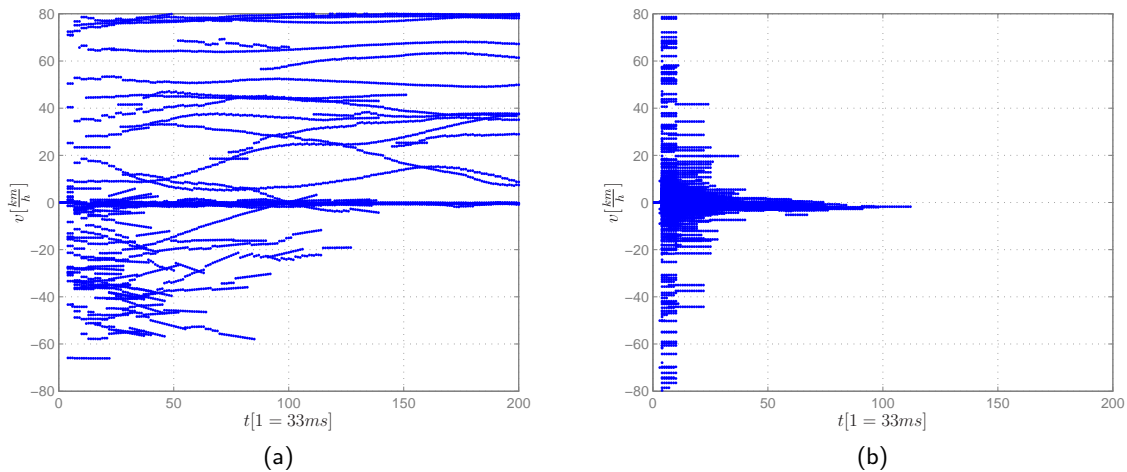


Abbildung 4.12: Aus dem Tracking gewonnene Merkmale Objektgeschwindigkeit v in km/h und Altersindex t von Fahrzeugen (a) und Nichtfahrzeugen (b). Die Daten basieren auf Katalog 6.

Eine zweckmäßige Merkmalsvektorzusammensetzung \mathbf{x}_D besteht somit aus den Komponenten Objektgeschwindigkeit und Altersindex. Um numerische Instabilitäten des Klassifikationsverfahren zu vermeiden, werden die Werte durch obere Schranken τ_p und τ_v begrenzt und auf einen Wertebereich zwischen 0 und 1 skaliert:

$$\mathbf{x}_D = \begin{pmatrix} \frac{1}{\tau_p} \min(t, \tau_p) \\ \frac{1}{\tau_v} \min(|v|, \tau_v) \end{pmatrix}$$

Als sinnvoller Zahlenwert für die Altersindexschranke τ_p bietet sich ein Wert von 100 ($1 \cong 33 \text{ ms}$) an, da nach 3300 ms Trackingzeit ausschließlich stationäre Hintergrundobjekte getrackt werden (vgl. Abbildung 4.12b). Um den Geschwindigkeitswert zu begrenzen, wird τ_v auf den Wert 250 km/h gesetzt. Auch wenn der Merkmalsraum wegen der Werteüberlagerungen durch keinen Klassifikator perfekt separierbar ist, kann eine fehlerminimale Unterteilung gefunden werden. Die Erkennungsraten sind stark abhängig von der spezifischen Fahrumgebung. In einem hohen Geschwindigkeitsbereich auf einer Autobahn sind die Erkennungsraten wesentlich

höher als im Stadtverkehr mit vielen stehenden bzw. parkenden Fahrzeugen. Durchschnittliche Erkennungsraten sind in Abbildung 4.13b zu sehen. Die Ergebnisse basieren auf Testkatalog 7.

Prinzipiell können über die Objektgeschwindigkeit und den Altersindex alle, sich signifikant bewegenden, Verkehrsteilnehmer von Hintergrundobjekten unterschieden werden. Diesen Aspekt machen sich aktuelle, auf Radar basierende Abstandhaltesysteme zu Nutze. Für die Erkennung von Fußgängern ist dieses Verfahren jedoch nur bedingt geeignet. Aufgrund von Rauschen, fehlerbehafteter Datenvorverarbeitung und ungenauer Eigenbewegungskompensationen weicht die Geschwindigkeitsschätzung von der realen Geschwindigkeit ab. Die Größenordnung dieser Abweichung liegt bei vielen Trackingverfahren im Bereich der Bewegungsgeschwindigkeit von Fußgängern. Allein anhand der Geschwindigkeit und des Altersindex kann somit keine eindeutige Unterscheidung zwischen stationären Hintergrundobjekten und Fußgängern getroffen werden.

4.6 Vergleich

Einen Überblick über die gesamten sensor-, merkmals- und objektspezifischen Erkennungsraten ist in Abbildung 4.13 zu sehen. Aus den beiden Übersichtsdarstellungen können verschiedene

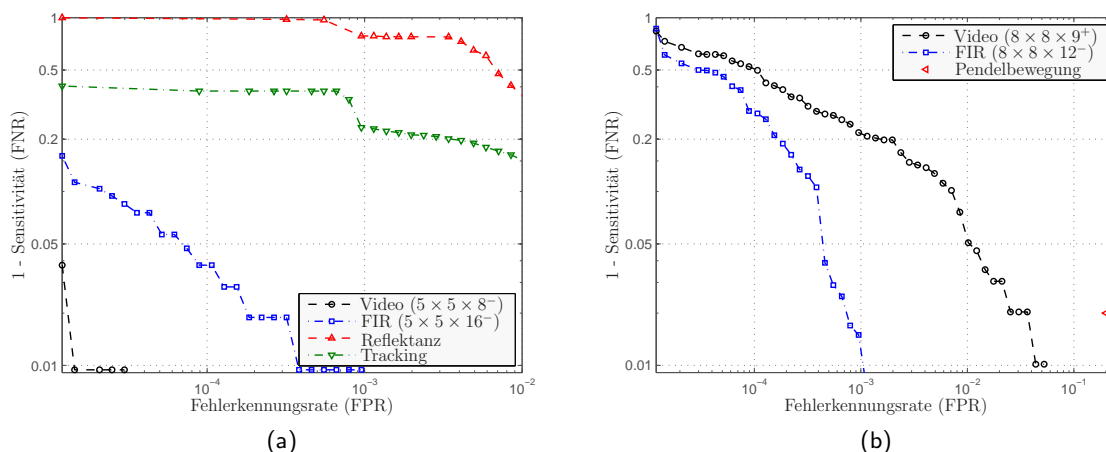


Abbildung 4.13: Merkmalsabhängige Erkennungsraten für Fahrzeuge (a) und Fußgänger (b). Für Fahrzeuge werden im Videobild $5 \times 5 \times 8^-$ Gradientenhistogramme und im Wärmebild $5 \times 5 \times 16^-$ Gradientenhistogramme verwendet. Zudem werden die Merkmale aus der Reflektanzsignatur und aus den Dynamikdaten herangezogen. Bei Fußgängern werden neben der Pendelbewegung, $8 \times 8 \times 9^+$ -Gradientenhistogramme im Videobild und $8 \times 8 \times 12^-$ -Gradientenhistogramme im Wärmebild verwendet.

Ergebnisse abgeleitet werden: Fahrzeugheckansichten, die mittels normierter Gradientenhistogramme kodiert werden, liefern, mit einer vergleichbaren Anzahl von Trainingsdaten, höhere Erkennungsraten im Video- als im Wärmebild. Bei Fußgängeransichten verhält es sich genau umgekehrt. Offensichtlich ist die Intraklassenvarianz bei Fahrzeugheckansichten im Videobild kleiner als bei Fahrzeugheckansichten im Wärmebild (z. B. schwarze Reifen auf heller Straße vs. unterschiedliche Verbaupositionen der Abgasanlage). Entsprechend ist bei Fußgängern die

Intraklassenvarianz im Videobild größer als im Wärmebild (z. B. unterschiedliche Kleidung vs. warmer Kopfbereich). Im Vergleich der Objekte untereinander sind Fahrzeugheckansichten um ca. eine Größenordnung (95 % Erkennungsrate bei 10^{-4} vs. 10^{-3} FPR) besser klassifizierbar als Fußgänger. Dies kann möglicherweise auf die nichtrigide und symmetrische Bauform von Fahrzeugen zurückgeführt werden.

4.7 Zusammenfassung

In diesem Kapitel wurden alle Merkmale vorgestellt, die für die hypothesengetriebene Fußgänger- und Fahrzeugerkennung dieser Arbeit Verwendung finden. Als Bildmerkmale werden für beide Objektklassen und für die bildgebenden Sensoren normierte Gradientenhistogramme genutzt. Diese stellen gegenwärtig eines der „state-of-the-art“-Merkmale zur Objekterkennung dar. Weitere Merkmale werden aus den Laserscannerrohdaten gewonnen: Für Fahrzeugheckansichten konnte gezeigt werden, dass die Rückstrahler- und Nummernschildanordnungen, anhand der Rückstreuintensitäten des Laserscanners detektiert werden können und als ein Merkmal verwendbar sind. Die Bein-Pendelbewegung, welche für gehende oder laufende Fußgänger spezifisch ist, kann aus den Laserscannerrohdaten extrahiert werden. Mittels einer Fourieranalyse wird die Lauffrequenz ermittelt und als Entscheidungskriterium herangezogen. Zusätzlich wurden zur Fahrzeugklassifikation, die Geschwindigkeit und der Altersindex von Fahrzeughypothese als charakteristisches Merkmal verwendet.

Es wurde zusätzlich untersucht, inwiefern sich die Distanzabnahme bzw. -zunahme von Objekten zur Kamera, auf die Erkennungsraten auswirkt. Als Skalierungsverfahren sollte für Bildausschnittsverkleinerungen eine Nearest Neighbour und für Bildausschnittsvergrößerungen eine bilineare Interpolation vorgenommen werden. Ein weiteres Ergebnis ist, dass bei einer Fehlerkennungsrate von 10^{-5} eine Klassifikationsreichweite für Fahrzeuge von ca. 26 m mit dem Videonormalobjektiv (hor. Öffnungswinkel $\pm 21.2^\circ$) und von ca. 60 m mit dem Teleobjektiv (hor. Öffnungswinkel $\pm 8.5^\circ$) möglich ist. Die Wärmebildkamera (hor. Öffnungswinkel $\pm 17.6^\circ$) kann Fahrzeuge bis zu einer Entfernung von ca. 40 m robust klassifizieren. Für Fußgänger bricht die Erkennungsrate im Videobild ab einer Entfernung von ca. 20 m (54 m mit Teleobjektiv) und im Wärmebild ab einer Entfernung von 30 m ein.

Des Weiteren wurde eine Methode vorgestellt und bewertet, wie durch eine vertikale Achsen spiegelung, der effektive Sichtbereich von bildgebenden Sensoren, zur Erkennung von teilverdeckten symmetrischen Objekten, vergrößert werden kann. Bei Fahrzeugen waren bei einem Verdeckungsgrad von bis zu 75 % lediglich geringe Einbrüche der Erkennungsraten zu beobachten. Vor allem im Nahbereich kann dadurch der effektive Klassifikationsbereich des Kamerasystems erheblich erweitert werden. Bei einer angenommenen Fahrzeugbreite von 1.5 m vergrößert sich der effektive horizontale Sichtbereich des bildgebenden Videosensors mit Normalobjektiv um $\pm 15^\circ$.

Sowohl die Reflektanzsignatur bei Fahrzeugen, als auch die Pendelbewegung bei Fußgängern erzielen, im Vergleich zu den Bildmerkmalen, lediglich geringe Erkennungsraten. Wie das nächste Kapitel zeigt, können diese Merkmale, im Zusammenspiel mit anderen Merkmalen, dennoch zur Gesamtleistung beitragen. Es werden dort Verfahren zur Hypothesenverifizierung untersucht, die eine gewinnbringende Fusion der vorgestellten Merkmale erlauben.



„Die fünf Sinne“ von Hans Makart (1840 - 1884).

5

Hypothesenverifizierung

Die Erzeugung von Hypothesen (siehe auch Kapitel 3) resultiert in einer Vielzahl von potentiellen Fahrumfeldobjekten. Das Ziel der Hypothesenverifizierung ist eine Einordnung dieser Objekte in vorgegebene Objektkategorien mittels geeigneter Klassifikationsverfahren. Von dieser Kategorisierung profitiert sowohl die Perzeptions- als auch die Anwendungsschicht. Anhand der Objektklasse kann beispielsweise auf das objektspezifische Dynamikverhalten geschlossen werden. Diese Informationen fließen in Spurmodelle, wodurch Objektverfolgungsverfahren an Genauigkeit gewinnen. Zudem können Hypothesen, die gesichert als Hintergrundobjekt identifiziert werden, verworfen und von künftigen Bearbeitungsschritten ausgeschlossen werden. Dadurch kann Rechenzeit eingespart werden. Auf der Applikationsschicht im Fahrzeug bietet eine Kategorisierung der Objekte gleichermaßen Vorteile. Neben gezielten (Warn-)Informationen für den Fahrer können Sicherheitssysteme wirkungsvoller agieren. Ist beispielsweise bei einem Auffahrunfall der Kollisionspartner dem System bekannt, können die Aktorikmaßnahmen an die Gefahrensituation angepasst werden (z. B. aktive Anhebung der Motorhaube bei Fußgängeranprall und Gurtstraffer bzw. Airbagauslösung bei Fahrzeuganprall).

Klassifikationsverfahren arbeiten auf dimensionsreduzierten Merkmalsvektoren, die aus Sensordaten gewonnen werden. In diesem Kapitel werden sämtliche Merkmale aus Kapitel 4 verschränkt betrachtet, um die Erkennungsleistung bezüglich Fahrzeugen und Fußgängern zu steigern. Neben einem bewährten monolithischen Klassifikationsansatz wird ein besonderer Fokus auf Ensembleverfahren und deren Einsatz im automotiven Umfeld gelegt. Dieser Forschungszweig im Bereich der Mustererkennung beschäftigt sich mit Methoden zur gewinnbringenden Zusammenführung von Basisklassifikatorausgaben. Eine gute Einführung findet sich in [VM02] und [RG00]. Theoretische Aspekte werden in [Kun02, KW03] behandelt. Ein umfassendes Werk zu diesem Themenkomplex wurde von Kuncheva [Kun04] verfasst. Es gibt einen Überblick zum aktuellen Stand der Forschung im Gebiet der Klassifikatorfusion.

Systeme zur Klassifikatorfusion werden erfolgreich in der medizinischen Diagnostik [Lac88], im militärischen Bereich z. B. zur Landminenerkennung [GBTC01], im Finanzwesen [WDQ05] und in der Biometrie zur Identifikation von Personen [RJ03, CBF03] eingesetzt. Erst seit kurzer Zeit sind hochintegrierte Schaltungen mit ausreichender Rechenleistung verfügbar, um heterogene Perzeptionssensorik auf einer mobilen Plattform in Echtzeit auszuwerten. Somit gibt es in der automotiven Forschungslandschaft erst vereinzelte Ansätze zur verschränkten Objekt-

klassifizierung: Krotosky [KT07] führt Farb-, Infrarot- und Stereobilder zur Erkennung von Fußgängern zusammen. Er verwendet einen monolithischen SVM-Klassifikationsansatz. Wender [WD05, WWFD06, WD07] nutzt monolithische neuronale Netze sowie einen regelbasierten Klassifikator, um Dynamik-, Bild- und Laserscannerdaten von Fahrzeugen zu kombinieren.

Derzeitig existiert kein uneingeschränkt optimaler Fusionsansatz. Vielmehr ist die Leistungsfähigkeit der Kombinationsverfahren stark von der Problemstellung und den verwendeten Merkmalen abhängig [Pol06]. Ein Ziel dieses Kapitels ist der Vergleich von verfügbaren Fusionsstrategien im Hinblick auf die Erkennung von Fahrzeugen und Fußgängern mit heterogener Sensorik. Dafür werden die Merkmale herangezogen, die in Kapitel 4 eingeführt wurden. Eine Übersicht mit den entsprechenden Parametrisierungen ist in Tabelle 5.1 aufgeführt.

Tabelle 5.1: Übersicht über die Merkmale, Merkmalsbezeichnungen und Parametrisierung, die für die Klassifikatorfusion verwendet werden. Genaue Beschreibungen der Merkmale, die verfügbaren Parametrisierungen sowie die zugehörigen Extraktionsverfahren finden sich in Kapitel 4.

Merkmal	Bezeichnung (Abkürzung)	Parametrisierung
Fahrzeuge		
Gradientenhistogramme Video	\mathbf{x}_V^V (V)	$20 \times 20 \times 4^-$
Gradientenhistogramme FIR	\mathbf{x}_W^V (W)	$20 \times 20 \times 4^-$
Reflektanzsignatur	\mathbf{x}_R^V (R)	$ \mathbf{x}_R^V = 10$
Dynamik	\mathbf{x}_D^V (D)	$ \mathbf{x}_D^V = 2$
Fußgänger		
Gradientenhistogramme Video	\mathbf{x}_V^P (V)	$8 \times 8 \times 15^+$
Gradientenhistogramme FIR	\mathbf{x}_W^P (W)	$8 \times 8 \times 15^-$
Pendelbewegung	x_P^P (P)	$x_P^P = \begin{cases} 1 & \text{vorhanden} \\ 0 & \text{sonst} \end{cases}$

Das Kapitel ist wie folgt gegliedert: Zunächst wird das Potential einer frühen Klassifikatorfusion auf Merkmalsebene untersucht. Dazu werden die Merkmale zu einem gemeinsamen Merkmalsvektor zusammengefasst und durch einen monolithischen Klassifikator bewertet. Im nächsten Abschnitt werden verschiedene Ensemblestrategien eingeführt und verglichen. Diese entscheidungsbasierte Fusion erfolgt hierbei zu einem späteren Zeitpunkt im Datenfluss, da die Verknüpfungen auf Basisklassifikatorausgaben durchgeführt werden. Im letzten Abschnitt wird auf die Problematik von fehlenden und unzulänglichen Merkmalen im Kontext von Fahrumfelderfassungssystemen eingegangen. Zudem werden die Auswirkungen auf die vorgestellten Fusionsmethoden untersucht. Um eine bessere Vergleichbarkeit der einzelnen Fusionsvorgehensweisen zu gewährleisten, werden als Basisklassifikatoren ausschließlich Support-Vektor-Maschinen mit linearem Kernel (siehe auch Abschnitt B.3 für eine Einführung) verwendet.

Im Vordergrund steht die Bewertung der Leistungsfähigkeit der Fusionsstrategien in Bezug auf die eingeführten Merkmale. Die Datenkataloge, die in diesem Kapitel zur Evaluierung herangezogen werden, sind im Anhang A auf Seite 179 zu finden.

5.1 Monolithische Klassifikation

Die Zusammenfassung von sensorspezifischen Merkmalsvektoren $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$ zu einem kombinierten Vektor $\mathbf{x} = (\mathbf{x}_1^T \ \mathbf{x}_2^T \ \dots)^T$ ist eine naheliegende Möglichkeit sensorübergreifend zu fusionieren. Bei dieser Fusion auf Merkmalsebene dient der aggregierte Vektor als Eingabe für einen monolithischen Klassifikator $d_{\text{mono}}(\mathbf{x})$ (siehe auch Abbildung 5.1). Durch diese Vorgehensweise steht dem Klassifikator ein erhöhter Informationsgehalt zur Verfügung. Zudem wird die Dimension des Merkmalsraums erweitert, wodurch die Separierbarkeit der Daten gegebenenfalls verbessert wird. Gleichzeitig sinkt jedoch die Flexibilität und die Robustheit gegenüber Rauschen und einzelnen fehlerbehafteten Merkmalen (siehe auch Abschnitt 5.3).

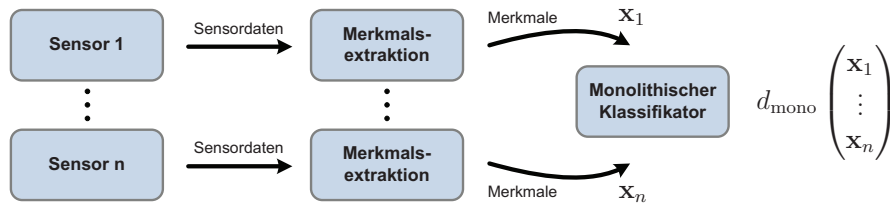


Abbildung 5.1: Datenfluss bei einer monolithischen Klassifikation. Die Merkmale werden für jeden Sensor gesondert extrahiert. Die fusionierte Entscheidung wird durch einen einzelnen monolithischen Klassifikator anhand von allen Merkmalen getroffen.

Um die Leistungsfähigkeit eines monolithischen Klassifikators zur Fahrzeug- und Fußgängererkennung zu erproben, wurden – basierend auf den Merkmalen aus Kapitel 4 – monolithische SVM-Klassifikatoren erstellt. Dabei wurden alle kombinatorisch möglichen Merkmalszusammensetzungen trainiert und mittels einer 10-fachen Kreuzvalidierung¹ auf ihre Leistungsfähigkeit hin untersucht. Die Erkennungsergebnisse für Fahrzeuge und Fußgänger sind in den beiden DET-Diagrammen in Abbildung 5.2 dargestellt.

Die Basisklassifikatoren d_V^V , d_W^V , d_R^V und d_D^V bzw. d_V^P , d_W^P und d_P^P , die nur auf einem Sensor operieren, erzielen Erkennungsraten von 0.29, 0.64, 0.15 und 0.78 bzw. 0.24, 0.90 und 0.1 bei einer FPR von 10^{-3} . Fließen mehr Merkmale in die monolithischen Klassifikatoren mit ein, sind sowohl für Fahrzeuge als auch für Fußgänger sukzessive Steigerungen der Erkennungsraten erkennbar. Dabei liegen die kombinierten Erkennungsraten stets über den Erkennungsraten des besten beteiligten Basisklassifikators. Die höchsten Erkennungsraten (0.953 bei Fahrzeugen und 0.963 bei Fußgängern) erzielen die monolithischen Klassifikatoren d_{WVRD}^V und d_{WVP}^P , die auf allen verfügbaren Merkmalen operieren.

¹ Die n -fache Kreuzvalidierung ist ein Schätzverfahren zur Approximation der Erkennungsleistung eines Klassifikators. Dazu werden sämtliche Daten in n Partitionen unterteilt. Eine Partition wird zum Testen und die restlichen $n-1$ Partitionen zum Training des Klassifikators verwendet. Wird dieser Vorgang n -mal mit jeweils anderen Trainings- und Testpartitionen wiederholt, ergibt die Mittelung der Raten eine gute Abschätzung der Erkennungsrate.

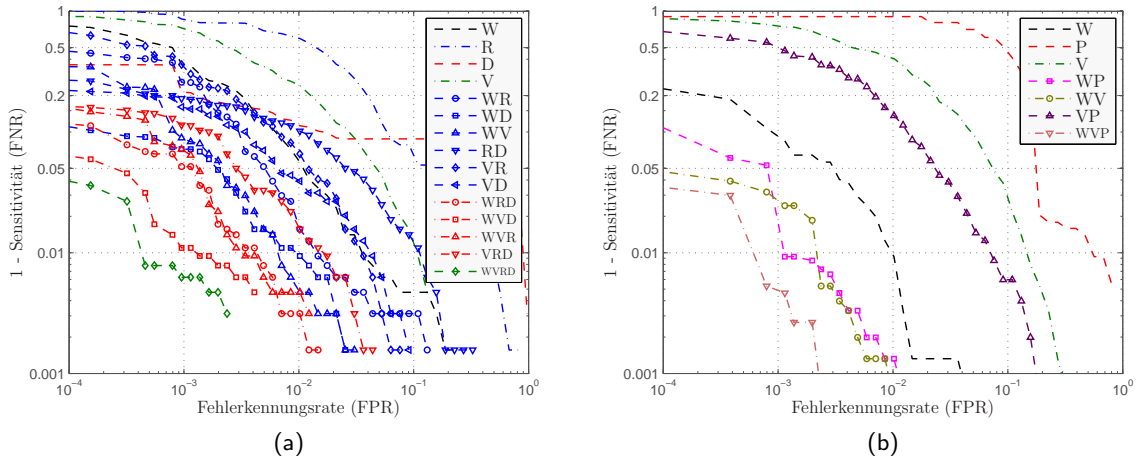


Abbildung 5.2: Erkennungsraten von PKWs (a) und Fußgängern (b) durch monolithische Klassifikatoren. Die Buchstabenkodierungen in der Legende geben die Merkmalszusammensetzung des jeweiligen monolithischen Klassifikators an (z. B. WV $\rightarrow d_{WV}(\mathbf{x}^V) = d_{WV} \left(\left(\mathbf{x}_W^T \quad \mathbf{x}_V^T \right)^T \right)$).

Ein Nachteil der monolithischen Klassifikatoren ist die geringe Transparenz der internen Abläufe. Ist ein Merkmal ungeeignet, kann sich dies auf den gesamten monolithischen Klassifikationsprozess und somit auf die Erkennungsraten auswirken, ohne dass das verursachende Merkmal unmittelbar identifiziert werden kann. Im nächsten Abschnitt wird auf Ensembleverfahren zur Verknüpfung von Klassifikatoren eingegangen. Ein Einblick in die Arbeitsweise von Ensembles ist leichter möglich, da die Gesamtentscheidung aus isolierten Einzelklassifikatoren gebildet wird.

5.2 Ensembleverfahren

Ein Ensemble ist ein Verbund von Basisklassifikatoren $\{d_1, \dots, d_n\}$ deren Entscheidungen zu einer fusionierten Gesamtentscheidung $\mathcal{K}(d_1, \dots, d_n)$ kombiniert wird (siehe auch Abbildung 5.3). Sind die Basisklassifikatoren divers und genau hinsichtlich ihrer Erkennungsleistung, so ist die Performanz des Verbundklassifikators besser als die des besten Basisklassifikators [HS90]. Divers oder verschiedenartig bedeutet dabei, dass die Basisklassifikatoren unterschiedliche Fehlerzuordnungen bezüglich der Testdaten treffen und somit nicht korreliert sind. Dadurch besteht die Möglichkeit, dass Fehlerkennungen wechselseitig kompensiert werden. Die Basisklassifikatoren sind genau, wenn die Performanz eines Basisklassifikators besser ist als die eines zufälligen bzw. ratenden Entscheiders.

Dietterich [Die00] gibt als Beispiel für die Leistungsfähigkeit eines Ensembles, welches aus L diversen und genauen Basisklassifikatoren besteht, ein binäres Entscheidungsproblem an, das durch einen Mehrheitsentscheid gelöst wird. Die Erkennungsrate p_{agg} dieses Ensembles berechnet sich aus der Summe der Kombinationsmöglichkeiten, die zu einem Mehrheitsbeschluss führen:

$$p_{\text{agg}} = \sum_{i=\lfloor L/2 \rfloor + 1}^L \binom{L}{i} p^i (1-p)^{L-i} \quad (5.2.1)$$

Schon für $L = 9$ Basisklassifikatoren mit einer Einzelerkennungsrate von $p = 0.7$ steigt die Gesamterkennungsrate des Ensembles auf $p_{\text{agg}} = 0.901$ an.

Fragestellungen zu Gruppenentscheidungen wurden bereits im 18. Jahrhundert von Marie Jean Antoine Nicolas Caritat, Marquis de Condorcet wissenschaftlich untersucht. Das nach ihm benannte Condorcet's Jury Theorem [dC85, Ber93] besagt, dass für eine Jury, die aus $L > 2$ Mitgliedern besteht (L ungerade), die Gesamtwahrscheinlichkeit für einen korrekten Urteilsspruch über den fehlerlosen Entscheidungswahrscheinlichkeiten der einzelnen Mitglieder liegt, sofern jedes Mitglied mit der Wahrscheinlichkeit $p > 0.5$ eine richtige Entscheidung trifft. Zudem konvergiert die Fehlerrate der Jury gegen 0, wenn die Anzahl der Mitglieder gegen unendlich geht. Das Theorem spielt noch heute eine Rolle bei der Bewertung von demokratischen Systemen oder Organisationsstrukturen in Unternehmen.

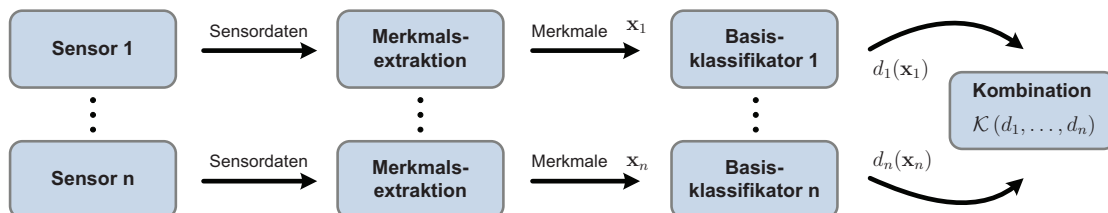


Abbildung 5.3: Datenfluss bei einer Ensembleklassifikation. Die Merkmale werden für jeden Sensor gesondert extrahiert und klassifiziert. Die Ausgaben der Basisklassifikatoren werden durch einen Kombinator, der wiederum ein Klassifikator sein kann, zusammengeführt und zu einer Gesamtentscheidung fusioniert.

Damit ein effektives Ensemble generiert werden kann, müssen die oben beschriebenen Vorbedingungen – diverse und genaue Basisklassifikatoren – eingehalten werden. Um ebensolche Klassifikatoren zu erhalten, gibt es verschiedene Vorgehensweisen [Pol06]. Ein beliebter Ansatz besteht darin, die Trainingsdaten durch Bagging- [Bre96] und Boostingverfahren [Sch90] zu partitionieren. Jeder Basisklassifikator wird hierbei auf einem unterschiedlichen Trainingsdatenausschnitt trainiert. Diversität kann auch erzeugt werden, indem die Basisklassifikatoren durch unterschiedliche Klassifikationsverfahren (z. B. Neuronale Netze, Entscheidungsbäumen, Support-Vektor-Maschinen) erstellt werden. Eine ähnliche Vorgehensweise verwendet instabile Klassifikationsverfahren und variiert die initialen Trainingsparameter. Schließlich können unterschiedliche Merkmale, die von heterogenen Sensoren gewonnen werden, zu einem Ensemble kombiniert werden. Sind die Messprinzipien der verwendeten Sensoren komplementär und divers, gilt das gleichermaßen für die abgeleiteten Merkmale.

In den folgenden Abschnitten werden Ensembles zur Fahrzeug und Fußgängererkennung erzeugt, indem komplementäre Sensorinformationen von Wärmebild, Videobild und Laserscanner zusammengeführt werden. Eine Voraussetzung für eine erfolgreiche Kombination der Basisklassifikatoren ist ein einheitliches numerisches Ausgabeintervall. Dieses kann mittels einer Kalibrierung der SVM-Basisklassifikatoren erzielt werden.

5.2.1 Ausgabenkalibrierung von SVM

Unkalibrierte Basisklassifikatoren erzeugen kontinuierliche Ausgabewerte in einem beliebigen Zahlenintervall. Diese sind oftmals nicht unmittelbar als Wahrscheinlichkeiten interpretierbar. In der Regel bedeuten jedoch höhere bzw. kleinere Zahlenwerte, ein größeres bzw. geringeres Vertrauen des Klassifikators in die entsprechende Klasse. Bei Support-Vektor-Maschinen repräsentiert deren Ausgabe den Abstand des Eingabemerkmalvektors von der – durch das Training ermittelten – Hyperebene (siehe auch Anhang B.3 für eine kurze Einführung). Das Vorzeichen der Ausgabe entscheidet über die Klassenzugehörigkeit und der orthogonale Distanzwert über den Zugehörigkeitsgrad. Die Verteilung dieser Zahlenwerte ist vom Trainingsdatensatz und der Merkmalszusammensetzung abhängig. Zur Verknüpfung von Basisklassifikatoren zu einem Ensemble ist es jedoch notwendig, dass deren Ausgaben numerisch vergleichbar sind und als posteriori Wahrscheinlichkeiten interpretiert werden können. In der Literatur finden sich eine Reihe von Verfahren, wie Klasseneinteilung [Dri01] und Ausgleichsfunktionen [Kwo99, Pla00, Rüp04], die die Ausgabewerte $d(\mathbf{x})$ von SVM's auf eine approximiertere Wahrscheinlichkeitsverteilung $p(\mathbf{x})$ abbilden und somit auf das Intervall $[0, 1]$ begrenzen.

Die *Softmax-Kalibrierung* [DHS00] bildet kontinuierliche Ausgabewerte, durch den Kehrwert einer Exponentialfunktion, auf ein Intervall zwischen 0 und 1 ab:

$$P(y = 1|d(\mathbf{x}))_{\text{soft}} \approx \frac{1}{1 + e^{-2 \cdot d(\mathbf{x})}} \quad (5.2.2)$$

Da ein SVM-Ausgabewert von 0 (Merkmalsvektor liegt auf der Hyperebene) auf 0.5 abgebildet wird und Werte < 0 bzw. > 0 auf Werte < 0.5 bzw. > 0.5 abgebildet werden, kann die Funktionsausgabe als Wahrscheinlichkeit interpretiert werden. Die Zuverlässigkeit der geschätzten Wahrscheinlichkeiten ist jedoch niedrig, da die Softmax-Kalibrierung nicht von Trainingsdaten gestützt wird.

Eine wesentlich bessere Schätzung der Wahrscheinlichkeiten erzielt Platt [Pla00] mit sigmoiden Funktionen, die anhand von Trainingsdaten in die Ausgabewerte gelegt werden. Die freien Funktionsparameter können mit einem Levenberg-Marquardt-Algorithmus approximiert werden. Eine ähnliche Genauigkeit erreicht die wesentlich einfachere *pp-Kalibrierung*, welche von Rüping [Rüp04] eingeführt wurde. Da außerhalb der Ränder einer SVM ($d(\mathbf{x}) > 1$ bzw. $d(\mathbf{x}) < -1$) keine eindeutige Wahrscheinlichkeitszuordnung getroffen werden kann [BT07], setzt Rüping in diesen Grenzbereichen die Wahrscheinlichkeiten auf die Anteile der positiven p^+ bzw. negativen p^- Trainingsdaten zu allen Trainingsdaten. Zwischen den Rändern der SVM werden die Wahrscheinlichkeiten linear interpoliert:

$$P(y = 1|d(\mathbf{x}))_{\text{pp}} \approx \begin{cases} p^+ & \text{für } d(\mathbf{x}) > 1 \\ \frac{1}{2}(d(\mathbf{x}) + 1) & \text{für } d(\mathbf{x}) \in [-1, 1] \\ p^- & \text{für } d(\mathbf{x}) < -1 \end{cases} \quad (5.2.3)$$

Sowohl die Softmax- als auch die pp-Kalibrierung wird in den folgenden Abschnitten zur Approximierung der Ausgabewahrscheinlichkeiten genutzt, indem die kalibrierten SVM-Ausgabewerte der Basisklassifikatoren durch funktionale Kombinatoren zu einem Ensemble zusammengefügt werden.

5.2.2 Funktionale Kombinatoren

Funktionale Kombinatoren [Dui02, Pol06] verknüpfen kalibrierte und kontinuierliche Basisklassifikatorausgaben $d_i(\mathbf{x})$ zu einer Ensembleausgabe. Sie werden nicht trainiert, sondern setzen sich aus statischen Funktionen $\mathfrak{F} : \mathbb{R}^n \rightarrow \mathbb{R}$ zusammen. Diese algebraischen Rechenvorschriften $\mathfrak{F}(d_1(\mathbf{x}), \dots, d_T(\mathbf{x}))$ erzeugen für eine Klasse ω_j eine fusionierte Gesamtkonfidenz u_j . Die zweckmäßigsten funktionalen Kombinatoren $\mathfrak{F}(\cdot)$ werden im Folgenden beschrieben.

Arithmetisches Mittel: Die fusionierte Konfidenz $u_j(\mathbf{x})$ berechnet sich über das arithmetische Mittel der Einzelkonfidenzen:

$$u_j(\mathbf{x}) = \mathfrak{F}_a(\cdot) = \frac{1}{T} \sum_{t=1}^T d_t(\mathbf{x}). \quad (5.2.4)$$

Gewichtetes Mittel: Zusätzlich zum arithmetischen Mittel werden bei dieser Regel die Einzelkonfidenzen mit Faktoren w_t gewichtet. Damit ergibt sich die Gesamtkonfidenz zu

$$u_j(\mathbf{x}) = \mathfrak{F}_g(\cdot) = \frac{1}{T} \sum_{t=1}^T w_t d_t(\mathbf{x}). \quad (5.2.5)$$

Die Gewichtevertelung kann statisch oder dynamisch erfolgen.

Beschnittenes Mittel: Bei dieser Regel werden zu optimistische bzw. pessimistische Konfidenzen vor der Berechnung des arithmetischen Mittelwertes entfernt. Dadurch kann der Einfluss von Ausreißern reduziert werden. Als zusätzlicher Parameter kann ein Prozentsatz angegeben werden, der bestimmt welcher Anteil der Basisklassifikatoren an den Enden des Wertebereichs ignoriert werden soll. Das beschnittene Mittel wird mit $\mathfrak{F}_b(\cdot)$ notiert.

Produktregel: Die Gesamtkonfidenz bildet sich aus dem Produkt der Einzelkonfidenzen:

$$u_j(\mathbf{x}) = \mathfrak{F}_p(\cdot) = \prod_{t=1}^T d_t(\mathbf{x}). \quad (5.2.6)$$

Die Produktregel ist die optimale Verknüpfung der Einzelkonfidenzen, sofern die Ausgabe-wahrscheinlichkeiten unverfälscht und die Basisklassifikatoren voneinander unabhängig sind. Dann gilt das probabilistische Produkt bzw. die naive Bayesregel [Kun04]:

$$P(\omega_k | \mathbf{x}) = \prod_{t=1}^T d_t(\mathbf{x}). \quad (5.2.7)$$

Eine vollständige Unabhängigkeit ist jedoch nur in seltenen Fällen gegeben. Bei Ausreißern von einzelnen Klassifikatoren wird die Gesamtwahrscheinlichkeit signifikant betroffen. Sinkt beispielsweise der pessimistischste Basisklassifikator fälschlicherweise auf einen Wert von 0, so sinkt das Gesamtergebnis unabhängig von den anderen Einzelergebnissen ebenfalls auf 0.

Maximum-, Minimum- und Medianregel: Durch diese Regeln wird die Einzelkonfidenz, deren Wert das Minimum, das Maximum bzw. den Median im Ensemble markiert, auf die Gesamtkonfidenz abgebildet. Es gilt somit

$$u_j(\mathbf{x}) = \mathfrak{F}_{max}(\cdot) = \max_{0 < t \leq T} \{d_t(\mathbf{x})\} \quad (5.2.8)$$

$$u_j(\mathbf{x}) = \mathfrak{F}_{min}(\cdot) = \min_{0 < t \leq T} \{d_t(\mathbf{x})\} \quad (5.2.9)$$

$$u_j(\mathbf{x}) = \mathfrak{F}_{med}(\cdot) = \text{med}_{0 < t \leq T} \{d_t(\mathbf{x})\} \quad (5.2.10)$$

Die Maximumregel scheint eine plausible Kombinationsstrategie zu sein, da der Basis-klassifikator mit der höchsten Konfidenz die Entscheidung fällt. Durch Überanpassungs-effekte können jedoch einzelne Klassifikatoren das Ensemble dominieren und dadurch die Ensembleleistung mindern. Ähnliches gilt für die Minimumregel.

Generalisiertes Mittel: Viele der oben genannten Regeln sind Spezialfälle des generalisierten Mittelwertes

$$u_j(\mathbf{x}, \alpha) = \left(\frac{1}{T} \sum_{t=1}^T d_t(\mathbf{x})^\alpha \right)^{\frac{1}{\alpha}} \quad (5.2.11)$$

und können durch entsprechende Werte von α gebildet werden können. Beispielsweise ergibt sich die Minimumregel für $\alpha \rightarrow -\infty$ und die Maximumregel für $\alpha \rightarrow \infty$. Das arithmetische Mittel kann über $\alpha = 1$ abgeleitet werden.

Im Folgenden werden mittels dieser funktionalen Kombinatoren Ensembles zur Fahrzeug- und Fußgängererkennung erzeugt und bewertet. Dafür werden die Basisklassifikatoren zunächst auf den einzelnen Merkmalsräumen gesondert trainiert (siehe auch Tabelle 5.1 für eine Übersicht). Dadurch dass die Merkmalsräume anhand unterschiedlicher Sensoren erzeugt werden, wird von einer Unkorreliertheit ausgegangen. Anschließend werden die Ausgaben der Basisklassifika-toren kalibriert und kombiniert. Die Erkennungsraten bei einer Fehlerkennungsrate von 10^{-4} wurden mit einer 10-fachen Kreuzvalidierung erstellt und sind für Fahrzeuge in den Tabellen 5.2 und 5.3, und für Fußgänger in Tabelle 5.4 aufgeführt. Die maximalen Erkennungsraten für eine Merkmalskonfiguration sind hervorgehoben.

Unabhängig von der Ausgabenkalibrierung erzielen das arithmetische Mittel $\mathfrak{F}_a(\cdot)$ und die Produktregel $\mathfrak{F}_p(\cdot)$ sowohl für Fahrzeuge als auch Fußgänger die höchsten Erkennungsraten. Bei den Basisklassifikatorausgaben, die vor der Fusion mit der Softmax-Kalibrierung in geschätzte Wahrscheinlichkeiten umgewandelt wurden, liegen die Erkennungsraten von $\mathfrak{F}_a(\cdot)$ höher als von $\mathfrak{F}_p(\cdot)$ (vergleiche Tabelle 5.2). Wird die pp-Kalibrierung, die im Allgemeinen eine bessere SVM-Kalibrierung erlaubt, auf die Basisklassifikatoren zur Wahrscheinlichkeitsapproximation angewendet, profitiert davon die Produktregel (siehe Tabelle 5.3), da diese bei korrekten Wahr-scheinlichkeiten eine optimale Klassifikatorfusion erzielt.

Das beschnittene arithmetische Mittel $\mathfrak{F}_b(\cdot)$ wurde so konfiguriert, dass das arithmetische Mittel des Ensembles gebildet wird, nachdem der optimistischste und pessimistischste Basis-klassifikator entfernt wurde. Durch die geringe Anzahl an verfügbaren Basisklassifikatoren konnte lediglich eine begrenzte Bewertung vorgenommen werden. Ab drei Basisklassifikatoren entspricht die Anwendung von $\mathfrak{F}_b(\cdot)$ der Medianregel $\mathfrak{F}_{med}(\cdot)$. Die Erkennungsraten nähern sich dabei erst ab vier Basisklassifikatoren den Erkennungsraten von $\mathfrak{F}_a(\cdot)$ an. Bei größeren Ensembles ist $\mathfrak{F}_b(\cdot)$ dem arithmetischen Mittel $\mathfrak{F}_a(\cdot)$ möglicherweise überlegen, da Ausreißer effektiv gefiltert werden.

Die funktionalen Kombinatoren $\mathfrak{F}_{min}(\cdot)$, $\mathfrak{F}_{max}(\cdot)$ und $\mathfrak{F}_{med}(\cdot)$ fungieren als Selektoren, da die Gesamtkonfidenz durch die Wahl eines einzelnen Basisklassifikators gebildet wird. Eine unmittelbare Verknüpfung der Basisklassifikatorausgaben findet nicht statt. Die Minimumregel selektiert stets das pessimistischste Basisklassifikatorergebnis und erzielt dadurch annehmbare Erkennungsraten, bei einer Fehlerrate von 10^{-4} . Dies kann darauf zurückgeführt werden, dass durch die pessimistische Basisklassifikatorwahl, Objekte der Hintergrundklasse vermehrt ausgesondert werden. Im Gegensatz zu $\mathfrak{F}_{min}(\cdot)$ selektiert die Maximumregel $\mathfrak{F}_{max}(\cdot)$ den optimistischsten Basisklassifikator. In der Auswertung wurde überwiegend der gleiche Basisklassifikator gewählt, da die Erkennungsraten von $\mathfrak{F}_{max}(\cdot)$ denen des besten Einzelklassifikators entsprechen. Die Medianregel $\mathfrak{F}_{med}(\cdot)$ ist ab einer Anzahl von drei Basisklassifikatoren sinnvoll einsetzbar und erzielt mit der pp-Kalibrierung vergleichbare Erkennungsraten wie das arithmetische Mittel. Die überwiegende Anzahl der betrachteten funktionalen Kombinatoren profitiert von der, auf Trainingsdaten basierenden, pp-Kalibrierung (vergleiche Tabelle 5.2 mit Tabelle 5.3).

Tabelle 5.2: Ensembleerkennungsraten bei einer FPR von 10^{-4} mittels funktionaler Kombinatoren für Fahrzeuge. Die ersten vier Zeilen beinhalten die Erkennungsraten für die isolierten Merkmalsvektoren \mathbf{x}_W^V , \mathbf{x}_R^V , \mathbf{x}_D^V und \mathbf{x}_V^V . In den restlichen Zeilen sind die Ergebnisse für eine kombinierte Betrachtung der angegebenen Merkmale aufgeführt. Die Wandlung der SVM-Ausgaben in Wahrscheinlichkeiten erfolgte mit der Softmax-Kalibrierung.

Merkmale	Funktionale Kombinatoren						
	Basiskl.	$\mathfrak{F}_a(\cdot)$	$\mathfrak{F}_{min}(\cdot)$	$\mathfrak{F}_{max}(\cdot)$	$\mathfrak{F}_{med}(\cdot)$	$\mathfrak{F}_p(\cdot)$	$\mathfrak{F}_b(\cdot)$
\mathbf{x}_W^V	0.200						
\mathbf{x}_R^V	0.000						
\mathbf{x}_D^V	0.640						
\mathbf{x}_V^V	0.091						
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V$	0.646	0.023	0.199	0.023	0.559		
$\mathbf{x}_W^V \wedge \mathbf{x}_D^V$	0.818	0.528	0.689	0.528	0.817		
$\mathbf{x}_W^V \wedge \mathbf{x}_V^V$	0.873	0.124	0.207	0.124	0.827		
$\mathbf{x}_R^V \wedge \mathbf{x}_D^V$	0.791	0.186	0.612	0.186	0.763		
$\mathbf{x}_R^V \wedge \mathbf{x}_V^V$	0.410	0.008	0.051	0.008	0.394		
$\mathbf{x}_D^V \wedge \mathbf{x}_V^V$	0.848	0.102	0.669	0.102	0.859		
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_D^V$	0.914	0.596	0.687	0.306	0.882	0.306	
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_V^V$	0.905	0.056	0.206	0.049	0.964	0.049	
$\mathbf{x}_W^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$	0.964	0.758	0.692	0.369	0.973	0.369	
$\mathbf{x}_R^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$	0.894	0.342	0.639	0.095	0.882	0.095	
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$	0.989	0.811	0.690	0.067	0.991	0.772	

Tabelle 5.3: Ensembleerkennungsraten bei einer FPR von 10^{-4} mittels verschiedener funktionaler Kombinatoren für Fahrzeuge. Die ersten vier Zeilen beinhalten die Erkennungsraten für die isolierten Merkmalsvektoren \mathbf{x}_W^V , \mathbf{x}_R^V , \mathbf{x}_D^V und \mathbf{x}_V^V . In den restlichen Zeilen sind die Ergebnisse für eine kombinierte Betrachtung der angegebenen Merkmale aufgeführt. Die Wandlung der SVM-Ausgaben in Wahrscheinlichkeiten erfolgte mit der pp-Kalibrierung.

Merkmale	Funktionale Kombinatoren						
	Basiskl.	$\mathfrak{F}_a(\cdot)$	$\mathfrak{F}_{min}(\cdot)$	$\mathfrak{F}_{max}(\cdot)$	$\mathfrak{F}_{med}(\cdot)$	$\mathfrak{F}_p(\cdot)$	$\mathfrak{F}_b(\cdot)$
\mathbf{x}_W^V	0.103						
\mathbf{x}_R^V	0.018						
\mathbf{x}_D^V	0.110						
\mathbf{x}_V^V	0.109						
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V$		0.509	0.340	0.103	0.340	0.633	
$\mathbf{x}_W^V \wedge \mathbf{x}_D^V$		0.797	0.774	0.110	0.774	0.808	
$\mathbf{x}_W^V \wedge \mathbf{x}_V^V$		0.674	0.731	0.103	0.731	0.825	
$\mathbf{x}_R^V \wedge \mathbf{x}_D^V$		0.713	0.442	0.110	0.442	0.763	
$\mathbf{x}_R^V \wedge \mathbf{x}_V^V$		0.164	0.351	0.110	0.351	0.385	
$\mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.775	0.761	0.110	0.761	0.826	
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_D^V$		0.808	0.565	0.110	0.774	0.856	0.774
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_V^V$		0.714	0.772	0.103	0.731	0.911	0.731
$\mathbf{x}_W^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.836	0.868	0.110	0.774	0.959	0.774
$\mathbf{x}_R^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.794	0.766	0.110	0.775	0.861	0.775
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.839	0.869	0.110	0.922	0.985	0.823

Tabelle 5.4: Ensembleerkennungsraten bei einer FPR von 10^{-4} mittels funktionaler Kombinatoren für Fußgänger. Die ersten drei Zeilen beinhalten die Erkennungsraten für die isolierten Merkmalsvektoren \mathbf{x}_W^P , \mathbf{x}_V^P und \mathbf{x}_P^P . In den restlichen Zeilen sind die Ergebnisse für eine kombinierte Betrachtung der angegebenen Merkmale aufgeführt. Die Wandlung der SVM-Ausgaben in Wahrscheinlichkeiten erfolgte mit der pp-Kalibrierung.

Merkmale	Funktionale Kombinatoren						
	Basiskl.	$\mathfrak{F}_a(\cdot)$	$\mathfrak{F}_{min}(\cdot)$	$\mathfrak{F}_{max}(\cdot)$	$\mathfrak{F}_{med}(\cdot)$	$\mathfrak{F}_p(\cdot)$	$\mathfrak{F}_b(\cdot)$
\mathbf{x}_W^P	0.866						
\mathbf{x}_V^P	0.064						
\mathbf{x}_P^P	0.097						

Tabelle 5.4: Ensembleerkennungsraten bei einer FPR von 10^{-4} mittels funktionaler Kombinatoren für Fußgänger. Die ersten drei Zeilen beinhalten die Erkennungsraten für die isolierten Merkmalsvektoren \mathbf{x}_W^P , \mathbf{x}_V^P und \mathbf{x}_P^P . In den restlichen Zeilen sind die Ergebnisse für eine kombinierte Betrachtung der angegebenen Merkmale aufgeführt. Die Wandlung der SVM-Ausgaben in Wahrscheinlichkeiten erfolgte mit der pp-Kalibrierung.

Merkmale	Funktionale Kombinatoren						
	Basiskl.	$\mathfrak{F}_a(\cdot)$	$\mathfrak{F}_{min}(\cdot)$	$\mathfrak{F}_{max}(\cdot)$	$\mathfrak{F}_{med}(\cdot)$	$\mathfrak{F}_p(\cdot)$	$\mathfrak{F}_b(\cdot)$
$\mathbf{x}_W^P \wedge \mathbf{x}_V^P$		0.950	0.943	0.540	0.943	0.945	0.099
$\mathbf{x}_W^P \wedge \mathbf{x}_P^P$		0.856	0.856	0.097	0.856	0.856	0.099
$\mathbf{x}_V^P \wedge \mathbf{x}_P^P$		0.123	0.123	0.097	0.123	0.123	0.099
$\mathbf{x}_W^P \wedge \mathbf{x}_V^P \wedge \mathbf{x}_P^P$		0.964	0.952	0.097	0.856	0.956	0.856

Von den untersuchten funktionalen Kombinatoren erzielen das (beschnittene) arithmetische Mittel, die Minimumregel und die Produktregel die höchsten Erkennungsraten. Für eine Klassifikatorfusion im Gesamtsystem hat sich jedoch $\mathfrak{F}_a(\cdot)$ bzw. $\mathfrak{F}_b(\cdot)$ als überlegen herausgestellt, da eine korrekte Wahrscheinlichkeitsapproximation für die Produktregel nicht gewährleistet werden kann, und das arithmetische Mittel zudem resistenter gegenüber verrauschten Merkmalen ist (siehe auch Abschnitt 5.4).

5.2.3 Metaklassifikation

Im Gegensatz zu den starren funktionalen Kombinatoren \mathcal{F} ist der Kombinator \mathcal{K} bei einer Metaklassifikation selbst ein Klassifikator. Dieser Metaklassifikator d_{meta} operiert auf den Ausgaben der Basisklassifikatoren $\{d_1, \dots, d_n\}$ und bildet daraus eine Gesamtausgabe

$$u = d_{\text{meta}}(d_1(\mathbf{x}_1), \dots, d_n(\mathbf{x}_n))$$

für das Ensemble [DHS00]. Durch das Training wird der Metaklassifikator flexibel an die Ausgaben der Basisklassifikatoren angepasst. Der Trainingsablauf gliedert sich in zwei Phasen, da zunächst die Basisklassifikatoren und im Anschluss daran der Metaklassifikator trainiert werden muss. Damit für beide Trainingsphasen ausreichend Trainingsdaten zur Verfügung stehen, werden die Trainingsdaten in zwei gleichgroße Teile partitioniert. Ein Teil wird zur Erzeugung der Basisklassifikatoren genutzt. Auf dem zweiten Teil werden anhand der Basisklassifikatoren die Klassifikatorausgaben generiert, die als Eingabe für das Training des Metaklassifikators dienen. Eine Kalibrierung der Basisklassifikatoren bewirkte keinen Anstieg der Erkennungsraten. Die Metaklassifikatoren wurden somit unmittelbar auf den SVM-Ausgaben der Basisklassifikatoren trainiert.

Die Fahrzeug und Fußgänger-Erkennungsraten, die mittels einer 10-fachen Kreuzvalidierung bestimmt wurden, sind in den Diagrammen in Abbildung 5.4 zu sehen. Es wird eine Staffelung der DET-Kurven in Abhängigkeit von der verfügbaren Merkmalsanzahl deutlich. Die höchsten Erkennungsraten werden mit den Metaklassifikatoren erzielt, die auf allen verfügbaren

Merkmale operieren. Sie liegen für Fahrzeuge bei 0.995 und für Fußgänger bei 0.984 (FPR von 10^{-4}).

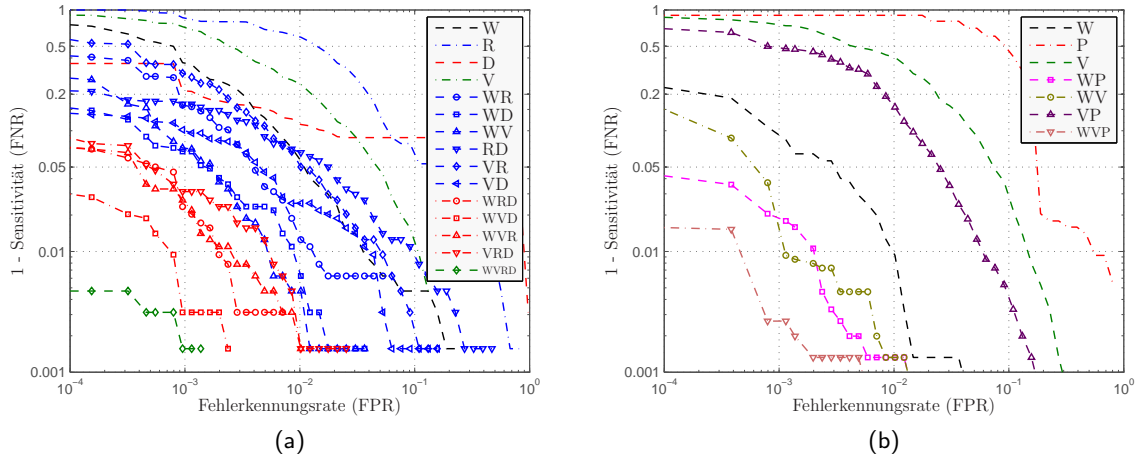


Abbildung 5.4: Erkennungsraten von Fahrzeugen und Fußgängern durch Metaklassifikatoren. Die Buchstabenkodierungen in der Legende geben die Merkmalszusammensetzung des jeweiligen Ensembleklassifikators an (z. B. $WV \rightarrow \mathbf{x}^V = \begin{pmatrix} \mathbf{x}_W^T & \mathbf{x}_V^T \end{pmatrix}^T$).

Vergleichbar zum monolithischen Klassifikationsansatz ist die Handhabung eines Metaklassifikators weniger flexibel als ein funktionaler Kombinator. Bei der Berücksichtigung eines diskontinuierlichen Merkmalsaufkommens, wie es im nächsten Abschnitt thematisiert wird, muss für jede Merkmalszusammensetzung ein dedizierter Metaklassifikator trainiert werden. Dies kann insbesondere bei einer hohen Anzahl von Merkmalen zu einem großen Aufwand bezüglich des Klassifikatortrainings und der -verwaltung führen.

5.3 Diskontinuierliches Merkmalsaufkommen

In Fahrumfelderfassungssystemen mit mehreren Sensoren kann es zu einem diskontinuierlichen Merkmalsaufkommen kommen. Das bedeutet, dass die Anzahl der für die Klassifikation zur Verfügung stehenden Merkmale entweder nicht konstant oder nicht repräsentativ ist. Grundsätzlich kann zwischen fehlenden und unzulänglichen Merkmalen unterschieden werden.

Fehlende Merkmale: Merkmale werden als fehlend bezeichnet, wenn sie für den aktuellen Klassifikationsvorgang nicht zur Verfügung stehen. Eine Ursache für fehlende Merkmale in Multisensorsystemen liegt z. B. in den unterschiedlichen Überlappungen der Sichtbereiche der einzelnen Sensoren. Dadurch entstehen Regionen, die nur von einer Untermenge der verfügbaren Sensorik eingesehen werden kann. Folglich kann in diesen Regionen nur eine Untermenge der verfügbaren Merkmale abgeleitet werden. Zudem gibt es Merkmale, die erst nach einer gewissen Beobachtungszeit deduziert werden können (z. B. Pendelbewegung bei Fußgängern). Sind Merkmale vorhanden, die unterschiedliche Beobachtungszeiten erfordern, ergeben sich Zeitintervalle mit verschiedenen Merkmalsverfügbarkeiten. Ein weiterer Grund für fehlende Merkmale können Sensorausfälle sein.

Unzulängliche Merkmale: Ursachen für unzulängliche Merkmale sind partielle Sensorausfälle (z. B. Übersteuerung einer Kamera), Messungen in nichtrepräsentativen Wertebereichen (z. B. Fußgängererkennung durch Wärmebildkamera im Heißland oder Fahrzeugerkennung im Videobild bei Dunkelheit oder starkem Regen), Verdeckungen, Fehlkalibrierungen oder ungenaue Vorverarbeitungs- und Merkmalsextraktionsroutinen. Werden diese Ursachen nicht detektiert, stehen die Merkmale dem Klassifikationsvorgang zwar zur Verfügung – repräsentieren aber in der Regel nicht die entsprechende Objektklasse. Kann ein unzulängliches Merkmal detektiert werden, sollte es demnach nicht in den Klassifikationsprozess eingebracht werden.

Im Folgenden wird thematisiert, welche Techniken für die Verarbeitung von diskontinuierlichen Merkmalsströmen zur Verfügung stehen. Gängige Klassifikationsverfahren sind auf eine vollständige Anzahl von Merkmalen angewiesen. Eine pragmatische Möglichkeit mit fehlenden Merkmalen umzugehen, ist die ausschließliche Betrachtung von vollständigen Merkmalsvektoren. Unvollständigen Merkmalsvektoren werden somit ignoriert. Diese Vorgehensweise ist in einem automotiven Umfeldersensungssystem jedoch suboptimal, da zum einen die Klassifikationsdauer für zeitkritische Sicherheitssysteme eine wesentliche Rolle spielt, und zum anderen (partielle) Sensorausfälle und somit Klassifikationsausfälle über einen längeren Zeitraum auftreten können.

Ein Ansatz, der auch unvollständige Merkmalszusammensetzungen behandelt, verwendet für jede Merkmalskonstellation einen dedizierten Klassifikator, der ausschließlich vorhandene Merkmale berücksichtigt. Unter der Annahme, dass aus einem Sensor ein Merkmal abgeleitet werden kann, gibt es bei einer Anzahl von $n > 1$ Sensoren

$$u = \sum_{k=1}^n \binom{n}{k} = 2^n - 1 \quad (5.3.1)$$

Kombinationsmöglichkeiten aus fehlenden und vorhandenen Merkmalsanordnungen. Für jede dieser u Kombinationen muss ein dedizierter Klassifikator verfügbar sein. Bei trainierten Ensembleklassifikatoren, wie dem monolithischen Klassifikator und dem Metaklassifikator müssen bei der Fahrzeugklasse mit vier Merkmalen 16 Teilensembles trainiert werden, die bei entsprechenden Fehlkonstellationen verwendet werden. Wesentlich einfacher ist die Handhabung von funktionalen Kombinatoren, da die algebraischen Verknüpfungen auf eine beliebige Anzahl von Basisklassifikatoren flexibel adaptiert werden können.

In der Literatur finden sich noch weitere Verfahren zum Umgang mit fehlenden Merkmalen, wie das Vervollständigen von unvollständigen Merkmalsvektoren auf Basis der verfügbaren Merkmale [HTF01, MSP06, STP07]. Auf diese Ansätze wird an dieser Stelle nicht näher eingegangen.

5.4 Bewertung und Vergleich

In diesem Abschnitt werden die in diesem Kapitel betrachteten Fusionsansätze zur multisensoriellen Erkennung von Fahrzeugen mit fehlerlosen und unzulänglichen Merkmalen gegenübergestellt. Als Objektkategorie werden ausschließlich Fahrzeuge betrachtet, da ein objektunabhängiges Performanzverhalten festgestellt wurde.

Abbildung 5.5 fasst die Ergebnisse der besten Fusionsansätze zusammen. Dabei wird von einem fehlerlosen Merkmalsaufkommen ausgegangen. Die Leistungsfähigkeit der unterschiedlichen Ansätze ist dabei weitgehend vergleichbar. Der monolithische Ansatz fällt im Vergleich zu den anderen Verfahren meist etwas ab, wodurch die These gestützt wird [Die00], dass ein einzelner Basisklassifikator heterogene Merkmale lediglich unzulänglich abbilden kann. Der Metaklassifikator hat überwiegend einen leichten Vorsprung gegenüber den funktionalen Kombinatoren, da diese nicht durch Trainingsdaten gestützt werden. Eine grundlegende verfahrens-unabhängige Tendenz ist hinsichtlich der Anzahl der verwendeten Merkmale zu beobachten. Merkmalshinzunahmen führen zu einem kontinuierlichen Anstieg der Erkennungsraten.

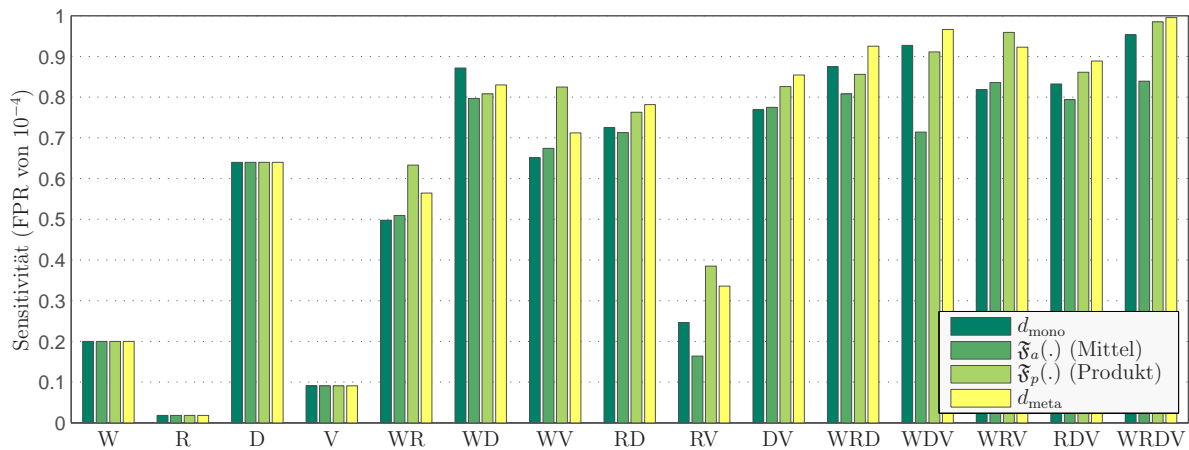


Abbildung 5.5: Erkennungsratenvergleich der verschiedenen Klassifikatorfusionsverfahren (monolithischer Klassifikator, funktionale Kombinatoren und Metaklassifikation) für Fahrzeuge.

Fließen unzulängliche Merkmale in den Klassifikationsprozess mit ein, zeichnet sich ein anderes Ergebnis als mit fehlerfreien Merkmalen ab. In Tabelle 5.5 wurden vor der Bewertung der Fusionsstrategien die Wärmebildmerkmale \mathbf{x}_W^V stark verrauscht. Dabei wurde sichergestellt, dass \mathbf{x}_W^V von der Einzelklassifikation auf dem Wärmebild als Hintergrundklasse ausgewiesen wird. Merkmalszusammensetzungen, die \mathbf{x}_W^V enthalten sind in der Tabelle grau hinterlegt. Die höchsten Erkennungsraten einer Tabellenzeile sind hervorgehoben. Die wesentlichen Resultate der Auswertung ändern sich lediglich marginal, wenn statt \mathbf{x}_W^V ein anderes Merkmal bzw. wenn mehrere Merkmale zeitgleich von Fehlern betroffen sind.

Sowohl bei dem monolithischen Klassifikationsansatz als auch bei der Metaklassifikation sinken die Erkennungsraten gegen Null, wenn eines der Merkmale bzw. wenn ein Basisklassifikator das verrauschte Wärmebildmerkmal enthält. Durch die hohe Merkmalsintegration in diesen Klassifikatormodellen, ist der gesamte Entscheidungsprozess durch das fehlerbehaftete Merkmal betroffen. Gleichmaßen ungeeignet ist der funktionale Kombinator $\mathfrak{F}_{\text{min}}(\cdot)$, da durch diesen stets der Basisklassifikator selektiert wird, der auf dem unzulänglichen Merkmal arbeitet. Die Maximumregel $\mathfrak{F}_{\text{max}}(\cdot)$ erzielt, entsprechend zur Verarbeitung von fehlerlosen Merkmalen, die Erkennungsleistung des besten Einzelklassifikators. Die höchsten Erkennungsraten und damit die beste Kompensation von unzulänglichen Merkmalen werden durch die Produktregel, die Medianregel und das (beschnittene) arithmetische Mittel erzielt.

Tabelle 5.5: Vergleich der Klassifikatorfusionsverfahren bei einem verrauschten Wärmebildmerkmal \mathbf{x}_W^V . In der Tabelle sind die Erkennungsraten der betrachteten Fusionsansätze bei einer FPR von 10^{-4} mit unterschiedlichen Merkmalszusammensetzungen aufgeführt. Tabellenzeilen, die \mathbf{x}_W^V enthalten sind grau hinterlegt.

Merkmale	Basiskl.	d_{mono}	d_{meta}	Funktionale Kombinatoren					
				$\mathfrak{F}_a(\cdot)$	$\mathfrak{F}_{\min}(\cdot)$	$\mathfrak{F}_{\max}(\cdot)$	$\mathfrak{F}_{\text{med}}(\cdot)$	$\mathfrak{F}_p(\cdot)$	$\mathfrak{F}_b(\cdot)$
\mathbf{x}_W^V	0.103								
\mathbf{x}_R^V	0.018								
\mathbf{x}_D^V	0.110								
\mathbf{x}_V^V	0.109								
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V$		0.000	0.000	0.005	0.000	0.018	0.000	0.011	
$\mathbf{x}_W^V \wedge \mathbf{x}_D^V$		0.000	0.000	0.034	0.000	0.110	0.000	0.036	
$\mathbf{x}_W^V \wedge \mathbf{x}_V^V$		0.000	0.000	0.071	0.000	0.110	0.000	0.010	
$\mathbf{x}_R^V \wedge \mathbf{x}_D^V$		0.725	0.782	0.713	0.442	0.110	0.442	0.763	
$\mathbf{x}_R^V \wedge \mathbf{x}_V^V$		0.247	0.336	0.164	0.351	0.110	0.351	0.385	
$\mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.769	0.854	0.775	0.761	0.110	0.761	0.826	
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_D^V$		0.000	0.000	0.707	0.000	0.110	0.442	0.748	0.442
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_V^V$		0.000	0.000	0.164	0.000	0.110	0.351	0.370	0.351
$\mathbf{x}_W^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.000	0.000	0.775	0.000	0.110	0.761	0.827	0.761
$\mathbf{x}_R^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.833	0.889	0.794	0.766	0.110	0.775	0.861	0.775
$\mathbf{x}_W^V \wedge \mathbf{x}_R^V \wedge \mathbf{x}_D^V \wedge \mathbf{x}_V^V$		0.000	0.000	0.794	0.000	0.110	0.766	0.861	0.816

Aus den Auswertungen können folgende Schlussfolgerungen gezogen werden: Treten keine unzulänglichen Merkmale auf, oder können diese robust detektiert und von der weiteren Betrachtung ausgeschlossen werden, bietet der Metaklassifikator die höchsten Erkennungsraten. Sind zudem die korrekten Wahrscheinlichkeiten bekannt, kann die Produktregel als Fusionsmethode in Erwägung gezogen werden. Im automotiven Einsatz hat sich jedoch gezeigt, dass eine zuverlässige Detektion unzulänglicher Merkmale, aufgrund der zahlreichen Ursachen und Erscheinungsformen, nicht ausnahmslos möglich ist. Zudem können die Ausgabewahrscheinlichkeiten der Basisklassifikatoren oftmals nur ungenau geschätzt werden. Vor diesem Hintergrund ist das (beschnittene) arithmetische Mittel als funktionaler Kombinator eine robuste Alternative zu den anderen betrachteten Verfahren.

5.5 Zusammenfassung

In diesem Kapitel wurden verschiedene Klassifikationsverfahren zur sensorübergreifenden Fusion von Merkmalen für die Erkennung von Fahrzeugen und Fußgängern betrachtet. Neben monolithischen Klassifikatoren wurden Klassifikatorensembles mittels funktionaler Kombina-

toren und Metaklassifikatoren eingeführt und bewertet. Als ein wesentlicher Aspekt für Klassifikationsverfahren im Kontext von multisensoriellen Fahrumfeldsystemen, wurde zudem die Leistungsfähigkeit der Ansätze bei fehlenden und unzulänglichen Merkmalen untersucht.

Die Güte der thematisierten Verfahren ist weitgehend unabhängig von der spezifischen Objektklasse. Fahrzeuge können mit den eingeführten Merkmalen jedoch zuverlässiger klassifiziert werden als Fußgänger. Von den untersuchten Verfahren erzielt die Metaklassifikation die höchsten Erkennungsraten. Allerdings ist das Verfahren lediglich unter fehlerlosen Merkmalszusammensetzungen geeignet. Ist die Güte einzelner Merkmale, etwa durch partielle Sensorausfälle beeinträchtigt, ist die Verwendung des arithmetischen Mittels als funktionaler Kombinator eine zweckmäßige Wahl. Grundsätzlich zeigt sich durch den Einsatz von mehr Sensoren ein signifikanter Anstieg in der Erkennungsleistung.

Im nächsten Kapitel werden die zielführenden Konzepte der Fahrzeug- und Fußgängerklassifikation aufgegriffen und in ein Gesamtsystem integriert. Dieses wird mittels geeigneter Evaluierungswerkzeuge auf repräsentativen Testszenarien bewertet.



„Montre molle au moment“ von Salvador Dalí (1904 - 1989).

6

Systembewertung

Eine Systembewertung bzw. -evaluierung generiert Gütemaße, die zur Systembeurteilung herangezogen werden können. Wesentliche Gütemaße im Kontext von Umfelderkennungssystemen sind Genauigkeiten, Objekterkennungsraten, Latenzen und Laufzeiten. Dadurch kann zum einen ein Vergleich zwischen unterschiedlichen Sensoren, Algorithmen und Parameterkonfigurationen durchgeführt, und zum anderen eine grundsätzliche Eignung der Fahrumfelderfassung für diverse Fahrerassistenzfunktionen überprüft werden. Damit in einem Fahrumfelderfassungssystem das Zusammenspiel zwischen Klassifikation und Tracking beurteilt werden kann, sind Evaluierungs- und Referenzdaten erforderlich. Evaluierungsdaten werden mittels Perzeptionsalgorithmen erzeugt, die auf Sensordaten arbeiten. Referenzdaten fungieren als Bewertungsgrundlage und ermöglichen, durch einen Abgleich mit der Systemausgabe, eine metrikgestützte Gewinnung der Gütemaße (siehe auch Abbildung 6.1).

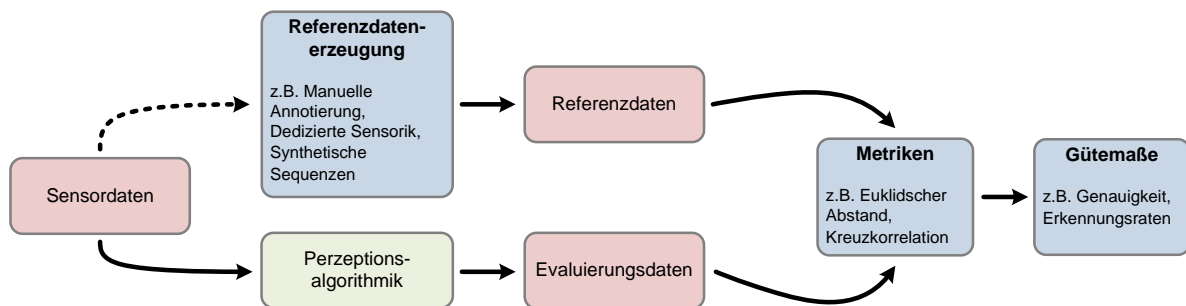


Abbildung 6.1: Übersichtsschaubild zur Systembewertung. Sensordaten sind die Ausgangsdaten zur Referenzdatenerzeugung und zur Evaluierungsdatengewinnung durch Perzeptionsalgorithmen. Durch den Abgleich der Referenz- und Evaluierungsdaten mit Hilfe von adäquaten Metriken sind Gütemaße ableitbar.

Das vorliegende Kapitel ist zweigeteilt. Zunächst werden semiautomatische (Abschnitt 6.1.1), algorithmusgetriebene (Abschnitt 6.1.2), sensorgestützte (Abschnitt 6.1.3) und hybride (Abschnitt 6.1.4) Verfahren zur Referenzdatengewinnung vorgestellt, die für eine Evaluierung von maschinellen Fahrumfelderfassungen geeignet sind. Aufgrund der speziellen Anforderungen

mussten zum Teil neue Vorgehensweisen entwickelt werden. Anschließend wird auf einige Metriken eingegangen mittels derer die Systemperformanz in Zahlen gefasst werden kann.

Im zweiten Teil des Kapitels wird die Systembewertung für das im Rahmen dieser Arbeit umgesetzte Fahrumfelderfassungssystem durchgeführt. Dabei wird die Systemleistung unter typischen Fahrsituationen und in verschiedenen Umgebungen, wie Stadt, Land und Autobahn, sowie unter verschiedenen Witterungsbedingungen ermittelt. Zudem werden die Synergieeffekte des multisensoriellen Sensorsystems bewertet.

6.1 Referenzdatengewinnung

Damit eine Systemevaluierung durchgeführt werden kann, sind Referenzdaten zwingend erforderlich. Diese ermöglichen, durch einen Abgleich mit der Systemausgabe, eine metrikgestützte Gewinnung der Gütemaße. Zur Erzeugung von Referenzdaten für die Evaluierung von Fahrumfelderfassungssystemen gibt es verschiedene Vorgehensweisen:

- Durch die rechnergestützte Simulation von Fahrscenen sowie den entsprechenden Sensordaten können nahezu beliebige Testsequenzen synthetisch erzeugt werden. Aufgrund der Erzeugungsmodelle, sind sämtliche Parameter des Fahrumfeldes bekannt. Die Umsetzung solcher Simulationsmodelle kann jedoch, insbesondere bei multisensoriellen Systemen, sehr aufwendig sein, da zusätzlich zu Objektgeometrien und -dynamiken ebenso alle Sensormodelle nachgebildet werden müssen. Des Weiteren kann durch diese Vorgehensweise eine reale Datenaufnahme lediglich approximiert werden.
- Von einem dedizierten Sensor können Referenzdaten abgeleitet werden. Dieser ist oftmals teuer und nur unter bestimmten Randbedingungen einsetzbar und somit kein permanenter Bestandteil des Fahrumfelderfassungssystems. Meist liefert solch ein Sensor, neben sehr hohen Zustandsgenauigkeiten, Objektklassenzugehörigkeiten und Latenzinformationen. Dadurch ist er, neben der Evaluierung von Klassifikationsalgorithmen, ebenso zur Bewertung von Trackingverfahren geeignet.
- Referenzdaten können zudem mittels einer manuellen Attributierung von Aufnahmesequenzen gewonnen werden. Dabei werden in den Sensordaten die Positionen und Klassenzugehörigkeiten aller relevanten Objekte per Hand markiert. Die dazu erforderlichen Arbeitsschritte können für den Anwender sehr monoton und zeitaufwendig sein. Zudem wird die erzielbare Positionsgenauigkeit durch die Genauigkeit der manuellen Markierungen begrenzt. Der Vorteil einer manuellen Attributierung von Testsequenzen liegt in der hohen Flexibilität hinsichtlich der zu markierenden Bereiche sowie in der uneingeschränkten Szenenauswahl.

In der Literatur finden sich vornehmlich bildgestützte Verfahren zur Referenzdatenerzeugung. Dörmann [DM00] und Jaynes [JWSX02] entwickelten die Systeme Viper und ODViS zur Leistungsbewertung von Erkennungssystemen, die ausschließlich auf Videosequenzen basieren. Diese Werkzeuge bieten Schnittstellen zur Referenzdatenerzeugung, Metriken zur Evaluierung und diverse Visualisierungskomponenten. Im Applikationsbereich der Videoüberwachung stellt Black et al. [BER03] ein System vor, das eine pseudosynthetische Bilddatenerzeugung ermöglicht. Dazu werden Bildsequenzen von Fußgängertrajektorien extrahiert und vor neue wechselnde Hintergrundansichten eingefügt. Dadurch kann eine Vielzahl von Szenen synthetisiert werden. Bertozzi et al. [BBG⁺04] beschreiben ein System zur Evaluierung von bildbasierten Fußgängererkennungsalgorithmen. Das vorgestellte Werkzeug ermöglicht manuelle und algo-

rithmusgetriebene Attributierungen von Bildsequenzen, sowie eine anschließende Analyse der Erkennungsraten. Ein Anwender erzielt etwa 100 Bildannotationen in der Stunde. Eine manuelle Bild-für-Bild-Annotierung von Sensordaten ist somit entsprechend zeitaufwändig.

Im nächsten Abschnitt wird ein Werkzeug vorgestellt, das eine semiautomatische Erzeugung von Referenzdaten ermöglicht. Es ist dabei nicht auf Bilddaten beschränkt, sondern erlaubt zudem eine effiziente Annotierung von räumlichen Sensordaten. Die Konzeption und Umsetzung des Werkzeugs zur Referenzdatenerzeugung erfolgte in Kooperation mit dem FORWISS¹ Institut der Universität Passau. Eine detaillierte Betrachtung findet sich in der gemeinsamen Veröffentlichung [6].

6.1.1 Semiautomatische Referenzdatenerzeugung

Zur Ableitung von räumlichen Informationen wird vorausgesetzt, dass an der Erfassungssensorik zumindest zwei Kameras oder alternativ mindestens ein bild- und ein entfernungsgebender Sensor beteiligt sind. Im hier exemplarisch verwendeten System wird für die Referenzdatenerzeugung eine Videokamera und ein Laserscanner eingesetzt. Neben den Sensoranforderungen wird von der Annahme ausgegangen, dass sich alle relevanten Objekte des Fahrumfeldes, wie beispielsweise Fußgänger und Fahrzeuge, durch einen umschließenden Quader (siehe Abbildung 6.2a) ausreichend genau approximieren lassen.

Ein Keyframe-Konzept, ist neben der Möglichkeit Daten in drei Dimensionen zu markieren, die wesentliche Neuerung des Annotierungswerkzeugs. Statt Objekte Bild-für-Bild zu markieren, werden die Objekttrajektorien durch einzelne Kontrollquader beschrieben. Eine lineare Interpolation, hinsichtlich der Position und Orientierung, zwischen den Kontrollquadern (siehe Abbildung 6.2b) stellt eine durchgängige Bild-für-Bild-Markierung sicher.

6.1.1.1 Vorgehensweise zur Attributierung einer Szene

Zwei Arbeitsschritte sind nötig, um einen Kontrollquader positionsgenau zu plazieren. Zunächst wird das zu attributierende Objekt im bildgebenden Sensor mit einem umschließenden Rechteck markiert, das die Rückseite des Kontrollquaders darstellt. Dieser Quader wird auf eine vorgegebene Initialdistanz (z. B. 3 m) zum Kamerasensor gesetzt. Dadurch ist nicht gewährleistet, dass die Abmessungen des Quaders den wirklichen Objektabmessungen entsprechen. Lediglich die Projektion in das Kamerabild ist korrekt (siehe Abbildung 6.3a). Im zweiten Arbeitsschritt wird die Entfernung des Objektes zur Kamera manuell korrigiert. Dazu wird der Kontrollquader entlang den Sichtstrahlen der Kamera verschoben. Zeitgleich wird er in einer Art und Weise skaliert daß die Projektionsgröße des Quaders im Bildbereich konstant bleibt. Der zweite Arbeitsschritt ist abgeschlossen, wenn die Rückfläche des Kontrollquaders auf den Entfernungsdaten zu liegen kommt (siehe Abbildung 6.3b), bzw. im Fall von zwei Kameras, die korrespondierenden Objektbereiche überdeckt. Anschließend kann noch die Tiefenabmessung und die Orientierung des Quaders angepasst werden.

Durch die beiden Arbeitsschritte wird ein einzelner Kontrollquader erzeugt, der anhand der Zustandsgrößen Position, Dimension und Orientierung beschrieben wird. Zur Evaluierung einer gesamten Sequenz ist es jedoch notwendig, für jeden diskreten Zeitschritt Referenzzustands-

¹ Institut für Softwaresysteme in technischen Anwendungen der Informatik.

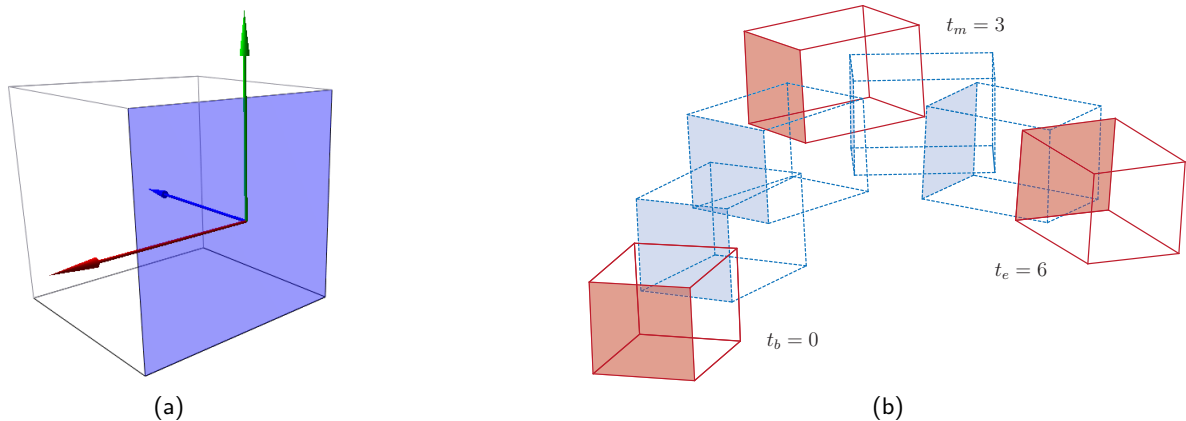


Abbildung 6.2: Quadermodell zur Approximierung von Verkehrsteilnehmern. In Abbildung (a) ist ein Quadermodell mit Ursprung und Ausrichtung des lokalen Koordinatensystems dargestellt. Abbildung (b) zeigt linear interpolierte Quader (gestrichelte Linien) zu den Zykluszeiten $t = 1, 2, 4, 5$ zwischen manuell platzierten Kontrollquadern (durchgezogene Linien) zu den Zykluszeiten $t_b = 0, t_m = 3$ und $t_e = 6$. Die Quaderfläche, in deren Zentrum der Ursprung des lokalen Koordinatensystems liegt, ist farblich hervorgehoben.

werte verfügbar zu haben. Dies kann erzielt werden, indem die obigen Arbeitsschritte für jeden Zeitschritt gesondert durchgeführt werden.

Dieses zeitaufwendige Vorgehen kann optimiert werden, indem für ein Objekt lediglich eine Folge von Kontrollquadern platziert wird und durch eine lineare Interpolation die Objektzustände zwischen den Kontrollquadern zum Zeitpunkt t_b und t_e approximiert werden (siehe Abbildung 6.3b). Je nach Fahrmanöver kann die Interpolation dabei sowohl in Fahrzeugkoordinaten als auch in Weltkoordinaten angestoßen werden. Sei $p : \mathcal{T} \rightarrow \mathbb{R}^3$ eine Funktion, die dem

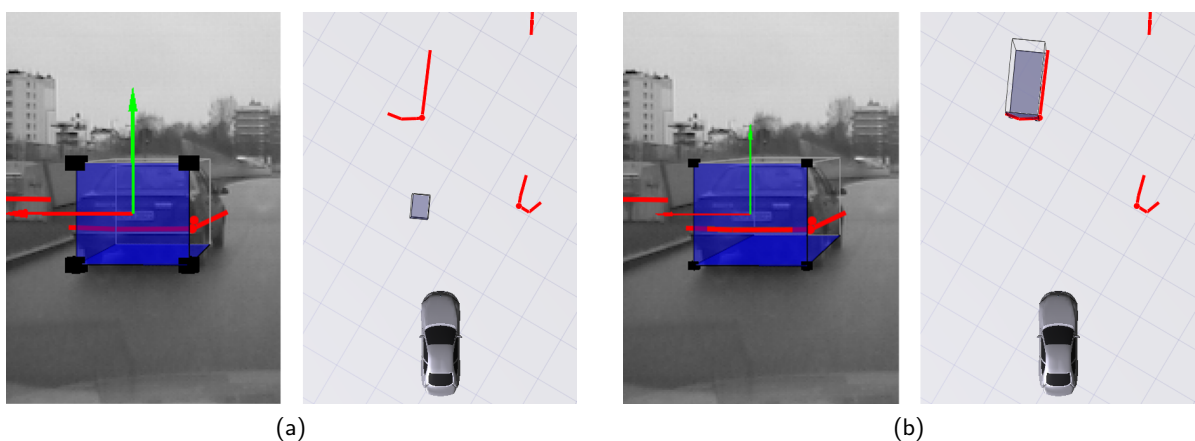


Abbildung 6.3: Arbeitsschritte zur Markierung eines Objektes in drei Dimensionen. Als ersten Schritt wird das Objekt im bildgebenden Sensor markiert (a). Anschließend wird die Entfernung, korrespondierend zu den Entfernungsdaten, justiert (b).

Objekt in jedem diskreten Zeitwert einen Positionswert zuordnet. So gilt für die Positionen der zu interpolierenden Zeitschritte t_i :

$$\forall t_i \in \mathcal{T}, t_b < t_i < t_e \quad p(t_i) = p(t_b) + \frac{p(t_e) - p(t_b)}{t_e - t_b} \cdot (t_i - t_b) \quad (6.1.1)$$

Die Objektausdehnungen und -orientierungen der Quader zwischen den Kontrollquadern werden analog zur Position interpoliert. Zur effizienten Annotierung von kompletten Sequenzen hat sich ein objektweises Vorgehen und eine „Divide and Conquer“-Strategie bewährt:

1. Zunächst werden zwei Kontrollquader plziert. Einer zum Zeitpunkt t_b , in dem das Objekt zum ersten Mal in der Sensorik sichtbar wird. Der zweite Kontrollquader zum Zeitpunkt t_e , sobald das Objekt den Sichtbereich der Sensorik verlässt. Das Zeitintervall $[t_b, t_e]$ wird als eine Sektion bezeichnet.
2. Falls die Interpolationsgenauigkeit in dieser Sektion $[t_b, t_e]$ nicht ausreichen sollte, kann ein weiterer Kontrollquader in der Sektionsmitte t_m eingefügt und repositioniert werden. Dadurch werden zwei neue Sektionen $[t_b, t_m]$ und $[t_m, t_e]$ erzeugt, für die jeweils eine separate lineare Interpolation angestossen wird. Insgesamt wird dadurch die Objekttrajektorie besser approximiert (siehe auch Abbildung 6.2b).
3. Der zweite Schritt wird mit den Untersektionen solange rekursiv wiederholt, bis die gewünschte Genauigkeit erreicht worden ist.

6.1.1.2 Bewertung

Mit der „Divide and Conquer“-Strategie kann eine komplexe Szene, wie jene in Abbildung 6.4a und 6.4b mit 12 Fußgängern und einem Fahrzeug, in weniger als 10 Minuten annotiert werden. Für die Autobahnscene in Abbildung 6.4c und 6.4d, die 20 Fahrzeuge enthält, wurden 15 Minuten Annotierungszeit benötigt. Bei beiden Szenarien bewegen sich die Objekte mit annähernd konstanter Geschwindigkeit bzw. sind weitgehend stationär. Dadurch wird eine geringe Anzahl von Kontrollquadern zur Approximierung der Trajektorie benötigt. In Szenen mit nichtlinearen Bewegungsphasen müssen entsprechend mehr Kontrollquader plziert werden, wodurch sich der Annotierungsaufwand erhöht. Eine zusätzliche Steigerung der Annotierungseffizienz kann durch nichtlineare (z. B. kubische) Interpolationsverfahren erzielt werden.

Um die zu erwartende Genauigkeit dieses Verfahrens zu bestimmen, wurden mehrere Szenen mit einem hochgenauen dedizierten Referenzsensor (siehe Abschnitt 6.1.3) vermessen und nachträglich mit dem semiautomatischen Keyframe-Verfahren attribuiert. In Tabelle 6.1 ist ein Vergleich der ermittelten Abweichungen zwischen Referenz und manueller Attributierung für Fußgänger sowie Fahrzeuge in zwei unterschiedliche Szenarien dargestellt. In dem ersten Szenario bewegen sich die Zielobjekte auf einfachen linearen Trajektorien. In dem zweiten Szenario verhalten sie sich hingegen hoch dynamisch. Die Qualität der Entfernungs- und Geschwindigkeitsbestimmung wird nur sehr gering von der linearen Interpolation beeinflusst. Für die laterale Ablage (Winkel ϕ) ist hingegen ein deutlicher Zusammenhang zu erkennen. Die Qualität des Attributs Geschwindigkeit konnte für Fußgänger nicht evaluiert werden, da der dedizierte Referenzsensor hierfür keine Messungen liefert.

Die Referenz, die durch diese manuelle Attributierung gewonnen werden kann, ist gut geeignet die Performanz von Hypothesengenerierungs- und Klassifikationsverfahren zu bestimmen. Des Weiteren können aus den markierten Sensordaten Trainingsdaten für einen Klassifikator gewonnen werden. Für die Evaluierung von Trackingverfahren, speziell zur Bestimmung von

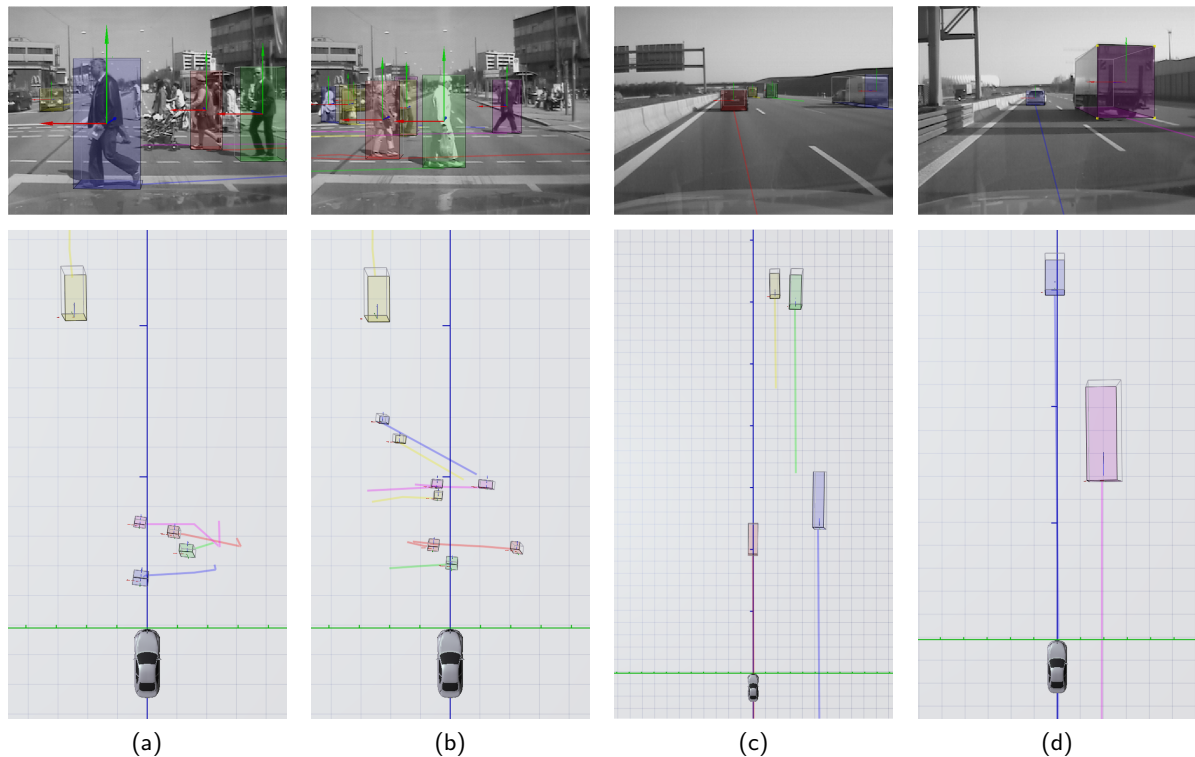


Abbildung 6.4: Zwei semiautomatisch attributierte Szenen: Die Abbildungen (a) und (b) zeigen Szenenschnappschüsse einer Stadtszene (Fußgängerüberweg) mit 12 Fußgängern und einem Fahrzeug, zu unterschiedlichen Zeitpunkten. In den Abbildungen (c) und (d) sind Ausschnitte einer Autobahnszene mit 20 Fahrzeugen abgebildet. Die obere Bildreihe gibt die Ansicht auf die jeweilige Szene durch eine Videokamera wieder. Die untere Bildreihe zeigt das Ergebnis der Attributierung aus einer Vogelperspektive.

Zustandsgenauigkeiten, ist ein manuelles Attributieren jedoch nur bedingt geeignet. Die Genauigkeit der Referenz ist stark von der Markierungsgenauigkeit und den dafür verwendeten Sensoren abhängig. Bei hohen Relativgeschwindigkeiten wird die Qualität der Referenz durch Latenzzeiten, die durch die Sensorik oder das Aufzeichnungsverfahren bedingt sind, negativ beeinflusst.

6.1.2 Algorithmusgestützte Referenzdatenerzeugung

Im vorherigen Abschnitt wurde eine Methode erläutert, wie räumliche Sensordaten sehr effizient annotiert werden können. Bei diesem Verfahren besteht die Hauptaufgabe des Anwenders darin, neue Objekte zu erzeugen, diese zu klassifizieren, die Qualität der Interpolation zwischen den einzelnen Keyframes zu überwachen und gegebenenfalls neue Keyframes einzufügen. Insbesondere bei Szenen mit vielen Objekten, die sich nicht linear bewegen, können diese Arbeitsschritte viel Zeit in Anspruch nehmen. Im Folgenden wird deshalb eine Erweiterung dieses Ansatzes vorgestellt, bei der durch eine Vorverarbeitung automatisch Referenzdaten erzeugt werden. Hierfür können Algorithmen gewählt werden, die nicht den Ansprüchen eines Echtzeit-

	Abstand r (RMS)	Winkel ϕ (RMS)	Geschwindigkeit v (RMS)
Fahrzeuge Szenario 1 (1443 Zyklen)	0.09 m	0.3°	0.37 $\frac{m}{s}$
Fahrzeuge Szenario 2 (1865 Zyklen)	0.10 m	0.5°	0.36 $\frac{m}{s}$
Fußgänger Szenario 1 (1250 Zyklen)	0.14 m	0.4°	nicht verfügbar
Fußgänger Szenario 2 (2680 Zyklen)	0.20 m	0.8°	nicht verfügbar

Tabelle 6.1: Vergleich der Genauigkeiten zwischen semiautomatisch erzeugten Referenzdaten für Fahrzeuge und Fußgänger mit den Genauigkeiten, die mit einem dedizierten Referenzsensor erreicht werden können. Es wurden jeweils zwei Szenarien untersucht: In Szenario 1 bewegen sich die Zielobjekte auf einfachen linearen Trajektorien. Im Szenario 2 verhalten sich die Zielobjekte hoch dynamisch.

systems genügen müssen. Bei diesem Ansatz arbeitet der Anwender nicht mehr ausschließlich auf den Rohdaten. Vielmehr bekommt er vom System sowohl für die Klassifikation, als auch für die Wahl der Keyframes Vorschläge unterbreitet. Dadurch lässt sich der Arbeitsaufwand auf ein Überwachen und Korrigieren der Referenzdaten aus der Vorverarbeitung reduzieren. Aus diesem Zusammenhang kann folgende Grundregel abgeleitet werden: Je besser die Vorverarbeitung, desto schneller und effizienter die Attributierung.

Im Folgenden wird ein Fahrumfelderfassungssystem als eine mögliche Art der Vorverarbeitung betrachtet. Das Fahrumfelderfassungssystem liefert zu jedem Zeitpunkt t eine Schätzung des Objektzustandes \mathbf{x} . Verglichen mit einer manuellen Vorgehensweise entspricht dies einer Bild-für-Bild Annotation der Szene. Im Gegensatz zu dem Keyframe Konzept ist ein nachträgliches manuelles Korrigieren der Daten sehr ineffizient, da statt weniger Stützstellen jeder Zeitpunkt t betrachtet werden muss. Deshalb ist es sinnvoll, die durch das Fahrumfelderfassungssystem erzeugten Daten $\mathcal{M} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ durch eine Menge an Keyframes $\mathcal{K} = \{\mathbf{k}_1, \dots, \mathbf{k}_m\}$ anzunähern. Hierzu wird ein iteratives Verfahren verwendet, das \mathcal{M} durch eine Menge an Regressionsgeraden beschreibt (siehe auch Algorithmus 1). Beginnend bei dem ersten Element von \mathcal{M} werden solange weitere Elemente $\mathbf{x}_i \in \mathcal{M}$ zu einer Regressionsgeraden hinzugefügt, bis die Standardabweichung dieser Approximation einen Schwellwert τ übersteigt. Die Menge \mathcal{K} an Keyframes wird um den aktuellen Punkt \mathbf{x}_i erweitert und eine neue Ausgleichsgerade begonnen.

In Abbildung 6.5 ist die Trajektorie eines vorausfahrenden Fahrzeugs auf einer kurvenreichen Strecke dargestellt. Die Trajektorie besteht aus insgesamt 789 einzelnen Punkten. Für diese Szene sind 63 Keyframes ausreichend, um die Daten bei einer Standardabweichung von maximal 3 cm anzunähern. Für Teilabschnitte der Szene, die Kurven enthalten, werden deutlich mehr Keyframes benötigt, als für lineare Abschnitte.

Das semiautomatische Annotieren in Verbindung mit der soeben vorgestellten Vorverarbeitung ermöglicht eine effiziente räumliche Attributierung von vielfältigen Fahrscenen. Für die Evaluierung von Trackingverfahren, speziell zur Bestimmung von Zustandsgenauigkeiten, ist ein manuelles Attributieren allerdings nur bedingt geeignet. So können zum Beispiel nichtli-

Algorithmus 1 : Iteratives Verfahren, um eine Trajektorie durch eine Menge von Keyframes zu approximieren. τ ist die maximal zulässige Standardabweichung der Approximation.

Eingabe : Menge an Positionen $\mathcal{M} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$
Ausgabe : Menge an Keyframes $\mathcal{K} = \{\mathbf{k}_1, \dots, \mathbf{k}_m\}$

- 1 Füge \mathbf{x}_1 zu \mathcal{K} hinzu;
- 2 $j = 1$;
- 3 **for** $i=1$ **to** n **do**
- 4 Füge \mathbf{x}_i zur aktuellen Regressionsgeraden g_j hinzu;
- 5 **if** Standardabweichung von $g_j \geq \tau$ **then**
- 6 Füge \mathbf{x}_i zu \mathcal{K} hinzu;
- 7 $j = j + 1$;
- 8 **end**
- 9 **end**
- 10 Füge \mathbf{x}_n zu \mathcal{K} hinzu;

neare Objekttrajektorien, nur ungenau durch einfache Linienzüge approximiert werden. Zum anderen ist die Genauigkeit von den direkt ableitbaren Attributen wie Position, Ausdehnung oder Orientierung abhängig von der Qualität der verwendeten Sensorik. Mit einer semiautomatischen Referenzdatenerzeugung ist es somit nicht möglich, Daten zu erzeugen, die präziser sind als die zugrundeliegende Sensorik. Dies ist im Allgemeinen ausreichend um Aussagen hinsichtlich der Klassifikationsgüte zu treffen. Für die Evaluierung von Trackingverfahren werden hingegen höhere Anforderungen an die Genauigkeit der Referenzdaten gestellt. Insbesondere wenn abgeleitete oder nicht direkt messbare Größen, wie die Beschleunigung oder die Gierrate untersucht werden sollen, ist es notwendig einen zusätzlichen, von der Fahrumfelderfassung unabhängigen Sensor zu verwenden, der die Attribute des Referenzobjektes präzise ermittelt.

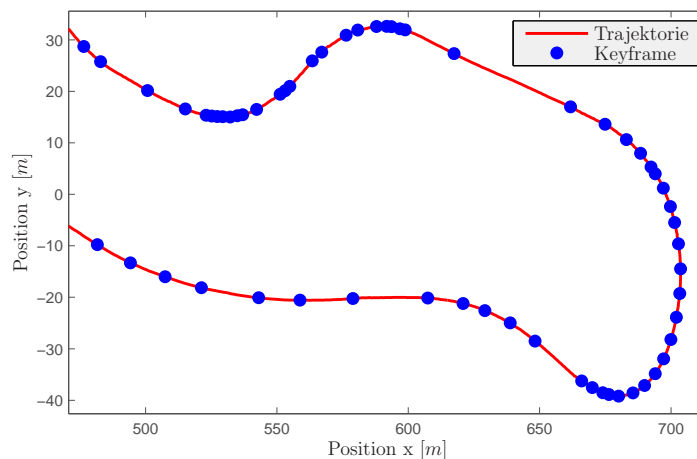


Abbildung 6.5: Die Abbildung zeigt die Trajektorie eines vorausfahrenden Fahrzeugs auf einer kurvenreichen Strecke. 63 Keyframes sind ausreichend, um den Fahrzeugverlauf mit einer durchschnittlichen Abweichung von 3 cm zu approximieren. Dies entspricht einer Datenreduktion um den Faktor 10 im Vergleich zu der Gesamttrajektorie.

6.1.3 Referenzsensor

Im Folgenden wird am Beispiel einer Inertialplattform mit GPS Empfänger erläutert, wie durch zusätzliche Sensorik hochgenaue Referenzdaten erzeugt werden können. Das vorgestellte Verfahren erweitert den Ansatz von Vogel [Vog07], mit dem Fußgänger zentimetergenau lokalisiert und getrackt werden können, um weitere Attribute wie Orientierung, Geschwindigkeit und Beschleunigung von beliebigen Referenzobjekten. Abschließend werden die Vor- und Nachteile dieser Technik diskutiert.

6.1.3.1 Inertialplattform mit Carrier Phase Differential GPS

Grundlage des Verfahrens sind zusätzliche zeitlich synchronisierte Sensoren, die sowohl im Eigenfahrzeug, als auch im Referenzobjekt verbaut sind. Im Eigenfahrzeug befindet sich eine Inertialplattform², die mittels Carrier Phase Differential GPS³ (CPD-GPS) die Ausrichtung nach Norden ϕ_{RT}^{GPS} , sowie die Position $\mathbf{p}_{RT}^{GPS} \in \mathbb{R}^3$ bezüglich eines globalen Weltkoordinatensystems bestimmt. Im Allgemeinen entspricht der Verbauort der Inertialplattform bzw. deren Koordinatensystem C_{RT} nicht dem Fahrzeugkoordinatensystem C_{ego} . In den nachfolgenden Berechnungen wird dies durch die zusätzliche Transformation $\mathbf{T}_{RT \rightarrow ego}$ ausgeglichen. Eine schematische Darstellung des Versuchsaufbaus bzw. der beteiligten Koordinatensysteme ist in Abbildung 6.6 zu sehen. Eine zweite Inertialplattform mit CPD-GPS Empfänger, die am Zielobjekt angebracht ist, liefert Daten über die Position $\mathbf{p}_{ref}^{GPS} \in \mathbb{R}^3$ und die Ausrichtung ϕ_{ref}^{GPS} des Referenzobjektes. Mit dem zusätzlichen Sensor können weitere Attribute wie die Geschwindigkeit v_{ref} , die Beschleunigung a_{ref} und die Gierrate $\dot{\phi}_{ref}$ gemessen werden. Durch die zeitliche Synchronisation zwischen dem Referenzobjekt und dem Eigenfahrzeug ist es letztendlich möglich aufgezeichnete Sensordaten mit den Messwerten des Referenzsensors zu attributieren.

Aus den Messwerten der dedizierten Sensorik können Zustandsgrößen wie die Orientierung, die Geschwindigkeit, die Beschleunigung und die Gierrate direkt abgeleitet werden. Die Position des Referenzobjektes \mathbf{p}_{ref}^{ego} , relativ zu dem Koordinatensystem des Eigenfahrzeug C_{ego} , lässt sich aus den Messwerten ϕ_{RT}^{GPS} , \mathbf{p}_{RT}^{GPS} und \mathbf{p}_{ref}^{GPS} wie folgt berechnen:

$$\mathbf{R} = \begin{pmatrix} \cos(\phi_{RT}^{GPS}) & -\sin(\phi_{RT}^{GPS}) & 0 \\ \sin(\phi_{RT}^{GPS}) & \cos(\phi_{RT}^{GPS}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (6.1.2)$$

$$\mathbf{T}_{RT \rightarrow GPS} = \begin{bmatrix} \mathbf{R} & \mathbf{p}_{RT}^{GPS} \\ 0 & 1 \end{bmatrix} \quad (6.1.3)$$

$$\mathbf{p}_{ref}^{ego} = \mathbf{T}_{RT \rightarrow ego} (\mathbf{T}_{RT \rightarrow GPS})^{-1} \mathbf{p}_{ref}^{GPS} \quad (6.1.4)$$

² Eine Inertialplattform besteht aus mehreren Gyroskopen und Beschleunigungssensoren. Durch eine Kombination mit einem CPD-GPS Empfänger können die Position, Geschwindigkeit, Beschleunigung sowie die Orientierung der Plattform sehr präzise gemessen werden. Als Inertialsystem wurde eine RT3003 [Oxf08] verwendet.

³ Satellitengestütztes System zur Bestimmung der Position auf der Erdoberfläche.

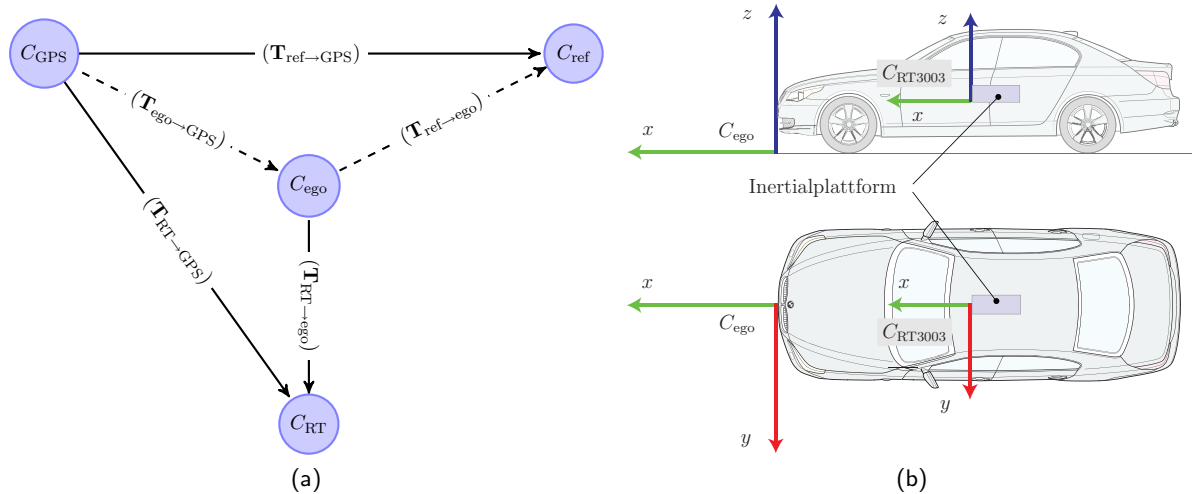


Abbildung 6.6: Referenzdatengewinnung mit einem dedizierten Sensor (Inertialplattform mit CPD-GPS Empfänger). In Abbildung (a) ist der Spatial-Relationship-Graph für den Versuchsaufbau dargestellt. Die beiden Kanten $(\mathbf{T}_{ego \rightarrow GPS})$ bzw. $(\mathbf{T}_{ref \rightarrow GPS})$ entsprechen den Messungen der Referenzsensorik für die Position und Orientierung des Eigenfahrzeugs bzw. des Zielobjektes. Über die Kante $(\mathbf{T}_{ref \rightarrow ego})$ lässt sich die Position des Zielobjektes bezüglich des Eigenfahrzeugs bestimmen. In Abbildung (b) ist die Verbauposition der dedizierten Sensorik im Eigenfahrzeug abgebildet.

6.1.3.2 Bewertung

Durch den Einsatz von dedizierter Sensorik erfolgt die Annotation von Szenen bereits zum Zeitpunkt der Aufnahme. Deshalb ist mit diesem Ansatz ein zeitaufwendiges manuelles Nachbearbeiten der aufgezeichneten Daten nicht mehr erforderlich. Vielmehr können Attribute wie die Position, die Geschwindigkeit oder die Beschleunigung direkt aus der Messung abgeleitet werden. Weitere Attribute wie zum Beispiel die Objektmaße oder die Klassifikation stehen als a-priori-Wissen zur Verfügung, wenn die Referenzobjekte mit zusätzlicher Sensorik ausgestattet werden. Darüber hinaus ist diese Art der Referenzdatenerzeugung unabhängig von der zu evaluierenden Sensorik. So können zum Beispiel höhere Reichweiten von mehr als 200 m sowie Erfassungsbereiche von bis zu 360° realisiert werden. Insbesondere bei der Bewertung von Trackingverfahren sind hochgenaue und unabhängige Referenzdaten entscheidend. Durch dedizierte Sensorik können zum Beispiel Attribute wie die Gierrate oder die Beschleunigung direkt gemessen werden, die sonst nur indirekt bestimmt werden könnten. Neben der Messgenauigkeit der dedizierten Sensorik beeinflussen die zeitliche Synchronisation und vorhandene Latenzzeiten die Qualität der Referenzdaten. Für das Fallbeispiel aus dem vorherigen Abschnitt sind in Tabelle 6.2 die durch den Hersteller angegebenen Genauigkeiten der jeweiligen Attribute dargestellt.

Mit diesem Verfahren können nur vordefinierte Szenarien mit wenigen Objekten attribuiert werden, da alle Objekte mit zusätzlicher Sensorik ausgestattet werden müssen. Realitätsnahe Aufnahmen, wie zum Beispiel in einer Stadt mit mehreren Fußgängern und Fahrzeugen sind, wenn überhaupt, nur mit sehr hohem organisatorischen und finanziellen Aufwand realisierbar.

	Position \mathbf{p} (RMS)	Geschwindigkeit v (RMS)	Beschleunigung a (RMS)	Orientierung ϕ (RMS)	Gierrate $\dot{\phi}$ (RMS)
Eigenfahrzeug	0.02 m	$0.02 \frac{m}{s}$	$0.01 \frac{m}{s^2}$	0.1°	$0.1 \frac{^\circ}{s}$
Referenz	0.02 m	$0.02 \frac{m}{s}$	$0.01 \frac{m}{s^2}$	0.1°	$0.1 \frac{^\circ}{s}$

Tabelle 6.2: Genauigkeiten der verwendeten Inertialplattform RT3003 für das Referenzobjekt sowie für das Eigenfahrzeug.

Für derartige Szenarien stellt die Kombination von dedizierter Sensorik und manueller Nachbearbeitung einen guten Kompromiss zwischen Genauigkeit und Kosten dar. Diese hybride Referenzdatenerzeugung wird im nächsten Abschnitt genauer betrachtet.

6.1.4 Hybride Referenzdatenerzeugung

In sehr komplexen und realitätsnahen Szenarien mit mehreren Fußgängern und Fahrzeugen, wie zum Beispiel in einer Stadt, ist es nicht praktikabel jeden Verkehrsteilnehmer mit einer eigenen zusätzlichen Referenzsensorik auszustatten. Die verfügbare Sensorik sollte auf die für diese Szene wichtigsten Teilnehmer beschränkt werden. Die restlichen Objekte können, wie in Abschnitt 6.1.1 beschrieben, nachträglich manuell attribuiert werden. Durch diese Kombination können sehr flexibel und effizient komplexe Szenarien bearbeitet werden. In Abbildung 6.7 trägt ein Fußgänger zum Beispiel einen GPS-Empfänger auf dem Rücken. Für ihn stehen somit hochgenaue Referenzdaten zur Verfügung. Die restlichen Objekte dieser Szene wurden nachträglich manuell attribuiert.

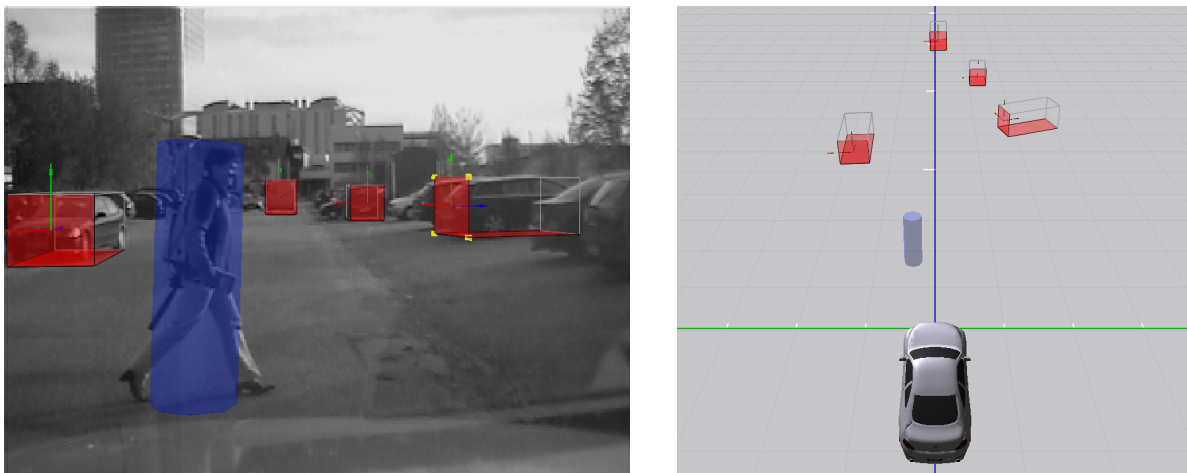


Abbildung 6.7: Kombination aus dedizierter Sensorik und manueller Nachbearbeitung (Hybride Referenzdatenerzeugung). Der Fußgänger trägt einen GPS-Empfänger auf dem Rücken. Für ihn stehen somit hochgenaue Referenzdaten (blauer Zylinder) zur Verfügung. Die vier Fahrzeuge in der Szene wurden nachträglich manuell attribuiert.

6.1.5 Zusammenfassung und Bewertung

In den vorherigen Abschnitten wurden mehrere Möglichkeiten aufgezeigt, wie für ein Fahrumfelderfassungssystem Referenzdaten erzeugt werden können. Die vorgestellten Verfahren adressieren im Wesentlichen die zwei größten Anforderungen, die im Allgemeinen an ein System zur Gewinnung von Referenzdaten gestellt werden: Zum einen sollte der gesamte Prozess so effizient wie möglich sein. Zum anderen werden insbesondere für die Evaluierung von Trackingverfahren sehr präzise und von der verwendeten Sensorik unabhängige Referenzdaten benötigt. Um aussagekräftige Daten zu erhalten, müssen sehr große Datenmengen unterschiedlichster Szenarien attribuiert werden. Hierzu wurden zwei Ansätze aufgezeigt, die die zu bearbeitende Datenmenge durch ein Keyframe-Konzept um den Faktor 5–10 reduziert. In Verbindung mit einer „Divide and Conquer“-Strategie können somit große Szenenkatologe sehr effizient attribuiert werden. Wird zusätzlich eine maschinelle Vorverarbeitung (Fahrumfelderfassung) eingesetzt, so muss der Anwender die Vorschläge der Vorverarbeitung nur noch überwachen und gegebenenfalls ergänzen.

Mit den semiautomatischen Verfahren können Attribute, wie die Position und die Klassifizierung von Objekten, präzise erfasst werden. Allerdings ist die Qualität der Referenzdaten stark mit der Genauigkeit der zugrundeliegenden Sensorik korreliert. Darüber hinaus können Attribute wie die Geschwindigkeit, die Beschleunigung oder die Gierrate nur sehr eingeschränkt und in schlechter Qualität ermittelt werden. Aus diesem Grund wurde ein Verfahren vorgestellt, das mit Hilfe eines dedizierten Referenzsensors sehr effizient hochgenaue und unabhängige Referenzdaten erzeugt. Allerdings ist die Attributierung von realitätsnahen Szenarien mit diesem Verfahren nur mit sehr hohem organisatorischen und finanziellen Aufwand möglich, da alle beteiligten Objekte mit zusätzlicher Sensorik ausgestattet werden müssen. Für derartige Szenarien ist die Kombination von dedizierter Sensorik und manueller Nachbearbeitung ein guter Kompromiss zwischen Genauigkeit und Kosten.

Die Gewinnung von Referenzdaten ist trotz der vorgestellten Verbesserungen ein arbeits- bzw. kostenintensiver Prozess. Eine pauschale Aussage, welches der vorgestellten Verfahren bevorzugt werden soll, kann nicht getroffen werden. Vielmehr ist die Wahl sehr stark von den Kriterien und Metriken der Evaluierung abhängig. So sind zum Beispiel für einen Vergleich der Genauigkeit von verschiedenen Trackingverfahren hochgenaue Referenzdaten erforderlich. Um Aussagen über die Erkennungsraten eines Klassifikators oder über die Gesamtperformanz eines Fahrumfelderfassungssystems treffen zu können, sind hingegen manuell attribuierte Daten ausreichend.

6.2 Evaluierung

In diesem Abschnitt wird das Gesamtsystem zur Fahrumfelderfassung hinsichtlich Erkennungsraten und Laufzeit anhand von repräsentativen Testszenerien bewertet. Nach einem Überblick über die verwendeten Szenarien und einer Beschreibung der Testsystemparametrisierung, werden Evaluierungsergebnisse und Problemkonstellationen aufgezeigt. Zunächst wird auf die Metriken eingegangen, die zur Leistungsbewertung herangezogen werden.

6.2.1 Metriken

Metriken sind in diesem Zusammenhang Indikatoren, anhand derer die Systemperformanz in Zahlen gefasst werden kann. Sie sollten möglichst intuitiv sein und alle signifikanten Leistungsfaktoren des Systems abdecken. Ein Abgleich der Ergebnisse einer maschinellen Fahrumfelderfassung mit einer, im Vorfeld generierten, Referenz bildet die Grundlage für die Berechnung der Metriken (siehe auch Anhang B für eine detaillierte Übersicht über die Bewertungskriterien zur Klassifikation).

Im Rahmen einer Fahrumfelderfassung können Metriken in drei Kategorien untergliedert werden. Die eine Gruppe spiegelt dabei vornehmlich die (Fehl-)Erkennungsrate wider, mit der Objekte klassifiziert werden. Die andere Gruppe beinhaltet die Genauigkeiten, mit der Objekte getrackt werden. Die dritte Kategorie umfasst applikationsspezifische Kriterien, wie die benötigte Rechenzeit oder dedizierte Sensorerfassungsbereiche. In den nächsten drei Abschnitten werden die für die Evaluierung des Gesamtsystems verwendeten Metriken beschrieben.

6.2.1.1 Metriken für die Klassifikation

Erkennungsrate (Richtig-Positiv-Rate): Ein Objekt gilt als erkannt, wenn durch die Klassifikation innerhalb der Objektlebenszeit die korrekte Objektklasse bestimmt wurde, und der Abstand zwischen Referenzobjekt und Evaluierungsobjekt unter einem Wert δ liegt. Die Erkennungsrate wird objektspezifisch für Fahrzeuge als RPR^V und für Fußgänger als RPR^P notiert. Diese Bewertungskriterien geben das Verhältnis von richtig klassifizierten Objekten zu allen verfügbaren Objekten an. In der Auswertung wurde $\delta = 1.5$ m gewählt.

Fehlerkennungsrate pro Stunde: Diese Metrik bildet die Fehlerkennungsrate (FER) auf eine zeitliche Basis ab und wird mit FER_h abgekürzt. Sie ist durch den Quotienten der Fehlerkennungen pro Stunde definiert. Die Erkennungsrate wird objektspezifisch für Fahrzeuge als FER_h^V und für Fußgänger als FER_h^P angegeben.

Matthews Correlation Coefficient: Der Matthews Correlation Coefficient [BBC⁺00] kombiniert die Erkennungs- und die Fehlerkennungsrate in einer gemeinsamen Metrik und ermöglicht somit einen direkten Vergleich zwischen zwei Klassifikatoren. Er ist wie folgt definiert:

$$\Phi_{MCC} = \frac{t_p t_n - f_p f_n}{\sqrt{(t_p + f_p)(t_p + f_n)(t_n + f_p)(t_n + f_n)}} \quad (6.2.1)$$

Hierbei bezeichnet t_p und t_n die Anzahl der korrekt bzw. nicht erkannten Fahrzeuge und Fußgänger. Analog dazu beschreiben f_p und f_n die korrekt klassifizierte Hintergrundobjekte bzw. die Anzahl der Fehlerkennungen. Der Wertebereich dieser Metrik liegt zwischen -1 und 1 , wobei ein perfekter Klassifikator durch einen Wert von 1 charakterisiert ist. Der Φ_{MCC} wird objektspezifisch für Fahrzeuge als Φ_{MCC}^V und für Fußgänger als Φ_{MCC}^P notiert.

6.2.1.2 Metriken für das Tracking

Zeitliche Abdeckung: Die zeitliche Abdeckung beschreibt den Prozentsatz in Bezug auf die Referenzlebenszeit, in der Objekte korrekt klassifiziert und verfolgt werden können. Sei t_λ die Zeit in der ein Objekt λ erfolgreich klassifiziert und verfolgt werden konnte und l_λ die zugehörige Referenzlebensdauer, dann berechnet sich die zeitliche Abdeckung mit

$\Phi_{ZA} = \frac{\sum_{v\lambda \in \mathcal{L}} t_\lambda}{\sum_{v\lambda \in \mathcal{L}} l_\lambda}$. Die zeitliche Abdeckung wird für Fahrzeuge als Φ_{ZA}^V und für Fußgänger als Φ_{ZA}^P angegeben.

MSE: Diese Metrik beschreibt die mittlere quadratische Abweichung eines Schätzwertes \hat{x}_t von der zu schätzenden Größe x_t und ist wie folgt definiert:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (\hat{x}_t - x_t)^2 \quad (6.2.2)$$

Für erwartungstreue⁴ Schätzverfahren entspricht der RMS der Varianz der Abweichung. Eine geringe mittlere quadratische Abweichung bedeutet, dass gleichzeitig Bias und Varianz des Schätzers klein sind, und das Verfahren unverfälschte Ergebnisse mit einer geringen Streuung liefert.

RMSE der Entfernung: Der RMSE ist als $\sqrt{\text{MSE}}$ definiert und bewahrt die Einheit der zu schätzenden Größen. Der RMSE der Entfernung (r) beschreibt die Qualität der Entfernungsschätzung und wird objektspezifisch für Fahrzeuge und Fußgänger als r^V bzw. r^P notiert.

6.2.1.3 Sonstige Evaluierungskriterien

Rechenzeit: Da die Fahrumfelderfassung als automotiv Anwendung konzipiert ist, sind Echtzeitansprüche von großer Bedeutung. Die Rechenzeit t_{Zyk} wird in Millisekunden angegeben und bezeichnet die Ausführungszeit, die für einen Zyklus (siehe auch Abbildung 2.9 auf Seite 30 für das zyklische Systemübersichtsschaubild) benötigt wird. Alternativ dazu kann die Rechenzeit indirekt als Frequenz f_{Zyk} (Zyklusdurchläufe pro Sekunde) angegeben werden. Überschreitet f_{Zyk} einen Wert von ca. 15, werden Echtzeitansprüche erfüllt.

Erfassungsbereiche: Damit eine aussagekräftige Performanzevaluierung einer Fahrumfelderfassung durchgeführt werden kann, müssen die Erfassungsbereiche der Sensoren mit berücksichtigt werden. Würden die Erfassungsbereiche nicht berücksichtigt werden, fließen Referenzobjekte, die außerhalb der Erfassungsbereiche liegen, in die Erkennungsratenberechnung mit ein. Diese wird erniedrigt und verfälscht somit die Systemleistung.

Es ist zu beachten, dass sich die vorgestellten Metriken zu einem gewissen Grad gegenseitig beeinflussen. So führt zum Beispiel eine Zunahme der Erkennungsrate zu einer veränderten Berechnungsgrundlage für die Metriken des Trackings. Bei der Bewertung sollte als weiterer Aspekt berücksichtigt werden, dass beim Abgleich mit der Referenz verfälschte Ergebnisse auftreten können. Beinhaltet die Referenz teilverdeckte Objekte, die von der Fahrumfelderfassung nicht erfasst werden, erniedrigt sich die Erkennungsrate (TPR). Objekte die in der Referenz hingegen nicht berücksichtigt werden, können zu erhöhten Fehlerkennungsraten (FER) führen. Die Trackingmetriken werden dadurch nur geringfügig betroffen. Grundsätzlich sollte eine möglichst umfassende Referenz erzeugt werden, um Vergleiche zwischen verschiedenen Algorithmusvarianten, Sensorkonfigurationen und Testszenarien durchführen zu können.

⁴ Weicht ein Schätzer im Mittel von der zu schätzenden Größe ab, so ist er verzerrt (engl. bias) und nicht erwartungstreu.

6.2.2 Beschreibung der Testszenerien

Zur Bewertung des Gesamtsystems wurde eine repräsentative Stichprobe aus allen aufgezeichneten Testszenerien ausgewählt. Tabelle 6.3 listet die verwendeten Testszenerien mit einer kurzen Beschreibung sowie einzelnen Szenenausschnitten auf. Die dargestellten Ausschnitte zeigen das Ergebnis der maschinellen Erkennung zu einem ausgewählten Zeitpunkt.

Die Szenen können hinsichtlich des Aufzeichnungsortes in Stadt-, Landstraßen- und Autobahnszenarien untergliedert werden. Zudem kann eine Unterteilung bezüglich der Witterungs- und Umgebungseinflüsse erfolgen. Es werden Schönwetter-, Schnee- und Regen- sowie Tag- und Nachtszenarien abgedeckt. In sämtlichen Szenen wurde mittels einer semiautomatischen algorithmusgestützten Referenzdatenerzeugung (siehe auch Abschnitt 6.1.2) alle relevanten Verkehrsteilnehmer annotiert. Dadurch wird für die nachfolgende Auswertung eine Bewertungsgrundlage geschaffen.

Tabelle 6.3: In dieser Tabelle wird ein Überblick über die Szenen gegeben, die zur Bewertung herangezogen wurden. Neben einer kurzen Szenenbeschreibung sind in der rechten Tabellenhälfte Szenenausschnitte mit maschinell erkannten Objekten zu sehen.

Szenenbeschreibung	Szenenausschnitte
Szenarien bei trockenen Witterungsbedingungen	
<p>Szene 1 (S_1): Stadtszene in München mit 17 Fahrzeugen und 10 Fußgängern. Starkes Eigenfahrzeugnicken wegen einer Brückenüberquerung. (Dauer 0:30 Minuten)</p>	
<p>Szene 2 (S_2): Szene in einem Wohngebiet in Garching mit 11 fahrenden und 24 parkenden Fahrzeugen sowie 13 Fußgängern. (Dauer 1:43 Minuten)</p>	
<p>Szene 3 (S_3): Landstraßenszene mit 20 entgegenkommenden Fahrzeugen und vier Fahrzeugen in gleicher Fahrtrichtung. (Dauer 1:39 Minuten)</p>	

Tabelle 6.3: In dieser Tabelle wird ein Überblick über die Szenen gegeben, die zur Bewertung herangezogen wurden. Neben einer kurzen Szenenbeschreibung sind in der rechten Tabellenhälfte Szenenausschnitte mit maschinell erkannten Objekten zu sehen.



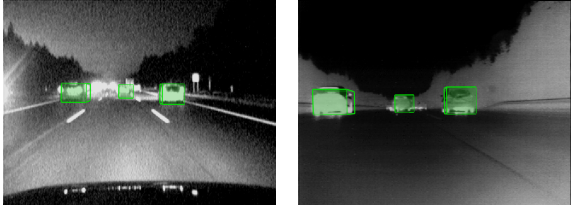

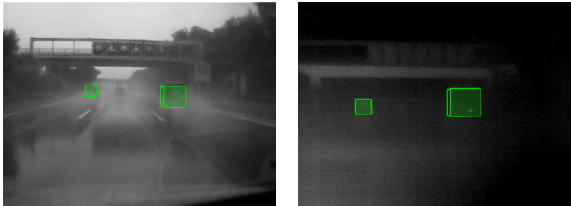
Szenenbeschreibung	Szenenausschnitte
<p>Szene 4 (S_4): Autobahnszene mit 86 Personenkraftwagen und 12 Lastkraftwagen bei trockener Fahrbahn. (Dauer 3:55 Minuten)</p>	
Nachtszenarien	
<p>Szene 5 (S_5): Stadtszene mit 15 entgegenkommenden, vier stationären, vier in Fahrtrichtung fahrenden Fahrzeugen sowie 11 Fußgängern bei Nacht und regennasser Fahrbahn. (Dauer 1:40 Minuten)</p>	
<p>Szene 6 (S_6): Autobahnszene mit 38 Fahrzeugen bei Nacht und regennasser Fahrbahn. (Dauer 4:04 Minuten)</p>	
Szenarien bei Niederschlag	
<p>Szene (S_7): Stadtszene in Dachau auf einer Ausfallstraße mit 12 entgegenkommenden Fahrzeugen, 13 Fahrzeugen in gleicher Fahrtrichtung und 7 Fußgängern. Mittelstarker Regen. (Dauer 2:36 Minuten)</p>	
<p>Szene 8 (S_8): Autobahnszene mit 9 Fahrzeugen bei starkem Regen und Gichtsleppen hinter den Fahrzeugen. (Dauer 0:54 Minuten)</p>	

Tabelle 6.3: In dieser Tabelle wird ein Überblick über die Szenen gegeben, die zur Bewertung herangezogen wurden. Neben einer kurzen Szenenbeschreibung sind in der rechten Tabellenhälfte Szenenausschnitte mit maschinell erkannten Objekten zu sehen.

Szenenbeschreibung	Szenenausschnitte
<p>Szene 9 (S_9): Stadtszene in München mit 17 Fahrzeugen und 2 Fußgängern bei starkem Schneefall. Der Schnee verursacht in der zweiten Szenenhälfte einen Totalausfall des Laserscanners. (Dauer 1:30 Minuten)</p>	

6.2.3 Konfiguration des Testsystems

Ein Überblickdiagramm, das die einzelnen Systemkomponenten sowie den Datenfluss innerhalb des multisensoriellen Fahrumfelderfassungssystems illustriert, ist im Grundlagenkapitel 2 in Abbildung 2.9 zu sehen. Dieses System wird für die Auswertung herangezogen. Die einzelnen Komponenten sind wie folgt parametrisiert:

Datenvorverarbeitung: Für die Datenvorverarbeitung der Laserscannerrohdaten wird die inkrementelle Linienaggregation verwendet (siehe auch Abschnitt 3.1.1.2). Im Video- und Wärmebild werden vertikale Kanten bestimmt, um den Extraktionsbereich der Bildmerkmale zu verfeinern (siehe Abschnitt 3.3.2.3 und 3.2.1).

Hypothesenerzeugung: Fahrzeughypothesen werden auf Basis von Radar- und Laserscannerrohdaten erzeugt. Fußgängerhypothesen werden ausschließlich auf Laserscannermessungen aufgesetzt (siehe auch Abschnitt 3.1.3).

Klassifikation: Der Fahrzeugklassifikator verwendet Gradientenhistogramme im Video- und Wärmebild sowie Dynamik- und Reflexionsmerkmale. Für die Erkennung von Fußgängern wird neben den Bildmerkmalen noch die Pendelbewegung genutzt (siehe Kapitel 4 für die Merkmalsdefinitionen und die entsprechenden Parametrisierungen). Das arithmetische Mittel als funktionaler Kombinator kommt als Klassifikatorfusionsverfahren zum Einsatz (siehe Abschnitt 5.2.2).

Tracking: Für das Verfolgen der Objekte wird ein erweiterter Kalman Filter verwendet. Für Fahrzeuge werden die Messwerte des Laserscanners und der Radar-Sensoren fusioniert. Fußgänger werden ausschließlich durch die Laserscannermessungen verfolgt [LWP07].

Für die Auswertungen wurden zwei Systemkonfigurationen untersucht. In der ersten Konfiguration (K_1) wurde ausschließlich der Laserscanner für das Tracking und die Klassifikation der Objekte verwendet. Alle restlichen Sensoren wurden deaktiviert. Als Merkmale fließen bei dieser Konfiguration für Fahrzeuge die Heckreflektanz und die Objektdynamik mit ein. Bei Fußgängern wird die Pendelbewegung als Merkmal herangezogen. Diese Konfiguration wurde gewählt, da durch den winkelauflösenden entfernungsgebenden Sensor, im Vergleich zu den anderen untersuchten Einzelsensorsystemen, die höchste Gesamtleistung erzielt werden konnte.

Um den Mehrwert des Fusionssystems zu evaluieren, wurde eine zweite Konfiguration K_2 erstellt, die alle Sensoren umfasst. Der Laserscanner und das Radar werden für das Objekttracking verwendet. Für die Klassifikation wird der Laserscanner, die Weitwinkel- und die Normalwinkelvideokamera, sowie die Wärmebildkamera herangezogen. Für Testszenarien, in denen eine Teilmenge aller Sensoren eine bessere Leistung als K_2 liefert, wird diese gesondert diskutiert. Die sich ergebenden Teilkonfigurationen von K_2 sind K_2^\dagger , das kein Wärmebild verwendet, sowie Konfiguration K_2^\ddagger , bei der alle bildgebenden Sensoren deaktiviert sind.

Für die Szenen S_1, S_3, S_4 und S_5 ist im Anhang D auf Seite 195 ein exemplarischer Durchlauf des Fahrumfelderfassungssystems zu sehen. Zur Erzeugung der Schnappschüsse wurde das System mit K_2 (bzw. K_2^\dagger für S_5) konfiguriert.

6.2.4 Ergebnisse

Die Szenarien, die in Abschnitt 6.2.2 eingeführt wurden, werden in diesem Abschnitt hinsichtlich der Metriken aus Abschnitt 6.2.1 bewertet. Dabei wird der Einfluss von unterschiedlichen Sensorkonfigurationen auf die Systemleistung untersucht. Charakteristische Fehler der multisensoriellen Fahrumfelderfassung werden gesondert beschrieben. Eine Laufzeitanalyse beschließt diesen Abschnitt.

6.2.4.1 Auswertung

Die Evaluierung der Szenen wurde durch den maschinellen Abgleich mit einer semiautomatisch generierten Referenz durchgeführt. Durch diesen Abgleich können sämtliche Metriken berechnet werden. In Tabelle 6.4 sind die ermittelten Ergebnisse der Szenenauswertung zusammengefasst. Die Tabellenspalten enthalten die untersuchten Szenen. Jeweils zwei Tabellenzeilen beinhalten die Werte einer Metrik für die unterschiedlichen Sensorkonfigurationen K_1 und K_2 . Für die Szenen S_5, S_6 und S_7, S_8 wurde Konfiguration K_2^\dagger und K_2^\ddagger verwendet (siehe auch Abschnitt 6.2.3 für eine Erläuterung zu den Konfigurationen). Im Folgenden wird ein Leistungsvergleich zwischen den beiden Sensorkonfigurationen sowie zwischen den einzelnen Szenen vorgenommen. Am Ende der Auswertung wird die Stadtszene S_1 exemplarisch herausgegriffen und detailliert betrachtet.

Tabelle 6.4: Auswertung der Fahrscenen S_1 bis S_9 (siehe Tabelle 6.3) mithilfe der Metriken aus Abschnitt 6.2.1. Die Konfiguration K_1 verwendet lediglich einen Laserscanner. Die Konfiguration K_2 verwendet alle verfügbaren Sensoren. Konfiguration K_2^\dagger ist eine Teilkonfiguration von K_2 , die kein Wärmebild verwendet. Analog dazu sind bei Konfiguration K_2^\ddagger alle bildgebenden Sensoren deaktiviert.

		S_1	S_2	S_3	S_4	S_5^\dagger	S_6^\dagger	S_7	S_8^\ddagger	S_9^\ddagger
RPR	(K_1)	0.24	0.32	0.78	0.91	0.24	0.91	0.46	0.57	0.53
RPR	(K_2)	0.76	0.80	0.96	0.87	0.62	0.82	0.79	0.71	0.67
RPR ^V	(K_1)	0.40	0.26	0.78	0.91	0.35	0.91	0.62	0.57	0.62
RPR ^V	(K_2)	0.93	0.81	0.96	0.87	0.61	0.82	0.90	0.71	0.69
RPR ^P	(K_1)	0.00	0.46	–	–	0.00	–	0.00	–	0.00
RPR ^P	(K_2)	0.50	0.77	–	–	0.64	–	0.43	–	0.50

Tabelle 6.4: Auswertung der Fahrscenen S_1 bis S_9 (siehe Tabelle 6.3) mithilfe der Metriken aus Abschnitt 6.2.1. Die Konfiguration K_1 verwendet lediglich einen Laserscanner. Die Konfiguration K_2 verwendet alle verfügbaren Sensoren. Konfiguration K_2^\dagger ist eine Teilkonfiguration von K_2 , die kein Wärmebild verwendet. Analog dazu sind bei Konfiguration K_2^\ddagger alle bildgebenden Sensoren deaktiviert.

		S_1	S_2	S_3	S_4	S_5^\dagger	S_6^\dagger	S_7	S_8^\ddagger	S_9^\ddagger
$FER_h(K_1)$	$[h^{-1}]$	240.55	380.71	51.02	12.97	45.24	26.65	183.31	0.00	0.00
$FER_h(K_2)$	$[h^{-1}]$	721.66	103.83	153.07	0.00	67.86	53.29	297.87	7.48	18.71
$FER_h^V(K_1)$	$[h^{-1}]$	120.28	34.61	51.02	12.97	45.24	26.65	183.31	0.00	0.00
$FER_h^V(K_2)$	$[h^{-1}]$	481.11	34.61	102.05	0.00	45.24	53.29	297.87	7.48	18.71
$FER_h^P(K_1)$	$[h^{-1}]$	120.28	346.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$FER_h^P(K_2)$	$[h^{-1}]$	240.55	69.22	51.02	0.00	22.62	0.00	0.00	0.00	0.00
$\Phi_{MCC}(K_1)$		0.42	0.42	0.84	0.94	0.43	0.93	0.54	0.76	0.73
$\Phi_{MCC}(K_2)$		0.76	0.86	0.87	0.93	0.73	0.85	0.70	0.77	0.78
$\Phi_{MCC}^V(K_1)$		0.58	0.48	0.84	0.94	0.53	0.93	0.62	0.76	0.78
$\Phi_{MCC}^V(K_2)$		0.85	0.88	0.90	0.93	0.73	0.85	0.73	0.77	0.79
$\Phi_{MCC}^P(K_1)$		–	0.42	–	–	0.00	0.00	–	–	0.00
$\Phi_{MCC}^P(K_2)$		0.60	0.80	–	–	0.75	0.00	0.65	–	0.71
$\Phi_{ZA}(K_1)$	[%]	0.49	0.36	0.75	0.95	0.64	0.95	0.81	0.40	0.42
$\Phi_{ZA}(K_2)$	[%]	0.69	0.74	0.84	0.92	0.75	0.71	0.95	0.82	0.76
$\Phi_{ZA}^V(K_1)$	[%]	0.61	0.39	0.75	0.95	0.78	0.95	0.85	0.40	0.42
$\Phi_{ZA}^V(K_2)$	[%]	0.82	0.74	0.84	0.92	0.81	0.71	0.97	0.82	0.77
$\Phi_{ZA}^P(K_1)$	[%]	–	0.30	–	–	0.00	–	–	–	0.00
$\Phi_{ZA}^P(K_2)$	[%]	0.16	0.74	–	–	0.49	–	0.45	–	0.42
$r(K_1)$	[m]	0.18	0.18	0.18	0.18	0.40	0.20	0.33	0.24	0.23
$r(K_2)$	[m]	0.29	0.18	0.21	0.19	0.38	0.22	0.31	0.47	0.62
$r^V(K_1)$	[m]	0.18	0.17	0.18	0.18	0.40	0.20	0.33	0.24	0.23
$r^V(K_2)$	[m]	0.28	0.19	0.21	0.19	0.41	0.22	0.32	0.47	0.62
$r^P(K_1)$	[m]	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$r^P(K_2)$	[m]	0.40	0.12	0.00	0.00	0.13	0.00	0.18	0.00	0.97

Aus dem Leistungsvergleich der beiden Sensorkonfigurationen können folgende Ergebnisse abgeleitet werden. Im Mittel über die Szenen werden Verkehrsteilnehmer durch die Konfiguration K_2 besser (vergleiche Φ_{MCC}) erkannt. Dies ist im Wesentlichen dadurch zu erklären, dass Fußgänger und parkende Fahrzeuge im städtischen Umfeld, allein durch den Laserscanner, lediglich sehr eingeschränkt klassifiziert werden können. Besser geeignet sind hierfür die Merkmale, die aus der bildgebenden Sensorik gewonnen werden. Aufgrund der höheren Erkennungsraten und der unveränderten Trackingleistung ergibt sich zudem eine verbesserte zeitliche Abdeckung (Φ_{ZA}). Ausnahmen sind die Autobahnscenarien S_4 und S_6 . Dort sind die Ergebnisse (Φ_{MCC} und Φ_{ZA}) der beiden Konfigurationen nahezu identisch, da durch die bildgebenden Sensoren kein zusätzlicher Beitrag zur dynamikgestützten Klassifikation aus K_1 geleistet wird. Der

leichte Abfall kann durch den negativen Einfluss der Videobildmerkmale bei weiter entfernten Objekten erklärt werden. Diese sind durch eine isolierte Dynamikklassifikation robuster separierbar.

Die Trackinggenauigkeit r ist in den Szenarien S_1 bis S_7 unabhängig von der jeweiligen Sensorkonfiguration. Da die evaluierten Szenen kaum hochdynamisches Objektverhalten beinhalten, schafft die zusätzliche Geschwindigkeitsbestimmung der Radar-Sensorik gegenüber dem Laserscanner, keinen signifikanten Mehrwert in Bezug auf die Entfernungsschätzung. In den Schlechtwetzerszenarien S_8 und S_9 , in denen die optische Sichtweite stark eingeschränkt ist, können durch den Laserscanner Objekte lediglich im Nahbereich detektiert und verfolgt werden. Dieses Defizit wird durch die Radarsensorik zwar ausgeglichen (vergleiche Φ_{ZA}), gleichzeitig sinkt jedoch die Genauigkeit der Entfernungsbestimmung, aufgrund der multiplen Reflexionsquellen an den detektierten Fahrzeugen.

Bei einem Vergleich der Szenen untereinander werden folgende Aspekte deutlich: Erwartungsgemäß ist die Systemperformanz sowohl bei Autobahn- als auch bei Landstraßenszenen besser, als bei den wesentlich komplexeren Stadtszenarien (vergleiche Metriken der Szenen S_1 und S_2 mit den Metriken der Szenen S_3 und S_4). Sowohl die Erkennungsrate, die zeitliche Abdeckung und die Φ_{MCC} -Metrik sind entsprechend höher. Insbesondere Fehlerkennungen (FER_h^P) von Fußgängern sind im urbanen Umfeld häufiger, da der Hintergrund im Vergleich zu den Außerstadtszenarien stärker strukturiert ist. Zudem erschweren wiederholt auftretende Verdeckungen von Verkehrsteilnehmern die Hypotheseninstantiierung und das Verfolgen von Objekten, wodurch sowohl die Qualität des Trackings (Φ_{ZA}) als auch die Klassifikation (RPR, Φ_{MCC}) negativ beeinflusst wird.

Die Auswertungen der Autobahnscene S_4 , die bei Tag aufgenommen wurde und der Autobahnscene S_6 , die bei Nacht aufgenommen wurde, sind für die Konfiguration K_1 weitgehend identisch (vergleiche Φ_{MCC} und Φ_{ZA}), da der Laserscanner beleuchtungsunabhängig arbeitet. Für den fusionierten Ansatz (K_2) bricht hingegen die Leistung für die Nachtszene stark ein, da aufgrund der eingeschränkten Videobildqualität lediglich verrauschte Merkmale extrahiert werden können. Unter der Annahme dass dieser „Sensorausfall“ detektiert werden kann, wurde statt Konfiguration K_2 , die um das Videobild reduzierte, Konfiguration K_2^\dagger für die Evaluierung gewählt. Mit dieser Konfiguration werden für Nachtszene S_6 etwas schlechtere Ergebnisse wie mit K_2 für die Tagszene S_4 erzielt, da der verwendete Wärmebildsensor eine geringere Entfernungsreichweite als der Videosensor aufweist. Im Vergleich der Stadtszene bei Nacht (Szene S_5) mit den Stadtszenen bei Tag (Szenen S_1 und S_2) ist eine reduzierte Erkennungsleistung für Fahrzeuge zu beobachten. Dies ist darauf zurückzuführen, dass die Wärmebildsignatur von entgegenkommenden Fahrzeugen stark von der Heckansicht abweicht. Bei den Tagszenen kann dies durch den Videosensor kompensiert werden.

Bei der Stadtszene S_6 mit mittelstarkem Regen sind im Vergleich zu den Stadtszenen S_1 und S_2 keine signifikanten Leistungsabfälle zu beobachten, da die Sensorik nur geringfügig beeinflusst wird. Autobahnscenen mit starkem Regen und/oder Gischtschleppen sowie Szenen mit starkem Schneefall (vergleiche Szene S_7 bzw. Szene S_8) reduzieren insbesondere die Sichtweiten der optischen Sensorsysteme. Dadurch nimmt für K_1 sowohl die Klassifikations- als auch die Trackingleistung erheblich ab. Für K_2 sinkt insbesondere die Klassifikationsleistung, da durch die witterungsbedingten Effekte sowohl die Video- als auch die Wärmebildmerkmale stark verrauscht sind. Unter der Annahme dass diese „Sensorausfälle“ detektiert werden können, wurde statt Konfiguration K_2 , Konfiguration K_2^\dagger verwendet, die ausschließlich die entfernungsgebende Sensorik beinhaltet. Da die Objektdetektion und -verfolgung im Nahbereich durch den Laser-

scanner und im Fernbereich durch den Radar-Sensor übernommen wird, ist ein erheblicher Anstieg der zeitlichen Abdeckung (Φ_{ZA}) im Vergleich zu K_1 zu beobachten.

Die Diagramme 6.8a und 6.8b visualisieren die detaillierte Erkennungsleistung für das Stadtszenario S_1 (siehe auch Tabelle 6.3) mit zwei unterschiedlichen Sensorkonfigurationen. Im Anhang D auf Seite 195 ist ein Durchlauf der Szene zu sehen. An der Abszisse ist der temporale Verlauf des Szenarios aufgetragen. An der Ordinate sind sämtliche Objekte (Abkürzung V für Fahrzeuge und P für Fußgänger) aufgelistet, die das Szenario beinhaltet. Dabei ist die Referenzlebenszeit der Objekte durch einen grau hinterlegten Balken gekennzeichnet. Durch die Hypothesengenerierung detektierte Objekte sind blau hinterlegt. Korrekt klassifizierte Objekte sind grün hinterlegt. Fehlerkennungen sind in einer dedizierten Zeile f_p ausgewiesen. Die Erkennungsraten in Diagramm 6.8a basieren auf Sensorkonfiguration K_1 . Als Merkmale fließen die Fahrzeugreflektanz und -dynamik sowie die Pendelbewegung der Fußgänger in den Klassifikationsprozess mit ein. Zur Erzeugung von Diagramm 6.8b wurden alle verfügbaren Sensoren und Merkmale zur Klassifikation genutzt (Sensorkonfiguration K_2).

Im direkten Vergleich der beiden Diagramme wird deutlich, dass durch die zusätzliche Sensorik eine höhere Erkennungsleistung erzielt werden kann. Bei einer vergleichbaren Anzahl von Fehlerkennungen werden nicht nur mehr Objekte korrekt klassifiziert, sondern auch zu einem früheren Zeitpunkt. Speziell die Fußgängererkennung profitiert von der bildgebenden Sensorik, da die Analyse der Pendelbewegung eine lange Beobachtungszeit und eine Bewegung des Fußgängers erfordert. Gleichermaßen werden insbesondere stehende Fahrzeuge durch die bildgebende Sensorik adressiert.

6.2.4.2 Charakteristische Fehlerfälle

In Bezug auf Fahrumfelderfassungssysteme werden Fehler als Abweichungen von den Referenzdaten definiert. Die Fehlerursachen sind vielfältig und liegen zum einen in der Algorithmik der einzelnen Systemkomponenten und zum anderen in äußeren Einflüssen, wie ungünstigen Witterungsbedingungen. Tabelle 6.5 fasst charakteristische Fehler sowie deren Ursachen anhand von Beispielen zusammen.

In der ersten Tabellenzeile sind Fehler zu sehen, die durch die Datenvorverarbeitung verursacht werden. Sinkt der Abstand zwischen Objekten unter die Sensorauflösung, kann keine Separierung mehr erfolgen. Fehlsegmentierungen treten insbesondere bei Objekten auf, die einen geringen Abstand zueinander aufweisen. Zudem können Fehlsegmentierungen bei Verdeckungen auftreten, da Objekte in diesem Fall meist nur partiell beobachtet werden können.

Bei der Klassifikation von Objekten treten falsch-positiv und falsch-negativ Erkennungen auf. Bei der ersten Fehlerart werden Hintergrundobjekte fälschlicherweise als relevante Verkehrsteilnehmer erkannt. Dies ist insbesondere der Fall, wenn Hintergrundobjekte eine ähnliche Merkmalssignatur wie Vordergrundobjekte aufweisen (vergleiche Straßenschilder und -barken in der zweiten Tabellenzeile). Bei der zweiten Fehlerart werden Vordergrundobjekte nicht als solche erkannt (vergleiche entgegenkommendes Fahrzeug in der zweiten Tabellenzeile), da die Merkmale entweder ungenau extrahiert wurden, oder das Objekterscheinungsbild zu starke Unterschiede zu den Objekten im Trainingsdatensatz aufweist.

Verlorene Objekte sind ein Hauptfehler beim Verfolgen von Objekten. Neben hochdynamischen Objekten, kann dies bei entgegenkommenden Fahrzeugen auftreten, da dort die Erfassungszeit kurz ist und damit lediglich eine begrenzte Einschwingzeit des Filters zur Verfügung

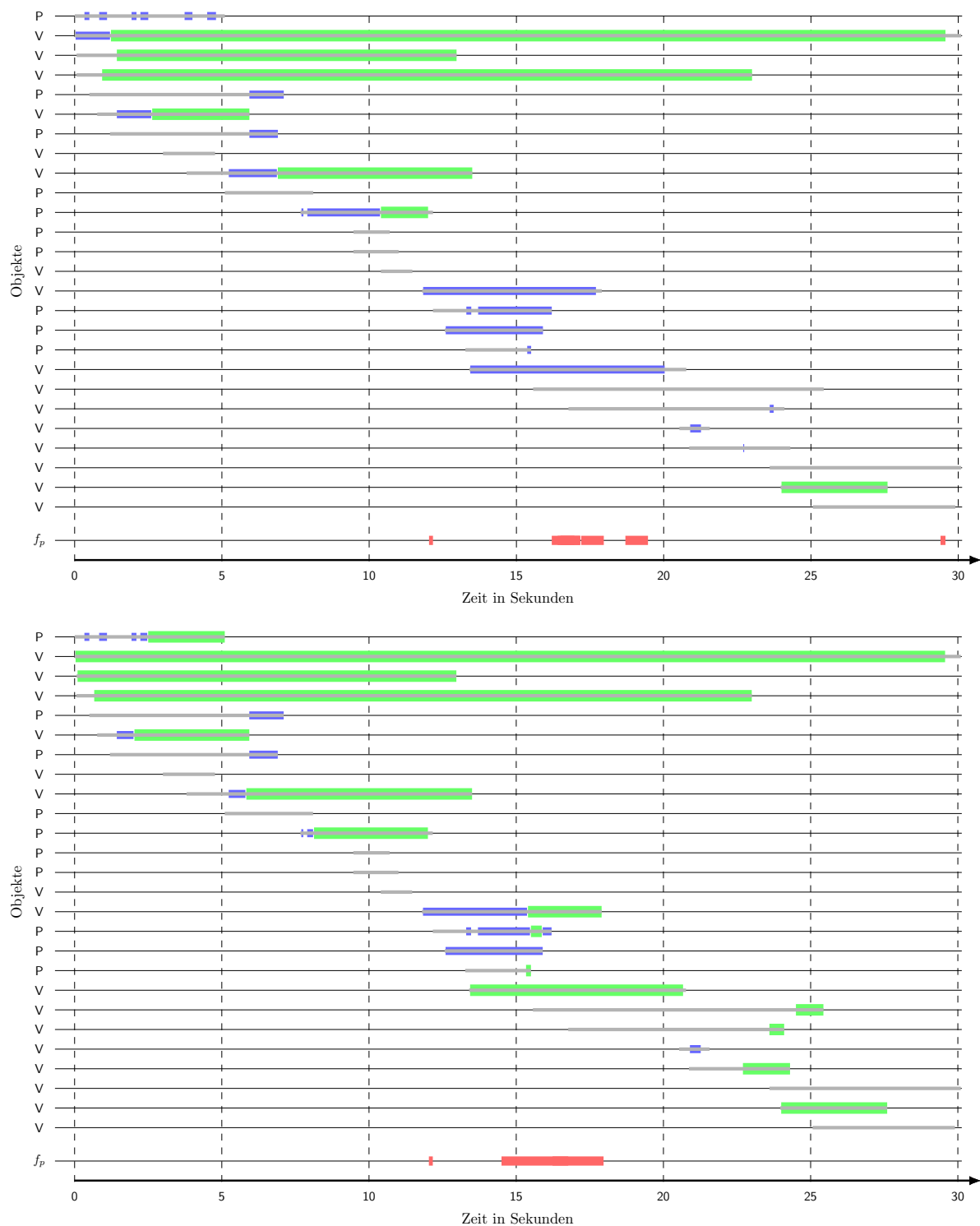


Abbildung 6.8: Zwei temporale Diagramme des Stadtszenarios S_1 zur Gegenüberstellung der Objektdetektionen (blau hinterlegt) und -klassifikationen (grün hinterlegt) mit der Objektreferenz (grau hinterlegt) bei unterschiedlichen Sensorkonfigurationen. Bei dem oberen Diagramm wurden Fahrzeuge (V) und Fußgänger (P) nur mit einem Laserscanner detektiert und klassifiziert. Bei dem unteren Diagramm wurde zusätzlich die bildgebende Sensorik zur Objekterkennung genutzt. Fehlerkennungen sind mit f_p gekennzeichnet.


steht (siehe linkes Bild in Tabellenzeile 3). Weitere Gründe sind Zuordnungsfehler, wie sie bei Objekten mit geringem Abstand zueinander oder multiplen Reflexionsquellen auftreten können (siehe Ampelpfosten im rechten Bild in Tabellenzeile 3). Eine weitere Fehlerquelle ist eine zu einfache Modellierung der zu verfolgenden Objekte oder falsche Annahmen in der Vorverarbeitung der Sensordaten. So können insbesondere bei LKW oder entgegenkommenden Fahrzeugen verfälschte Messwerte entstehen, da ihr Erscheinungsbild im Laserscanner von der Annahme eines Quadermodells abweicht. Diese systematischen Fehler können nur sehr begrenzt durch den verwendeten Kalman Filter kompensiert werden.

Grundsätzlich können widrige Umweltbedingungen eine Ursache für die oben angegebenen Fehler sein. Durch verrauschte Sensordaten, partielle Sensorausfälle und zusätzliche Störquellen wie beispielsweise Wassertropfen und Schneeflocken können Geisterziele, Fehlklassifikationen und fehlerbehaftete Datenvorverarbeitungen verursacht werden (vergleiche auch Szenen S_8 und S_9 sowie Tabellenzeile 4).

Tabelle 6.5: Charakteristische Fehler des multisensoriellen Fahrumfelderfassungssystems mit kurzer Beschreibung und Beispielen.

Problembeschreibung	Beispiele
Fehler von Komponenten des Fahrumfelderfassungssystems	
Datenvorverarbeitung: Verschmelzung von Objekten, mit zu geringem Abstand. Beispiele dafür sind Fußgängergruppen (linkes Bild) und parkende Fahrzeuge (rechtes Bild).	
Klassifizierung: Das entgegenkommende Fahrzeug im linken Bild ist ein Beispiel für eine falsch-positiv Erkennung. Das Schild im linken Bild und die Barke im rechten Bild sind Beispiele für falsch-negativ Erkennungen.	
Tracking: Abfliegender Track bei einem entgegenkommenden Fahrzeug (linkes Bild). An einem Ampelpfosten „hängen gebliebener“ Fußgängertrack durch einen Zuordnungsfehler beim nahen Passieren eines Fußgängers (rechtes Bild).	

Tabelle 6.5: Charakteristische Fehler des multisensoriellen Fahrumfelderfassungssystems mit kurzer Beschreibung und Beispielen.

Problembeschreibung	Beispiele
Umweltverursachte Fehler	
<p>Eingeschränkte Sensorsichtbereiche durch widrige Witterungsbedingungen wie starkem Regen (linkes Bild), Nebel oder Schneefall (rechtes Bild).</p>	

6.2.4.3 Laufzeitanalyse

Da die Fahrumfelderfassung als automotiv Anwendung konzipiert ist, sind Echtzeitanprüche von großer Bedeutung. Die Komplexität der einzelnen Szenen ist entscheidend für die durchschnittliche Gesamtlaufzeit der Fahrumfelderfassung. In Abbildung 6.9 sind für ein typisches Autobahn- und Stadtszenario (siehe auch Szenarien S_4 und S_1 in Tabelle 6.3) die zeitlichen Verläufe der durchschnittlichen Zykluslaufzeiten t_{Zyk} für das Tracking, die Klassifikation sowie die Vorverarbeitung dargestellt. Darüber hinaus ist die Anzahl an Objekten aufgetragen, die in jedem Zeitschritt verarbeitet werden müssen. Sie beinhaltet sowohl bereits klassifizierte Fahrzeuge und Fußgänger, als auch noch nicht klassifizierte Objekthypothesen und dient als Indikator für die Komplexität der Szene. Die folgende Laufzeitanalyse basiert auf einem Core2Duo-Prozessor mit 2 GHz, wobei nur ein Prozessorkern verwendet wurde. Die Skalierbarkeit der vorgestellten Fahrumfelderfassung auf aktuelle Multiprozessorarchitekturen wird im zweiten Teil dieses Abschnitts thematisiert.

Die Vorverarbeitung für den Laserscanner und die Wärmebildkamera ist weitgehend unabhängig von der Anzahl der Objekte und benötigt für die beiden Szenarien im Schnitt 13 ms. Die durchschnittlichen Laufzeiten für die Klassifikation und das Tracking sind hingegen direkt proportional zur Komplexität der Szene. So benötigt der gesamte Klassifikationsablauf mit bis zu 25 Objekten in der Autobahnscene 5 ms. Bei bis zu 50 gleichzeitig zu verarbeitenden Objekten, wie es für eine Stadtszene typisch ist, steigt die durchschnittliche Laufzeit für die Klassifikation hingegen auf 12 ms. Für das Tracking der Objekthypothesen werden 2.5 bzw. 5 ms benötigt. Zusammengefasst lässt sich eine echtzeitfähige Fahrumfelderfassung für Autobahn- und Landstraßenszenarien mit einer Systemrate von $f_{Zyk} = 30$ Hz realisieren. Für komplexe Stadtszenarien kann unter Verwendung eines Einprozessorsystems ebenfalls Echtzeitfähigkeit bei $f_{Zyk} = 25$ Hz erreicht werden. An dieser Stelle bieten aktuelle Multiprozessorarchitekturen weiteres Optimierungspotenzial.

Durch eine parallele Abarbeitung von Programmroutinen kann auf Multiprozessorsystemen Laufzeit eingespart werden. In der Architekturentscheidung für das Fahrumfelderfassungssystem wurde eine parallele Ausführung speziell berücksichtigt. So werden die Objekte des Fahrum-

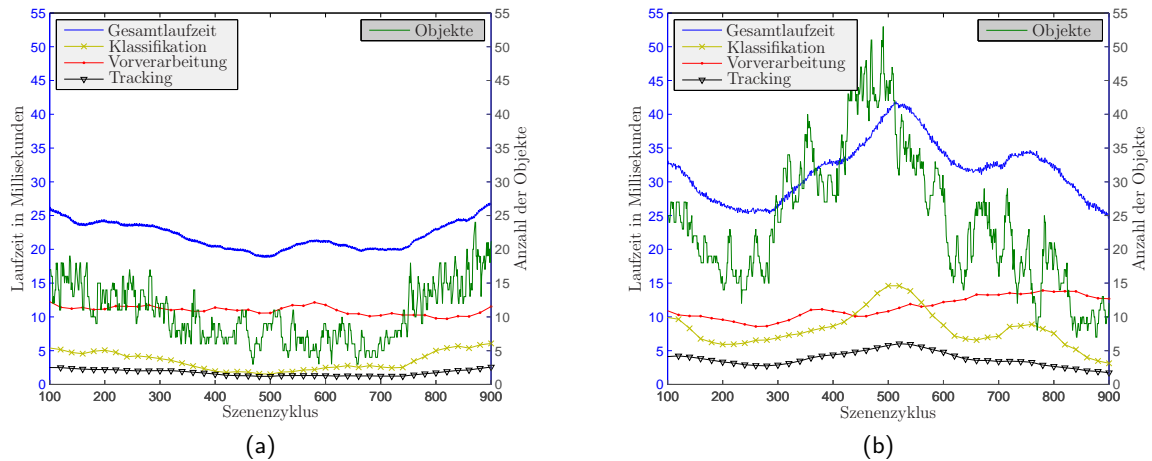


Abbildung 6.9: Vergleich der System- und Komponentenlaufzeiten für das Autobahnszenario S_4 (a) und das Stadtszenario S_1 (b) abhängig vom Szenenzeitpunkt. Es ist ein deutlicher Zusammenhang zwischen Zykluslaufzeiten t_{Zyk} und Objektaufkommen erkennbar.

feldes (Hypothesen, Verkehrsteilnehmer, Sensordaten) in einer hierarchischen Baumstruktur verwaltet. Dieser kreisfreie Graph kann durch ein Visitor-Designmuster⁵ [GHJV95] effizient durchlaufen werden, um Operationen auf den einzelnen Knoten auszuführen. Für die Fahrumfelderfassung werden alle Operationen, die auf dieser Datenstruktur operieren, als Visitor angelegt. Eine parallele Ausführung kann demnach erzielt werden, indem die Baumstruktur durch mehrere Prozesse zeitgleich durchlaufen wird. Dabei müssen Serialisierungsanforderungen berücksichtigt werden.

Alle Rechenoperationen, die per Objekt oder per Sensor angewendet werden sind parallelisierbar. Dazu zählen die Module Datenvorverarbeitung, Messwertaktualisierung, Zustandsprädiktion, Merkmalsgewinnung und Klassifikation. Die Datenvorverarbeitung wird für alle beteiligten Sensoren durchgeführt und kann demnach pro Sensor parallelisiert werden. Die restlichen Module werden objektweise ausgeführt und sind dadurch gleichermaßen parallelisierbar. Die Laufzeitmessungen mit und ohne Parallelisierung sind in Tabelle 6.6 aufgeführt. Dabei wurden die durchschnittlichen Laufzeiten für Szenario S_4 und Szenario S_1 (siehe auch Abschnitt 6.2.2) mit einem Core2Duo-Prozessor unter Verwendung beider Prozessoren ausgewertet.

Die Komponenten Prädiktion und Messwertaktualisierung, Merkmalsgewinnung und Klassifizierung, die für jedes Objekt durchgeführt werden, erreichen einen Speedup⁶ von nahezu 0.5. Die Datenvorverarbeitung erzielt lediglich einen geringen Speedup, da die Vorverarbeitung für den Laserscanner deutlich mehr Zeit beansprucht, als die Vorverarbeitung für die bildgebende Sensorik. Insgesamt kann durch die parallele Ausführung eine Laufzeitsteigerung um 20 % im Autobahnszenario und um 25 % im Stadtszenario erzielt werden.

⁵ Das Visitor-Designmuster ist ein Entwurfsmuster aus der Softwareentwicklung wodurch Operationen auf Datenstrukturen gekapselt werden können.

⁶ Speedup (Beschleunigung) definiert das Laufzeitverhältnis eines parallelisierten Programms auf einem Prozessor, verglichen mit der Laufzeit desselben Programms auf p Prozessoren.

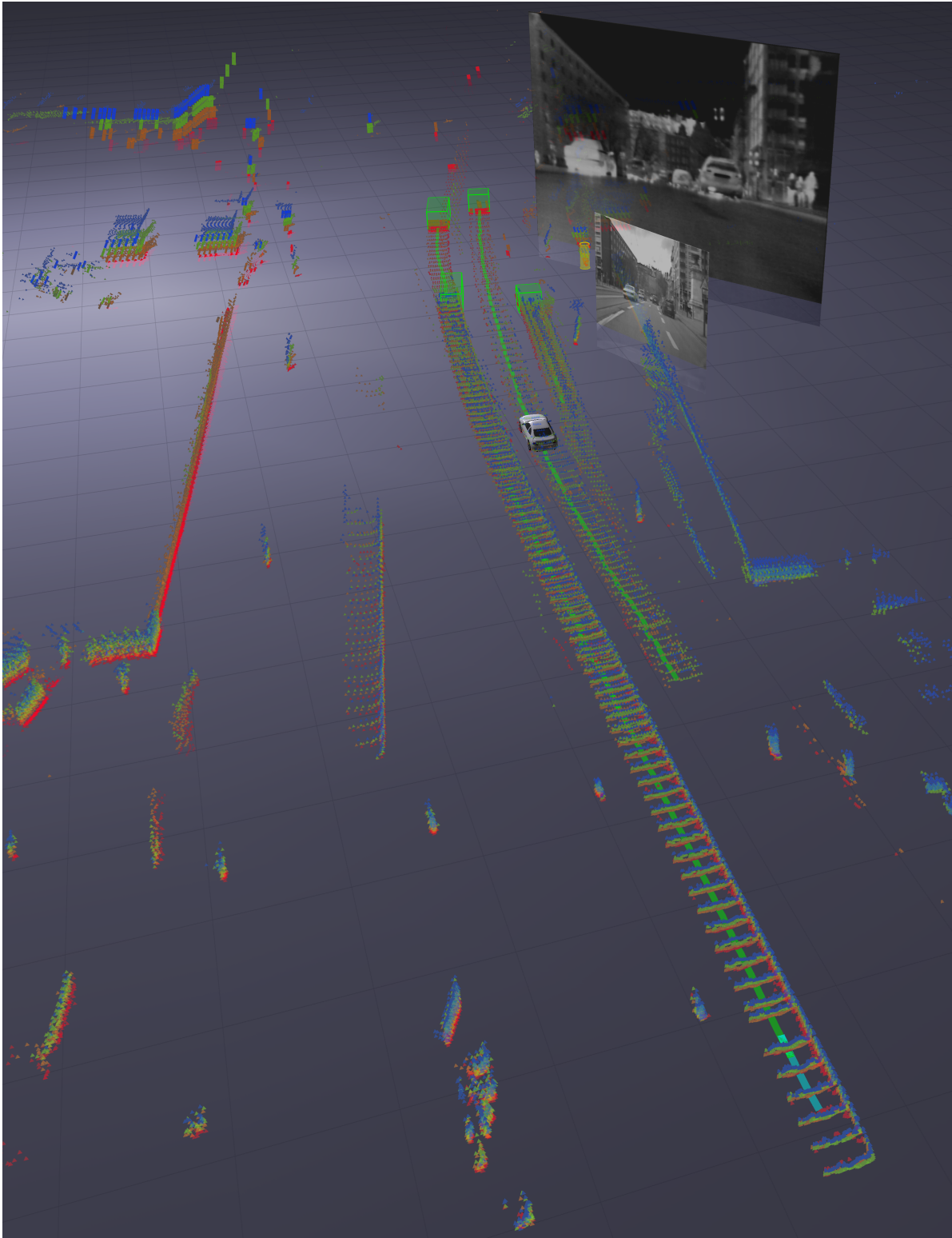
Parallelisierung	Systemkomponenten				
	Datenvorverarbeitung	Prädiktion und Messwertaktualisierung	Merkmalsgewinnung	Klassifikation	Gesamtlaufzeit
Autobahnszene (Szenario S ₄)					
1 CPU	12.89 ms	0.71 ms	0.98 ms	3.19 ms	22.96 ms
2 CPUs	10.16 ms	0.48 ms	0.57 ms	1.75 ms	18.43 ms
Stadtszene (Szenario S ₁)					
1 CPU	15.34 ms	0.98 ms	3.13 ms	6.46 ms	38.04 ms
2 CPUs	12.29 ms	0.5 ms	1.62 ms	3.75 ms	29.46 ms

Tabelle 6.6: Vergleich der durchschnittlichen Zykluslaufzeiten der einzelnen Systemkomponenten bei einer parallelen Abarbeitung. Als Testsystem kam ein Core2Duo-Prozessor zum Einsatz.

6.3 Zusammenfassung

Im ersten Teil des Kapitels wurden mehrere Möglichkeiten aufgezeigt, wie für ein Fahrumfelderfassungssystem Referenzdaten erzeugt werden können. Die vorgestellten Verfahren adressieren im Wesentlichen eine effiziente und präzise Gewinnung dieser Daten. Das semiautomatische Verfahren bzw. die algorithmusgetriebene Vorgehensweise optimiert den gesamten Annotierungsprozess in Bezug auf den Arbeitsaufwand. Durch den Einsatz eines dedizierten Sensors können für die Evaluierung von Trackingverfahren sehr genaue und von der verwendeten Sensorik unabhängige Referenzdaten generiert werden.

Im zweiten Teil des Kapitels wurde die Systembewertung für das im Rahmen dieser Arbeit umgesetzte multisensorielle Fahrumfelderfassungssystem durchgeführt. Dabei wurde die Systemleistung unter typischen Fahrsituationen in Umgebungen, wie Stadt, Land und Autobahn, sowie unter verschiedenen Witterungsbedingungen ermittelt. Zudem wurden die Synergieeffekte des multisensoriellen Sensorsystems bewertet. Mittels eines winkelauflösenden entfernungsgebenden Sensors kann bei Autobahn- und Landstraßenszenen unter normalen Witterungsbedingungen eine gute Erkennungsleistung erzielt werden. Durch die Hinzunahme von weiteren Sensoren konnte in dieser Domäne kein Leistungszuwachs beobachtet werden. Erst in komplexen Stadtszenarien sowie unter widrigen Umweltbedingungen zeigt sich ein deutlicher Nutzen einer multisensoriellen Fahrumfelderfassung. Die durchgeführte Laufzeitanalyse belegt, dass das damit verbundene erhöhte Datenaufkommen mit aktuellen Rechensystemen in Echtzeit bewältigt werden kann.



Visualisierte Sensordaten und Verkehrsteilnehmer auf der Zweibrückenstraße in München.

7

Visualisierung von heterogenen Sensordaten

Die Visualisierung von Daten kann als eine Berechnungsmethode zur Überführung von symbolischer in geometrische Information definiert werden [MDB87]. Damit werden relevante Aspekte abstrakter Datenmengen in eine bildlich veranschaulichte Form gewandelt, um einen umfassenden Überblick zu gewähren, verborgene Datenzusammenhänge aufzuschlüsseln und sowohl Erkenntnis als auch Kommunikation zu erleichtern [SM00]. Shneiderman sieht die Essenz von Informationsvisualisierung in der Beschleunigung menschlichen Denkens durch Werkzeuge, die die menschliche Intelligenz verstärken [Che04].

Speziell für die Entwicklung von maschinellen Umfeld erfassungssystemen, die aufgrund zahlreicher Sensoren ein hohes Datenaufkommen aufweisen, spielen geeignete Darstellungsstrategien eine große Rolle. Für die Umsetzung eines Perzeptionssystems benötigt der Entwickler eine Veranschaulichung der Sensordaten in allen Ausprägungsstufen. Angefangen bei Rohdaten über aggregierte Daten, bis hin zu fusionierten und interpretierten Daten, sollten adäquate Visualisierungsmechanismen verfügbar sein. Diese erleichtern sowohl das Auffinden, als auch das Beseitigen von Programmfehlern und Falschparametrisierungen. Fehlerbehaftete Sensorkalibrierungen oder ungenaue zeitliche Synchronisierungen beispielsweise können gewissermaßen auf einen Blick erkannt und behoben werden.

Eine adäquate Sensordatenvisualisierung bietet des Weiteren eine stete und unmittelbare Performanzüberwachung des Systems. Zum einen lassen sich aus dieser sinnvolle Verbesserungsmöglichkeiten und Erweiterungen erkennen und ableiten, zum anderen kann diese der Ausgangspunkt für gänzlich neuartige Einblicke und Ideen sein.

Eine weitere wichtige Zielsetzung einer Datenveranschaulichung ist die Vereinfachung von Präsentation und Kommunikation. Eine bildliche Aufbereitung von Sensordaten kann Experten als Diskussionsgrundlage dienen, aber auch fachfremden Entwicklern eine gemeinsame Schnittstelle bereitstellen. Kommunikationslücken zwischen Sensorikern und automotiven HMI-Designern¹ können dadurch überbrückt, und der Ideen- und Informationsaustausch vereinfacht und gefördert werden.

¹ HMI ist die Abkürzung für Human Machine Interface und bezeichnet die Schnittstelle zwischen Mensch und Maschine.

In den folgenden Abschnitten werden Visualisierungskonzepte für das im Rahmen dieser Arbeit umgesetzte Perzeptionssystem vorgestellt und diskutiert. Dabei wird in Abschnitt 7.1 zunächst ein Überblick über den Datenfluss innerhalb eines automotiven Umfelderkennungssystems gegeben, woraus auf grundsätzliche Darstellungsrichtlinien basierende allgemeine Visualisierungsanforderungen abgeleitet werden. Speziell für die Anforderungen während des Entwicklungsprozesses werden im nachfolgenden Abschnitt 7.2 Darstellungslösungen vorgestellt und exemplarisch illustriert. Abschnitt 7.3 widmet sich Darstellungsformen die sowohl die Systempräsentation vor unbedarften Dritten, als auch die Kommunikation zwischen Entwicklern erleichtern. Abschließend wird in Abschnitt 7.4 als Beispiel für eine visualisierungsgetriebene Konzeptidee eine Fahrbahnerkennung mittels der Flottenbewegung vorausfahrender Fahrzeuge vorgestellt.

7.1 Grundlagen der Visualisierung

Visualisierungsmechanismen beziehen sich stets auf eine Datenbasis. Dieser Abschnitt gibt somit zunächst einen Überblick über den Datenbestand und -fluss eines Umfelderkennungssystems und behandelt anschließend grundlegende Visualisierungsrichtlinien, die beispielhaft auf die Belange von Perzeptionssystemen umgesetzt werden.

7.1.1 Daten und Informationsfluss

In einem Umfelderkennungssystem liegen Daten in unterschiedlichen Verarbeitungszuständen vor, welche durch geeignete Algorithmik ineinander überführt werden. Den Anfang des Informationsflusses (siehe Abbildung 7.1) bilden Rohdaten, die unmittelbar von der verwendeten Sensorik stammen. Durch Merkmalsextraktion und Zusammenführung werden diese Basisdaten in vorverarbeitete Daten umgewandelt. In einem letzten Schritt werden die Daten durch wissensbasierte und modellgetriebene Verfahren interpretiert. Dabei nimmt das Datenvolumen von links nach rechts ab, wobei zugleich der Abstraktionsgrad ansteigt. Die drei Ausprägungsstufen der Daten liegen in der Regel zeitgleich vor und werden meist zyklisch im Takt der Sensoraktualisierungsraten aufgefrischt. Eine geeignete dynamische Visualisierung sollte Schritt halten mit den Datenaktualisierungsraten der Sensorik und dem Entwickler Darstellungsmechanismen für alle Datenausprägungen und -überführungen zur Verfügung stellen.

7.1.2 Designrichtlinien

Im vorangegangenen Abschnitt wurden die Daten einer Fahrumfelderfassung sowie deren Entstehung betrachtet. Neben der Analyse dieser Daten in Bezug auf Art und Struktur ist die Wahl geeigneter Darstellungsformen ein wichtiger Aspekt während des Designs einer Visualisierung wissenschaftlicher Daten. Schumann [SM00] beschreibt die folgenden drei grundsätzlichen Anforderungen, die jede Visualisierung erfüllen sollte:

Expressivität: Eine gute Visualisierung darf ausschließlich die in den Daten enthaltenen bzw. daraus abgeleiteten Informationen verwenden und muss sie möglichst unverfälscht wiedergeben. Auf diese Weise werden Fehlentscheidungen bzw. Falschinterpretationen vermieden, die darauf zurückzuführen sind, dass die Visualisierung dem Anwender Fakten suggeriert, die nicht in den Daten enthalten sind.

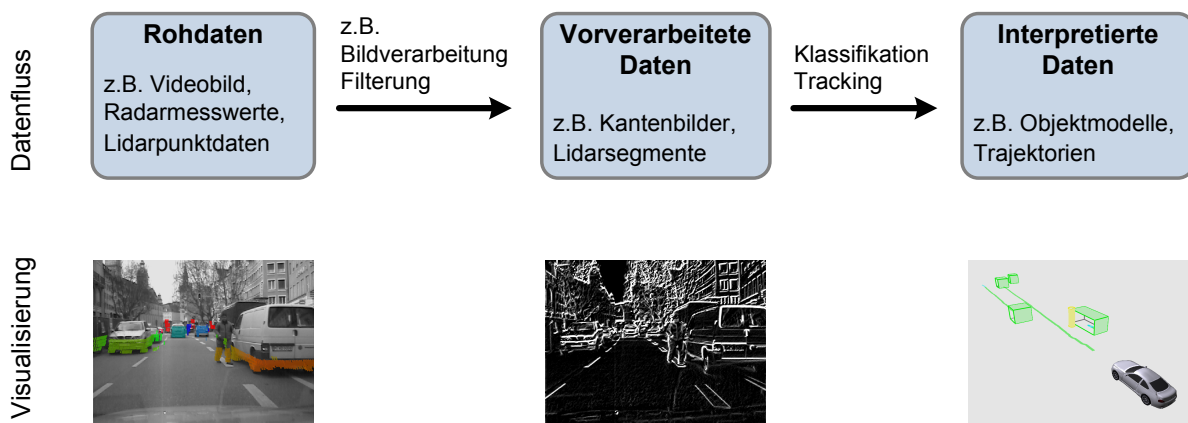


Abbildung 7.1: Datenausprägungen und Informationsfluss innerhalb eines Umfeldersfassungssystems mit Visualisierungsbeispielen.

Effektivität: Um dem Anwender Informationen in intuitiver und ergonomischer Form zu präsentieren, versucht eine effektive Visualisierung die Möglichkeiten des Ausgabemediums, sowie die Fähigkeiten der menschlichen Wahrnehmung, optimal auszunützen [Mac86]. Informationen können zum Beispiel durch Farben effektiv kodiert werden, da sie ähnlich zu Texturen in den frühen Phasen des visuellen Wahrnehmungsprozesses ausgewertet werden und somit zu einer geringen kognitiven Belastung des Anwenders führen [SM00].

Angemessenheit: Wenn Nutzen und Aufwand einer Visualisierung in einem vertretbaren Verhältnis zueinander stehen, spricht man von einer angemessenen Darstellung. Unter Aufwand sind in diesem Zusammenhang nicht nur die Kosten zur Erstellung der Visualisierung, sondern auch der zeitliche Aufwand zur Generierung der grafischen Darstellung zu verstehen. Eine zu komplexe und somit langsame Visualisierung ist nicht mehr ergonomisch und verliert folglich ihre Effektivität.

Eine expressive und effektive Darstellung ist eine notwendige, aber in vielen Fällen nicht ausreichende Voraussetzung für eine adäquate Visualisierung von wissenschaftlichen Daten. Oft können erst durch Manipulation bzw. Interaktion mit den Daten neue Erkenntnisse gewonnen oder Unstimmigkeiten geklärt werden. In der Theorie existieren unzählige Interaktionsmöglichkeiten für den Anwender. In der Praxis lassen sich hierfür allerdings grundlegende Designrichtlinien finden, die Shneidermann [Shn96] als „Information Seeking Mantra“ bezeichnet. Im Folgenden werden diese Richtlinien am Beispiel der Visualisierung eines Fahrumfeldersfassungssystems weiter erläutert:

Kontext: Der Anwender benötigt einen Überblick über die zu visualisierenden Daten, um grobe Zusammenhänge zu erfassen bzw. interessante Bereiche identifizieren zu können. Während der Anwender in Details vertieft ist, trägt eine Übersichtsdarstellung zusätzlich dazu bei, den Überblick zu behalten und Daten in ihrem globalen Kontext zu betrachten.

Für ein Fahrumfeldersfassungssystem sind räumliche Daten aus dem Bereich von wenigen Zentimetern bis zu mehreren hundert Metern typisch. Vordefinierte Kamerapositionen wie beispielsweise die Vogelperspektive helfen dem Anwender, den räumlichen Überblick innerhalb der virtuellen Darstellung des Fahrumfeldes zu behalten. Darüber hinaus kön-

nen ein fahrzeugfestes Koordinatensystem oder die Visualisierung der Position des Eigenfahrzeugs als zusätzliche Orientierungshilfe dienen.

Die Grundlage für die Visualisierung eines Fahrumfelderfassungssystems bilden Szenenkataloge, die sich aus hunderten Aufnahmen aus mehreren Jahren zusammensetzen können. Um in diesen großen Datenmengen den Überblick nicht zu verlieren, werden die einzelnen Aufnahmesequenzen nach Aufzeichnungsdatum und -zeit gruppiert bzw. sortiert. Je nach Anforderung können andere Sortierungs- bzw. Gruppierungskriterien, wie z. B. Kreuzungsszenarien oder Aufzeichnungen mit Gegenverkehr von Interesse sein. Ein Navigationsbalken informiert über die aktuelle Position innerhalb einer Aufnahmesequenz.

Fokus: Nachdem sich der Anwender einen groben Überblick über die Daten verschafft hat, bietet eine vergrößerte Darstellung die Möglichkeit, interessante Daten bzw. Bereiche genauer zu untersuchen. Erfolgt die Vergrößerung in mehreren Stufen, verliert der Benutzer weder die Orientierung noch den Blick für Zusammenhänge.

In einer Fahrumfelderfassung lassen sich einige Vorgänge, wie z. B. die Vorverarbeitung oder bestimmte Messwerte, nur aus der Nähe oder unter einem bestimmten Blickwinkel untersuchen oder beobachten. Mit Hilfe einer virtuellen Kamera hat der Benutzer die Möglichkeit jede beliebige Position und jeden Blickwinkel innerhalb einer virtuellen dreidimensionalen Repräsentation des Fahrumfeldes einzunehmen (siehe Abschnitt. 7.2.5.1). Für die Untersuchung von stationären Objekten wie zum Beispiel Ampeln, Verkehrsschildern oder sonstiger Randbebauung kann der Wechsel zu einer ortsfesten Kameraeinstellung, die nicht an die Bewegung des Eigenfahrzeugs gekoppelt ist, hilfreich sein.

Filterung: Neben der vergrößerten Darstellung bietet das Ausblenden bzw. Herausfiltern von irrelevanten oder störenden Daten dem Anwender die Möglichkeit sich auf ausgewählte Daten zu konzentrieren oder bestimmte Aspekte hervorzuheben.

Insbesondere in einer multisensoriellen Fahrumfelderfassung kann der Blick für das Wesentliche von Messwerten verschiedener Sensoren oder Daten, unterschiedlichen Abstraktionsgrades, verdeckt oder beeinträchtigt sein. Durch das Ein- bzw. Ausblenden von Informationen, wie zum Beispiel bestimmter Sensoren oder Objektklassen, hat der Anwender die Kontrolle darüber, welche Daten im Augenblick relevant sind.

Es gibt allerdings Situationen, in denen weder das Ein- noch das Ausblenden ein zufriedenstellendes Ergebnis liefert. Beispielsweise können Messwerte im Augenblick von Interesse sein, obwohl sie andere Informationen verdecken und somit als störend empfunden werden. Dieser Zielkonflikt kann durch verschiedene Detail- oder Transluzenzstufen² entschärft werden.

Optionale Details (details-on-demand): Sobald die zu visualisierende Datenmenge durch Zoom- und Filteroperationen auf ein überschaubares Maß reduziert worden ist, sollten Detailinformationen zu einzelnen Datensätzen oder Gruppen bei Bedarf eingeblendet werden können.

In einer Fahrumfelderfassung gibt es viele Informationen wie z. B. Beschleunigungen, Gierraten oder Kovarianzmatrizen, die entweder selten benötigt oder schwer grafisch darzustellen sind. Um die Visualisierung nicht unnötig zu überfrachten, werden diese Daten erst bei Bedarf und meist in textueller Form präsentiert.

² Transluzenz ist die partielle Lichtdurchlässigkeit eines Körpers.

In Beziehung setzen (relate): Im Allgemeinen existieren Beziehungen und Zusammenhänge zwischen den zu visualisierenden Daten. Dem Anwender muss die Möglichkeit gegeben werden, bekannte Abhängigkeiten hervorzuheben und unbekannte Zusammenhänge neu zu erschließen.

Insbesondere bei entfernungsgebender Sensorik (Laserscanner oder Radar) ist es schwierig, einen Bezug zwischen einzelnen räumlichen Messwerten und realen Objekten herzustellen: Wurde ein Radarecho von einem Baum oder einem Menschen generiert? Diese Frage ist ohne zusätzliches Kontextwissen in Form weiterer Verarbeitungsergebnisse oder ergänzender Sensorik meist nicht zu beantworten. Eine einfache Art den Anwender in diesem Prozess zu unterstützen, ist die Projektion von 3D Daten in ein Videobild. Abstrakte dreidimensionale Daten können so mit „Bildausschnitten“ attribuiert werden, was sowohl zu einem besseren Verständnis der Daten, als auch zu einer besseren Orientierung innerhalb der virtuellen Umgebung führt.

Jede Stufe im Datenflussdiagramm (siehe Abbildung 7.1) erhöht den Abstraktionsgrad der Daten und erzeugt somit neue Abhängigkeiten. So kann es zum Beispiel von Interesse sein, welche Rohdaten zu welchem Ergebnis der Vorverarbeitung gehören oder welche Messwerte in der Datenassoziation welchen Objekten zugeordnet wurden. Derartige Beziehungen können zum Beispiel durch Verbindungslinien zwischen den beteiligten Objekten hervorgehoben werden.

7.2 Visualisierung des Informationsflusses

Ein Softwareentwickler, der ein automotives Perceptionssystem umsetzt, arbeitet an komplexen Algorithmen, die Sensordaten in weiterverarbeitete und diese in interpretierte Daten wandeln. Um solch eine anspruchsvolle Tätigkeit ausführen zu können, benötigt ein Experte Einblick in die Abläufe und Gespür für die Daten und Werkzeuge zur Fehlererkennung und -suche. Eine rein textuelle Darstellung der internen Mechanismen leistet diese Anforderungen nur unzulänglich und würde schnell die kognitiven Fähigkeiten eines Menschen übersteigen. Wesentlich effektiver und somit hilfreicher ist eine grafische Aufbereitung der Daten und Strukturen, wodurch der Entwickler Zusammenhänge schneller erschließen kann und ein Kontrollinstrument an die Hand bekommt. Durch eine geeignete visuelle Aufbereitung steht dem Entwickler des Weiteren eine Plattform zur Verfügung, die es erlaubt schwierige Sachverhalte zu veranschaulichen und einem fachfremden Dritten darzulegen.

In den folgenden Abschnitten werden die wesentlichen Daten und Strukturen des Informationsflusses eines Perceptionssystems herausgegriffen und mögliche Darstellungskonzepte beschrieben. Zudem werden diese Konzepte, die zum Teil auf pragmatischen Visualisierungsansätzen aus der Sensordomäne beruhen, einer systematischen Ordnung zugeführt. Im praktischen Umsetzungsprozess, der im Rahmen dieser Arbeit entstandenen Fahrumfelderfassung, haben sich die aufgezeigten Visualisierungsmethoden als unentbehrliches Entwicklungswerkzeug erwiesen. Eine fundierte Aussage bezüglich des quantitativen Nutzens kann erst durch aufwendige Benutzerstudien getroffen werden. Gleichwohl ist, durch die konsequente Berücksichtigung der in Abschnitt 7.1.2 eingeführten Designrichtlinien sowie durch die gewonnenen empirischen Werte, ein positiver Effekt auf den Entwicklungsprozess sehr wahrscheinlich.

7.2.1 Sensorverbauorte und -erfassungsbereiche

Die Sensoren eines multisensoriellen Perzeptionssystem sind in der Regel an unterschiedlichen Orten verbaut. Um die Messwerte korrekt in ein gemeinsames Koordinatensystem zu transformieren, ist eine exakte Kalibrierung notwendig, die den Verbauort, die Orientierung sowie interne Abbildungseigenschaften jedes Sensors bereitstellt. Die Güte einer durchgeführten Kalibrierung wird indirekt an den Messwertüberlagerungen sichtbar. Sie kann aber auch direkt über die Visualisierung der Sensorverbauorte und -orientierungen plausibilisiert werden. Im Falle eines automotiven Umfelderkennungssystems wird das Sensorsystem direkt am Fahrzeug befestigt. Steht nun ein virtuelles dreidimensionales Modell des Fahrzeugs zur Verfügung, können die Verbauorte der Sensoren direkt an diesem Modell augmentiert werden, was einen unmittelbaren Rückschluss auf die Qualität der Kalibrierung zulässt (siehe Abbildung 7.2a).

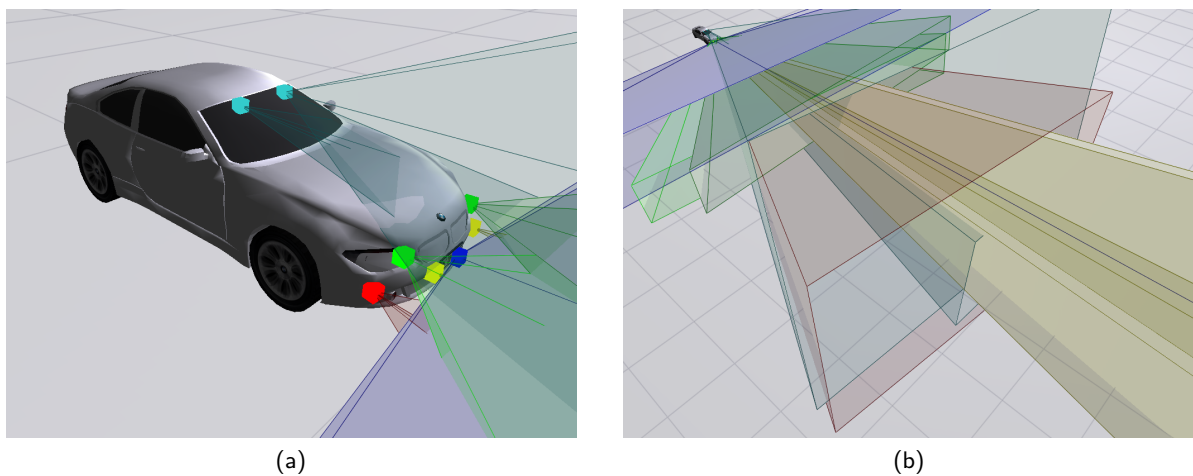


Abbildung 7.2: Visualisierte Sensorverbauorte und -erfassungsbereiche. Abbildung (a) zeigt die Verbauorte von acht Sensoren in einem Versuchsfahrzeug anhand unterschiedlich eingefärbter Würfel. Die jeweiligen approximierten Sichtbereiche sind in Abbildung (b) anhand von transluzenten Pyramiden dargestellt.

Von großer Relevanz für die sensorbasierten Anwendungen, aber auch für die Algorithmmik, sind des Weiteren die räumlichen Erfassungsbereiche der einzelnen Sensoren und deren gegenseitige Überlagerungen. Ein Sensorerfassungsbereich definiert sich über Öffnungswinkel und Sensorreichweite. In der virtuellen Umgebung eignen sich somit entsprechend ausgerichtete Pyramiden oder Kegel, um die jeweiligen Sensorsichtbereiche approximiert darzustellen (siehe Abbildung 7.2b). Die exakte Nachbildung vieler Sensorsichtbereiche ist jedoch nicht trivial, da die Erfassungsbereiche nur ungenau bestimmbar sind, die Grenzen nicht scharf definiert sind oder die Struktur der Bereiche komplexe Formen annehmen. Der Sichtbereich von Mehrstrahlradaren beispielsweise setzt sich messprinzipbedingt aus einzelnen sich überlagernden Radarkeulen zusammen, wobei deren Ränder nicht scharf begrenzt sind.

7.2.2 Rohdaten und vorverarbeitete Sensordaten

Neben Rohdaten, die unmittelbar von der Sensorik geliefert werden, sollte der Entwickler visuellen Zugriff auf weiterverarbeitete Daten und möglichst alle relevanten Zwischenzustände haben.

Grundsätzlich unterteilen sich Daten in Werte und Relationen. Zu Ersteren zählen beispielsweise bildgebende Datenquellen und metrische Messdaten. Letztere sind logische Verknüpfungen zwischen Daten, wie beispielsweise räumliche Messwertüberlagerungen von unterschiedlichen Sensoren an identischen Objekten.

7.2.2.1 Mehrdimensionale Messdaten

Zweidimensionale Messdaten werden häufig durch Streudiagramme dargestellt. Dabei werden Wertepaare, die durch grafische Symbole gekennzeichnet sind, in ein kartesisches Koordinatensystem eingetragen. Ein Beispiel dafür sind aus einer Vogelperspektive betrachtete Radardaten mit den Attributen Entfernung und Winkel, die in kartesische Koordinaten transformiert wurden (siehe Abbildung 7.3). Die unterschiedliche Farb- und Formgebung der Symbole kodiert die Sensorzugehörigkeit des Messwertes. Eine symbolische Darstellung des Versuchsfahrzeugs als Verbauort der Sensoren gibt einen Anhaltspunkt für die Orientierung des Streudiagramms. Das Streudiagramm erlaubt Distanzschätzungen und gibt Aufschluss über Datenhäufungen.

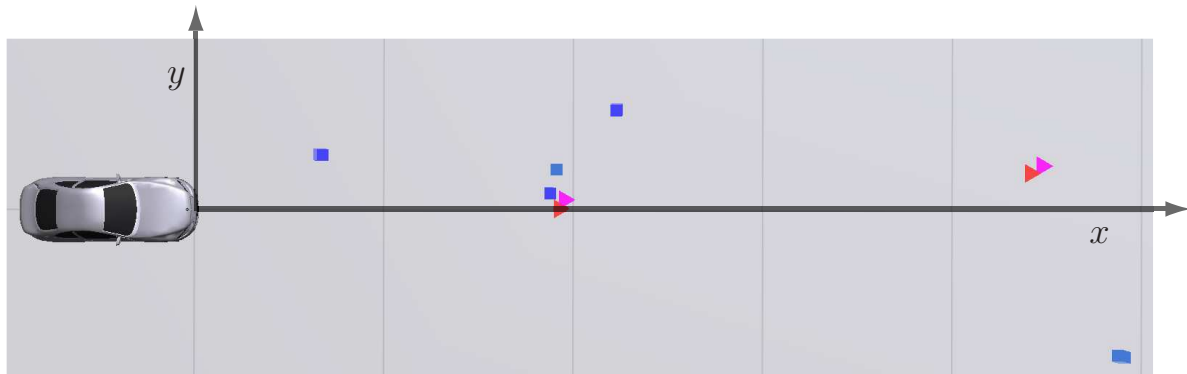


Abbildung 7.3: Zweidimensionale Entfernungsdaten von vier Radarsensoren aus einer Vogelperspektive betrachtet und als Streudiagramm dargestellt. Die unterschiedlichen Sensorzugehörigkeiten sind durch Farb- und Formkodierungen der Symbole gekennzeichnet.

Aktuelle Sensoren, die ein immer detaillierteres Abbild der Realität erstellen, generieren zunehmend dreidimensionale oder sogar mehrdimensionale Messwerte. Eine mögliche Lösung [Spe00], diese Daten zu visualisieren, sind multiple Streudiagramme, die jeweils zwei Dimensionen in einem Standardstreudiagramm darstellen. Damit ergeben sich für r Dimensionen insgesamt $r(r - 1)$ Streudiagramme. Dadurch dass jeder Messwert in jedem Streudiagramm existiert, ergeben sich r analoge Repräsentationen des selben Datenpunktes. Dies erhöht die kognitive Last des Benutzers und verringert zudem den freien Platz für Beschriftungen.

Eine isolierte dreidimensionale Darstellung von räumlichen Sensordaten bietet eine platzsparende Datenvisualisierung, da ein einzelnes Visualisierungsdiagramm ausreichend ist. Zudem ist die Verknüpfung zwischen Realität und deren Abbild für den Betrachter meist intuitiver. Im Gegensatz zu multiplen Streudiagrammen sind mit der Projektion von dreidimensionalen Umgebungen Verdeckungseffekte und Zuordnungsschwierigkeiten verbunden. Diese inhärenten Nachteile einer räumlichen Datenvisualisierung können durch eine interaktive Raumnavigation (siehe auch Abschnitt 7.2.5) aufgelöst werden [Shn03]. Höherwertige Dimensionen von räumli-

chen Messwerten lassen sich über Farbe, Gestalt und Größe der Datenpunkte kodieren (siehe Abbildungen 7.4 und 7.5).

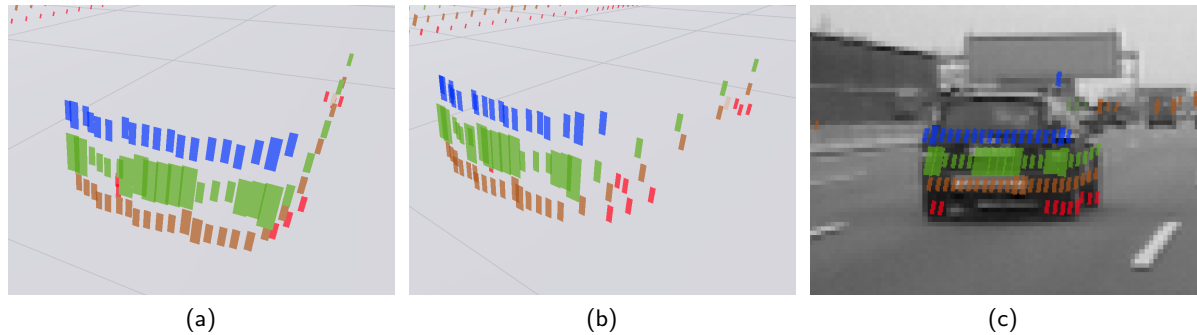


Abbildung 7.4: Multidimensionale Laserscannerdaten unter zwei verschiedenen räumlichen Ansichten (a) und (b) sowie als Projektion in das Videobild (c). Neben der Position im Raum wird durch die Visualisierung die Reflexionsstärke über die Größe, die Abtastebene über die Farbe, und der Reflexionskanal über den Grad der Transparenz dargestellt. Die typische L-Form des Fahrzeugs wird ebenso sichtbar wie die starken Reflexionen des Nummernschildes und der Rückstrahler.

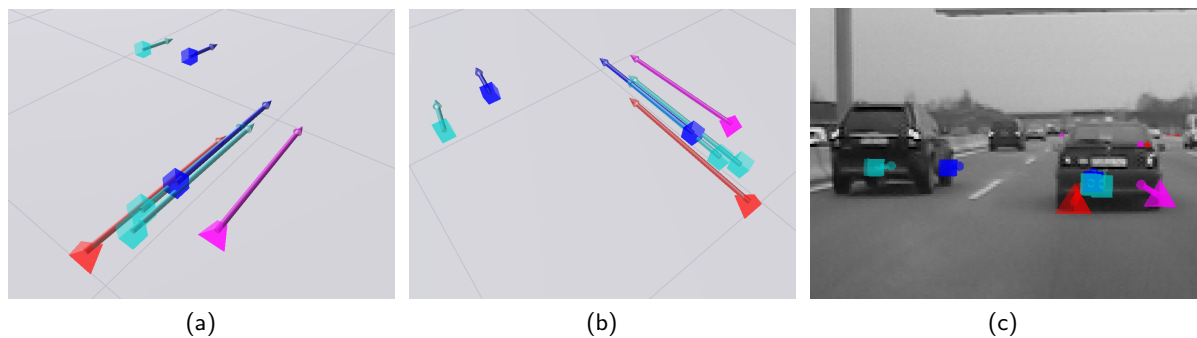


Abbildung 7.5: Multidimensionale Radardaten (a) und (b) sowie deren Projektion in das korrespondierende Videobild (c). Neben der Position des Radarechos im Raum wird die, über den Dopplereffekt gemessene, radiale Relativgeschwindigkeit über die Lage und Länge eines Pfeils sichtbar gemacht.

7.2.2.2 Dynamische Bilddaten

Ein digitalisiertes Bild wird in der Regel als Matrix abgelegt, deren Elemente diskrete Werte (z. B. schwarz-weiß Kamera), dreidimensionale oder mehrdimensionale Vektoren (z. B. RGB Farbkamera bzw. hyperspektrale Sensoren) oder kontinuierliche Werte (z. B. Spektrometer) annehmen können [FP03]. Gewöhnlich werden diese Matrixeinträge direkt auf die Bildschirmspeichermatrix abgebildet und somit dargestellt. Die Abbildungsfunktion ist abhängig von der bildgebenden Sensorik und der Anwendung. Während z. B. RGB-Farbkamerabilder unverändert übertragen und dargestellt werden können, muss bei mehrdimensionalen Bilddaten eine Be-

schneidung erfolgen oder eine Zuordnungsvorschrift (z. B. Falschfarbdarstellung) angewendet werden.

Von der dreidimensionalen Welt können zweidimensionale Bilder erzeugt werden, indem sie durch ein optisches System auf eine Bildebene projiziert wird. Perspektivische Kameras setzen dabei diese Abbildung um, wobei sich deren Projektionseigenschaften über mathematische Modelle beschreiben lassen. Eine detaillierte Einführung in Sensormodelle findet sich in Abschnitt 2.1. Hier sei lediglich das grobe Prinzip erklärt: Ein dreidimensionaler Punkt \mathbf{x}^W , der in einem Koordinatensystem W vorliegt, wird zunächst über eine Transformationsmatrix $\mathbf{T}_{W \rightarrow C} \in \mathbb{R}^{4 \times 4}$ in einen Punkt $\tilde{\mathbf{x}}^C$ in Kamerakoordinaten C überführt und anschließend über eine Projektionsmatrix $\mathbf{P} \in \mathbb{R}^{4 \times 4}$ auf einen zweidimensionalen Punkt \mathbf{x} in der Bildebene projiziert:

$$\mathbf{x} = \mathbf{P} \mathbf{T}_{W \rightarrow C} \mathbf{x}^W \quad (7.2.1)$$

Über diese Rechenvorschrift können Messwerte von entfernungsgebenden Sensoren, aber auch interpretierte Daten, direkt in das Kamerabild projiziert und damit Bildinhalte überlagert werden. Dadurch wird es dem Betrachter möglich, Zusammenhänge zwischen Daten unterschiedlicher Sensoren herzustellen, und auf den Ursprung der Messungen zu schließen (vergleiche Abbildungen 7.4c, 7.5c und 7.8b).

Eine weitere Möglichkeit, Bilddaten in Verbindung mit anderen Sensordaten darzustellen, ist eine texturbasierte Anzeige in einer virtuellen dreidimensionalen Umgebung. Dazu werden die Bilder über ein Musterabbildungsverfahren³ als Textur auf eine virtuelle Ebene im Raum gelegt [Fol95]. Um ein ansprechendes Ergebnis zu erhalten, kann die Darstellung durch ein geeignetes Interpolationsverfahren geglättet werden (vergleiche Abbildung 7.6a). Liegt diese Ebene nun parallel zur Bildebene der entsprechenden Kamera und wird zusätzlich das Modell der virtuellen Szenenkamera (siehe auch Abschnitt 7.2.5) an das der Kamera angeglichen, sind analog zur obigen Bildprojektion passgenaue Überlagerung erzielbar (vergleiche Abbildung 7.6b). Mit Hilfe dieser Vorgehensweise können die Bilder mehrerer Kameras zur Überlagerung gebracht werden [HZ03]. Über den Grad der Transluzenz und über die Ausdehnung der Ebenen wird eine Aufmerksamkeitssteuerung realisierbar.

7.2.2.3 Vorverarbeitete Daten und Relationen

Die Datenvorverarbeitung bündelt Daten mit ähnlichen Eigenschaften (Aggregation), extrahiert durch algorithmische Umformungen markante Merkmale (Merkmalsextraktion) oder filtert die Datenströme anhand ausgewählter Attribute (Filterung). Alle relevanten Vorverarbeitungsergebnisse sollten sichtbar gemacht werden.

Die einfachste Art, Beziehungen bzw. Relationen von aggregierten Datenverbänden darzustellen, sind Verbindungslinien. Über die Farbgebung und die Linienstärke kann die Relation mit weiteren Eigenschaften attributiert werden. Die Abbildung 7.7a zeigt das Ergebnis einer Aggregation von Laserscannerdaten, die mittels Linienzügen visualisiert wird. Fehler in der Segmentierung in Form von unzugeordneten Rohdaten werden unmittelbar sichtbar.

Die Visualisierung von Filterungsvorgängen lässt sich über eine vollständige Darstellungsunterdrückung oder durch eine adäquate Transluzenz- bzw. Farbgebung erreichen. In Abbildung

³ engl. texture mapping

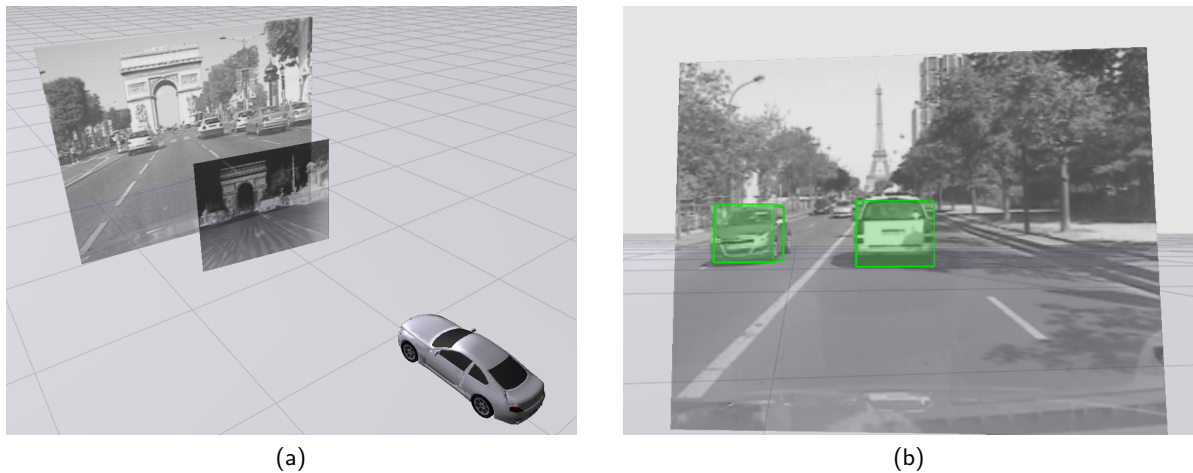


Abbildung 7.6: Visualisierung von Bilddaten in einer dreidimensionalen Umgebung. Abbildung (a) zeigt das Umgebungsbild von einer Videokamera und einer Wärmebildkamera in einer virtuellen Umgebung als Texturen auf zwei Ebenen parallel zu den jeweiligen Bildebenen der Kameramodelle. In Abbildung (b) wurde die Lage der virtuellen Szenenkamera mit der Lage der realen Videokamera in Einklang gebracht. Dadurch ist eine passgenaue Projektion der erkannten Objekte in das Bild möglich.

7.7b werden Daten, die eine bestimmte Eigenschaft nicht erfüllen, stark durchscheinend gezeichnet und damit aus dem Wahrnehmungsfokus entfernt. Gleichzeitig sind sie noch sichtbar, um Rückschlüsse auf die Arbeitsweise der Filteringsroutine zuzulassen.

7.2.3 Interpretierte Daten

Am Ende des Datenflusses einer Fahrumfelderfassung liegen interpretierte Daten in Form von Objektlisten vor. Dazu zählen alle für die nachfolgende Applikationsschicht relevanten Objekte, wie zum Beispiel Fahrbahnmarkierungen und -begrenzungen, Lichtsignalanlagen und Verkehrsteilnehmer. Jedes dieser Objekte wird über Attribute wie Objekttyp, Ausdehnung, Orientierung, Konfidenzen, zeitlicher Verlauf, Position, Geschwindigkeit und Beschleunigung definiert, die zu einem Zustandsvektor zusammengefasst werden. Aufgabe der Visualisierung ist es, diese Objekte in Verbindung mit ihrem Zustand darzustellen.

Die Veranschaulichung von Position, Ausdehnung, Orientierung und Typ der Objekte kann dabei über die Platzierung von primitiven geometrischen Modellen erreicht werden. Fahrzeuge können durch Quader, Fußgänger und Lichtsignalanlagen durch Zylinder und Fahrbahnmarkierungen durch Linien- oder Bezierkurven repräsentiert werden (siehe auch Abbildung 7.8). Sofern eine detailliertere Perzeption notwendig wird, wie beispielsweise der Radstand von Fahrzeugen, die Fußstellung von Fußgängern oder eine exakte Geometrie der Signalanlagen, kann eine sukzessive geometrische Verfeinerung dieser groben Basismodelle vorgenommen werden. Weitere Zustandsinformationen der Objekte, wie Geschwindigkeiten, aber auch Beschleunigungswerte, können anhand von Pfeilen wiedergegeben werden. Mittels der Zeichenfarbe und

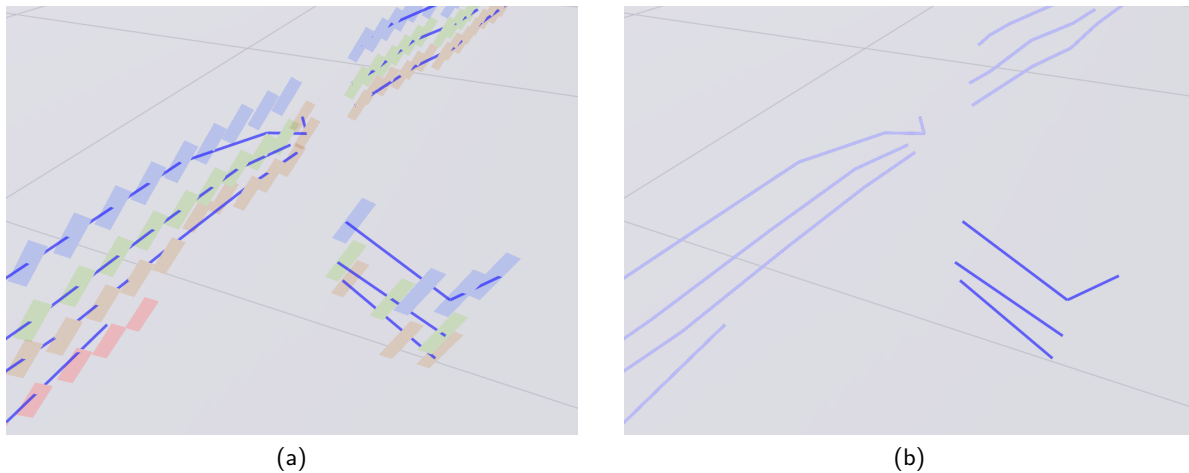


Abbildung 7.7: Abbildung (a) zeigt durch Linienzüge visualisierte Relationen zwischen Rohdaten. In Abbildung (b) wird die Transparenz von Linienzügen erhöht, die eine Länge von 4.0m überschreiten. Dadurch wird ein möglicher Filtervorgang zwischen Fahrzeughypothesen und Randbebauung veranschaulicht.

-transparenz lassen sich sowohl kontinuierliche als auch diskrete Zustandsattribute wie beispielsweise Konfidenz und Objekttyp sichtbar machen.

7.2.4 Zeitlicher Verlauf der Messungen

Erst vielfache Messungen pro Sekunde machen Sensoren tauglich für eine automotive Umfelderkennung. Diese zeitliche Komponente oder vierte Dimension eines Messwertes kann in der Regel durch eine mit der Messung Schritt haltende Visualisierung dargestellt werden. Eine benutzergetriebene Unterbrechung des zeitlichen Ablaufs durch Einzelschritt-, Zeitlupen- und Pausenmodus, gewährt dem Anwender die volle Kontrolle über die Datensequenzen. Der Benutzer kann durch diese Vorgehensweise jedoch lediglich eine Momentaufnahme der Datenaufnahme einsehen, was gegebenenfalls von Nachteil ist.

Eine weitere Möglichkeit, den zeitlichen Verlauf von Messwerten darzustellen, ist in Abbildung 7.9a zu sehen. Vergleichbar zu den Trajektorienverläufen in Abbildung 7.8 werden zeitgleich mit den aktuellen Daten alte Daten an ihren jeweiligen absoluten Messorten angezeigt. Wenn neuere von älteren Messwerten unterscheidbar sein sollen, kann eine adäquate Farb- oder Transparenzgebung (siehe auch Abschnitt 7.2.2.1) der Messwerte von Nutzen sein. Diese verschränkte Darstellung ist durchführbar, solange sich die Messwerte nicht gegenseitig überlagern und damit eine zeitliche Differenzierung erschweren. Vor Allem bei Bilddaten ist es dann sinnvoller, für jeden Zeitschritt eine gesonderte Anzeige zu verwenden.

7.2.5 Navigation und Interaktion

Mehrdimensionale Messwerte sowie weiterverarbeitete Daten werden, wie in Abschnitt 7.2.2.1 beschrieben, in einer dreidimensionalen Darstellung (virtuellen Welt) visualisiert. Hierbei können bedingt durch den Informationsverlust der Projektion von dreidimensionalen Daten auf

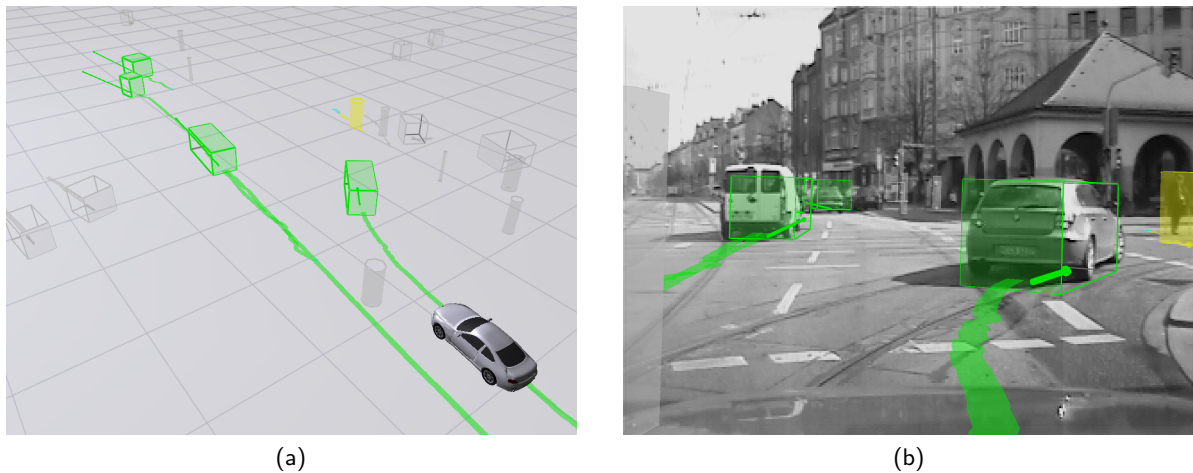


Abbildung 7.8: Visualisierte Objektdaten, deren Trajektorien und Bewegungsrichtungen. Abbildung (a) zeigt eine dreidimensionale Schrägansicht auf eine Straßenszene mit eingblendeten Objekten. Grau eingefärbte Objekte sind Objekthypothesen, grüne Quader stellen erkannte Fahrzeuge und gelbe Zylinder Fußgänger dar. Das eigene Fahrzeug ist durch ein Modell visualisiert. Abbildung (b) zeigt die Projektion der Szenerie in das Bild einer Videokamera.

einen zweidimensionalen Bildschirm Verdeckungen bzw. Mehrdeutigkeiten entstehen. Eine Möglichkeit, dieses Problem zu lösen, sind drei vordefinierte orthogonale Projektionen der Szene, wie sie häufig in Programmen zur 3D Modellierung verwendet werden. Diese Darstellungsart ist besonders dann vorteilhaft, wenn Objekte sehr genau manipuliert werden müssen, da keine perspektivischen Verzerrungen auftreten. Allerdings muss der Anwender gleichzeitig drei unterschiedliche Repräsentationen von einem Objekt verarbeiten, was sich in einer erhöhten kognitiven Belastung niederschlägt [Spe00].

Eine weitere Möglichkeit Verdeckungen und Mehrdeutigkeiten aufzulösen, ist die Navigation, da der Anwender selbst bestimmt unter welchem Blickwinkel Daten untersucht werden sollen.⁴ Insbesondere in einer Fahrumfelderfassung, in der die Manipulation von Objekten eine untergeordnete Rolle spielt, stellt eine einzige dreidimensionale Ansicht der virtuellen Welt eine platzsparende, und den Anwender kognitiv wenig belastende, Alternative zu multiplen Streudiagrammen dar. Allerdings muss hierfür eine intuitive und effektive Navigationsmöglichkeit gefunden werden, die im Folgenden näher beschrieben wird.

7.2.5.1 Navigation

Je nach verfügbaren Eingabegeräten bzw. Aufgabenstellung existieren mehrere Möglichkeiten in einer virtuellen Welt zu navigieren: Ware und Osborne [WO90] benötigen zum Beispiel für die von ihnen beschriebenen Metaphern der „Szene in der Hand“, des „Augapfels in der Hand“ oder des „fliegenden Fahrzeugs“ ein Eingabegerät mit sechs Freiheitsgraden. Hierbei steuert das Eingabegerät direkt eine virtuelle Kamera, manipuliert die gesamte Szene bei fixiertem Blickwinkel oder kontrolliert die Bewegung eines virtuellen Fahrzeugs. Während der Entwick-

⁴ ‘for 3D to be useful, you’ve got to be able to move it’ [Spe00]

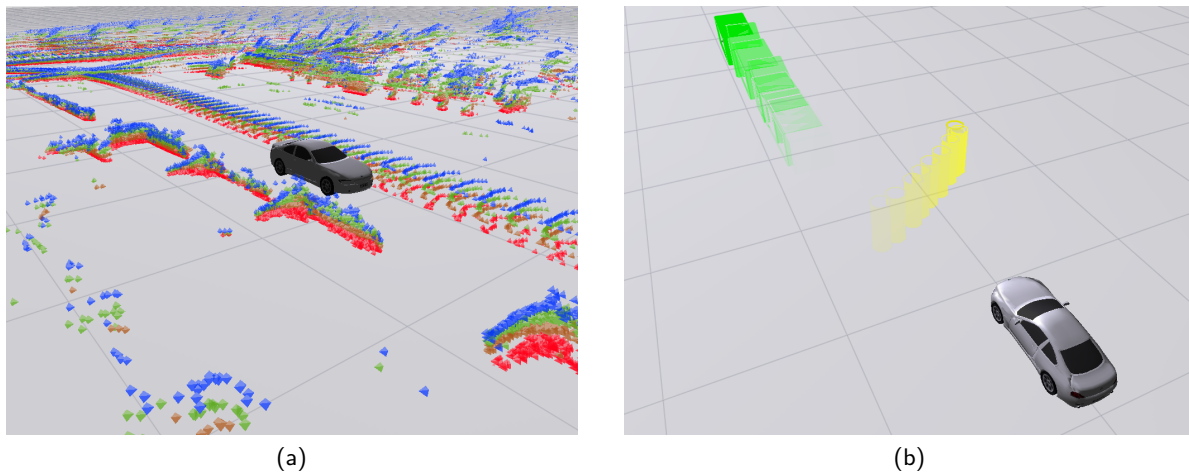


Abbildung 7.9: Visualisierung von zeitlichen Messwertverläufen. Abbildung (a) zeigt Laserscannerdaten in einem Weltkoordinatensystem, die während eines Abbiegevorgangs an einer Kreuzung aufgezeichnet wurden. Am Straßenrand sind parkende Fahrzeuge zu erkennen, und auch die Datenspur des vorausfahrenden Fahrzeugs ist sichtbar. Abbildung (b) zeigt Objektdaten von einem Fahrzeug (grün) und einem Fußgänger (gelb). Weiter in der Zeit zurückliegende Objektpositionen werden dabei durchscheinender gezeichnet.

lung einer Fahrumfelderfassung sind allerdings konventionelle Eingabegeräte wie Maus oder Tastatur vorzuziehen, da sie weit verbreitet und sowohl am Arbeitsplatz, als auch im Fahrzeug problemlos verwendet werden können. Für die Maus wird von Nielson [NO87] eine Technik beschrieben, die zwei- in dreidimensionale Bewegungen überführt, indem der Bewegungsvektor der Maus mit den Projektionen der Koordinatenachsen verglichen wird, und die resultierende Bewegung entlang der Achsen mit der größten Korrelation durchgeführt wird. Diese Art der Navigation ist nicht sehr intuitiv, da bestimmte Bewegungsrichtungen nur über Umwege realisiert werden können. Im Arcball Modell von Shoemaker [Sho92] werden zweidimensionale Mausbewegungen als Kreisbögen auf einer Kugeloberfläche interpretiert. Dadurch können beliebige dreidimensionale rotatorische Eingaben vorgenommen werden. Zu Navigationszwecken im automotiven Einsatz ist dieses Interaktionsmodell zu mächtig, da nicht alle rotatorischen Freiheitsgrade benötigt werden. Eine ähnliche Navigationsmöglichkeit, die zudem translatorische Eingaben ermöglicht, ist die Metapher eines Orbits von Tan [TRC01] bzw. Phillips und Badler [PB88], die wie folgt an die Bedürfnisse einer Fahrumfelderfassung angepasst wurde (vergleiche auch Abbildung 7.10):

Der Betrachter befindet sich in dem Punkt P auf der Oberfläche einer Kugel mit Radius r und blickt in Richtung des Kugelmittelpunktes C , der sich immer auf der xy -Ebene des Weltkoordinatensystems befindet (siehe auch Abbildung 2.7 auf Seite 29). Ausgehend von dem Punkt P hat der Betrachter nur zwei Möglichkeiten seine Position auf der Oberfläche zu verändern: Er kann sich entweder auf einer Kreisbahn bewegen, die parallel zur xy -Ebene liegt und durch P verläuft, oder er läuft die Umlaufbahn entlang, die Orthogonal zur xy -Ebene ist und durch P verläuft. Dadurch können Objekte, die sich innerhalb der Kugel befinden,

von allen Seiten begutachtet werden. Beide Varianten entsprechen in Kugelkoordinaten⁵ einer Änderung des Azimutwinkels ϕ bzw. des Polarwinkels θ . Um ein „Überschlagen“ der Kamera an den Polen zu vermeiden, ist θ auf das Intervall $[-90^\circ; +90^\circ]$ beschränkt. Sollen Details vergrößert betrachtet werden, muss der Benutzer bedingt durch sein starres Sichtfeld näher an das Objekt gelangen. Dies lässt sich dadurch erreichen, dass der Radius r und somit die Höhe der Umlaufbahn variiert wird. Letztendlich erhält der Betrachter die Möglichkeit, sich frei in der virtuellen Welt zu bewegen, indem der Kugelmittelpunkt translatiert wird.

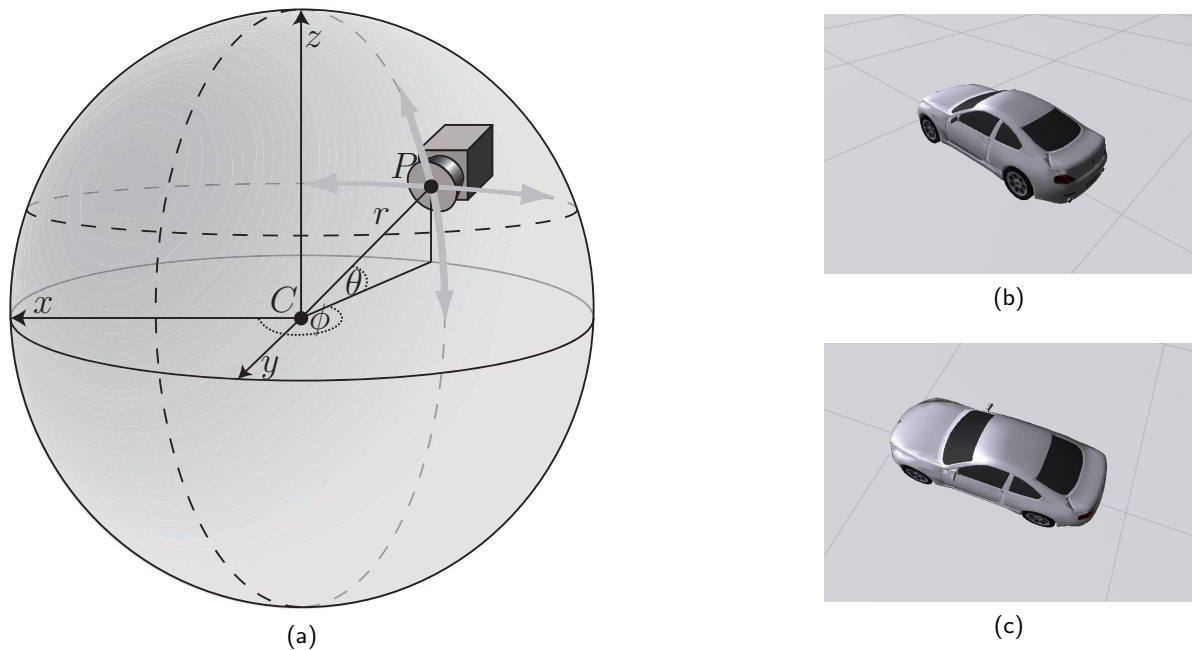


Abbildung 7.10: Orbit-Navigation innerhalb einer virtuellen Umgebung: Der Radius r , der Azimutwinkel ϕ sowie der Polarwinkel θ bestimmen eindeutig die Position P einer virtuellen Kamera auf einer Kugeloberfläche (a). Bewegt sich die Kamera auf einer Umlaufbahn, können Objekte, die sich innerhalb der Kugel befinden, von allen Seiten betrachtet werden (siehe (b) und (c)).

Im Folgenden muss geklärt werden, wie konventionelle Eingabemedien wie Maus oder Tastatur auf die oben beschriebene Navigationsmetapher abgebildet werden können. Da diese weniger Freiheitsgrade als die Orbit-Metapher besitzen, können nicht alle Parameter gleichzeitig verändert werden. Deshalb wird die Navigation in folgende drei Abschnitte unterteilt, die einzeln gesteuert werden:

Translation: Um eine intuitive Bedienung zu gewährleisten, erfolgt die Translation in der egozentrischen Sicht des Betrachters. Eine Bewegung der Maus um Δu nach links bzw. rechts hat eine laterale Verschiebung der Szene aus Sicht des Anwenders zur Folge. Analog wird eine Vor- bzw. Zurück-Bewegung um Δv auf eine longitudinale Änderung abgebildet. Die Granularität der Translation lässt sich über den Skalierungsfaktor τ steuern. Schließlich berechnet sich der neue Kugelmittelpunkt $\tilde{\mathbf{c}}$ aus dem Azimutwinkel ϕ , Δu , Δv , τ sowie dem alten Kugelmittelpunkt \mathbf{c} wie folgt:

⁵ In Kugelkoordinaten ist ein Punkt P über seinen Abstand r zum Kugelmittelpunkt C , den Azimutwinkel ϕ sowie den Polarwinkel θ eindeutig bestimmt.

$$\tilde{\mathbf{c}} = \mathbf{c} + \begin{bmatrix} \cos \phi & -\sin \phi & 0 & 0 \\ \sin \phi & \cos \phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tau \Delta u + \begin{bmatrix} \cos \phi & -\sin \phi & 0 & 0 \\ \sin \phi & \cos \phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \tau \Delta v \quad (7.2.2)$$

$$= \mathbf{c} + \begin{pmatrix} \cos \phi \\ \sin \phi \\ 0 \\ 1 \end{pmatrix} \tau \Delta u + \begin{pmatrix} -\sin \phi \\ \cos \phi \\ 0 \\ 1 \end{pmatrix} \tau \Delta v \quad (7.2.3)$$

Änderung des Blickwinkels: Ausgehend vom Punkt P hat der Betrachter nur zwei Möglichkeiten seine Position auf der Kugeloberfläche und damit seinen Blickwinkel zu verändern: Er kann sich entweder auf einer Umlaufbahn bewegen, die parallel zur xy -Ebene liegt und durch P verläuft, oder er verwendet die Umlaufbahn, die orthogonal zur xy -Ebene ist und ebenfalls durch P verläuft. Jede der beiden Varianten wird eigenständig mit der Maus kontrolliert. Im ersten Fall ergibt sich der neue Azimutwinkel $\tilde{\phi}$ aus der Links- bzw. Rechtsbewegung der Maus um Δu sowie dem alten Azimutwinkel ϕ wie folgt:

$$\tilde{\phi} = \phi + \frac{\pi}{180} \Delta u \quad (7.2.4)$$

Im zweiten Fall berechnet sich der neue Polarwinkel $\tilde{\theta}$ aus der Bewegung der Maus um Δu und dem alten Polarwinkel θ :

$$\tilde{\theta} = \theta + \frac{\pi}{180} \Delta u \quad (7.2.5)$$

Zoom In vielen Anwendungen wird das „Scroll-Rad“ der Maus dazu verwendet, um den Zoomfaktor zu bestimmen. Analog wird der Radius r direkt durch das „Scroll-Rad“ variiert.

Die Tastatur lässt sich analog zur Maus auf die Orbitnavigation abbilden, da Bewegungen der Maus mit unterschiedlichen Tasten sehr gut simuliert werden können. Der neue Punkt P auf der Kugeloberfläche berechnet sich als

$$\mathbf{p} = \mathbf{c} + \begin{pmatrix} r \sin \theta \cos \phi \\ r \sin \theta \sin \phi \\ r \cos \theta \\ 1 \end{pmatrix} \quad (7.2.6)$$

Das Referenzkoordinatensystem, in dem die Navigation durchgeführt wird kann frei gewählt werden. Mit einer virtuellen Kamera in Weltkoordinaten beispielsweise lassen sich stationäre Objekte wie Ampeln, Verkehrsschilder oder sonstige Randbebauung gut beobachten. Bewegte Objekte hingegen verschwinden sehr schnell aus dem Blickfeld des Betrachters. Um dennoch den Fokus nicht zu verlieren, müsste die Position der Kamera in Weltkoordinaten ständig manuell korrigiert werden. In dieser Situation ist die Koppelung der Position und Orientierung der virtuellen Kamera an andere Objekte, wie zum Beispiel dem Eigenfahrzeug, deutlich effektiver. In diesem Betrachtungsmodus ist die Navigation auf das Koordinatensystem des Zielobjektes bezogen. Dadurch befindet sich das fokussierte Objekt fortwährend im Blickfeld der virtuellen Kamera, da jede Positions- und Orientierungsänderung des Zielobjektes automatisch übernommen wird.

7.2.5.2 Interaktion

Sobald die zu visualisierende Datenmenge durch Navigation und Zoom sowie optionaler Filteroperationen auf ein überschaubares Maß reduziert worden ist, benötigt der Anwender weitere Interaktionsmöglichkeiten, um Beziehungen zwischen Daten zu definieren oder Detailinformationen zu einzelnen Datensätzen zu erhalten. Der Anwender hat, wie in Abbildung 7.11a dargestellt, die Möglichkeit die Entfernung zwischen ausgewählten Objekten oder Messwerten zu ermitteln. Diese Art der Beziehung ist beispielsweise hilfreich, um Fehlfunktionen in der Vorverarbeitung oder Datenassoziation zu untersuchen.

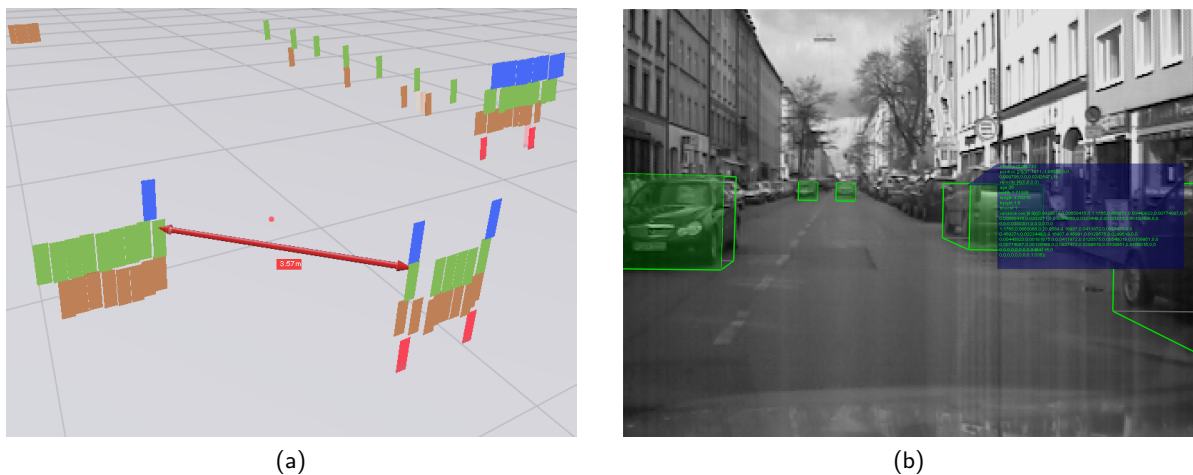


Abbildung 7.11: Unterschiedliche Interaktionsmöglichkeiten: In Abbildung (a) wird die Entfernung zwischen zwei Messwerten des Laserscanners bestimmt. In Abbildung (b) werden verfügbare Detailinformation zu einem ausgewählten Objekt in textueller Form dargestellt.

In einer Fahrumfelderfassung sind einzelne Objekte wie Fahrzeuge oder Fußgänger mit zusätzlichen Informationen wie z. B. Klassifikationsgüten, Merkmalsvektoren, Beschleunigungen oder Kovarianzmatrizen attribuiert, die selten benötigt oder schwer darzustellen sind. Um die Visualisierung nicht unnötig zu überfrachten, kann der Anwender entscheiden, zu welchem Objekt diese Daten in textueller Form präsentiert werden sollen (siehe Abbildung 7.11b).

7.2.6 Werkzeuge der Implementierung

Eine aufwändige Visualisierung von mehrdimensionalen Daten mit Transluzenzdarstellungen, Kantenglättung, Lichtberechnungen und Interaktionsmöglichkeiten benötigt ein hohes Maß an Rechenleistung. Um dennoch den Echtzeitanforderungen gerecht zu werden, ist es sinnvoll alle Darstellungsberechnungen auf einem dedizierten Grafikprozessor auszuführen. Damit kümmert sich ein spezialisierter Rechenkern um die Visualisierung, und der Hauptprozessor steht vollständig für die Sensordatenverarbeitung zur Verfügung. Die Spezifikation OpenGL [SWND05] bietet hierfür eine plattform- und programmiersprachenunabhängige Schnittstelle zur Anwendungsprogrammierung von 3D-Grafik.

7.3 Präsentation und Kommunikation

Die Präsentation und Kommunikation von erzielten Ergebnissen ist eine wichtige Visualisierungsanforderung. Sie schafft eine Diskussionsgrundlage für Experten und veranschaulicht Fakten und Zusammenhänge, so dass sie auch für fachfremde Entwickler leicht verständlich werden. Kommunikationslücken zwischen Sensorikern und automotiven HMI-Designern können dadurch überbrückt, und der Ideen- und Informationsaustausch vereinfacht bzw. gefördert werden. Selbst Laien können durch bildstarke Darstellungen einen Einblick in die Arbeitsweise einer Fahrumfelderfassung gewinnen.

Im Folgenden werden unterschiedliche Präsentationsmedien und -szenarien beschrieben, um Informationen einer Fahrumfelderfassung zu visualisieren. Eine Live-Darstellung des Perzeptionssystems während des Fahrbetriebs stellt dabei andere Anforderungen an die Wahl des Ausgabemediums und des Visualisierungskonzeptes als eine expertengeführte Diskussionsrunde über interne Systemabläufe. Zunächst wird auf die konventionelle Visualisierungsumsetzung im Fahrzeug eingegangen, die eine Systemüberwachung zur Laufzeit ermöglicht. Anschließend werden zukunftsweisende Darstellungskonzepte vorgestellt, die mittels erweiterter Realität gänzlich neuartige Einblicke in die Daten erlauben.

7.3.1 Konventionelle Ausgabemedien

Die Erkennungsgüte und Performanz von automotiven Umfelderfassungssystemen kann besonders eindrucksvoll im laufenden Betrieb im Fahrzeug vermittelt werden. Voraussetzung für eine intuitive Wahrnehmung ist eine echtzeitfähige Visualisierung, die dem Betrachter den aktuellen Zustand des Perzeptionssystems widerspiegelt. Dabei sollte die Art und Weise der Darstellung den Betrachter in der Zuordnung von maschinell erkannten Objekten (virtuelle Objekte) zu realen Umgebungsobjekten unterstützen. Um diese Anforderungen umzusetzen, haben wir in unserem Versuchsträger die Anzeige in eine Kamera- und Übersichtsansicht aufgeteilt (siehe Abbildung 7.12). Auf der linken Bildschirmhälfte wird die vertraute Sicht auf die Straße durch eine den Perzeptionsbereich überblickenden Videokamera dargestellt, deren Bild mit allen virtuellen Objekten halbtransparent überlagert wird. Durch die ähnlichen Blickrichtungen von Kamera und Betrachter wird die Zuordnung von realem zu virtuellem Objekt erleichtert. Die Vogelperspektivenansicht auf der rechten Bildschirmhälfte erlaubt einen Szenerieüberblick und gibt exakten Aufschluss über Positionierung und Ausrichtung der virtuellen Objekte.

7.3.2 Erweiterte Realität

Unter Erweiterter Realität, engl. Augmented Reality (AR), versteht man die Erweiterung der realen Welt durch virtuelle Informationen. So kann zum Beispiel mit Hilfe der AR einem Chirurgen während einer Operation die optimale Position einer Prothese visualisiert oder ein Pilot bei schlechter Sicht sicher zum Flughafen geleitet werden. Im Allgemeinen ist die Erweiterung der Realität nicht nur auf eine Sinnesmodalität beschränkt. Vielmehr sind neben optischer Überlagerungen auch akkustische Anreicherungen denkbar. Für eine detaillierte Einführung in die erweiterte Realität sei auf Azuma [Azu97, ABB⁺01] verwiesen.

Eine ökonomische und realitätsnahe Evaluierung neuer Konzepte ist ein wichtiger Bestandteil in der Entwicklung zukünftiger Fahrerassistenzsysteme. Auf diesem Gebiet gewinnen Methoden der erweiterten Realität neben klassischen Fahrsimulatoren immer mehr an Bedeutung.

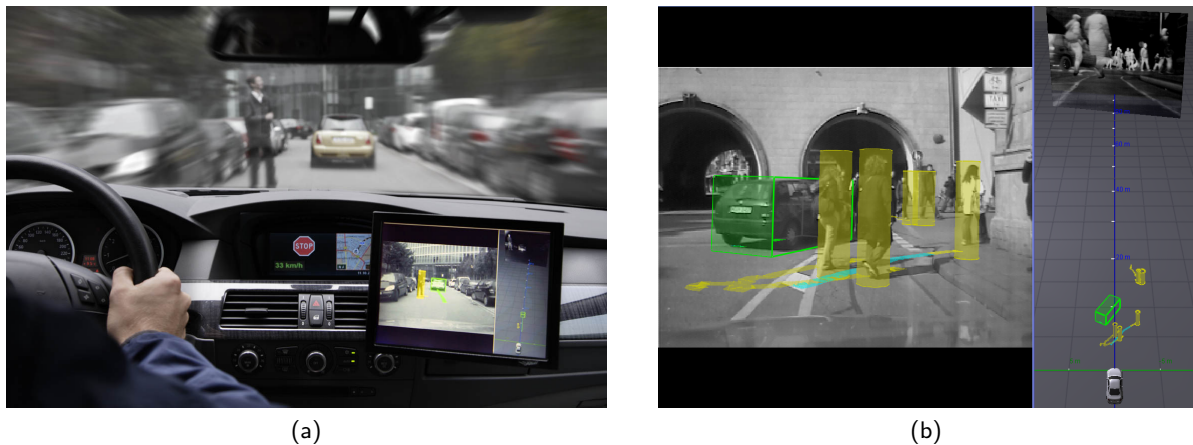


Abbildung 7.12: Visualisierung der Fahrumfelderfassung im Fahrzeug. Abbildung (a) zeigt rechts unten den zu Präsentations- und Überwachungszwecken integrierten Bordmonitor mit der Anzeige der aktuellen Fahrszene. In Abbildung (b) ist die Anzeigeaufteilung des Bordmonitors gesondert illustriert. Links ist das aktuelle Videokamerabild mit einprojizierten maschinell erkannten Umfeldobjekten dargestellt. Die rechte Hälfte des Monitorbildes zeigt eine Vogelperspektivenansicht der erfassten Szenerie mit einem weiteren räumlich eingebetteten Sensorbild (in der Abbildung ein Wärmebild).

So können Benutzeroberflächendesigner zum Beispiel neuartige Fahrerassistenzsysteme, bereits bevor sie technisch realisierbar sind, erleben und evaluieren. Bock et al. [BM06, BMF07] bilden, mit Hilfe der AR, gefährliche Situationen in einer realen Umgebung nach, indem für den Fahrer virtueller Straßenverkehr auf einem Testgelände simuliert wird. Ein ähnliches Prinzip wird von Regenbrecht et al. [RBW05] für ein erweitertes Fahrsicherheitstraining verwendet: Die Realität wird hierbei sowohl für den Schüler als auch den Lehrer mit virtuellen Verkehrsteilnehmern und künstlichen Wetterbedingungen überlagert. Auf diese Weise lernt der Schüler in einer sorgfältig definierten Umgebung, kritische Situationen wie Unfälle, Hindernisse oder schlechte Wetter- bzw. Straßenbedingungen zu meistern.

Im Vergleich zu den starren vordefinierten Anzeigekonzepten, wie sie im vorherigen Abschnitt 7.3.1 beschrieben wurden, kann durch die Verwendung der erweiterten Realität ein sehr individueller Eindruck von den Ergebnissen einer Fahrumfelderfassung vermittelt werden. Dies verschafft fachfremden Entwicklern mehr Verständnis für Messwerte bzw. Daten und die daraus resultierenden Zusammenhänge und Probleme. Je nach verwendetem Anzeigegerät kann die AR für Experten auch als Arbeits- bzw. Diskussionsgrundlage dienen und somit die Zusammenarbeit während der Entwicklung der Fahrumfelderfassung fördern.

In den folgenden Abschnitten wird exemplarisch an zwei unterschiedlichen Szenarien (Labor und Fahrzeug) beschrieben, wie sich Konzepte der erweiterten Realität als Präsentationsmittel in einer Fahrumfelderfassung nutzen lassen. Neben technischen Aspekten in der Realisierung wird in einer abschließenden Diskussion über Erfahrungen berichtet, die während der Erprobung im Labor bzw. Fahrzeug gesammelt wurden. Die Umsetzung und die Auswertung der Ergebnisse erfolgte in Kooperation mit dem Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality der TU München. Eine detaillierte Betrachtung findet sich in der gemeinsamen Veröffentlichung [7].

7.3.2.1 Trackingsystem

In der erweiterten Realität können virtuelle Informationen nur dann korrekt visualisiert werden, wenn der aktuelle Blickwinkel und die Position des Betrachters in Bezug auf die reale Welt bekannt sind. Sobald sich der Benutzer allerdings frei bewegen kann, ist diese Beziehung nicht mehr statisch und die Visualisierung muss jeder Positions- und Blickwinkeländerung folgen.

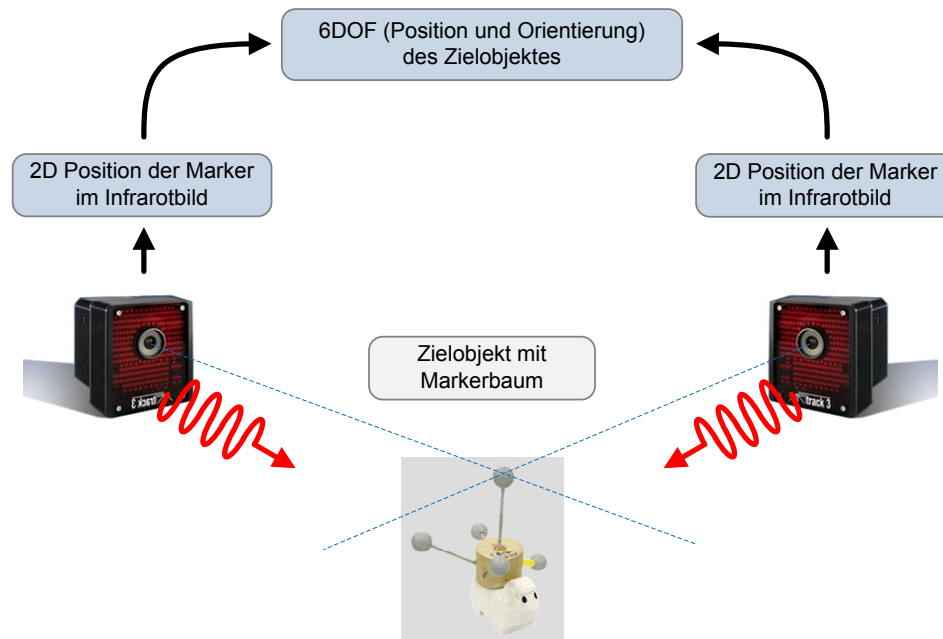


Abbildung 7.13: Schematischer Aufbau eines optischen Trackingsystems: Zwei Infrarotkameras bestimmen die Position der retroreflektierenden Kugeln des Zielobjektes im Bild. Über Epipolargeometrie und Triangulierung wird aus der Bildposition in beiden Kameras die Position der Kugeln im Raum ermittelt. Schließlich kann aus der Lage der einzelnen Marker zueinander die Orientierung des Zielobjektes bestimmt werden.

Für das automatische Verfolgen dieser Parameter werden in der erweiterten Realität optische, magnetische oder akustische Trackingverfahren verwendet. Hierbei bieten optische Verfahren im Allgemeinen die höchste Genauigkeit [Sch04]. Allerdings ist eine Sichtverbindung zwischen den Kameras und den zu verfolgenden Objekten erforderlich. In der technischen Umsetzung wurde ein kommerzielles optisches Trackingsystem⁶ verwendet. In diesem System werden die zu verfolgenden Objekte mit mehreren retroreflektierenden⁷ Kugeln (Marker) versehen. Infrarotkameras erkennen die Marker im Bild und bestimmen mittels Epipolargeometrie und Triangulierung deren Position im Raum. Aus der bekannten räumlichen Anordnung der Marker zueinander kann die Position und Orientierung der Zielobjekte im Raum bestimmt werden. Um die zugrundeliegende maschinelle Bildverarbeitung zu vereinfachen und optimale Lichtbedingungen zu gewährleisten, wird die komplette Szene mit zeitlich synchronisierten Infrarot-

⁶ DTrack von Advanced Realtime Tracking GmbH (A. R. T.).

⁷ Als Retroreflektoren werden Reflektoren bezeichnet, die weitgehend unabhängig von ihrer Ausrichtung, den größten Teil der eingehenden Strahlung zurück in Richtung der Strahlungsquelle reflektieren.

blitzen beleuchtet. Der schematische Aufbau des verwendeten optischen Trackingsystems ist in Abbildung 7.13 dargestellt. Eine detaillierte Einführung sowie eine Beschreibung weiterer Trackingverfahren für die erweiterte Realität wird in [RDB01] gegeben.

7.3.2.2 Anzeigergeräte

Neben dem Tracking sind geeignete Anzeigergeräte (siehe Abbildung 7.14) ein wichtiger Aspekt in der erweiterten Realität. Sie sind dafür verantwortlich, dass Betrachter künstliche Objekte als Teil der realen Welt wahrnehmen. Im Folgenden werden die Anzeigergeräte erläutert, die in der auf AR basierenden Visualisierung eines automotiven Perzeptionssystems verwendet wurden.

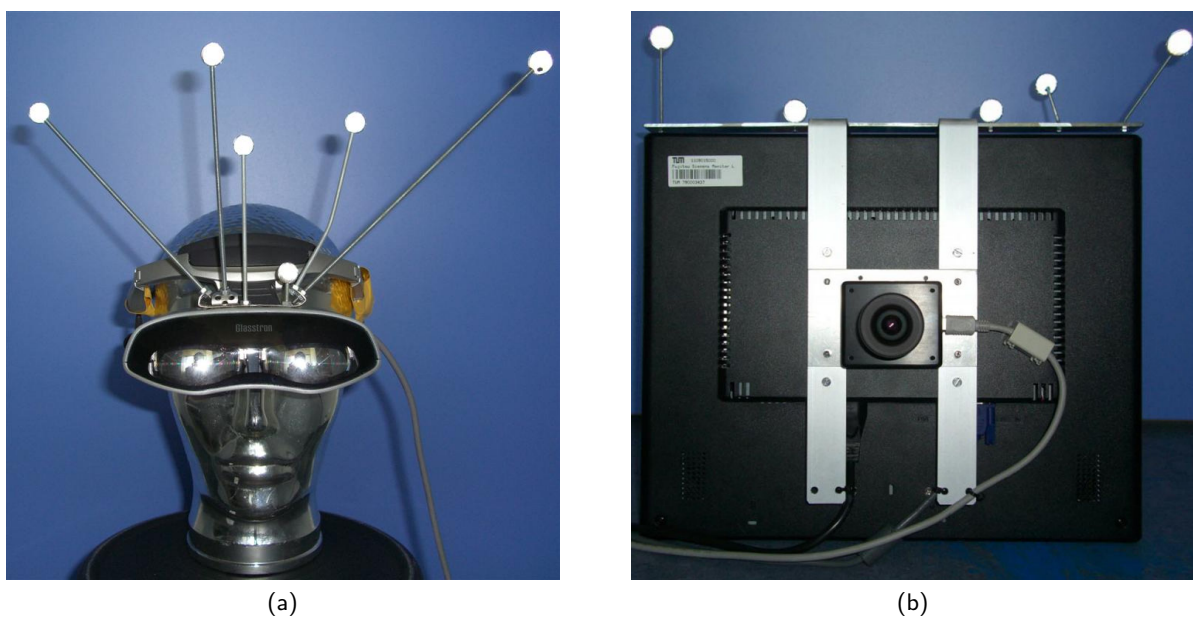


Abbildung 7.14: Unterschiedliche Ausgabegeräte für die erweiterte Realität: Abbildung (a) zeigt ein HMD (Sony Glasstron PLM-S700). Um einen durchsichtigen Bildschirm zu simulieren wurde an die Rückseite eines TFT Monitors (b) eine Kamera installiert. Die zusätzlich angebrachten Markierungskugeln in Abbildung (a) und (b) dienen der robusten Verfolgung der Position und Orientierung der jeweiligen Anzeigergeräte.

Head-Mounted Display: Head-Mounted Displays (HMD) sind visuelle Ausgabemedien, die auf dem Kopf getragen werden. Am Computer erzeugte Bilder werden auf augennahen Bildschirmen dargestellt oder unmittelbar auf die Netzhaut projiziert. In der AR werden HMDs dazu verwendet, um die Realität mit computererzeugten Bildern zu überlagern [RR91]. Dieser Effekt lässt sich mit einem halbtransparenten Bildschirm oder mit zusätzlichen Kameras erreichen. Bei letzteren wird das Videobild der Kamera als Hintergrund für die Augmentierung verwendet und auf einem undurchsichtigen Bildschirm dargestellt. Für die Visualisierung wurde ein „Sony Glasstron PLM-S700“ verwendet, das über zwei semitransparente Bildschirme mit einer Auflösung von 800×600 Bildpunkten verfügt.

Mit diesem HMD kann allerdings kein dreidimensionaler Scheindruck generiert werden, da beide Bildschirme nicht unabhängig voneinander angesteuert werden können. Sechs optische Marker, die, wie in Abbildung 7.14a dargestellt, an der Oberseite angebracht sind, ermöglichen ein robustes und präzises Verfolgen der Position und Orientierung [Bau07].

Tragbarer „Video-see-through“ Bildschirm: Neben halbtransparenten Bildschirmen, wie sie in vielen HMD's zum Einsatz kommen, kann das von einer Kamera erzeugte Abbild der Umwelt dazu verwendet werden, um die Realität mit virtuellen Informationen zu überlagern und dadurch einen transparenten Bildschirm zu simulieren. Hierfür wird eine Kamera an der Rückseite eines Bildschirms angebracht und das von ihr generierte Bild auf der Vorderseite dargestellt (siehe Abbildung 7.14b). Somit lässt sich für den Benutzer der Eindruck eines durchsichtigen Bildschirms erzeugen. Für die Visualisierung wird ein auf diese Weise modifizierter 19" Bildschirm verwendet, auf dessen Oberseite zusätzlich fünf optische Marker angebracht wurden, um ein robustes Verfolgen der Position und Orientierung zu ermöglichen.

7.3.2.3 Visualisierung im Labor

Für die Realisierung des Systems wurde im ersten Schritt eine Laborumgebung gewählt, da hier eine Evaluierung und Optimierung unter vordefinierten Bedingungen und unter dem Ausschluss von externen Störgrößen erfolgen kann. Der gesamte Versuchsaufbau ist in Abbildung 7.15a schematisch dargestellt.

In dieser Umgebung werden aufgezeichnete Daten gleichzeitig auf einem Rückprojektionstisch⁸ und einem tragbaren Bildschirm (siehe Abbildungen 7.14b und 7.15b) dargestellt. Der Tisch dient zum einen als Übersichtsdarstellung und trägt somit zu einer besseren Orientierung innerhalb der Szene bei. Zum anderen bietet er eine Diskussionsgrundlage, wenn mehrere Benutzer die Daten bzw. Ergebnisse begutachten. Nachdem mit Hilfe des Tisches interessante Bereiche ermittelt wurden, ermöglicht der tragbare Bildschirm eine Lupen- bzw. Detailvisualisierung beliebiger Szenenausschnitte (siehe Abbildung 7.15b). Der Anwender kann somit durch die Bewegung des Bildschirms intuitiv den Blickwinkel auf die Daten bestimmen. Hierzu wird durch ein Trackingsystem fortlaufend die Position und Orientierung der Anzeigegeräte bestimmt. Die „Video-see-through“-Funktionalität des tragbaren Bildschirms ist in diesem Szenario nicht sinnvoll und wurde deaktiviert, da die zu visualisierenden Daten außerhalb des Labors aufgezeichneten wurden.

7.3.2.4 Kontaktanaloge Visualisierung im Fahrzeug

Im Fahrzeug werden die Daten der Fahrumfelderfassung sowohl auf einem HMD als auch auf einem tragbaren videobasierten Bildschirm visualisiert. Besonderes Interesse gilt hierbei innerstädtischen Szenen bzw. Fahrten auf der Autobahn. Um den Fahrer nicht unnötig von seiner primären Aufgabe abzulenken, wurde der Versuchsaufbau (siehe Abbildung 7.16a) so gewählt, dass beide Anzeigemodalitäten ausschließlich dem Beifahrer zur Verfügung stehen. Durch die kopfabhängige Perspektivenwahl soll mit beiden Geräten eine möglichst individuelle Visualisierung erreicht werden. Um eine derartige AR realisieren zu können, wurde ein Tracking-System

⁸ Ein Rückprojektionstisch ist ein Tisch, bei dem das Bild von unten auf eine halbtransparente Tischplatte projiziert wird.

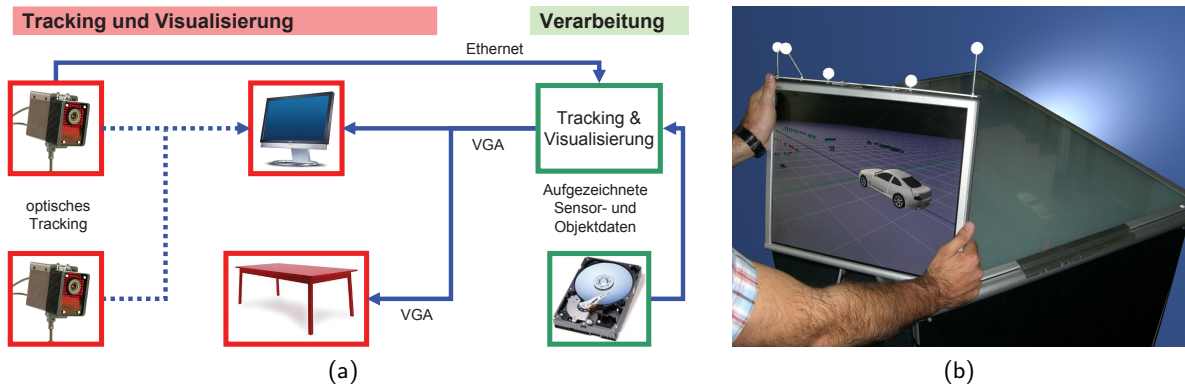


Abbildung 7.15: Visualisierung im Labor: Abbildung (a) zeigt eine Übersichtsdarstellung für das Visualisierungssystem in der Laborumgebung. Der tragbare TFT Monitor in Abbildung (b) dient in der Laborumgebung als Lupen- bzw. Detailvisualisierung für die Daten, die auf dem Rückprojektionstisch in einer Vogelperspektive dargestellt werden.

in das Fahrzeug integriert, das die Position und Orientierung der Anzeigeräte robust verfolgt und darüber hinaus jede Ablenkung und Störung des Fahrers und Anwenders vermeidet. Aus diesem Grund wurden die beiden Kameras des Trackingsystems, wie in Abbildung 7.16b zu sehen, auf der Rücksitzbank mit Blick auf den Beifahrer verbaut. Der maximal mögliche Interaktionsbereich beschränkt sich somit auf den Beifahrer.

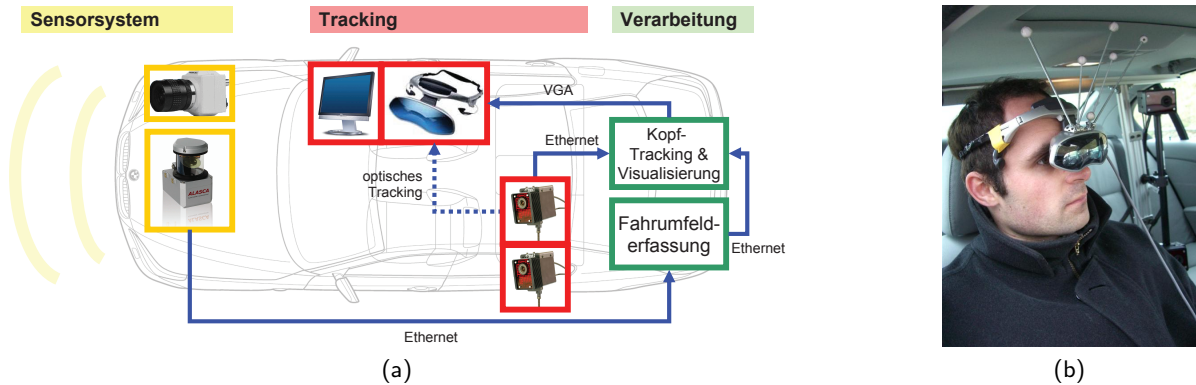


Abbildung 7.16: Kontaktanaloge Visualisierung im Fahrzeug: Abbildung (a) zeigt eine Übersichtsdarstellung für die AR-Visualisierung im Fahrzeug. Ein HMD und ein tragbarer Bildschirm werden verwendet, um Daten der Fahrumfelderfassung zu visualisieren. Für eine korrekte Überlagerung der realen Welt mit Sensor- und Objektdaten verfolgt ein Trackingsystem die Position und Orientierung der Anzeigeräte. Abbildung (b) zeigt einen Nutzer mit HMD auf dem Beifahrersitz. Im Hintergrund ist das Trackingsystem erkennbar.

7.3.2.5 Spatial Relationship Graph

Das Visualisierungssystem muss kontinuierlich die 3D Positionen und Orientierungen (Pose) von Sensordaten, Objekten und Anzeigegeräten zueinander in Bezug setzen. Diese Beziehungen, die sich aus Translationen, Rotationen, Skalierungen oder Projektionen zusammensetzen, lassen sich durch eine Transformationsmatrix \mathbf{T} beschreiben. In Abbildung 7.17 sind die Beziehungen zwischen allen in der AR-Visualisierung beteiligten Komponenten in einem so genannten „Spatial Relationship Graph“ (SRG) [NWB04, PHBK06] dargestellt. Hierbei werden Komponenten durch Knoten und räumlichen Beziehungen zwischen den einzelnen Knoten durch gerichtete Kanten symbolisiert.

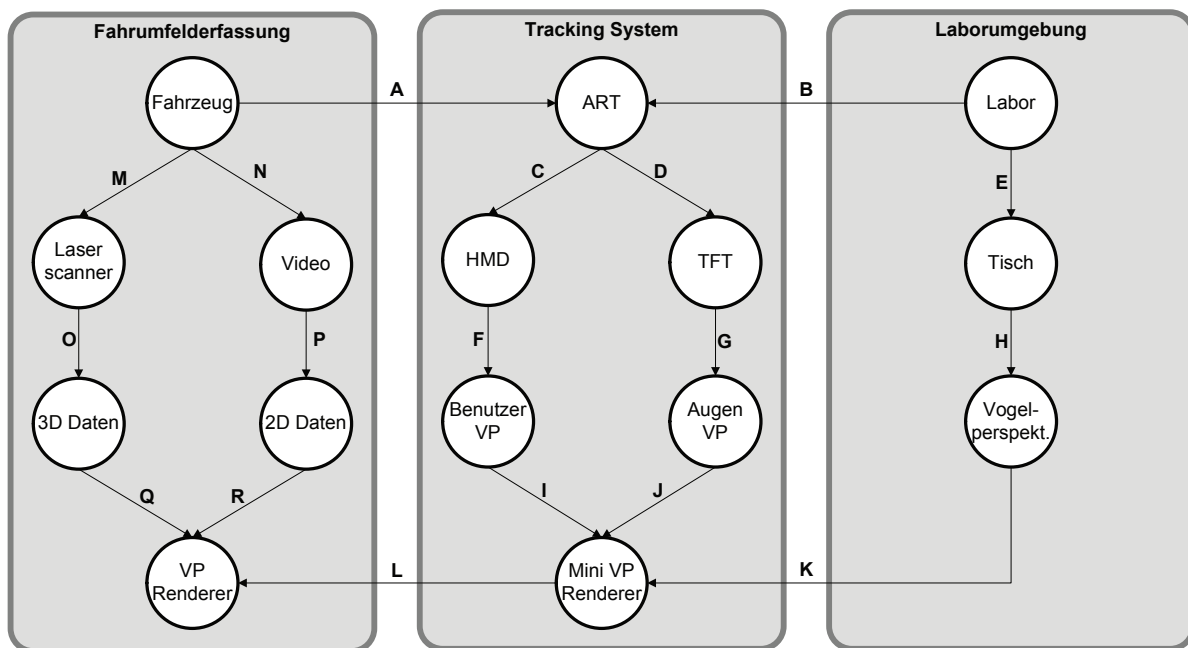


Abbildung 7.17: Der „Spatial Relationship Graph“ des Visualisierungssystems für die Labor- und Fahrzeugumgebung. Er beschreibt die Beziehung zwischen Sensordaten, Objekten und Anzeigegeräten in Bezug auf ihre Position und Orientierung und ihren Einfluss auf die Darstellung der Daten.

Der linke Teil des SRGs beschreibt den internen Aufbau der Fahrumfelderfassung, die aus mehreren Sensoren besteht. Messwerte werden im Allgemeinen im Koordinatensystem des zugehörigen Sensors beschrieben (Kanten **O** und **P**), wohingegen die Position und Orientierung der Sensoren selbst, relativ zum Fahrzeugkoordinatensystem angegeben werden (Kanten **M** und **N**). Innerhalb des SRGs bezeichnet der „VP Renderer“ Knoten den virtuellen Viewport, bezüglich dessen die Szene dargestellt wird. In einer eigenständigen Visualisierung wird dieser Viewport, wie in Abschnitt 7.2.5.1 beschrieben, mit der Maus oder Tastatur kontrolliert (Kanten **Q** und **R**). In der AR-Visualisierung hingegen bestimmen die getrackten Anzeigegeräte über die Kanten **I**, **J** bzw. **K**, wie die Szene dargestellt werden soll.

Der mittlere Abschnitt des SRGs beschreibt den Aufbau des Trackings, das benötigt wird, um innerhalb der Fahrzeugumgebung die Position und Orientierung der mobilen Anzeigegeräte (TFT und HMD) zu ermitteln. Als Trackingsystem wird ein optischer Tracker verwendet (siehe

Abschnitt 7.3.2.1). Der Ursprung des Trackerkoordinatensystems ist abhängig vom Fahrzeugkoordinatensystem (siehe Kante **A**). Für beide Anzeigeräte ermittelt das Trackingsystem die Position und Orientierung der optischen Marker (Kante **C** bzw. **D**). Im Allgemeinen stimmt die Pose der Marker nicht mit der vordefinierten Blickrichtung der Anzeigeräte überein, was mit den zusätzlichen Kanten **F** und **G** kompensiert wird. Das HMD bzw. TFT kontrollieren letztendlich über die Kanten **J** und **I** die eigentliche Darstellung der Szene. Für die Laborumgebung wird die Darstellung um den Faktor $\frac{1}{10}$ skaliert (Kante **L**).

Im rechten Teil des SRG ist die Pose des Rückprojektionstisches innerhalb der Laborumgebung, sowie die Position der Vogelperspektive relativ zum Tisch durch die Kanten **E** bzw. **H** definiert. Über die Kante **K** wird bestimmt, in welcher Lage die Visualisierungsdaten auf dem Tisch dargestellt werden.

7.3.2.6 Kalibrierung der Anzeigeräte

Um Daten auf dem HMD, TFT oder dem Rückprojektionstisch darstellen zu können, muss der Blickwinkel des Anwenders relativ zu den Anzeigeräten beschrieben werden (Kanten **F**, **G** und **H** aus Abbildung 7.17). Das Trackingsystem definiert anhand der Position der optischen Marker zueinander die Position und Lage des Zielkoordinatensystems. Im Allgemeinen befindet sich dieses Koordinatensystem innerhalb der Marker und nicht auf oder innerhalb des Anzeigerätes. Deshalb sind zusätzliche Kalibrierungsschritte notwendig, um die Transformation zwischen dem Zielkoordinatensystem und dem Ausgabegerät bzw. dem Betrachter zu ermitteln. Diese Schritte werden im Folgenden näher erläutert.

Für das TFT (Kante **F** aus Abbildung 7.17) muss die Transformation zwischen Zielkoordinatensystem und Mittelpunkt des Bildschirms gefunden werden. Mit einem getrackten Zeigestab werden hierzu alle vier Ecken (P_1, P_2, P_3, P_4) der sichtbaren Anzeigefläche des TFTs markiert. Durch diese vier Punkte ist der Mittelpunkt $\mathbf{c} = \frac{\mathbf{p}_3 + \mathbf{p}_1}{2}$ und das Koordinatensystem des Bildschirms mit den Basisvektoren $\mathbf{b}_1 = \mathbf{p}_1 - \mathbf{p}_2$, $\mathbf{b}_2 = \mathbf{p}_2 - \mathbf{p}_3$ und $\mathbf{b}_3 = \mathbf{b}_1 \times \mathbf{b}_2$ eindeutig bestimmt.

Für die Kalibrierung des HMDs (Kante **G** aus Abbildung 7.17) wurde die „Single Point Active Alignment Method“ (SPAAM) [TN00] verwendet. Hierbei muss der Anwender ein Fadenzkreuz, das auf dem Bildschirm des HMDs dargestellt ist, mit mehreren vordefinierten Punkten im dreidimensionalen Raum zur Deckung bringen. Der Algorithmus berechnet anschließend die Projektionseigenschaften sowie die Transformation zwischen dem Ziel- und Anzeigekoordinatensystem.

Der Rückprojektionstisch befindet sich an einer vordefinierten Position innerhalb der Laborumgebung. Eine Kalibrierung für die Kante **H** aus Abbildung 7.17 ist deshalb nicht erforderlich.

7.3.2.7 Aspekte der Umsetzung

Die AR Visualisierung wurde im ersten Schritt in der Laborumgebung entwickelt bzw. getestet und anschließend ins Fahrzeug integriert. Die in diesem Prozess gesammelten Erfahrungen sowie geschlossenen Kompromisse werden im Folgenden beschrieben und diskutiert.

Laborumgebung: Um aufgezeichnete Sensordaten in einer Laborumgebung zu visualisieren, mussten folgende Anpassungen vorgenommen werden:

In einer Fahrumfelderfassung sind Sensordaten mit einer Entfernung von über 100 m keine Seltenheit. Damit dennoch eine vollständige Visualisierung in der räumlich beschränkten Laborumgebung erfolgen kann, müssen die original Sensordaten um den Faktor $\frac{1}{10}$ skaliert werden. Darüber hinaus, lässt sich ein skaliertes Abbild der Umgebung einfacher aus verschiedenen Blickwinkeln außer- oder innerhalb des Fahrzeugs untersuchen.

Im Allgemeinen bewegt sich das Fahrzeug in der realen Welt. Eine Darstellung der Sensordaten in Weltkoordinaten erscheint deshalb nicht sinnvoll, da das Fahrzeug während einer Testfahrt schnell die Laborumgebung verlassen würde. Stattdessen wird das Koordinatensystem des Fahrzeugs auf den Ursprung des Trackingkoordinatensystems fixiert (siehe Kanten **B** und **E** in Abbildung 7.17). Die Daten werden somit relativ zum Fahrzeugkoordinatensystem visualisiert.

Die Laborumgebung besteht aus einer vollständigen virtuellen Umgebung, in der keine realen Objekte wie das Eigenfahrzeug oder Sensordaten als Orientierungshilfen existieren. Für einen intuitiven Umgang mit den Daten, muss der Benutzer jedoch wissen, wo und aus welchem Blickwinkel die Szene betrachtet werden soll. Hierfür hat sich die Übersichtsdarstellung auf dem Rückprojektionstisch als hilfreich erwiesen. Darüber hinaus ist die Position des Tisches an den Ursprung des Trackingkoordinatensystems ausgerichtet. Dies dient als weitere Orientierungshilfe für den Anwender.

Tracking im Fahrzeug: Das nötige Trackingvolumen innerhalb des Fahrzeugs kann durch einen Quader beschrieben werden (40 cm vertikal, 80 cm horizontal und 60 cm vor den Beifahrer). In diesem Bereich wird gewährleistet, dass dem Anwender ausreichend Raum für die Interaktion mit den beiden Anzeigegeräten zur Verfügung steht. Um dieses Areal überwachen zu können, wurde ein optisches Trackingsystem verwendet, das über zwei Weitwinkelkameras mit einer Basisbreite von 50 cm verfügt. Die beiden optischen Achsen treffen sich in einer Entfernung von 160 cm, und das Trackingvolumen beginnt in einem Abstand von 30 cm, wo sich die Grenzen der Sichtbereiche nahezu orthogonal schneiden. Mit diesem Aufbau kann für ein Standardziel, abhängig von dessen Entfernung (ca. 50–100 cm) zu den beiden Kameras, eine Auflösung von 0.25 mm bis zu 0.5 mm erreicht werden [Bau07].

Bedingt durch die Größe des Trackingsystems und den minimalen Abstand des Trackingvolumens, war eine Installation des Systems vor dem Beifahrer über dem Armaturenbrett nicht möglich. Die Kameras konnten auch nicht weiter auf die Seite des Fahrers verschoben werden, da jede Ablenkung und Störung des Fahrers vermieden werden sollte. Aus diesem Grund mussten die Kameras hinter den Vordersitzen angebracht werden. Der Versuch das System horizontal hinter der Kopfstütze des Beifahrers oder des Fahrers anzubringen scheiterte, da entweder der Kopf des Anwenders oder eine der beiden Kopfstützen die Sicht der Kameras zu stark beeinträchtigten. Schließlich ermöglichte eine vertikale Installation des Trackingsystems, wie sie in Abbildung 7.18c dargestellt ist, eine ausreichende Sicht auf das geforderte Trackingvolumen. Allerdings können unter bestimmten Voraussetzungen immer noch Verdeckungen entstehen. Zum einen kann der rechte Ellbogen des Fahrers, wenn er auf der Armlehne zu weit nach hinten bewegt wird, die Sicht der Kameras auf den unteren Teil des Volumens beeinträchtigen. Zum anderen werden die Marker des tragbaren „Video-see-through“-Bildschirm, wenn er zu weit nach rechts bewegt wird, teilweise vom linken Arm oder der linken Schulter des Beifahrers verdeckt.

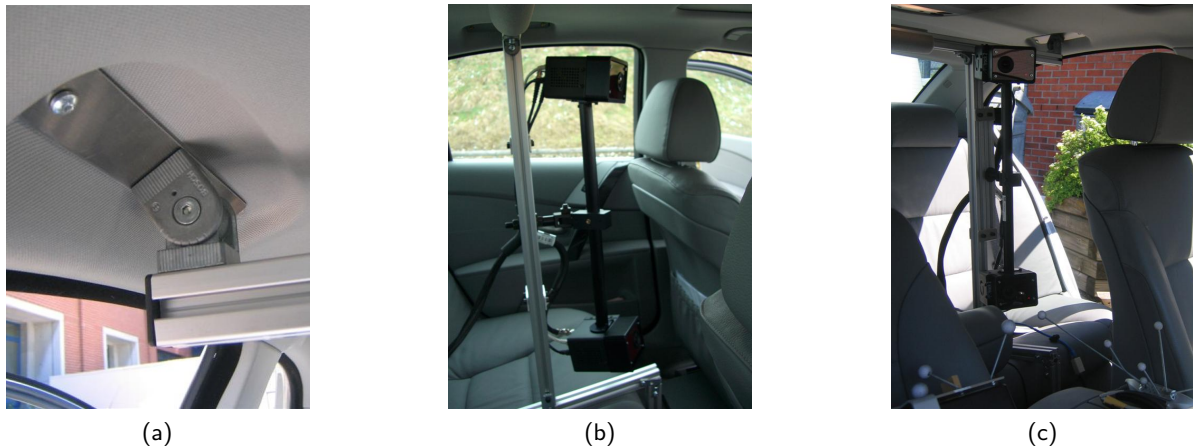


Abbildung 7.18: Installation und Ausrichtung des optischen Trackingsystems im Fahrzeug. In Abbildung (a) ist die Befestigung an den Handgriffen zu sehen. Die vertikale Verbauposition über der Rücksitzbank zwischen den Vordersitzen und die Ausrichtung der Kameras auf das Trackingvolumen (Beifahrer) ist in Abbildung (b) und (c) zu sehen.

Nachdem ein geeigneter Verbauort für das Trackingsystem gefunden wurde, musste sichergestellt werden, dass keine Infrarotlichtquelle innerhalb des Fahrzeugs die Qualität des Trackings negativ beeinflusst. Untersuchungen haben ergeben, dass der Rückspiegel je nach Einstellung unerwünschte Reflexionen der infraroten Beleuchtung erzeugt. Als weitere Störquelle konnte die Nachtbeleuchtung, die im Himmel des Fahrzeugs über dem Schalthebel integriert ist, identifiziert werden. Diese statischen Lichtquellen können allerdings vernachlässigt werden, indem die zugehörigen Bereiche auf dem Kamerabild ignoriert werden. Dieses Verfahren, das auch unter dem Namen „blind spot detection“ bekannt ist, lässt sich jedoch nicht auf externe Störeinflüsse, wie zum Beispiel Sonneneinstrahlung anwenden. Die Thermoverglasung der Windschutzscheibe absorbiert den infraroten Anteil des Sonnenlichts nicht zu 100%, deshalb kann es je nach Einfallswinkel zum Teil zu erheblichen Einschränkungen der Trackingqualität kommen.

Ein letzter Punkt in Bezug auf das Tracking betrifft die stabile und vibrationsfreie Befestigung des Trackingsystems im Fahrzeug. Um sicherzustellen, dass sich das Trackingkoordinatensystem durch Vibrationen des Fahrzeugs während der Fahrt nicht verändert, muss eine starre Verbindung zwischen dem Fahrzeug und dem Trackingsystem erreicht werden. Eine Konstruktion, die die Isofix Kindersitzhalterung und die beiden Handgriffe über den Rücksitztüren verwendet, gewährt ausreichend Stabilität, um den Einfluss lateraler und longitudinaler Vibrationen zu minimieren. Vertikale Schwingungen wurden mit einer zusätzlichen Stütze im Fußraum gedämpft. Die komplette Konstruktion ist in Abbildung 7.18 zu sehen. Um den verbleibenden Einfluss von Schwingungen auf die Trackingqualität zu ermitteln, wurde ein Zielobjekt in einer Entfernung von 120 cm zu den Kameras starr mit der Karosserie verbunden. Das Schwingungsverhalten der Marker ist hierbei zu vernachlässigen, da sie im Vergleich zu den Kameras sehr leicht sind. Basierend auf aufgezeichneten Daten wurde für ein stillstehendes Fahrzeug eine Standardabweichung für die Position von unter 0.1 mm und für die Orientierung von unter 0.2 Grad ermittelt. Während der Fahrt erhöhte sich die Positionsungenauigkeit für alle drei Achsen auf 4 mm. Die Standardabweichung der Orientierung blieb unter einem Grad.

Systemperformanz: Für eine korrekte Überlagerung der Realität mit virtuellen Informationen sind kurze Systemantwortzeiten außerordentlich wichtig. Deshalb wurde besonderes Augenmerk auf eine echtzeitfähige Implementierung gelegt. Die Systemantwortzeit setzt sich aus den Latenzen für das Tracking, für die Kamera des tragbaren Bildschirms, für die Fahrumfelderfassung, für die Darstellung der Szene und für die Kommunikation der Subsysteme untereinander zusammen. Das Visualisierungssystem wurde in OpenGL implementiert und wird von einer 3D Grafikkarte beschleunigt ausgeführt. Dies führt zu einer Darstellungszeit von ca. 20 ms. Die Fahrumfelderfassung benötigt weitere 20 bis 30 ms. Zusammen mit der durchschnittlichen Latenz des Trackingsystems von ca. 20 ms ergibt sich eine durchschnittliche Systemantwortzeit von ca. 70 ms. Der Beitrag der Kommunikationswege ist vernachlässigbar.



Abbildung 7.19: Im Fahrzeug können Sensor- und Objektdaten auf einem tragbaren Bildschirm dargestellt. Abbildung (a) zeigt die Daten des Laserscanners auf einem Parkplatz. Die Beine des Fußgängers werden von den Sensormesswerten überlagert. In Abbildung (b) wird das Ergebnis der Fahrumfelderfassung während der Fahrt auf dem tragbaren TFT visualisiert. Das Bild der Kamera wurde mit drei erkannten Fahrzeugen (grüne Quader) überlagert.

Anzeigegeräte: Während der Fahrt können unterschiedlichste Lichtverhältnisse vorherrschen: Von annähernder Dunkelheit in der Nacht auf Landstraßen bis zu direkter Sonneneinstrahlung am Tag. Unter derartigen Bedingungen kann die Wahrnehmung der erweiterten Realität geschmälert werden, wenn Anzeigegeräte nicht in der Lage sind einen ausreichenden Kontrast im Vergleich zur realen Welt zu erzeugen. Darüber hinaus erhöht sich bei starken Helligkeitsschwankungen die Fokussierungszeit der Augen, wodurch der Anwender schneller ermüdet. Insbesondere bei optisch durchsichtigen Anzeigegeräten wie dem HMD kann der Kontrast in hellen Umgebungen derart abnehmen, dass die Visualisierung für den Menschen nicht mehr sichtbar ist.

Die Darstellung von Sensor- und Objektdaten erscheint auf dem tragbaren Bildschirm in einem guten Kontrastverhältnis zur realen Welt, da das TFT eine Helligkeit von $1000 \frac{cd}{m^2}$ aufweist. Zusätzlich regelt die Treibersoftware der Kamera über die Blende die Helligkeit des Bildes. Abbildung 7.19 zeigt die Visualisierung an einem sonnigen Tag. Allerdings existieren

bei der Verwendung des TFTs einige Einschränkungen: Aufgrund der zentralen Position der Kamera muss der Bildschirm sehr hoch gehalten werden, um einen ausreichenden Blick auf die Szene zu erlangen. So belegt das Armaturenbrett selbst unter maximaler Ausnutzung des zur Verfügung stehenden Raumes innerhalb des Fahrzeugs noch immer ein Drittel des Bildes. Darüber hinaus ist der Bildschirm verhältnismäßig schwer und lässt sich deshalb nur sehr unkomfortabel über längere Zeit verwenden. Ein letzter Punkt betrifft die Position und Anordnung der optischen Marker. Der Anwender muss stets darauf achten, dass die Marker nicht den Fahrzeughimmel berühren und deformiert werden. Da alle Marker auf der Oberseite des Bildschirms annähernd auf einer Achse angebracht sind, kann der Tracker Rotationen um diese Achse schlecht verfolgen.



Abbildung 7.20: Im Fahrzeug werden Sensor- und Objektdaten auf einem HMD dargestellt. Abbildung (a) zeigt die Daten des Laserscanners auf einem Parkplatz. Die Beine der Fußgänger werden von den Messwerten überlagert. Um einen guten Kontrast zu erreichen, wurde in Abbildung (b) das Ergebnis der Fahrumfelderfassung in einer Tiefgarage auf dem HMD visualisiert. Die Realität wird mit dem detektierten Fußgänger (gelber Zylinder) und dem erkannten Fahrzeug (grüner Quader) überlagert.

Beim HMD (Sony Glasstron) wird die reale Welt durch einen halbtransparenten Bildschirm mit virtuellen Objekten überlagert. Der Kontrast der Darstellung ist hier deshalb von besonderer Bedeutung. An einem sonnigen Tag ist der Kontrast des HMDs im Vergleich zum dominanten Umgebungslicht zu gering, und die Visualisierung ist kaum noch sichtbar. An bedeckten Tagen oder nachts kann hingegen ein gutes Ergebnis, wie in Abbildung 7.20 dargestellt, erreicht werden.

7.4 Fahrbahnerkennung mittels Flottenbewegung

Eine geeignete Visualisierung eines Fahrumfelderfassungssystems macht Zusammenhänge zwischen den Daten transparenter und liefert gänzlich neue Einblicke in die Sensordaten. Somit ist sie ein Ideenlieferant für Verbesserungsmöglichkeiten oder gänzlich neuartige Einfälle. Am Beispiel einer umgesetzten Anwendung zur Ermittlung der befahrbaren Fläche durch die Flottenbewegung vorausfahrender Fahrzeuge soll dieses Potential verdeutlicht werden. Die Anregung für die Umsetzung dieses Systems stammt von der Visualisierung der erkannten Fahrzeugflotte und deren Trajektorien (vergleiche auch Abbildung 7.8 auf Seite 150).

Viele Automobilhersteller haben schon Systeme zur Spurverlassenswarnung auf dem Markt. Bisher funktioniert dieses Fahrerassistenzsystem jedoch lediglich auf gut ausgebauten Straßen mit entsprechend vorhandenen Begrenzungsmarkierungen. Fehlen diese Markierungen jedoch, ist eine Lokalisierung der Straße durch diese meist kamerabasierten Ansätze nicht mehr möglich. Neuere Vorgehensweisen versuchen dieses Defizit durch eine direkte, meist texturbasierte Fahrbahnerkennung auszugleichen.

Einen anderen Ansatz verfolgt das im Folgenden detailliert beschriebene System. Dieses ist allerdings nicht als Konkurrenz zu bereits verfügbaren Fahrspurerkennungssystemen zu sehen. Vielmehr liegt das Potential dieser Vorgehensweise in der Informationsfusion, indem die Qualität und Zuverlässigkeit durch zusätzliche Evidenz erhöht wird.

7.4.0.8 Umsetzung

Die einzelnen Trajektorienverläufe der vorausfahrenden Fahrzeugflotte werden gemäß ihrer geschätzten Querausdehnung verbreitert und zu einer aus Dreiecken bestehenden Trajektorienfläche zusammengesetzt. Anschließend werden alle Trajektorienflächen zu einem oder mehreren Bereichspolygonen aggregiert. Diese Polygone stellen den befahrbaren Bereich dar. Durch den Flottenverlauf können zusätzlich Straßenverlauf und -krümmung abgeleitet werden (siehe Abbildung 7.21).

Der Trajektorienverlauf $T_f = (V_1, V_2, \dots, V_n)$ eines einzelnen Fahrzeugs f in der Flotte \mathcal{F} (siehe Abbildung 7.22a) ist das Grundgerüst für das Dreiecksnetz. Es ergibt sich ein schlauchähnliches Konstrukt indem durch jeden zweidimensionalen Verlaufsvektor V_i mit $i \in \{1, \dots, n-1\}$ eine zur Strecke $\overline{V_i V_{i+1}}$ orthogonale Begrenzungslinie $\overline{V_i^l V_i^r}$ gelegt wird. Durch den aktuellen Fahrzeugstandort v_n wird dabei eine Begrenzungslinie parallel zu $\overline{V_{n-1}^l V_{n-1}^r}$ gelegt. Dabei entspricht die Länge von $\overline{V_i^l V_i^r}$ bzw. von $\overline{V_i^r V_i^l}$ der Hälfte der gemessenen Fahrzeugbreite γ_f , vergrößert um eine durchschnittliche Seitenabstandsweite δ zu anderen Objekten. Erfahrungsgemäß hat sich ein Wert von $\delta = 0.5$ m als praktikabel erwiesen. Besser wäre allerdings eine geschwindigkeitsabhängige bzw. kontextsensitive Seitenabstandsweitenadaptation. Das Dreiecksnetz eines einzelnen Trajektorienverlaufs besteht nun aus der Dreiecksmenge

$$\mathcal{D}_f = \{[V_i V_{i+1} V_i^l], [V_i V_{i+1} V_i^r], [V_i^l V_{i+1} V_i^l], [V_i^r V_{i+1} V_i^r]\} \quad \forall i \in \{1, \dots, n-1\} \quad (7.4.1)$$

Nachdem dieses Dreiecksnetz \mathcal{D}_f für jedes Fahrzeug $f \in \mathcal{F}$ erzeugt wurde, werden alle diese Netze zu einem Gesamtpolygon $\mathcal{G} = \bigcup_{f \in \mathcal{F}} \mathcal{D}_f$ aggregiert. Überlappende Bereiche werden mittels Tessellierung⁹ in eine flache Dreieckstruktur umgewandelt. Die Begrenzung des befahrbaren Bereichs bildet die äußere Kontur von \mathcal{G} (siehe Abbildung 7.22b für ein Beispiel mit zwei Flottenfahrzeugen).

7.4.0.9 Bewertung

Eine kontinuierliche Ermittlung der befahrbaren Bereiche ist mit dem vorgestellten Verfahren nicht möglich. Im Wesentlichen müssen hier zwei Vorbedingungen erfüllt sein, damit die Erkennung funktioniert: Es müssen vorausfahrende Fahrzeuge vorhanden sein, und diese müssen über das Perzeptionssystem auch erkannt werden. Erst dann kann über die Fahrzeugtrajektorien und die Breitenschätzung der Fahrzeuge ein befahrbarer Bereich abgeleitet werden. Selbst

⁹ Tessellierung ist eine Technik, die sich mit der Zerlegung von Polygonen beschäftigt.

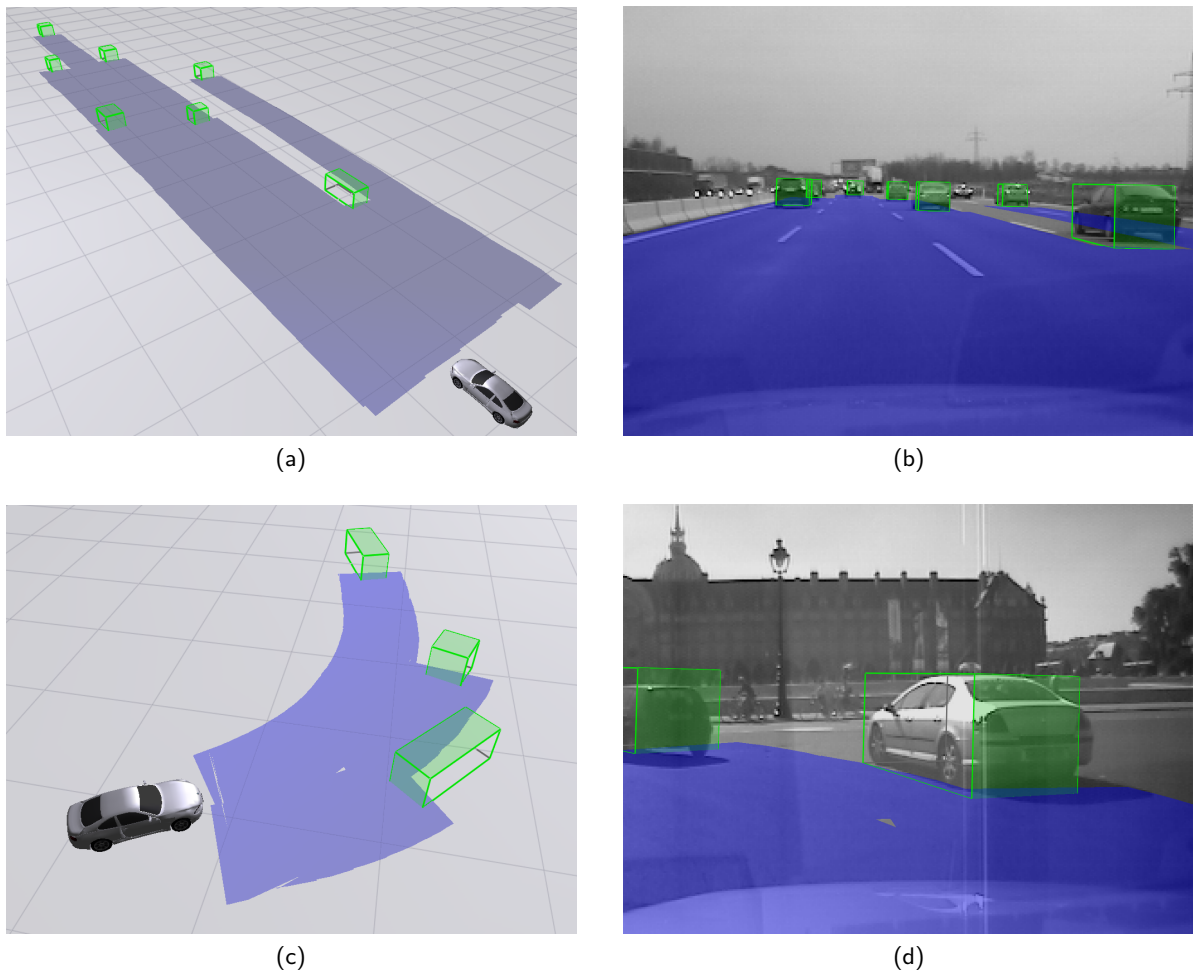


Abbildung 7.21: Befahrbarer Bereich (blaues Polygon), der durch die Flottenbewegung vorausfahrender Fahrzeuge (grüne Quader) ermittelt wurde. Die Abbildungen (a) und (b) bzw. (c) und (d) zeigen dabei jeweils die gleiche Szenerie (Autobahn bzw. Kreisverkehr) aus unterschiedlichen Ansichten (Vogelperspektive und Projektion in fahrzeugintegrierte Videokamera).

wenn diese Bedingungen erfüllt sind, deckt die ermittelte Region, abhängig von der Flottengröße und -verteilung, selten die gesamte zur Verfügung stehende Fahrbahn ab. Meist wird lediglich eine Untermenge detektiert (siehe Abbildung 7.21b). Des Weiteren ist die Qualität der Bereichserkennung direkt an die Qualität der Fahrumfelderfassung gekoppelt. Ist beispielsweise die Breitenschätzung der Flottenfahrzeuge ungenau, wirkt sich das unmittelbar auf die Ausdehnung der einzelnen Spuren aus. Eine Krümmungsschätzung der Fahrbahn aufgrund der Flottenbewegung ist ebenfalls mit Einschränkungen verbunden. Nur wenn die detektierte Flotte eine ausreichend große Zahl an Fahrzeugen beinhaltet, können Ausreißer, die den globalen Krümmungswert verfälschen würden (z. B. durch einen Spurwechsel), ausfindig gemacht werden. Ist die beobachtbare Flotte sehr klein oder besteht lediglich aus einem einzelnen Fahrzeug, sind die generierten Spurdaten, selbst bei einer fehlerlosen Fahrumfelderfassung, mit einer verminderten Verlässlichkeit zu betrachten. Das Vorderfahrzeug kann sich beispielsweise

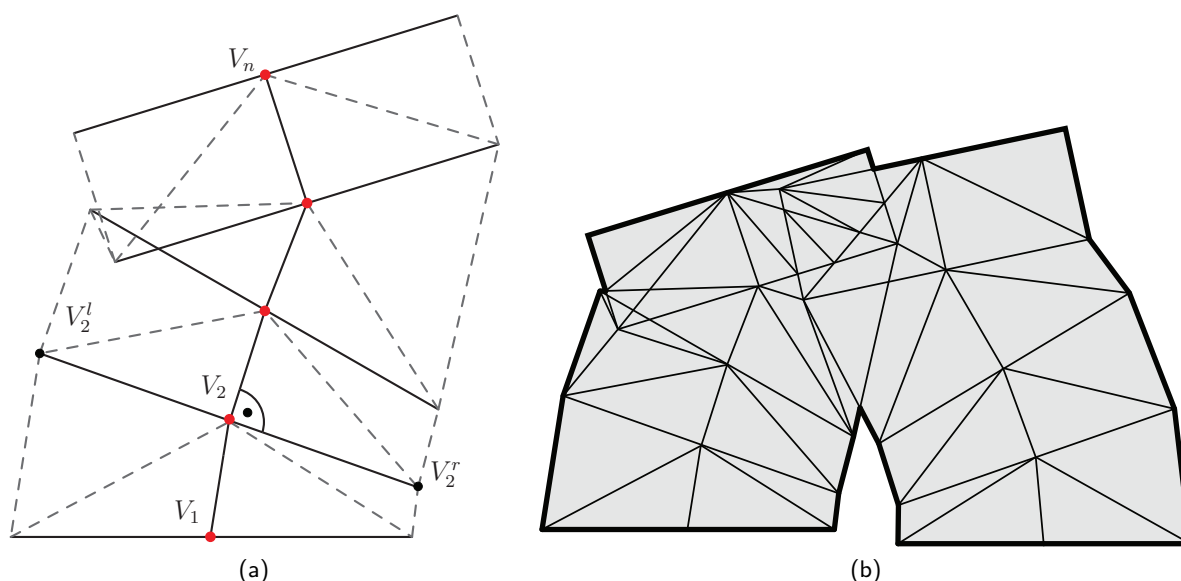


Abbildung 7.22: Polygonerzeugung aus den Flottentrajektorien. Abbildung (a) zeigt die diskrete Trajektorie eines einzelnen Fahrzeugs (rote Punkte) und das daraus erzeugte Dreiecksnetz. In Abbildung (b) werden die Dreiecksnetze von zwei Fahrzeugen mittels Tesselierung zu einem gemeinsamen Polygon aggregiert.

regelwidrig verhalten, oder durch einen Fahrfehler oder äußere Einwirkungen von der Fahrbahn abkommen.

Trotz dieser signifikanten Nachteile hat das umgesetzte Fahrbahnerkennungssystem mittels der Flottenbewegung einige entscheidende Vorteile. Ist eine Fahrumfelderfassung im Fahrzeug, wie sie für eine steigende Anzahl an Fahrerassistenzsysteme (z. B. Abstandshaltung, Gefahrenbremsung, Spurwechselassistent) benötigt wird, ohnehin vorhanden, verursacht diese Art der Fahrbahnerkennung keine zusätzlichen Hardwarekosten, sofern die nötige Software mit den verfügbaren Ressourcen auskommt. Selbst wenn dieses Verfahren aufgrund der signifikanten Nachteile nicht als exklusives Spurerkennungssystem betreibbar ist, liegt das Potential in der Informationsfusion. Eine herkömmliche markierungsbasierte Spurerkennung kann nämlich in vielen Fällen von zusätzlichen Informationen profitieren. Sind beispielsweise die Spurmarkierungen durch Verdeckung, schlechte Wetterverhältnisse oder an Baustellen nicht oder nur unzulänglich erkennbar, ist eine robuste Detektion der Markierungsverläufe nicht gewährleistet. Zudem ist die Entfernungreichweite der visuellen Spurerkennung durch die Auflösung der bildgebenden Sensorik begrenzt und könnte über die Flottenbewegung, die mittels entfernungsgebender Sensorik erfasst wird, vergrößert werden.

Zukünftige Spurerkennungssysteme werden von den Informationen aus markierungsbasierter Spurerkennung, Straßenkartendaten, Fahrzeug-zu-Fahrzeug-Kommunikation und Flottenbewegung profitieren. Wie eine solche Informationsfusion gewinnbringend realisiert werden kann, ist Gegenstand kommender Forschungsaktivitäten.

7.5 Diskussion

Als Darstellungsmöglichkeiten für heterogene Sensordaten wurden Streudiagramme, eine virtuelle dreidimensionale Fahrzeugumgebung und Messwertprojektionen in das Videobild vorgestellt. Eine von diesen Möglichkeiten alleine genügt jedoch nicht vollständig den Visualisierungsanforderungen wie Expressivität, Effektivität und Angemessenheit. Streudiagramme erlauben ähnlich wie Vogelperspektivendarstellungen eine gute Repräsentation von Entfernungen und Datenhäufungen. Lagebeziehungen von räumlichen Daten können aber nur unzureichend veranschaulicht werden. Einen besseren Einblick und eine flexiblere Darstellung gewährt eine interaktiv dreidimensionale Umgebungsmodellierung des Fahrzeugs und der entfernungsbeinhaltenden Sensormesswerte. Eine Benutzernavigation ist dabei unabdingbar, damit sowohl Verdeckungen der Daten untereinander als auch Mehrdeutigkeiten durch den fehlenden stereoskopischen Raumeindruck aufgelöst werden können. Gleichzeitig erhöht eine adäquate Navigationsmöglichkeit die Flexibilität der Darstellung. Sowohl vogelperspektivische als auch isometrische Ansichten sind dadurch denkbar.

Für Bilddaten ist eine räumliche Visualisierungsumgebung weniger geeignet. Es können zwar Kamerabilder als Textur auf Polygone aufgebracht und dadurch dargestellt werden, eine Zuordnung von räumlichen Messwerten zu zweidimensionalen Bilddaten ist dadurch aber nicht unmittelbar möglich. Erst durch eine Übertragung der intrinsischen und extrinsischen Kamerageometrie auf eine virtuelle Szenenkamera ist eine korrekte Anreicherung der Bilddaten mit Entfernungsdaten durchführbar. Der Gesamtüberblick über die räumlichen Daten geht aber meist verloren, da der Kameraerfassungsbereich oft nur einen Ausschnitt der Szene überdeckt. Aufgrund der unterschiedlichen Stärken und Schwächen dieser Darstellungsmöglichkeiten hat sich eine Verbunddarstellung aus „Projektion in das Bild“ und dreidimensionaler virtueller Szenenvisualisierung sowohl im Fahrzeug als auch in der Entwicklungsumgebung bewährt. Nachteile der Darstellungsformen wie der mangelhafte Gesamtüberblick in der Bildprojektion und die unzulängliche Bilddarstellung in der virtuellen Umgebung werden dadurch wechselseitig kompensiert.

Ein möglicher Ansatz, die starren Visualisierungsmethoden aufzulockern, bietet der Einsatz von Erweiterter Realität (AR). Durch kontaktanaloge Visualisierungen mittels getrackter Anzeigemedien können Messwerte in Verbindung mit der Realität erlebbar gemacht werden. Dazu wurden optische Trackingsysteme verwendet, um sowohl in der Labor- als auch in der Fahrzeugumgebung Head-Mounted Displays und tragbaren „Video-see-through“-Bildschirmen als AR-Darstellungsgeräte nutzbar zu machen. Die verwendeten optischen markerbasierten Trackingsysteme liefern zwar eine hohe Positionsgenauigkeit, sind aber relativ anfällig für Verdeckungen und ungünstige Beleuchtungsbedingungen. Speziell in der Fahrzeugumgebung wird die Trackingleistung durch starke Sonneneinstrahlung negativ beeinflusst. Ein weiterer Nachteil bedingt sich durch die beschränkten Raumverhältnisse im Fahrzeug. Der dadurch reduzierte Interaktionsradius wird zusätzlich durch die erhabenen Markerbäume begrenzt. Vibrationen der Trackingkameras konnten durch einen robusten Einbau minimiert werden. Latenzzeiten wurden mittels einer mit einem 3D-Prozessor beschleunigten Anzeige und kurzen Kommunikationspfaden gering gehalten. Trotz der bestehenden technischen Probleme, die möglicherweise mit alternativen Trackingansätzen (magnetisch) beseitigt werden könnten, fördert die aktuelle Umsetzung von AR im Kontext von Perzeptionssystemen das Sensordatenverständnis von fachfremden Entwicklern und die Gruppenarbeit während der Entwicklungsphase eines Umfelderkennungssystems.

Zudem bieten sowohl die beschriebenen herkömmlichen Visualisierungsansätze als auch die Ansätze der Erweiterten Realität nicht nur eine Plattform zur Fehlererkennung und -beseitigung, sondern auch eine Inspirationsquelle für kreatives Denken und innovative Ideen. Dabei ist ein quantitativer Performanzvergleich der einzelnen Visualisierungsmöglichkeiten für automotiv Perzeptionssysteme nicht einfach. Selbst wenn sinnvolle Vergleichsmetriken zur Verfügung stünden, könnten erst aufwendige Nutzerstudien eine abgesicherte Aussage treffen. Die in diesem Kapitel aufgestellten Thesen stützen sich deshalb vorwiegend auf wissenschaftlich fundierten Visualisierungsrichtlinien und die im Rahmen dieser Arbeit gewonnene Erfahrung bei der Umsetzung der Fahrumfelderfassung.

7.6 Zusammenfassung

In diesem Kapitel wurden Visualisierungsmöglichkeiten von heterogenen Sensordaten eines automotiven Umfeldfassungssystems vorgestellt. Die Gründe, warum eine geeignete Informationsrepräsentation von Vorteil ist, sind vielfältig: Der Systementwickler wird in der Umsetzung unterstützt, indem ihm durch eine adäquate Visualisierung ein Instrument an die Hand gegeben wird, das interne Abläufe und Zusammenhänge transparenter macht, die Fehler-Erkennung und Suche erleichtert und ein Gespür für die Daten und Datenbeziehungen fördert. Durch eine geeignete visuelle Aufbereitung steht dem Entwickler des Weiteren eine Plattform zur Verfügung, die es erlaubt schwierige Sachverhalte zu veranschaulichen und einem unbedarften Dritten darzulegen. Dadurch werden Kommunikationslücken zwischen fachfremden Experten geschlossen und der Ideen- und Informationsaustausch vereinfacht und gefördert. Die Präsentations- und Kommunikationswirkung ist ein weiteres wichtiges Ziel einer Visualisierung. Durch bildstarke Informationsrepräsentationen können verborgene Prozesse sichtbar gemacht und medienwirksam dargestellt werden.

Um dieses Visualisierungspotential ausschöpfen zu können, wurde im Rahmen dieser Arbeit eine dreidimensionale virtuelle Fahrumgebungsdarstellung entwickelt, in der alle relevanten Daten des Informationsflusses eines automotiven Perzeptionssystems repräsentiert werden können. Zudem wurde eine Navigations- und Interaktionsmöglichkeit in die virtuelle Umgebung integriert, die dem Benutzer einen Überblick über die Daten, eine Auflösung von Verdeckungen und Mehrdeutigkeiten und eine Betrachtung von Details ermöglicht. Die verwendeten Konzepte, die zum Teil auf pragmatischen Visualisierungsansätzen aus der Sensordomäne beruhen, wurden einer systematischen Ordnung zugeführt.

Das entwickelte Darstellungssystem wird mittels Zeichenfilter und frei konfigurierbarer Ansichtsmodi vielfältig und modular eingesetzt. Zum einen dient es als Offline-Entwicklungswerkzeug am Arbeitsplatz. Zum anderen demonstriert es die aktuelle Perzeptionsleistung während der Fahrt im Versuchsträger. Zusätzlich wurde die gleiche Visualisierung verwendet, um über Anzeigeräte wie Head-Mounted Displays, „Video-see-through“-Bildschirmen und Rückprojektionstische die Sensordaten in einer erweiterten Realität erlebbar zu machen. Das Fallbeispiel „Fahrbahnerkennung mittels Flottenbewegung“ am Ende des Kapitels macht deutlich, wie Visualisierungen den Ideenentstehungsprozess anstoßen können.

8

Schluss

Im Rahmen der vorliegenden Dissertation wurde das Potential multisensorieller Sensorik zur Erkennung von Fahrzeugen und Fußgängern im automotiven Kontext untersucht. Durch die Integration von heterogenen Sensordaten konnte ein leistungsfähiges Umfelderkassungssystem realisiert werden, das die Detektion und Klassifikation von Verkehrsteilnehmern in Echtzeit ermöglicht. In diesem Kapitel werden die erzielten Ergebnisse der vorliegenden Arbeit zusammengefasst. Zudem wird ein Ausblick hinsichtlich weiterführender Themengebiete gegeben.

8.1 Wissenschaftlicher Beitrag

Der wissenschaftliche Beitrag ist weit gefächert, da aufgrund der Zielsetzungen der Dissertation (siehe auch Abschnitt 1.2 auf Seite 3), ein breites Aufgabenspektrum abgedeckt wurde. Ergebnisse, die in Gemeinschaftsarbeit erzielt wurden, sind gesondert ausgewiesen.

Aufmerksamkeitssteuerung: Es wurde experimentell nachgewiesen, dass durch eine kombinierte Aufmerksamkeitssteuerung mit einem entfernungs- und einem bildgebenden Sensor, die Ausfallsicherheit und die Erkennungsleistung gesteigert werden kann (siehe Abschnitt 3.3.1). Durch eine verschränkte und gezielte Erzeugung der Hypothesen kann im Vergleich zu einer erschöpfenden Objektsuche, sowohl eine geringere Laufzeitkomplexität als auch eine höhere Erkennungsrate erzielt werden (siehe Abschnitt 3.3.2). Zudem wurde demonstriert, dass eine multikategorische Klassifizierung effizient realisierbar wird, da viele Multiklassenprobleme durch eine geometriebasierte Vorklassifikation in binäre Entscheidungsprobleme umgewandelt werden können.

Merkmalsgewinnung: Zur Klassifikation von Fahrzeugen und Fußgängern wurden Merkmale aus Videokameras, Wärmebildkameras und Laserscannersensoren extrahiert. Es wurden die optimalen Parametrisierungen für Gradientenhistogramme im Wärme- und Videobild ermittelt, um maximale Erkennungsraten zu erzielen (siehe Abschnitt 4.2.1).

Für die Bildmerkmale wurde zudem untersucht, wie sich Distanzänderungen von Fahrzeugen und Fußgängern auf die Erkennungsraten auswirken. Erst ab einer Halbierung der Bandbreite ist ein signifikanter Performanzrückgang zu beobachten. Des Weiteren wurde ein neues Verfahren für eine Sichtbereichsvergrößerung des bildgebenden Sensors, durch eine Achsenspiegelung von symmetrischen Objekten, vorgestellt und bewertet. Dadurch können einsicherende Fahrzeuge frühzeitiger klassifiziert werden (siehe Abschnitt 4.2.3).

Klassifikatorfusion: Es wurde das Potential verschiedener Ansätze zur Klassifikatorfusion untersucht, um heterogene Merkmale zur Erkennung von Fahrzeugen und Fußgängern einzubeziehen. Die Güte der thematisierten Verfahren ist weitgehend unabhängig von der gewählten Objektklasse.

Von den betrachteten Verfahren erzielt die Metaklassifikation die höchsten Erkennungsraten. Allerdings ist das Verfahren lediglich bei fehlerlosen Merkmalszusammensetzungen geeignet. Wird die Güte einzelner Merkmale, etwa durch partielle Sensorausfälle beeinträchtigt, ist die Verwendung des arithmetischen Mittels als funktionaler Kombinator eine zweckmäßige Wahl (siehe Abschnitt 5.4). Insgesamt bewirkt das Miteinbeziehen von zusätzlicher Sensorik einen signifikanten Anstieg der Erkennungsleistung.

Systembewertung: Zur Systemevaluierung mussten zunächst adäquate Werkzeuge geschaffen werden. In Zusammenarbeit mit dem FORWISS Institut der Universität Passau wurde ein Evaluierungswerkzeug für die Referenzdatengewinnung im automotiven Umfeld konzipiert und umgesetzt [6]. Dieses ermöglicht eine räumliche manuelle Attributierung von Verkehrsteilnehmern in beliebigen Fahrscenarien (siehe Abschnitt 6.1.1). Eine Evaluierung des Gesamtsystems bzw. eine Untersuchung des Zusammenspiels zwischen Klassifikation und Tracking wurde dadurch realisierbar.

Mittels eines winkelauflösenden entfernungsgebenden Sensors wird bei Autobahn- und Landstraßenszenen unter normalen Witterungsbedingungen eine gute Erkennungsleistung erzielt. Durch die Hinzunahme weiterer Sensoren konnte jedoch in dieser Domäne kein signifikanter Leistungszuwachs beobachtet werden. Erst in komplexen Stadtszenarien, sowie unter widrigen Umweltbedingungen, zeigt sich ein deutlicher Nutzen einer multisensoriellen Fahrumfelderfassung. Die durchgeführte Laufzeitanalyse belegt, dass das damit verbundene erhöhte Datenaufkommen mit aktuellen Rechensystemen in Echtzeit bewältigt werden kann (siehe Abschnitt 6.2.4).

Sensordatenvisualisierung: Im Rahmen der hier vorliegenden Dissertation und der Arbeit von Rudi Lindl [Lin08] wurde eine dreidimensionale virtuelle Fahrumgebungsdarstellung entwickelt, in der alle relevanten Daten eines automotiven Perceptionssystems repräsentiert werden können. Zudem wurde eine Navigations- und Interaktionsmöglichkeit in die virtuelle Umgebung integriert, die dem Benutzer einen Überblick über die Daten und eine Auflösung von Verdeckungen und Mehrdeutigkeiten ermöglicht. Die verwendeten Konzepte, die zum Teil auf pragmatischen Visualisierungsansätzen aus der Sensordomäne beruhen, wurden einer systematischen Ordnung zugeführt (siehe Abschnitt 7.2).

Um starre Visualisierungsmethoden aufzulockern, wurden optische Trackingsysteme aus dem AR-Bereich genutzt, wodurch sowohl in der Labor- als auch in der Fahrzeugumgebung eine kontaktanaloge Nutzung von Head-Mounted Displays und tragbaren „Video-see-through“-Bildschirmen ermöglicht wurde. Die Umsetzung und Auswertung der AR-Ergebnisse [7] erfolgte in Kooperation mit dem Lehrstuhl für Informatikanwendungen in der Medizin & Augmented Reality der TU München (siehe Abschnitt 7.3.2).

8.2 Ausblick

Die vorliegende Arbeit bietet Lösungsansätze zur echtzeitfähigen multisensoriellen Klassifikation im automotiven Umfeld. Dabei wurde jedoch lediglich eine Teilmenge der Techniken betrachtet, durch die eine möglichst robuste Perzeption erzielt werden kann. Zahlreiche Herausforderungen bleiben bestehen.

Klassifikatorfusion

Das Gebiet der Klassifikatorfusion ist eine verhältnismäßig junge Disziplin. Es gibt bislang kaum eine fundierte theoretische Basis, und die Erkenntnisse beschränken sich vorwiegend auf empirische Untersuchungen. In der vorliegenden Arbeit wurde eruiert, welche Fusionsstrategien bei einer vorgegebenen Merkmals- und Objektzusammensetzung optimal sind. Von Interesse wären jedoch allgemeingültige Aussagen zu Fusionsvorgehensweisen, die jedoch erst durch ein besseres Verständnis der internen Zusammenhänge abgeleitet werden können. Dafür ist eine Fortführung der Grundlagenforschung im Bereich der Klassifikatorfusion notwendig.

Integration von Kontextwissen

Neben multisensoriellen Sensordaten ist es zudem denkbar, dass zusätzliche Informationsquellen, wie beispielsweise Kartendaten, Fahrzeugzustandsinformationen und telematische Daten, kontextabhängig in den Objekterkennungsprozess einfließen. Beispielsweise könnten für den Fußgängerklassifikator in Autobahn- und Stadtszenarios entsprechende Sensitivitätsparametrisierungen getroffen werden. Durch diese Integration von Kontextwissen steht dem Klassifikationsprozess eine breitere Wissensbasis zur Verfügung, wodurch mittels adäquater Algorithmen eine zuverlässigere Entscheidungsfindung möglich wird.

Multikategorische Klassifikation

In dieser Arbeit wurden bivariate Klassifikationsprobleme betrachtet. Das bedeutet, dass ausschließlich zwischen einer Vorder- und Hintergrundklasse unterschieden wurde. Eine multikategorische Klassifikation wurde umgangen, indem mittels der Hypothesengenerierung eine Vorklassifizierung vorgenommen wurde. Für viele Fahrerassistenzsysteme ist es relevant, dass eine Vielzahl von Objektklassen unterschieden werden, wie beispielsweise PKW, Lieferwagen, LKW, Motorradfahrer, Fahrradfahrer, Fußgänger und Verkehrszeichen. Dafür müssen spezielle Klassifikationsansätze, wie beispielsweise Multi-Klassen-SVM, herangezogen werden. Eine multisensorielle Betrachtung von derartigen Konzepten im automotiven Einsatz, würde ein neues Forschungsfeld eröffnen.

Zeitliche Konfidenzbildung

Die Entscheidungsfindung in dieser Arbeit wurde anhand zeitlich lokaler Merkmalsinformationen getroffen. Eine Robustifizierung der Klassifikation kann womöglich erzielt werden, wenn Konfidenzinformationen über eine gewisse Zeitspanne in den Klassifikationsprozess einfließen. Durch den temporalen Verlauf der Konfidenzen kann oftmals eine besser fundierte Aussage

über die Klassenzugehörigkeit von Objekten getroffen werden, als durch eine auf einen Zeitpunkt beschränkte Entscheidungsfindung.

8.3 Fazit

Unter idealen Voraussetzungen werden durch eine multisensorielle Klassifikatorfusion, im Vergleich zu einem Einzelklassifikator, erhebliche Leistungssteigerungen erzielt. Im automotiven Umfeld müssen allerdings spezielle Aspekte, wie (partielle) Sensorausfälle und ungenaue Sensorkalibrierungen berücksichtigt werden. Dadurch ist die theoretisch zu realisierende Erkennungsleistung nicht unmittelbar auf den automotiven Bereich übertragbar. Im Gegenteil: Werden automotiv Gesichtspunkte außer Acht gelassen, können die multisensoriellen Erkennungsraten unter den möglichen Leistungswerten eines einzelnen Sensors liegen.

Bei einfachen Fahrszenarien (beispielsweise Fahrt auf Autobahn ohne Niederschlag) ist durch ein Multisensorsystem kein wesentlicher Mehrwert beobachtbar. Aufgrund des unstrukturierten Hintergrunds und der einfachen Klassifikation von sich signifikant bewegenden Fahrzeugen, kann diese Szene allein durch einen winkelauflösenden entfernungsgebenden Sensor sehr gut interpretiert werden. Erheblich bessere Leistungsmerkmale, im Vergleich zu einem Einzelsensorsystem, werden in komplexen Stadtszenarien erzielt. Dort profitiert die Klassifikation erheblich von der Aufmerksamkeitssteuerung der entfernungsgebenden und den heterogenen Merkmalen der bildgebenden Sensorik – insbesondere bei der Erkennung von Fußgängern. Ebenso unter widrigen Witterungsbedingungen werden die Vorzüge eines Fusionssystems deutlich.

Mit den untersuchten Fusionsalgorithmen konnte, trotz der hohen Sensorintegration, keine ausreichende Erkennungszuverlässigkeit erreicht werden, wie es für massiv intervenierende Fahrerassistenzsysteme wünschenswert wäre. Auf dem Weg zu einer optimalen Fahrumfelderfassung müssen somit noch zahlreiche Herausforderungen bewältigt werden. Eine verschränkte algorithmische Auswertung von heterogenen Merkmalen aus bild- und entfernungsgebenden Sensoren ist jedoch ein entscheidender Schritt in die richtige Richtung.

A

Datenkataloge

Zum Training und Testen der Klassifikatoren, sowie für die Parametrisierung der Merkmalsextraktionsverfahren wurden verschiedene Datenkataloge zusammengestellt. Die Kataloge wurden überwiegend manuell (siehe auch Abschnitt 6.1.1) erstellt und beinhalten Video-, Wärmebild- und Laserscannerdaten von Fahrzeugen und Fußgängern. Um die Anzahl der Daten in den Bilddatenkatalogen zu verdoppeln, wurden zusätzlich vertikal gespiegelte Bildansichten hinzugenommen. Eine Übersicht über alle Datenkataloge sowie deren Umfang und Beschreibung ist in Tabelle A.1 aufgelistet.

Tabelle A.1: Übersicht über die Datenkataloge, die zum Training und Testen, der in dieser Arbeit untersuchten Verfahren, herangezogen wurden. Mit positiv wird die abgebildete Objektklasse bezeichnet. Negative Daten beinhalten Hintergrundobjekte, die von der Objektklasse verschieden sind.

Katalog	Klasse	Sensor	Umfang	Beschreibung
Katalog 1	Fahrzeugheck	Video	1500 positive und 6000 negative Bilder	Die Bildgrößen betragen 60×60 Bildpunkte.
Katalog 2	Fahrzeugheck	FIR	1500 positive und 6000 negative Bilder	Die Bildgrößen betragen 60×60 Bildpunkte.
Katalog 3	Fußgänger	Video	3000 positive und 6000 negative Bilder	Die Bildgrößen betragen 64×128 Bildpunkte (siehe auch Abbildung A.1).
Katalog 4	Fußgänger	FIR	3000 positive und 6000 negative Bilder	Die Bildgrößen betragen 64×128 Bildpunkte (siehe auch Abbildung A.2).

Tabelle A.1: Übersicht über die Datenkataloge, die zum Training und Testen, der in dieser Arbeit untersuchten Verfahren, herangezogen wurden. Mit positiv wird die abgebildete Objektklasse bezeichnet. Negative Daten beinhalten Hintergrundobjekte, die von der Objektklasse verschieden sind.

Katalog	Klasse	Sensor	Umfang	Beschreibung
Katalog 5	Fahrzeugheck	Laser-scanner	630 Heckansichten und 7600 Hintergrundansichten	Der Merkmalsvektor hat eine Kardinalität von 10 und enthält die normierte vertikale Reflexionsverteilung von Fahrzeughecks.
Katalog 6	Fußgänger	Laser-scanner	Frequenzanalyse von 80 Fußgängern und 72 fußgängerförmigen Objekten (Ampeln, Bäume)	Das Merkmal ist binär und kodiert das Vorhandensein einer Pendelbewegung.
Katalog 7	Fahrzeuge	Laser-scanner und RADAR	12700 positive und 16800 negative Dynamikdaten	Der Merkmalsvektor beinhaltet Geschwindigkeit und Alter der Objekte.



Abbildung A.1: Beispielbilder aus Datenkatalog 3, der Fußgänger- (erste und zweite Zeile) und Hintergrundansichten (dritte und vierte Zeile) im Videobild beinhaltet.

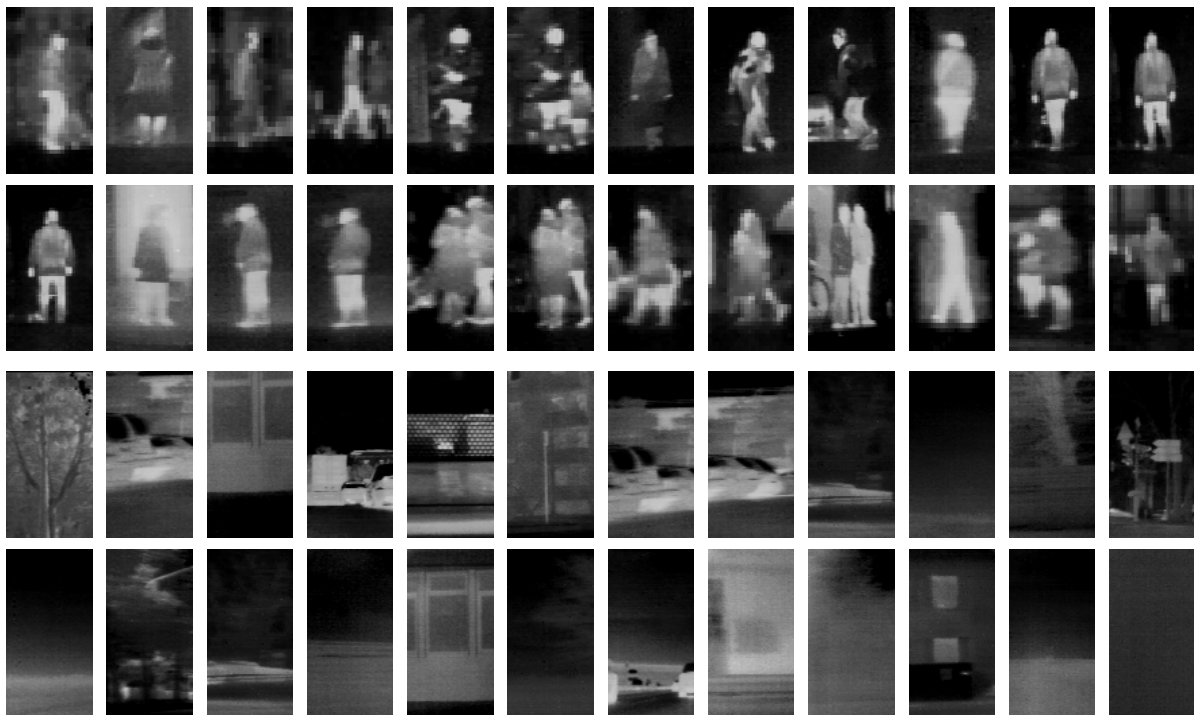


Abbildung A.2: Beispielbilder aus Datenkatalog 4, der Fußgänger- (erste und zweite Zeile) und Hintergrundansichten (dritte und vierte Zeile) im Wärmebild beinhaltet.

B

Bewertungskriterien für Klassifikatoren

B.1 Kontingenztafel

Werden die Ausgaben eines binären Klassifikators als positiv (p) und negativ (n) bezeichnet, können vier mögliche Resultate unterschieden werden. Ist die Ausgabe des Klassifikators p und der tatsächliche Klassenwert ebenfalls p ist die Ausgabe „richtig-positiv“ (r_p). Ist der tatsächliche Klassenwert allerdings n , handelt es sich um ein „falsch-positiv“ (f_p) Resultat oder um einen Fehler 1. Art. Analog dazu wird bei einer n Ausgabe des Klassifikators und einer tatsächlichen Klasse n , die Ausgabe als „richtig-negativ“ (r_n) bezeichnet. Stimmt die Vorhersage des Klassifikators nicht überein (p) dann ist es ein „falsch-negativ“ (f_n) Resultat bzw. ein Fehler 2. Art. Bei einem binären Klassifikationsprozess mit positiven und negativen Testinstanzen können die vier möglichen Ausgaben in einer 2×2 Kontingenztafel zusammengefasst werden (siehe Tabelle B.1). Für k -Klassenprobleme werden $k \times k$ -große Kontingenztafeln herangezogen [Faw04].

	Referenz	
Vorhersage	ω_1 ($r_p + f_n$)	ω_2 ($f_p + r_n$)
ω'_1 ($r_p + f_p$)	richtig positiv (r_p)	falsch positiv (f_p)
ω'_2 ($f_n + r_n$)	falsch negativ (f_n)	richtig negativ (r_n)

Tabelle B.1: 2×2 -Kontingenztafel zur Übersicht über die vier möglichen Ausgaben eines Klassifikators für ein Entscheidungsproblem mit zwei Klassen ω_1 und ω_2 .

Anhand der Kontingenztafel können verschiedene Metriken abgeleitet werden. Diese sind in Tabelle B.2 zusammengefasst. Für diese Arbeit wird insbesondere die Sensitivität (RPR) und

die Falsch-Positiv-Rate (FPR) zur Beurteilung der Klassifikatoren herangezogen. Beispielsweise ist bei einem binären Klassifikator, der Fußgänger von Hintergrundobjekten unterscheiden kann, die Sensitivität die Wahrscheinlichkeit, dass ein vorhandener Fußgänger erkannt wird. Die Falsch-Positiv-Rate ist die Wahrscheinlichkeit, unter der ein Hintergrundobjekt fälschlicherweise als Fußgänger eingeordnet wird. Die Sensitivität wird auch als Erkennungsrate oder Richtig-Positiv-Rate bezeichnet. Die Falsch-Positiv-Rate (bzw. englisch „false-positive-rate“) wird auch als Fehlerrate oder Fehlerkennungsrate bezeichnet. Die Werte liegen in einem Intervall zwischen 0 und 1.

Bezeichnungen	Abkürzung	Berechnungsvorschrift
Sensitivität (Richtig-Positiv-Rate, Erkennungsrate, hit rate, recall, sensitivity)	RPR	$\frac{r_p}{(r_p + f_n)}$
Spezifität (Richtig-Negativ-Rate)	RNR	$\frac{r_n}{(f_p + r_n)}$
Falsch-Positiv-Rate (Fehlerrate, Fehlerkennungsrate)	FPR	$\frac{f_p}{(f_p + r_n)}$
Falsch-Negativ-Rate	FNR	$\frac{f_n}{(r_p + f_n)}$
Relevanz		$\frac{r_p}{(r_p + f_p)}$
Segreganz		$\frac{r_n}{(f_n + r_n)}$
Korrektklassifikationsrate (Exaktheit, accuracy)		$\frac{r_p + r_n}{(r_p + f_p + f_n + r_n)}$
Falschklassifikationsrate		$\frac{f_p + f_n}{(r_p + f_p + f_n + r_n)}$

Tabelle B.2: Von der 2×2 -Kontingenztafel abgeleitete Bewertungskriterien und die in dieser Arbeit verwendeten Abkürzungen.

B.2 ROC- und DET-Kurve

Mittels einzelner Bewertungskriterien kann oftmals keine gesicherte Aussage über die Güte eines Klassifikators getroffen werden. Entscheidend ist das Zusammenspiel zwischen Detektions- und Fehlerkennungsrate, das in ROC- und DET-Kurven berücksichtigt wird. Die ROC-Kurve (Receiver Operating Characteristic) kommt aus dem militärischen Bereich und wurde in den 40er Jahren entwickelt, um die Güte von Empfängern zu beschreiben, die Funkwellen aus dem Hintergrundrauschen herausfiltern sollten. Heute wird sie hauptsächlich in der medizinischen Forschung und Diagnostik [Cri02], sowie zur Beurteilung von Algorithmen aus dem Bereich des maschinellen Lernens verwendet [Fla04].

In einem zweidimensionalen ROC-Diagramm repräsentiert ein Punkt die Erkennungs- und Fehlerkennungsrate eines Klassifikators. Wird nun eine Stellschraube gewählt (z. B. Schwellwert) mittels der die Erkennungsleistung des zu testenden Klassifikators variiert werden kann, führt dies zu einem Kurvenverlauf. Abbildung B.1a zeigt ein solches ROC-Diagramm. Auf der Ordinate wird die Erkennungsrate (RPR) und auf der Abszisse die Fehlerkennungsrate (FPR) aufgetragen. Die in der Abbildung eingezeichnete Diagonale dient der Orientierung und würde die Leistung eines Klassifikators darstellen, der weniger nach einem bestimmten Prinzip klassifiziert, als vielmehr die richtige Klasse zufällig zu erraten versucht, da die Fehlerkennungsrate gleich der Detektionsrate entspricht.

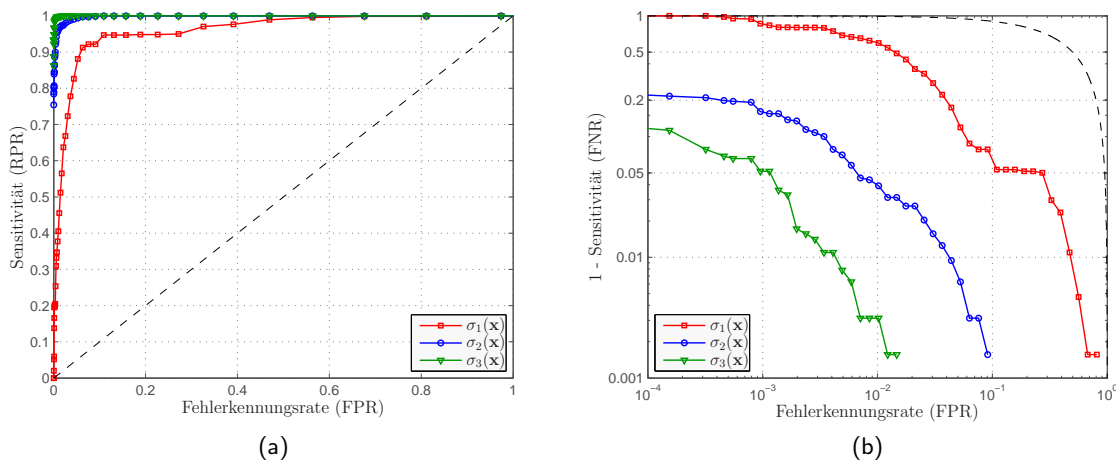


Abbildung B.1: Beispieldiagramme für ROC- (a) und DET-Kurven (b), die die Erkennungsleistung der Klassifikatoren $\sigma_1(\mathbf{x})$, $\sigma_2(\mathbf{x})$ und $\sigma_3(\mathbf{x})$ darstellen. Die gestrichelte Kurve markiert einen Klassifikator mit schlechtmöglicher Leistung.

Die folgenden Punkte (FPR,RPR) eines ROC-Diagramms repräsentieren besondere Klassifikatorgüten:

- Punkt (0,0) stellt einen Klassifikator dar, der Objekte stets zurückweist. Somit kommt es auf der einen Seite zu keinerlei Fehlklassifikationen, auf der anderen Seite ist die Erkennungsleistung ebenfalls gleich Null.
- Punkt (1,1) wäre hingegen ein Klassifikator, der alle Objekte zuverlässig erkennt. Dem steht jedoch eine 100%ige Fehlerrate für alle anderen Klassen gegenüber.
- Punkt (0,1) steht für den optimalen Klassifikator. Er erkennt alle Objekte korrekt und weist keinerlei Fehlerkennungen bezüglich dieser Klasse auf.
- Punkt (1,0): Ein solcher Klassifikator liegt mit seiner Entscheidung stets falsch und stellt den Gegenpol zum perfekten Klassifikator dar. Theoretisch würde die Negation jeder seiner Entscheidungen zu einem perfekten Klassifikator führen. D.h. der Punkt würde an der Diagonalen gespiegelt.

Die ROC-Kurve verläuft stets vom Punkt (0,0) zu Punkt (1,1), da jeder Klassifikator durch geeignete Parametrisierung auf diese Sonderleistungen eingestellt werden kann. Von Interesse ist vornehmlich der Kurvenverlauf zwischen diesen beiden Punkten und der minimale Abstand der Kurve zum Optimum (0,1). Der Bereich unterhalb der gestrichelten Linie bleibt dabei unbelegt, da eine systematisch schlechte Erkennungsleistung eines Klassifikators durch Negation

der Ausgabe invertiert werden kann. Dies entspricht somit einer Spiegelung an der gestrichelten Linie. Der beste Klassifikator in Diagramm B.1a wäre somit Klassifikator $\sigma_3(\mathbf{x})$.

Oftmals müssen viele Klassifikatoren mit sehr guter Erkennungsleistung verglichen werden. Um den Darstellungsbereich der optimalen Klassifikatoren im ROC-Diagramm etwas zu entzerren, stellen DET-Diagramme (Detection Error Tradeoff) [MDK⁺97] ROC-Kurven in einem doppelt logarithmischen Maßstab dar. Die Erzeugung der Kurven wird hierbei analog zu den ROC-Kurven vorgenommen. Die Fehlerraten werden jetzt allerdings auf beiden Achsen des Koordinatensystems aufgetragen. Die Ordinate steht für die Falsch-Negativ-Rate ($\text{FNR} = 1 - \frac{r_p}{(r_p + f_n)} = 1 - \text{RPR}$), die Abszisse, wie bei der ROC-Kurve, für die Falsch-Positiv-Rate (FPR). Abbildung B.1b zeigt die identischen ROC-Kurven aus Abbildung B.1a. Allerdings werden die Kurven $\sigma_2(\mathbf{x})$ und $\sigma_3(\mathbf{x})$ durch die logarithmische Darstellung besser entzerrt und können dadurch leichter verglichen werden.

Das Erzeugen einer ROC bzw. DET-Kurve kann ein aufwändiger Prozess sein, da für jede Einstellung der Stellschraube jeweils die RPR und die FPR ermittelt werden müssen. Dafür müssen wiederum die Maßzahlen r_p , r_n , f_p und f_n für alle Evaluierungsdaten neu bestimmt werden. Um das schrittweise Verschieben der Grenze zu umgehen, wird in dieser Arbeit der Algorithmus aus [Faw04] angewendet. Hierbei handelt es sich um ein effizientes Verfahren, mit dem es möglich ist, ROC-Punkte eines Klassifikators ohne manuelles Verschieben der Grenze zu erzeugen. Jede Eingabe, die bei einer bestimmten Grenze als positiv klassifiziert wird, wird bei einer kleineren Schwelle analog klassifiziert. Die ausgegebenen Werte des Klassifikators für den Testdatensatz werden somit absteigend sortiert, um beim schrittweisen Durchgehen dieser Liste die richtig-positiven und falsch-positiven Erkennungen kontinuierlich zu aktualisieren. Durch diese Vorgehensweise kann eine ROC-Kurve mit linearem Berechnungsaufwand erzeugt werden.

B.3 Grundlagen der Support-Vektor-Maschine

Eine Support-Vektor-Maschine (SVM) [Bur98, Vap00] ist ein Klassifikationsverfahren, das für die Entscheidungsfindung verwendet werden kann ob ein Objekt, das durch einen n -dimensionalen Merkmalsvektor $\mathbf{x}_i \in \mathbb{R}^n$ beschrieben ist, einer Objektklasse angehört oder nicht. Ausgangsbasis für das Training ist eine Menge von Vektoren, von denen bekannt ist, zu welcher Objektklasse sie gehören. Durch das Lösen eines Minimierungsproblems kann zwischen den Vektoren der beiden Trainingsklassen eine n -dimensionale Hyperebene bestimmt werden, die beide Klassen bestmöglich voneinander trennt. Die Ebene wird so in den Raum eingepasst, dass der Abstand zu den Vektoren, die ihr am nächsten liegen, maximal ist. Diese Vektoren werden auch als Support-Vektoren bezeichnet, da sie die Ebene eindeutig beschreiben. Durch diese *max-margin*-Trennung wird eine größere Generalisierung des Klassifikators erzielt. Im Erkennungsfall wird nun für einen beliebigen Merkmalsvektor – insbesondere Vektoren die nicht Teil der Trainingsmenge sind – der Abstand zur Hyperebene bestimmt. Ist dieser positiv, liegt der Vektor auf der Seite der gesuchten Objektklasse.

C

Multispektrale Kamerakalibrierung

Das Ziel einer Kamerakalibrierung ist die Berechnung der Abbildungseigenschaften von Kameras, sowie die Bestimmung der relativen Position und Orientierung der einzelnen Kameras zueinander. Im Folgenden wird zunächst das, dem Kalibrierungsprozess zugrundeliegende, mathematische Modell einer Kamera eingeführt. Anschließend werden zwei Verfahren diskutiert, die den besonderen Anforderungen von multispektralen Sensorsystemen entsprechen.

C.1 Projektives Kameramodell

In diesem Abschnitt wird das häufig verwendete Modell einer Lochkamera [HZ03] vorgestellt. Dieses Modell ist zum einen die Grundlage für die intrinsische Kalibrierung der verwendeten Kameras, und stellt zum anderen das mathematische Handwerkszeug zur Verfügung, um Messwerte anderer Sensoren mit den Bilddaten der Kamera zu fusionieren. Zunächst wird von einem einfachen Modell ausgegangen, das dann schrittweise verfeinert wird, um den Besonderheiten aktueller Kameras gerecht zu werden.

Für eine projektive Lochkamera ist die Abbildung von einem 3D Punkt $\mathbf{x} = (x, y, z)^T$ aus dem Kamerakoordinatensystem, auf einen 2D Punkt $\tilde{\mathbf{x}} = (u, v)^T$ in der Bildebene, wie folgt durch die Brennweite f bestimmt:

$$(x, y, z)^T \mapsto (fx/z, fy/z)^T \tag{C.1.1}$$

Das einfache Modell einer Lochkamera muss für reale digitale Kameras um zwei Erweiterungen ergänzt werden: Zum einen sind bei digitalen Kameras im Allgemeinen die Höhe und Breite eines Pixels auf der Bildebene nicht identisch. Deshalb wird die Brennweite f durch eine für u und v spezifische Brennweite f_u bzw. f_v ersetzt. Zum anderen wird ein Offset p_u bzw. p_v eingeführt, der den Abstand angibt, um den das Koordinatensystem eines Bildsensors vom Ursprung der Bildebene der einfachen Lochkamera abweicht. Rein theoretisch entspricht dieser Offset bei den meisten Kameras der Hälfte von Bildbreite und Bildhöhe in Pixeln. Durch Fertigungstoleranzen von Sensoren und Objektiven, ist dies in der Praxis jedoch selten der Fall. Zusammen ergibt sich folgende erweiterte Abbildungsvorschrift:

$$(x, y, z)^T \mapsto \left(f_u \frac{x}{z} + p_u, f_v \frac{y}{z} + p_v \right)^T \quad (\text{C.1.2})$$

In homogenen Koordinaten lässt sich die Abbildung eines 3D Punktes $\mathbf{x} = (x, y, z, 1)^T$ aus dem Kamerakoordinatensystem auf einen 2D Punkt $\tilde{\mathbf{x}} = (u, v, 1)^T$ in der Bildebene wie folgt beschreiben, wobei $\mathbf{K} \in \mathbb{R}^{3 \times 4}$ die Kamerakalibrierungsmatrix bezeichnet:

$$\tilde{\mathbf{x}} = \mathbf{K} (1/z \mathbf{x}) \quad (\text{C.1.3})$$

$$\mathbf{K} = \begin{bmatrix} f_u & 0 & p_u & 0 \\ 0 & f_v & p_v & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (\text{C.1.4})$$

Die Bestimmung der Matrix \mathbf{K} wird auch als *intrinsische* Kalibrierung bezeichnet. Im bisherigen Kameramodell wurden Verzerrungen, die auf das Objektiv der Kamera zurückzuführen sind, nicht betrachtet. Für die meisten Kameras sind hierbei vor allem radiale Verzerrungen erster und zweiter Ordnung dominant [Zha99]. Komplexere Modelle wie die tangential- oder Prisma-Verzerrung [WCH92] werden an dieser Stelle nicht weiter betrachtet, da sie im Rahmen der in dieser Arbeit geforderten Genauigkeit vernachlässigt werden können. Im Folgenden repräsentieren $\mathbf{x} = (x, y, 1, 1)^T$ bzw. $\mathbf{x}_d = (x_d, y_d, 1, 1)^T$ die verzerrungsfreien bzw. verzerrten normalisierten¹ Bildkoordinate. Zusammen mit den Koeffizienten k_1 und k_2 ist die radiale Verzerrung dementsprechend wie folgt definiert:

$$\mathbf{x}_d = \left(1 + k_1 (x^2 + y^2) + k_2 (x^2 + y^2)^2 \right) \mathbf{x} \quad (\text{C.1.5})$$

Der 2D Punkt $\tilde{\mathbf{x}}$ auf der Bildebene ergibt sich analog zu C.1.3:

$$\tilde{\mathbf{x}} = \mathbf{K} \mathbf{x}_d \quad (\text{C.1.6})$$

Das bisher beschriebene Modell einer Kamera ist noch nicht vollständig, da in der Abbildung die Lage und Orientierung der Kamera in Bezug auf ein Weltkoordinatensystem, bzw. in Bezug auf andere Sensoren, nicht berücksichtigt wird. Diese fehlenden Größen werden als *extrinsische* Parameter bezeichnet und setzen sich aus einer Rotation $\mathbf{R} \in \mathbb{R}^{3 \times 3}$, sowie einer Translation um den Vektor $\mathbf{t} \in \mathbb{R}^3$ zusammen. In der Darstellung als Szenegraph [PHBK06] entspricht dies genau der Kante von der Kamera zur Wurzel des Graphen. Die Transformation von Welt- nach Kamerakoordinaten lässt sich in homogenen Koordinaten mit einer Matrix darstellen. Zusammen mit der Kamerakalibrierungsmatrix $\mathbf{K} \in \mathbb{R}^{3 \times 4}$ ergibt sich schließlich für einen 3D Punkt \mathbf{x} aus dem Weltkoordinatensystem, folgende Abbildung $\tilde{\mathbf{x}}$ auf der Bildebene.

$$\tilde{\mathbf{x}} = \overbrace{\begin{bmatrix} f_u & p_u & 0 \\ & f_v & p_v & 0 \\ & & 1 & 0 \end{bmatrix}}^{\text{intrinsische Parameter}} \overbrace{\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}}^{\text{extrinsische Parameter}} \mathbf{x} = \mathbf{P} \mathbf{x} \quad (\text{C.1.7})$$

$$(\text{C.1.8})$$

¹ $(x, y, z, 1)^T \mapsto (x/z, y/z, 1, 1)^T$

Im den folgenden Abschnitten des Kapitels wird die vollständige Projektionsmatrix $\mathbf{P} \in \mathbb{R}^{3 \times 4}$ verwendet, die sich aus einer Komposition von intrinsischen und extrinsischen Parametern zusammensetzt.

C.2 Kalibrierungsverfahren

Nachdem im vorherigen Abschnitt das Kameramodell vorgestellt wurde, das dem Kalibrierungsprozess zugrunde liegt, werden im Folgenden die Verfahren beschrieben, um die 12 Freiheitsgrade dieses Modells bzw. der Projektionsmatrix \mathbf{P} zu bestimmen. Darüber hinaus werden zwei Kalibrierkörper² eingeführt, die den Bedürfnissen eines multispektralen Kamerasystems³ gerecht werden.

Einige intrinsischen Werte könnten durchaus aus den Herstellerangaben des verwendeten Kameratyps entnommen werden. In der Praxis jedoch sind diese Werte starken Schwankungen unterworfen, so dass, bedingt durch Fertigungstoleranzen, zwei baugleiche Kameras unterschiedliche Werte aufweisen können. Ähnliches gilt für die extrinsischen Parameter. Hier kann zwar in erster Näherung der Verbauort der Kameras aus CAD Modellen übernommen oder mit einfachen Hilfsmitteln vermessen werden. Für eine präzise Bestimmung der extrinsischen Parameter, insbesondere der Orientierung ist dieses Verfahren allerdings nicht geeignet. Hierfür ist ein gesondertes Kalibrierungsverfahren notwendig.

Schachbrett aus Peltierelementen

In dieser Arbeit wurde die „Camera Calibration Toolbox“ [Bou07] verwendet, die eine komfortable Implementierung der Verfahren von Heikkila [HS97] und Zang [Zha99] zur Kalibrierung eines Multikamerasystems bietet. Der Kalibrierkörper ist ein planares Schachbrett, welches in $m \times n$ Felder mit gut sichtbarem Kontrast aufgeteilt ist. Bei zwei Kameras aus unterschiedlichen Spektralbereichen kommt erschwerend hinzu, dass der Kalibrierkörper in beiden Spektren sichtbar sein sollte. Für die in dieser Arbeit verwendete Sensorik bedeutet dies, dass er im Videobild einen Kontrastunterschied und im Ferninfrarotbild einen Temperaturunterschied aufweisen muss.

Der Aufbau eines solchen Kalibrierkörpers aus Peltierelementen ist in [Nev05] ausführlich beschrieben und wird als gegeben vorausgesetzt. An dieser Stelle werden nur die wichtigsten Eigenschaften beschrieben: Das Schachbrett besteht aus 64 Aluminium Quadern, die abwechselnd schwarz und weiß lackiert sind (siehe Abb. C.1a). Im Videobild ergibt sich somit ein guter Kontrast. Um auch im Ferninfrarotbild einen guten Kontrast zu erreichen, wurden an den Rückseiten der schwarzen Quader heizende Peltierelemente angebracht und analog dazu an die Rückseite der weißen Flächen kühlende Elemente befestigt. Abbildung C.1 zeigt eine Szene einer Kalibrierung mit einem Schachbrett aus Peltierelementen. Im Ferninfrarot- wie auch im Videobild sind die wichtigen Kanten zwischen den einzelnen Feldern deutlich sichtbar.

² Ein Kalibrierkörper ist ein Gegenstand, der durch vordefinierte Eigenschaften wie Form oder Textur, charakteristische Messwerte im Sensor erzeugt. Durch die Beziehung aus Messwert und dem Zusatzwissen der vordefinierten Eigenschaften des Körpers können Rückschlüsse auf interne Parameter des jeweiligen Sensors gezogen werden. Dieser Vorgang wird als Kalibrierung bezeichnet.

³ Ein multispektrales Kamerasystem besteht aus mindestens zwei Kameras, die in unterschiedlichen Spektren operieren, wie zum Beispiel sichtbares Licht und Wärmestrahlung.

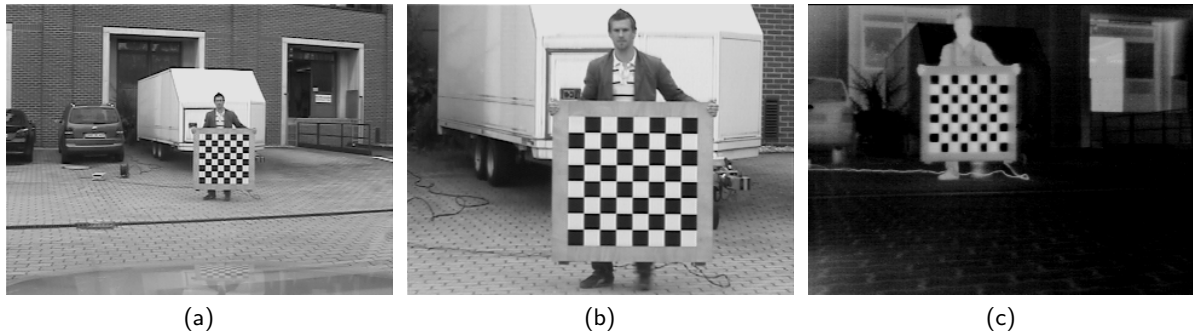


Abbildung C.1: Kalibrierung einer multispektralen Kamera mit einem Schachbrett aus Peltierelementen. Die Kanten des Schachbretts sind im Videobild (a) bzw. (b) gut zu erkennen. Durch Peltierelemente, die abwechselnd die Kacheln des Schachbrettes beheizen bzw. kühlen, ist auch im Bild der Ferninfrarotkamera (c) ein guter Kontrast gewährleistet.

Mit der „Calibration Toolbox“ werden zuerst für jede Kamera einzeln die intrinsischen Parameter bestimmt. In einem zweiten Schritt werden die extrinsischen Parameter ermittelt. Die verwendete Toolbox ist allerdings auf zwei Kameras beschränkt. Um dennoch mehrere Kameras extrinsisch kalibrieren zu können, werden für alle Kameras Paare mit einem festen Partner, der später als Referenzkoordinatensystem C_1 dient, gebildet. Die extrinsische Kalibrierung berechnet anschließend für jedes Paar die Rotationsmatrix \mathbf{R} und den Translationsvektor \mathbf{t} in Bezug auf das Referenzkoordinatensystem C_1 . Alle Kameras sind nun in Bezug auf C_1 registriert. Die Schätzung der intrinsischen Parameter wird im Allgemeinen durch die zusätzlichen Informationen aus der extrinsischen Kalibrierung verbessert.

Punktlichtquelle

Verfahren zur Punktkalibrierung benötigen keine geometrischen Kalibrierkörper (z. B. Schachbrettkörper wie in Abschnitt C.2), sondern lediglich eine Vorrichtung, wodurch einzelne korrespondierende Punkte in den beteiligten Kameras erzeugt werden können. Diese Verfahren kommen vor Allem dann zum Einsatz, wenn ein geometrischer Kalibrierkörper nicht zeitgleich sichtbar in allen Kamerabildern postiert werden kann, oder viele Kameras zueinander kalibriert werden sollen. In einer Publikation von Svoboda [SMP05] wird ein automatisches Punktkalibrierverfahren vorgestellt, womit die intrinsischen und extrinsischen Parameter von einer unbegrenzten Menge (mindestens drei) visueller Kameras ermittelt werden können. Über einen modifizierten Laserpointer werden durch den Benutzer korrespondierende Punkte in allen Kamerabildbereichen erzeugt, und durch Bildverarbeitungsalgorithmen automatisch extrahiert. Alle Punkte werden anhand von Epipolarbedingungen validiert und einander zugeordnet. Dabei ist es nicht notwendig, dass die Punkte stets in allen Kameras sichtbar sind. Vielmehr sollte das Sichtvolumen der Kameras ausgefüllt und eine ausreichende Überlappung vorhanden sein.

In den folgenden Abschnitten wird versucht, die Vorgehensweise von Svoboda auf multispektrale bildgebende Sensoren zu übertragen. Dafür musste eine neue Kalibriervorrichtung erstellt werden, um Korrespondenzen ebenso in Wärmebildkameras zu generieren. Der Kali-

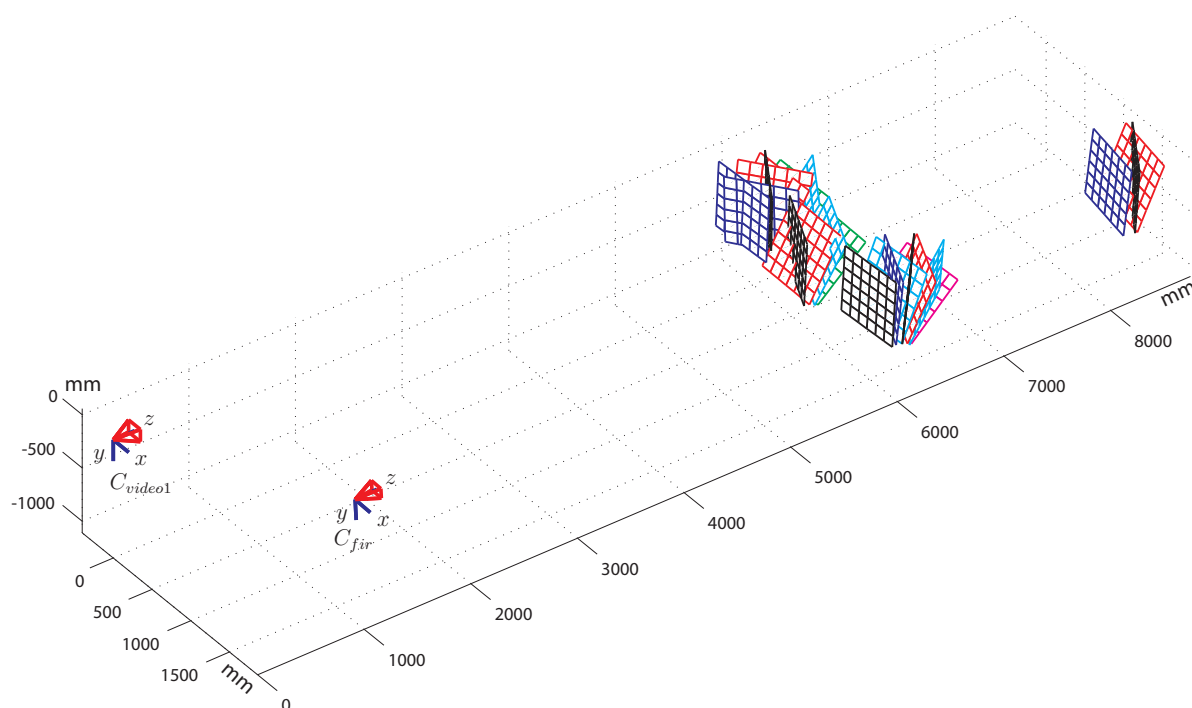


Abbildung C.2: Extrinsische Kalibrierung der Versuchskameras C_{video1} (Videokamera) und C_{fir} (Wärmebildkamera) mittels einem Schachbrett aus Peltierelementen. Die Abbildung gibt zum einen die räumliche Anordnung (extrinsische Kalibrierung) der beiden Kameras zueinander wieder und zum anderen die unterschiedlichen Positionen des Schachbrettes während der Durchführung der Kalibrierung.

brierungsvorgang selbst wurde nach der Nutzeranleitung von Svoboda [Svo03] und mit Hilfe des zugehörigen, leicht modifizierten und öffentlich zugänglichen Quellcodes durchgeführt.

Kalibriervorrichtung Bei der Kalibrierung von multispektralen Kameras muss darauf geachtet werden, dass korrespondierende Punkte in allen Kamerabildern zur Verfügung stehen. Sind im bildgebenden Sensorsystem Wärmebildkameras integriert, ist ein Laserpointer ungeeignet zur Korrespondenzerzeugung, da von diesem nicht ausreichend Wärmeenergie abgestrahlt wird. Deshalb wurde eine neuartige prototypische Kalibriervorrichtung erstellt, die in Abbildung C.3a zu sehen ist. Durch einen von innen beleuchteten Kunststoffdiffusor wird sowohl Licht als auch ausreichend Wärmestrahlung emittiert. Der Diffusor wurde kugelförmig gewählt, um in allen Kamerabildern eine elliptische Abbildung zu gewährleisten. Dadurch kann mittels einer subpixelgenauen Ellipsenextraktion im Bildbereich, das Zentrum des Diffusors als korrespondierender Punkt verwendet werden. In den Abbildungen C.3b und C.3c ist zu sehen, wie sich die Kalibriervorrichtung in den multispektralen Kamerabildern darstellt.

Durchführung und Ergebnis Mit der vorgestellten Kalibriervorrichtung können korrespondierende Punkte in multispektralen Kamerabildern generiert werden. Dieser Arbeitsschritt muss manuell durchgeführt werden, erfordert aber lediglich einen geringen Zeitaufwand. In einer

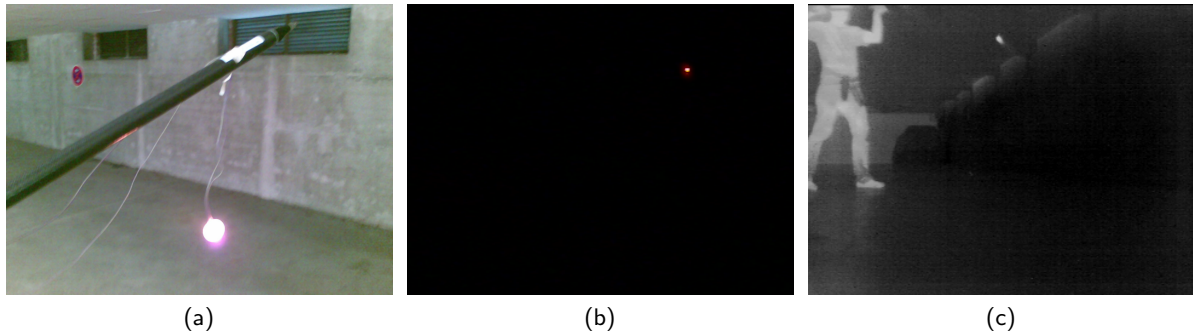


Abbildung C.3: (a) Prototypische Kalibriervorrichtung für die Punktkalibrierung von multispektralen Kamerasystemen. An einem Stab ist eine aktive Leuchtquelle befestigt, die von einem kugelförmigen Kunststoffdiffusor umgeben ist. Es wird sowohl Licht- als auch Wärmestrahlung emittiert. (b) Abbildung der Kalibriervorrichtung in der Videokamera und (c) Wärmebildkamera. Im Wärmebild ist deutlich der Benutzer zu erkennen.

dunklen Umgebung oder bei entsprechenden Kameraeinstellungen (verminderte Verschlusszeiten) können diese Punkte für visuelle Kameras automatisch segmentiert werden. Im Wärmebild ist die Segmentierung aufwändiger, da der Benutzer, der die Kalibrierung vornimmt, zusätzliche Intensitätswerte erzeugt (siehe Abbildung C.3c).

In Abbildung C.4 ist das Ergebnis einer durchgeführten Punktkalibrierung zu sehen. Die erzielten Genauigkeiten sind allerdings unzureichend. Dies liegt jedoch nicht an der Kalibriervorrichtung oder an den Segmentierungsschritten, sondern vielmehr an der ungünstigen Verbaukonstellation der Kameras. So verlaufen die optischen Achsen der drei Versuchsträgerkameras nahezu parallel zueinander (siehe auch Abbildung 2.5). Diese Konstellation kann zu numerischen Instabilitäten führen und sollte für eine Punktkalibrierung vermieden werden [KTÅ00]. Für andere multispektrale Kameraanordnungen (z. B. im Fahrzeuginnenraum) ist die Kalibriervorrichtung möglicherweise gut geeignet. Für die vorliegende Verbaukonstellation ist das Verfahren jedoch nicht zweckmäßig.

C.3 Bewertung

In einem multispektralen Kamerasystem werden besonderen Anforderungen an die Kalibrierungsverfahren gestellt. So müssen zum Beispiel spezielle auf die jeweiligen Spektralbereiche abgestimmte Kalibrierkörper verwendet werden, um die intrinsischen Parameter der einzelnen Kameras bestimmen zu können. Für eine extrinsische Kalibrierung ist es zusätzlich erforderlich, dass der Kalibrierkörper in beiden Spektralbereichen gleichzeitig verwendet werden kann. Für ein System bestehend aus Wärmebild- und Videokamera wurden zwei Verfahren vorgestellt, die diesen Ansprüchen genügen. Ein Schachbrett aus Peltierelementen überträgt das Grundprinzip der weitverbreiteten Kalibrierung mit einem Schachbrett in den ferninfraroten Spektralbereich. Bestehende Verfahren zur intrinsischen und extrinsischen Kalibrierung können somit direkt angewendet werden. Die erzielten Genauigkeiten sind für die Ansprüche einer Fahrumfelderfassung ausreichend. Eine präzise Fertigung eines derartigen Kalibrierkörpers ist jedoch sowohl zeit- als auch kostenintensiv. Darüber hinaus müssen die Daten manuell seg-

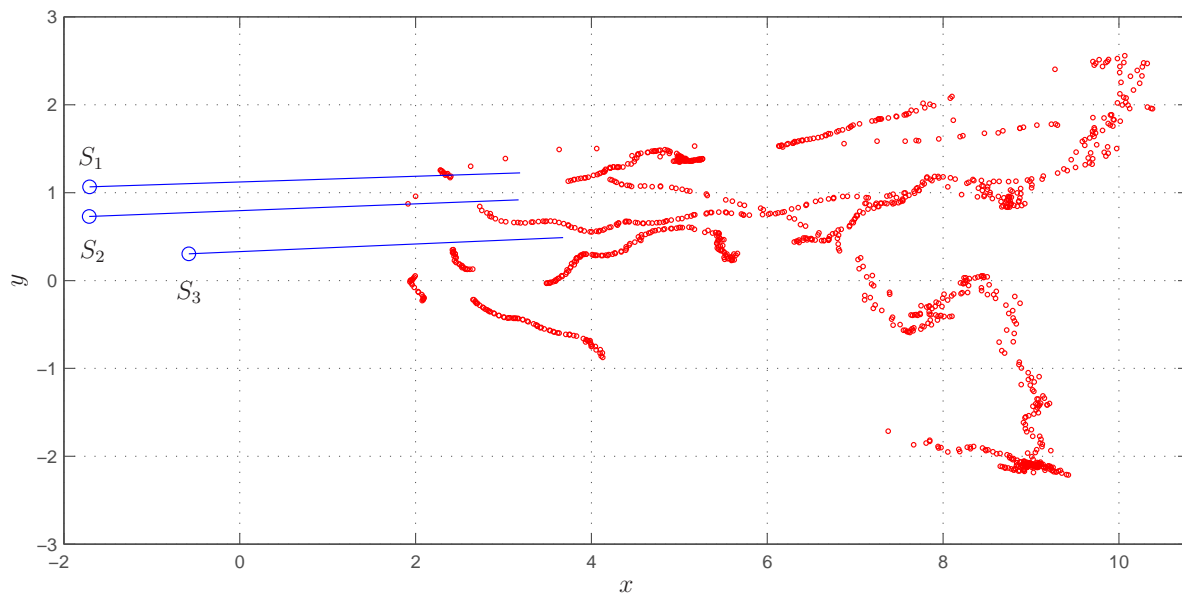


Abbildung C.4: Beispielkalibrierung der drei Versuchskameras S_1 (Videokamera), S_2 (Videokamera mit Teleobjektiv) und S_3 (Wärmebildkamera) mittels der Punktkalibrierung. Die Datenpunkte spiegeln die rückprojizierten Positionen des Kalibriergeräts wider.

mentiert werden und in einem mehrstufigen Prozess extrinsisch kalibriert werden. Das zweite Verfahren mit einer punktförmigen Lichtquelle bietet in diesen Punkten entscheidende Vorteile. In einem vollautomatischen Prozess können beliebig viele Kameras kalibriert werden. Einzige Anforderung an den Kalibrierkörper ist, dass sowohl Licht- als auch Wärmestrahlung emittiert wird, die ein punktförmiges Bild erzeugen. Dies wurde durch einen kugelförmigen Kunststoffdiffusor und einer handelsüblichen Lichtquelle erreicht. Durch diesen einfachen Aufbau ist eine mobile und einfache multispektrale Kamerakalibrierung möglich.

Allerdings können im Gegensatz zu dem Schachbrettkalibrierkörper keine zusätzlichen geometrischen Eigenschaften (Zellenbreite und planarer Körper) herangezogen werden. Bei ungünstigen Verbaukonstellation der Kameras, kann dies zu Mehrdeutigkeiten und somit zu numerischen Instabilitäten führen [KTÅ00]. Nahezu parallele Kamerasichtachsen, wie sie im Versuchsträgerfahrzeug dieser Arbeit zum Einsatz kommen, ist eine dieser ungünstigen Kameraanordnungen. Somit war mit dem Kunststoffdiffusor eine erheblich weniger genaue Kalibrierung als mit dem Schachbrettkalibrierkörper möglich.

D

Exemplarische Systemleistung

Im Folgenden sind einige Schnappschüsse von Szenen aufgeführt, die mit der multisensoriellen Fahrumfelderfassung dieser Arbeit analysiert wurden. Die Einzelbilder sind zeitlich angeordnet (zunächst von links nach rechts) und zeigen die multisensoriell erkannten Objekte im Videobild. Beschreibungen zu den dargestellten Szenen finden sich in Tabelle 6.3 auf Seite 125 (Stadtscene S_1 , Landstraßenszene S_3 , Autobahnszene S_4 und Nachtszene S_5).

Tabelle D.1: Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Stadtscene (Szene S_1). Das Zeitintervall zwischen den Schnappschüssen beträgt ca. eine halbe Sekunde.



Tabelle D.1: Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfeld-
erfassung auf einer Stadtszene (Szene S_1). Das Zeitintervall zwischen den Schnapp-
schüssen beträgt ca. eine halbe Sekunde.



Tabelle D.2: Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfeld-
erfassung auf einer Autobahnszene (Szene S_4). Das Zeitintervall zwischen den Schnapp-
schüssen beträgt ca. eine Sekunde.



Tabelle D.2: Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfeld-
erfassung auf einer Autobahnszene (Szene S₄). Das Zeitintervall zwischen den
Schnappschüssen beträgt ca. eine Sekunde.



Tabelle D.3: Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfeld-
erfassung auf einer Landstraßenszene (Szene S₃). Das Zeitintervall zwischen den
Schnappschüssen beträgt ca. eine Sekunde.

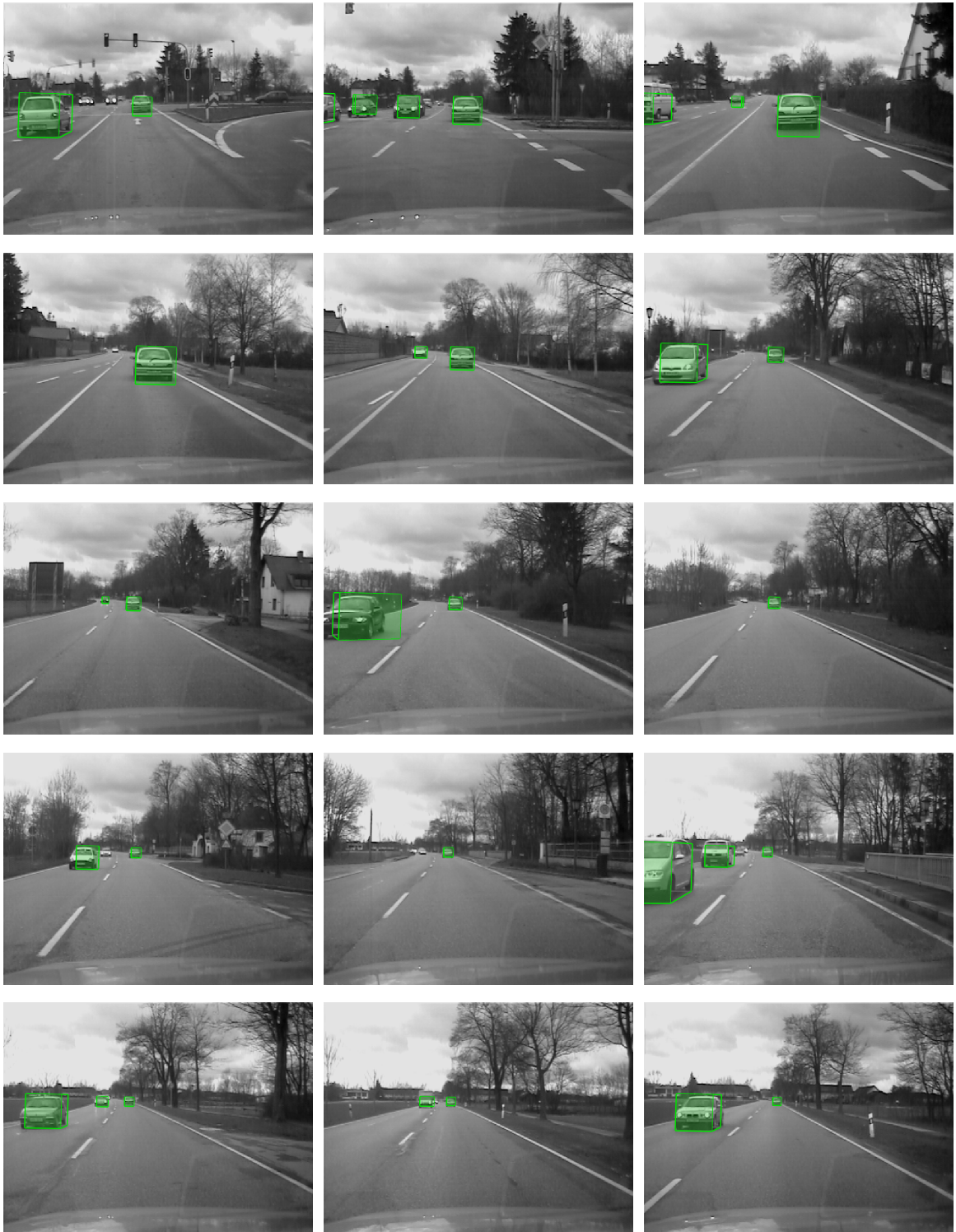


Tabelle D.4: Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Nachtszene (Szene S₅). Das Zeitintervall zwischen den Schnappschüssen beträgt ca. sieben Sekunden.



Tabellenverzeichnis

2.1	Vergleich automotiver Sensorsysteme zur Fahrumfelderfassung.	19
2.2	Details zur Sensorik des Versuchsfahrzeugs.	25
2.2	Details zur Sensorik des Versuchsfahrzeugs.	26
3.1	Bewertung der Hypothesenerzeugung für Fußgänger.	52
3.2	Übersicht über die entfernungsbasierte Hypothesenerzeugung im Bild.	53
3.2	Übersicht über die entfernungsbasierte Hypothesenerzeugung im Bild.	54
4.1	Verwendete Merkmale bezüglich der Objektklassen und Sensoren.	64
4.2	Entfernungsabhängige Erkennungsraten mit Gradientenhistogrammen für Fahrzeugheckansichten und Fußgänger.	72
4.2	Entfernungsabhängige Erkennungsraten mit Gradientenhistogrammen für Fahrzeugheckansichten und Fußgänger.	73
5.1	Übersicht über Merkmale, Merkmalsbezeichnungen und Aufbau.	90
5.2	Ensembleerkennungsraten mittels funktionaler Kombinatoren für Fahrzeuge (Softmax-Kalibrierung).	97
5.3	Ensembleerkennungsraten mit funktionalen Kombinatoren für Fahrzeuge (pp-Kalibrierung).	98
5.4	Ensembleerkennungsraten für Fußgänger mit funktionalen Kombinatoren.	98
5.4	Ensembleerkennungsraten für Fußgänger mit funktionalen Kombinatoren.	99
5.5	Vergleich der Klassifikatorfusionsverfahren bei verrauschten Merkmalen.	103
6.1	Vergleich der Genauigkeiten zwischen semiautomatisch erzeugten Referenzdaten für Fahrzeuge und Fußgänger.	115
6.2	Genauigkeiten der verwendeten Inertialplattform.	119
6.3	Überblick über die Evaluierungsszenarien.	123
6.3	Überblick über die Evaluierungsszenarien.	124
6.3	Überblick über die Evaluierungsszenarien.	125
6.4	Auswertung von Fahr Szenen.	126
6.4	Auswertung von Fahr Szenen.	127
6.5	Charakteristische Fehler des multisensoriellen Fahrumfelderfassungssystems.	131
6.5	Charakteristische Fehler des multisensoriellen Fahrumfelderfassungssystems.	132

6.6	Durchschnittliche Laufzeiten der einzelnen Systemkomponenten bei paralleler Abarbeitung.	134
A.1	Übersicht über die Datenkataloge.	179
A.1	Übersicht über die Datenkataloge.	180
B.1	2×2 -Kontingenztafel zur Übersicht über die vier möglichen Ausgaben eines Klassifikators.	183
B.2	Bewertungskriterien und Abkürzungen.	184
D.1	Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Stadtszene.	195
D.1	Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Stadtszene.	196
D.2	Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Autobahnscene.	196
D.2	Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Autobahnscene.	197
D.3	Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Landstraßenszene.	198
D.4	Schnappschüsse von den Erkennungsergebnissen der multisensoriellen Fahrumfelderfassung auf einer Nachtszene.	199

Abbildungsverzeichnis

1.1	Illustration eines automotiven Sensorsystems und einer aktiven Gefahrenbremsung.	2
1.2	Aufgabenteilung.	4
1.3	Ergebnisvorschau des entwickelten Umfelderkennungssystems.	5
2.1	Einteilung von ausgewählten Fahrerassistenzsystemen nach Komfort und Sicherheit bzw. nach der Art ihrer Unterstützung.	13
2.2	Einteilung von ausgewählten Fahrerassistenzsystemen nach dem Zeitpunkt ihres Eingriffs.	14
2.3	Kategorisierung automotiver Umfelderkennungssensorik hinsichtlich Bild- und Entfernungswiedergabe.	18
2.4	Omnibus Prozessmodell für Fusionssysteme.	22
2.5	Sensorverbaupositionen und -sichtbereiche am Versuchsfahrzeug.	27
2.6	Momentaufnahme der Sensordaten des Versuchsfahrzeugs.	28
2.7	Fahrzeug-, Sensor- und Objektkoordinatensysteme.	29
2.8	Spatial Relationship Graph (SRG) einer automotiven Fahrumfelderfassung. . .	29
2.9	Systemüberblick.	30
3.1	Beispielszene mit Laserscannerrohdaten.	36
3.2	Verlängerung, Konkatenation und Aggregation für Laserscanner-Segmente. . .	38
3.3	Exemplarischer Vergleich zwischen dem Region-Growing Algorithmus und der erweiterten inkrementellen Liniextraktion für Laserscannerrohdaten.	40
3.4	Verschiedene Laserscanner-Segmenttypen.	41
3.5	Verteilung der Laserscanner-Segmentlängen von Fahrzeugheckansichten, Fußgängern, und Hintergrundobjekten.	43
3.6	Laserscanner-Segmente zur Generierung von Hypothesen.	44
3.7	Hypothesenerzeugung anhand von Laserscanner-Segmenten.	45
3.8	Segmentierungsprozess von „U“-Formen zur Fußgängerhypothetisierung.	48
3.9	Beispielaufnahmen für die Fußgängersegmentierung anhand von Wärmebildkanal-ten.	49
3.10	Verbunderzeugung von Fußgängerhypothesen mittels Laserscanner und Wärmebildkamera.	51

3.11	Korrektur der Bildprojektion von Fahrzeughypothesen durch kantenbasierte Bildhypothesen.	55
3.12	Bewertung der verschiedenen Aufmerksamkeitssteuerungsverfahren zur Fahrzeugdetektion.	56
3.13	Klassifikation mit und ohne Aufmerksamkeitssteuerung.	57
4.1	Datensegmentierung zur Merkmalsgewinnung.	65
4.2	Unterteilung der Kantenrichtungen in Intervalle.	68
4.3	Erkennungsraten von PKW unter verschiedenen Parametrisierungen.	69
4.4	Erkennungsraten unter verschiedenen Skalierungsstufen.	72
4.5	Bildmerkmalsergänzung durch Achsenspiegelung.	74
4.6	Beispielansichten von gespiegelten Fahrzeughecks.	76
4.7	DET-Kurven zum Vergleich der Erkennungsraten von gespiegelten mit ungespiegelten Fahrzeughecks.	77
4.8	Mittels eines Laserscanners detektierte Retroreflektoren auf der Heckansicht eines Fahrzeugs.	79
4.9	Durch einen Laserscanner ermittelte Durchschnittsreflexionsstärken von Fahrzeughecks.	80
4.10	Pendelbewegung eines Fußgängers.	81
4.11	Frequenzanalyse der Pendelbewegung eines Fußgängers.	82
4.12	Aus dem Tracking gewonnene Merkmale Geschwindigkeit und Altersindex von Fahrzeugen und Nichtfahrzeugen.	83
4.13	Merkmalsabhängige Erkennungsraten für Fahrzeuge und Fußgänger.	84
5.1	Datenfluss bei einer monolithischen Klassifikation.	91
5.2	Erkennungsraten von PKW und Fußgängern durch monolithische Klassifikatoren.	92
5.3	Datenfluss bei einer Ensembleklassifikation.	93
5.4	Erkennungsraten von Fahrzeugen und Fußgängern durch Metaklassifikatoren.	100
5.5	Erkennungsratenvergleich der verschiedenen Klassifikatorfusionsverfahren für Fahrzeuge.	102
6.1	Übersichtsschaubild zur Systembewertung.	109
6.2	Quadermodell und -interpolation.	112
6.3	Arbeitsschritte zur Markierung eines Objektes in drei Dimensionen.	112
6.4	Semiautomatisch attributierte Szenen.	114
6.5	Approximierte Trajektorie eines vorausfahrenden Fahrzeugs auf einer kurvenreichen Strecke.	116
6.6	Referenzdatengewinnung mit einem dedizierten Sensor.	118
6.7	Kombination aus dedizierter Sensorik und manueller Nachbearbeitung.	119
6.8	Temporales Diagramm zum Erkennungsratenvergleich von unterschiedlichen Sensorsystemen.	130
6.9	Vergleich der Systemlaufzeiten für ein Stadt- und ein Autobahnscenario.	133
7.1	Datenausprägungen und Informationsfluss innerhalb eines Umfeld erfassungssystems.	141
7.2	Visualisierung der Sensorverbauorte und -erfassungsbereiche.	144
7.3	Streudiagramm von Entfernungsdaten.	145
7.4	Multidimensionale Laserscannerdaten.	146

7.5	Multidimensionale Radardaten.	146
7.6	Visualisierung von Bilddaten in einer dreidimensionalen Umgebung.	148
7.7	Datenrelationen und Filtervorgänge.	149
7.8	Visualisierung von interpretierten Daten.	150
7.9	Visualisierung von zeitlichen Messwertverläufen.	151
7.10	Orbit-Navigation innerhalb einer virtuellen Umgebung.	152
7.11	Interaktionsmöglichkeiten.	154
7.12	Visualisierung der Fahrumfelderfassung im Fahrzeug.	156
7.13	Schematischer Aufbau eines optischen Trackingsystems.	157
7.14	Ausgabegeräte für die erweiterte Realität.	158
7.15	Visualisierung im Labor.	160
7.16	Kontaktanaloge Visualisierung im Fahrzeug.	160
7.17	„Spatial Relationship Graph“ des Visualisierungssystems für die Labor- und Fahrzeugumgebung.	161
7.18	Installation des Trackingsystems.	164
7.19	Visualisierung von Sensor- und Objektdaten auf einem tragbaren TFT.	165
7.20	Visualisierung von Sensor- und Objektdaten auf einem HMD.	166
7.21	Ermittlung von befahrbaren Bereichen mittels der Flottenbewegung vorausfahrender Fahrzeuge.	168
7.22	Polygonerzeugung aus den Flottentrajektorien.	169
A.1	Beispielbilder aus Datenkatalog 3.	180
A.2	Beispielbilder aus Datenkatalog 4.	181
B.1	Beispieldiagramme für ROC- und DET-Kurven.	185
C.1	Schachbrett aus Peltierelementen als Kalibrierkörper.	190
C.2	Extrinsische multispektrale Kalibrierung der Versuchskameras mittels einem Schachbrett aus Peltierelementen.	191
C.3	Erfassung der Kalibriervorrichtung in den Kamerasensoren.	192
C.4	Beispielkalibrierung der multispektralen Versuchskameras mittels der Punktkalibrierung.	193

Eigene Veröffentlichungen und betreute Diplomarbeiten

- [1] BAUER, G. ; HOMM, F. ; WALCHSHÄUSL, L. ; BURSCHKA, D.: Multi Spectral Pedestrian Detection and Localisation. In: *Proceedings of the 12th International Forum on Advanced Microsystems for Automotive Applications*. Berlin : VDI/VDE/IT, March 2008, S. 21–28
- [2] GACKSTATTER, C.: *Klassifikation von Fußgängern im Video- und Ferninfrarotbild*, Eberhard-Karls-Universität Tübingen, Wilhelm Schickard Institut für Informatik, Diplomarbeit, März 2008
- [3] HOMM, F.: *Fusion von multispektralen Videoquellen zur robusten Fußgängererkennung*, Technische Universität München, Fakultät für Informatik, Diplomarbeit, November 2007
- [4] LEBSACK, V.: *Multisensorielle Klassifikation im Kontext automotiver Umfelderkennungssysteme*, Fachhochschule Landshut, Fachbereich Informatik, Diplomarbeit, März 2007
- [5] LINDL, R. ; WALCHSHÄUSL, L.: Three-Level Early Fusion for Road User Detection. In: *PReVENT Fusion Forum e-Journal 1* (2006), September, S. 19–24
- [6] TATSCHKE, T. ; FÄRBER, F. ; FUCHS, E. ; WALCHSHÄUSL, L. ; LINDL, R.: Semi-autonomous reference data generation for perception performance evaluation. In: *Proceedings of the 10th International Conference on Information Fusion (FUSION)*. Quebec, Kanada, July 2007, S. 1–7
- [7] TÖNNIS, M. ; LINDL, R. ; WALCHSHÄUSL, L. ; KLINKER, G.: Visualization of Spatial Sensor Data in the Context of Automotive Environment Perception Systems. In: *Proceedings of the 6th International Symposium on Mixed and Augmented Reality (ISMAR)*. Nara, Japan, November 2007
- [8] WALCHSHÄUSL, L. ; LINDL, R.: Multi-Sensor Classification using a boosted Cascade Detector. In: *Proceedings of the 2007 IEEE Intelligent Vehicles Symposium*. Istanbul : IEEE, June 2007, S. 1045–1049
- [9] WALCHSHÄUSL, L. ; LINDL, R. ; VOGEL, K. ; TATSCHKE, T.: Detection of Road Users in Fused Sensor Data streams for Collision Mitigation. In: *Proceedings of the 10th International Forum on Advanced Microsystems for Automotive Applications*. Berlin : VDI/VDE/IT, April 2006

Literaturverzeichnis

- [ABB⁺01] AZUMA, R.T., Y. BAILLOT, R. BEHRINGER, S. FEINER, S. JULIER und B. MACINTYRE: *Recent advances in augmented reality*. Computer Graphics and Applications, IEEE, 21(6):34–47, 2001.
- [ABW01] ALLEN, B.D., G. BISHOP und G. WELCH: *Tracking: Beyond 15 Minutes of Thought*. Chapel Hill, NC, Seiten 27599–3175, 2001. (Zitiert auf Seite 28.)
- [AS04] ABRAMSON, Y. und B. STEUX: *Hardware-friendly pedestrian detection and impact prediction*. In: *Proceedings of the 2004 IEEE Intelligent Vehicles Symposium*, Seiten 590–595, 2004. (Zitiert auf Seite 34.)
- [Azu97] AZUMA, R.T.: *A survey of augmented reality*. Presence: Teleoperators and Virtual Environments(1054-7460), 6(4):355–385, 1997.
- [Bau07] BAUER, M.A.: *Tracking Errors in Augmented Reality*. Doktorarbeit, Technische Universität München, 2007.
- [BBC⁺00] BALDI, P., S. BRUNAK, Y. CHAUVIN, C.A.F. ANDERSEN und H. NIELSEN: *Assessing the accuracy of prediction algorithms for classification: an overview*. Bioinformatics, 16(5):412–424, 2000.
- [BBC⁺02] BERTOZZI, M., A. BROGGI, M. CELLARIO, A. FASCIOLI, P. LOMBARDI und M. PORTA: *Artificial vision in road vehicles*. In: *Proceedings of the 2004 IEEE Intelligent Vehicles Symposium*, Band 7, Seiten 1258–1271, 2002. (Zitiert auf Seite 63.)
- [BBG⁺04] BERTOZZI, M., A. BROGGI, P. GRISLERI, A. TIBALDI und M.D. ROSE: *A tool for vision based pedestrian detection performance evaluation*. In: *Proceedings of the 2004 IEEE Intelligent Vehicles Symposium*, Seiten 784–789, 2004. (Zitiert auf Seite 110.)
- [BBK04] BROADHURST, A., S. BAKER und T. KANADE: *A Prediction and Planning Framework for Road Safety Analysis, Obstacle Avoidance and Driver Information*. In: *Proceedings of the 11th World Congress on Intelligent Transportation Systems*, October 2004. (Zitiert auf Seite 15.)
- [BBLR05] BERTOZZI, M., A. BROGGI, A. LASAGNI und M.D. ROSE: *Infrared stereo vision-based pedestrian detection*. In: *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, Seiten 24–29, 2005. (Zitiert auf den Seiten 34 und 63.)
- [BD98] BULUSWAR, S.D. und B.A. DRAPER: *Color machine vision for autonomous ve-*

- hicles*. Engineering Applications of Artificial Intelligence, 11(2):245–256, 1998. (Zitiert auf Seite 34.)
- [Bed92] BEDWORTH, M.: *Probability Moderation for Multilevel Information Processing*. Personal Communication, 1992. (Zitiert auf Seite 22.)
- [Ber93] BERG, S.: *Condorcet's jury theorem, dependency among jurors*. Social Choice and Welfare, 10(1):87–95, 1993. (Zitiert auf Seite 93.)
- [BER03] BLACK, J., T. ELLIS und P. ROSIN: *A novel method for video tracking performance evaluation*. Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS), Seiten 125–132, 2003. (Zitiert auf Seite 110.)
- [BFC⁺04] BROGGI, A., A. FASCIOLI, M. CARLETTI, T. GRAF und M. MEINECKE: *A multi-resolution approach for infrared vision-based pedestrian detection*. In: *Proceedings of the 2004 IEEE Intelligent Vehicles Symposium*, Seiten 7–12, 2004. (Zitiert auf den Seiten 34 und 63.)
- [BHD00] BETKE, M., E. HARITAOGLU und L.S. DAVIS: *Real-time multiple vehicle detection and tracking from a moving vehicle*. Machine Vision and Applications, 12(2):69–83, 2000. (Zitiert auf Seite 34.)
- [BM06] BOCK, T. und M. MAURER: *Vehicle in the Loop Simulation*. In: *4. Workshop Fahrerassistenzsysteme*, Seiten 125–134, Löwenstein / Hößlinsülz, October 2006.
- [BMF07] BOCK, T., M. MAURER und G. FÄRBER: *Validation of the Vehicle in the Loop (VIL) – A milestone for the simulation of driver assistance systems*. In: *Proceedings of the 2007 IEEE Intelligent Vehicles Symposium*, Istanbul, June 2007. IEEE.
- [BO00] BEDWORTH, M. und J. O'BRIEN: *The Omnibus model: a new model of data fusion?* IEEE Aerospace and Electronic Systems Magazine, 15(4):30–36, 2000. (Zitiert auf Seite 22.)
- [Bor06] BORNER, A.: *Optische Sensoren für Fahrzeuge*. VDI BERICHTE, 1944:37, 2006. (Zitiert auf Seite 16.)
- [Bou07] BOUGUET, J.Y.: *Camera Calibration Toolbox for Matlab*. http://www.vision.caltech.edu/bouguetj/calib_doc/, 2007.
- [Boy87] BOYD, J.R.: *A Discourse on Winning and Losing*. Maxwell AFB lecture, 1987. (Zitiert auf den Seiten 21 und 22.)
- [BPT97] BATAVIA, P.H., D.E. POMERLEAU und C.E. THORPE: *Overtaking vehicle detection using implicit optical flow*. In: *Proceedings of the IEEE Conference on Intelligent Transportation System (ITSC 1997)*, Seiten 729–734, 1997. (Zitiert auf Seite 34.)
- [Bre96] BREIMAN, L.: *Bagging Predictors*. Machine Learning, 24(2):123–140, 1996. (Zitiert auf Seite 93.)
- [BT07] BARTLETT, P.L. und A. TEWARI: *Sparseness vs Estimating Conditional Probabilities: Some Asymptotic Results*. The Journal of Machine Learning Research, 8:775–790, 2007. (Zitiert auf Seite 94.)
- [Bun07] BUNDESMINISTERIUM FÜR BILDUNG UND FORSCHUNG: *INVENT*. <http://www.invent-online.de/index.html>, 2007. (Zitiert auf Seite 24.)

-
- [Bur98] BURGESS, C.J.C.: *A Tutorial on Support Vector Machines for Pattern Recognition*. Data Mining and Knowledge Discovery, 2(2):121–167, 1998.
- [CBF03] CHANG, K., K. BOWYER und P. FLYNN: *Face recognition using 2D and 3D facial data*. ACM Workshop on Multimodal User Authentication, Seiten 25–32, 2003. (Zitiert auf Seite 89.)
- [Che04] CHEN, C.: *Information Visualization: Beyond the Horizon*. Springer, 2004.
- [CLS⁺07] CUI, X., Y. LIU, S. SHAN, X. CHEN und W. GAO: *3D Haar-Like Features for Pedestrian Detection*. In: *Proceedings of the IEEE International Conference on Multimedia and Expo*, Seiten 1263–1266, 2007. (Zitiert auf Seite 64.)
- [CP99] CUCCHIARA, R. und M. PICCARDI: *Vehicle Detection under Day and Night Illumination*. In: *Proceedings of the 3rd International ICSC Symposia on Intelligent Industrial Automation and Soft computing*, Seiten 1–4, June 1999. (Zitiert auf den Seiten 34 und 64.)
- [Cri02] CRICHTON, N.: *Receiver Operating Characteristic (ROC) curves*. Journal of Clinical Nursing, 11:134–136, 2002.
- [CT88] CRISMAN, J. und C. THORPE: *Color Vision for Road Following*. In: *Proceedings of SPIE*, 1988. (Zitiert auf Seite 34.)
- [CZQ05] CHENG, H., N. ZHENG und J. QIN: *Pedestrian detection using sparse gabor filter and support vector machine*. In: *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, Seiten 583–587, 2005. (Zitiert auf Seite 64.)
- [CZS06] CHENG, H., N. ZHENG und C. SUN: *Boosted Gabor Features Applied to Vehicle Detection*. In: *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Band 1, Seiten 662–666. IEEE Computer Society Washington, DC, USA, 2006. (Zitiert auf Seite 64.)
- [DAR07] DARPA: *Grand Challenge*. <http://www.darpa.mil/grandchallenge/>, 2007. (Zitiert auf Seite 14.)
- [Das97] DASARATHY, B.V.: *Sensor fusion potential exploitation-innovative architectures and illustrative applications*. Proceedings of the IEEE, 85(1):24–38, 1997. (Zitiert auf Seite 22.)
- [Das00] DASARATHY, B.V.: *More the merrier...or is it? Sensor suite augmentation benefits assessment*. Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on, 2:WEC3/20–WEC3/25 vol.2, 2000. (Zitiert auf Seite 20.)
- [dC85] CONDORCET, M. DE: *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris, 1785. (Zitiert auf Seite 93.)
- [DHS00] DUDA, R.O., P.E. HART und D.G. STORK: *Pattern Classification*. Wiley-Interscience, 2000. (Zitiert auf den Seiten 63, 94 und 99.)
- [Dic02] DICKMANN, E.D.: *The development of machine vision for road vehicles in the last decade*. In: *Proceedings of the 2002 IEEE Intelligent Vehicles Symposium*, Band 1, Seiten 268–281 vol.1, 2002. (Zitiert auf Seite 24.)
- [Die00] DIETTERICH, T.G.: *Ensemble methods in machine learning*. Lecture Notes in Computer Science, 1857:1–15, 2000. (Zitiert auf den Seiten 92 und 102.)

- [DM00] DOERMANN, D. und D. MIHALCIK: *Tools and techniques for video performance evaluation*. Pattern Recognition, 2000. Proceedings. 15th International Conference on, 4, 2000. (Zitiert auf Seite 110.)
- [Dri01] DRISH, J.: *Obtaining calibrated probability estimates from support vector machines*. San Diego, 2001. (Zitiert auf Seite 94.)
- [DT05] DALAL, N. und B. TRIGGS: *Histograms of oriented gradients for human detection*. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1:886–893, June 2005. (Zitiert auf den Seiten 64, 67, 68 und 69.)
- [Dui02] DUIN, R.P.W.: *The combining classifier: to train or not to train*. In: *Proceedings of the 16th International Conference on Pattern Recognition*, Band 2, Seiten 765–770, 2002. (Zitiert auf Seite 95.)
- [DW88] DURRANT-WHYTE, H.F.: *Sensor models and multisensor integration*. International Journal of Robotics Research, 7(6):97–113, 1988. (Zitiert auf Seite 23.)
- [EBMM03] EFROS, A.A., A.C. BERG, G. MORI und J. MALIK: *Recognizing action at a distance*. In: *Proceedings of the 9th IEEE International Conference on Computer Vision*, Seiten 726–733, 2003. (Zitiert auf Seite 80.)
- [Eur07a] EUROPEAN COMMISSION: *PReVENT*. <http://www.prevent-ip.org/>, 2007. (Zitiert auf Seite 24.)
- [Eur07b] EUROPEAN COMMISSION: *Radar Net*. <http://www.radarnet.org/>, 2007. (Zitiert auf Seite 24.)
- [EW00] EWALD, A. und V. WILLHOEFT: *Laser scanners for obstacle detection in automotive applications*. In: *Proceedings of the 2000 IEEE Intelligent Vehicles Symposium*, Seiten 682–687, 2000. (Zitiert auf Seite 34.)
- [Faw04] FAWCETT, T.: *ROC Graphs: Notes and Practical Considerations for Researchers*. Machine Learning, 31, 2004.
- [FD04] FÜRSTENBERG, K.C. und K.C.J. DIETMAYER: *Fahrzeugumfelderfassung mit mehrzeiligen Laserscannern*. tm - Technisches Messen, 71(3):164–172, 2004. (Zitiert auf Seite 77.)
- [FDW02] FÜRSTENBERG, K.C., K.C.J. DIETMAYER und V. WILLHOEFT: *Pedestrian recognition in urban traffic using a vehicle based multilayer laserscanner*. In: *Proceedings of the 2002 IEEE Intelligent Vehicles Symposium*, Band 1, 2002. (Zitiert auf Seite 64.)
- [Fla04] FLACH, P.: *The many faces of ROC analysis in machine learning*. ICML Tutorial, 2004.
- [FLD03] FÜRSTENBERG, K.C., D.T. LINZMEIER und K.C.J. DIETMAYER: *Pedestrian Recognition and Tracking of Vehicles using a vehicle based Multilayer Laserscanner*. In: *Proceedings of International Conference on Transportation Systems*, 2003. (Zitiert auf den Seiten 34 und 80.)
- [Fol95] FOLEY, J.D.: *Computer Graphics: Principles and Practice*. Addison-Wesley Professional, 1995.
- [FP03] FORSYTH, D.A. und J. PONCE: *Computer Vision: A Modern Approach*. Prentice Hall, 2003. (Zitiert auf Seite 37.)
- [Fre04] FREYMAN, R.: *Möglichkeiten und Grenzen von Fahrerassistenz- und Aktiven Si-*

- cherheitssystemen. In: *Tagung Aktive Sicherheit durch Fahrerassistenz*, Garching, Germany, März 2004. (Zitiert auf den Seiten 11, 12, 13 und 14.)
- [FS05] FÜRSTENBERG, K.C. und J. SCHOLZ: *Reliable Pedestrian Protection using Laserscanners*. In: *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, Las Vegas, Nevada, U.S.A., June 2005. (Zitiert auf Seite 15.)
- [FSWG06] FARDI, B., I. SEIFERT, G. WANIELIK und J. GAYKO: *Motion-based pedestrian recognition from a moving vehicle*. In: *Proceedings of the 2006 IEEE Intelligent Vehicles Symposium*, Seiten 219–224, 2006. (Zitiert auf Seite 80.)
- [FTSW01] FLEMING, W.J., S. TECHNOL, T.R.W.O.S. SYST und M.I. WASHINGTON: *Overview of automotive sensors*. *Sensors Journal, IEEE*, 1(4):296–308, 2001. (Zitiert auf Seite 16.)
- [FW01] FÜRSTENBERG, K.C. und V. WILLHOEFT: *Pedestrian Recognition in urban traffic using Laserscanners*. In: *Proceedings of the 8th World Congress on Intelligent Transport Systems*, 2001. (Zitiert auf den Seiten 34 und 64.)
- [FYN⁺03] FANG, Y., K. YAMADA, Y. NINOMIYA, B. HORN und I. MASAKI: *Comparison between infrared-image-based and visible-image-based approaches for pedestrian detection*. In: *Proceedings of the 2003 IEEE Intelligent Vehicles Symposium*, Seiten 505–510, 2003. (Zitiert auf Seite 34.)
- [FYN⁺04] FANG, Y., K. YAMADA, Y. NINOMIYA, B.K.P. HORN und I. MASAKI: *A shape-independent method for pedestrian detection with far-infrared images*. *IEEE Transactions on Vehicular Technology*, 53(6):1679–1697, 2004. (Zitiert auf den Seiten 34 und 63.)
- [Gav99] GAVRILA, D.M.: *Visual analysis of human movement: A survey*. *Computer Vision and Image Understanding*, 73(1):82–98, 1999. (Zitiert auf Seite 80.)
- [Gav01] GAVRILA, D.M.: *Sensor-based pedestrian protection*. *IEEE Intelligent Systems*, 16(6):77–81, 2001. (Zitiert auf Seite 63.)
- [GBTC01] GUNATILAKA, A.H., B.A. BAERTLEIN, L. TECHNOL und O.H. COLUMBUS: *Feature-level and decision-level fusion of noncoincidentally sampled sensors for land mine detection*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):577–589, 2001. (Zitiert auf Seite 89.)
- [GCTC98] GIACHETTI, A., M. CAMPANI, V. TORRE und C. CRS: *The use of optical flow for road navigation*. *IEEE Transactions on Robotics and Automation*, 14(1):34–48, 1998. (Zitiert auf Seite 34.)
- [GGM⁺04] GAVRILA, D.M., J. GIEBEL, S. MUNDER, D.C. RES und G. ULM: *Vision-based pedestrian detection: the PROTECTOR system*. In: *Proceedings of the 2004 IEEE Intelligent Vehicles Symposium*, Seiten 13–18, 2004. (Zitiert auf Seite 34.)
- [GHJV95] GAMMA, E., R. HELM, R. JOHNSON und J. VLISSIDES: *Design patterns: elements of reusable object-oriented software*. Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1995.
- [GJE⁺04] GRESHAM, I., A. JENKINS, R. EGRI, C. ESWARAPPA, N. KINAYMAN, N. JAIN, R. ANDERSON, F. KOLAK, R. WOHLERT, S.P. BAWELL et al.: *Ultra-wideband radar sensors for short-range vehicular applications*. *IEEE Transactions on Microwave Theory and Techniques*, 52(9):2105–2122, 2004. (Zitiert auf Seite 17.)

- [GM07] GAVRILA, D.M. und S. MUNDER: *Multi-cue Pedestrian Detection and Tracking from a Moving Vehicle*. International Journal of Computer Vision, 73(1):41–59, 2007. (Zitiert auf Seite 34.)
- [GT06] GANDHI, T. und M.M. TRIVEDI: *Pedestrian Collision Avoidance Systems: A Survey of Computer Vision Based Recent Studies*. In: *Proceedings of the IEEE International Intelligent Transport Systems Conference*, Seiten 976–981, 2006. (Zitiert auf Seite 34.)
- [GT07] GANDHI, T. und M.M. TRIVEDI: *Pedestrian Protection Systems: Issues, Survey, and Challenges*. IEEE Transactions on Intelligent Transportation Systems, 8(3):413–430, 2007. (Zitiert auf den Seiten 33 und 63.)
- [HGS07] HARATI, A., S. GACHTER und R. SIEGWART: *Fast range image segmentation for indoor 3D-SLAM*. 6th IFAC Symposium on Intelligent Autonomous Vehicles, Toulouse, France, 2007. (Zitiert auf Seite 35.)
- [HHK02] HOTTA, K., Y. HONMA und T. KURITA: *Pedestrian Detection Method Using Contrast Gabor Features and Improvement of Generalization Ability by Feature Selection*. Joho Shori Gakkai Shinpojiumu Ronbunshu, 11, 2002. (Zitiert auf Seite 64.)
- [HKT⁺00] HANDMANN, U., T. KALINKE, C. TZOMAKAS, M. WERNER und W. SEELEN: *An image processing system for driver assistance*. Image and Vision Computing, 18(5):367–376, 2000. (Zitiert auf Seite 64.)
- [HMYO03] HASHIYAMA, T., D. MOCHIZUKI, Y. YANO und S. OKUMA: *Active frame subtraction for pedestrian detection from images of moving camera*. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Band 1, 2003. (Zitiert auf Seite 34.)
- [How82] HOWARD, I.P.: *Human visual orientation*. J. Wiley New York, 1982. (Zitiert auf Seite 1.)
- [HS81] HORN, B.K.P. und B.G. SCHUNCK: *Determining Optical Flow*. Artificial Intelligence, 17(1-3):185–203, 1981. (Zitiert auf Seite 34.)
- [HS90] HANSEN, L.K. und P. SALAMON: *Neural network ensembles*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(10):993–1001, 1990. (Zitiert auf Seite 92.)
- [HS97] HEIKKILA, J. und O. SILVEN: *A Four-Step Camera Calibration Procedure with Implicit Image Correction*. In: *Proceedings of IEEE Computer Vision and Pattern Recognition*, Seiten 1106–1112, 1997.
- [HSS04] HAVASI, L., Z. SZLÁVIK und T. SZIRÁNYI: „*Pedestrian detection using derived third-order symmetry of legs*”. In: *Proceedings of the International Conference on Computer Vision and Graphics*. Springer, 2004. (Zitiert auf Seite 34.)
- [HTF01] HASTIE, T., R. TIBSHIRANI und J. FRIEDMAN: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2001. (Zitiert auf Seite 101.)
- [HZ03] HARTLEY, R. und A. ZISSERMAN: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. (Zitiert auf Seite 26.)
- [INR02] INRIA: *CARSENSE*. <http://www.carsense.org/en/>, 2002. (Zitiert auf Sei-

- te 24.)
- [JJ02] JANSSON, J. und J. JOHANSSON: *Decision Making for Collision Avoidance Systems*. Technischer Bericht 2002-01-0403, Society of Automotive Engineers, Detroit, 2002. (Zitiert auf Seite 15.)
- [JWSX02] JAYNES, C., S. WEBB, R.M. STEELE und Q. XIONG: *An Open Development Environment for Evaluation of Video Surveillance Systems*. In: *Proceedings of the Third International Workshop on Performance Evaluation of Tracking and Surveillance (PETS'2002)*, Copenhagen, June, 2002. (Zitiert auf Seite 110.)
- [JZ07] JIA, H.X. und Y.J. ZHANG: *Fast Human Detection by Boosting Histograms of Oriented Gradients*. In: *Proceedings of the 4th International Conference on Image and Graphics (ICIG 2007)*, Seiten 683–688, 2007. (Zitiert auf Seite 64.)
- [KBD05] KÄMPCHEN, N., M. BÜHLER und K. DIETMAYER: *Feature-level fusion for free-form object tracking using laserscanner and video*. In: *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, Seiten 453–458, 2005. (Zitiert auf Seite 46.)
- [KER95] KRUGER, W., W. ENKELMANN und S. ROSSLE: *Real-time estimation and tracking of optical flow vectors for obstacle detection*. In: *Proceedings of the 1995 IEEE Intelligent Vehicles Symposium*, Seiten 304–309, 1995. (Zitiert auf Seite 34.)
- [KJG04] KARLSSON, R., J. JANSSON und F. GUSTAFSSON: *Model-based statistical tracking and decision making for collision avoidance application*. In: *Proceedings of the 2004 American Control Conference*, Band 4, Seiten 3435–3440 vol.4, 2004. (Zitiert auf Seite 15.)
- [KKNK03] KLAPPER, S., R. KYLE, R. NICKLIN und A. KORMOS: *Night Vision—Changing the Way We Drive [White Paper] Texas*. USA: Raytheon Electronic Systems, 2003. (Zitiert auf Seite 17.)
- [Käm07] KÄMPCHEN, N.: *Feature-level fusion of laser scanner and video data for advanced driver assistance systems*. Doktorarbeit, Universität Ulm. Fakultät für Ingenieurwissenschaften und Informatik, June 2007. (Zitiert auf Seite 15.)
- [KT07] KROTOSKY, S.J. und M.M. TRIVEDI: *On Color-, Infrared-, and Multimodal-Stereo Approaches to Pedestrian Detection*. *IEEE Transactions on Intelligent Transportation Systems*, 8(4):619–629, 2007. (Zitiert auf Seite 90.)
- [KTÅ00] KAHL, F., B. TRIGGS und K. ÅSTRÖM: *Critical Motions for Auto-Calibration When Some Intrinsic Parameters Can Vary*. *Journal of Mathematical Imaging and Vision*, 13(2):131–146, 2000.
- [KTvS98] KALINKE, T., C. TZOMAKAS und W. VON SEELEN: *A Texture-based Object Detection and an Adaptive Model-based Classification*. In: *Proceedings of the 1998 IEEE Intelligent Vehicles Symposium*, Seiten 341–346, 1998. (Zitiert auf Seite 34.)
- [Kue91] KUEHNLE, A.: *Symmetry-based recognition of vehicle rears*. *Pattern Recognition Letters*, 12(4):249–258, 1991. (Zitiert auf Seite 34.)
- [Kun02] KUNCHEVA, L.I.: *A theoretical study on six classifier fusion strategies*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):281–286, 2002. (Zitiert auf Seite 89.)
- [Kun04] KUNCHEVA, L.I.: *Combining Pattern Classifiers: Methods and Algorithms*. Wiley-Interscience, 2004. (Zitiert auf den Seiten 89 und 95.)

- [KW94] KONITZER, H. und J. WEHRMEISTER: *StVZO*. Straßenverkehrs-Zulassungs-Ordnung Bonn: Kirschbaum, Bonn, 1994. (Zitiert auf Seite 78.)
- [KW03] KUNCHEVA, L.I. und C.J. WHITAKER: *Measures of diversity in classifier ensembles*. Machine Learning, 51(2):181–207, 2003. (Zitiert auf Seite 89.)
- [Kwo99] KWOK, J.T.Y.: *Moderating the outputs of support vector machine classifiers*. IEEE Transactions on Neural Networks, 10(5):1018–1031, 1999. (Zitiert auf Seite 94.)
- [Lac88] LACHENBRUCH, P.A.: *Multiple reading procedures: the performance of diagnostic tests*. Statistics in Medicine, 7:549–557, 1988. (Zitiert auf Seite 89.)
- [LB98] LITTLE, J. und J. BOYD: *Recognizing people by their gait: the shape of motion*. Videre: Journal of Computer Vision Research, 1(2):1–32, 1998. (Zitiert auf Seite 80.)
- [LF04] LIU, X. und K. FUJIMURA: *Pedestrian detection using stereo night vision*. IEEE Transactions on Vehicular Technology, 53(6):1657–1665, 2004. (Zitiert auf Seite 34.)
- [Lin08] LINDL, R.: *Tracking von Verkehrsteilnehmern im Kontext von Multisensor-Systemen*. Doktorarbeit, Technische Universität München, Institut für Informatik, Munich, Germany, 2008. (Zitiert auf den Seiten 4, 46, 50 und 82.)
- [LK89] LUO, R.C. und M.G. KAY: *Multisensor integration and fusion in intelligent systems*. IEEE Transactions on Systems, Man and Cybernetics, 19(5):901–931, 1989. (Zitiert auf Seite 1.)
- [Low99] LOWE, D.G.: *Object Recognition from Local Scale-Invariant Features*. In: *Proceedings of the International Conference on Computer Vision*, Band 02, Seite 1150, Los Alamitos, CA, USA, 1999. IEEE Computer Society. (Zitiert auf Seite 67.)
- [Low04] LOWE, D.G.: *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, 60(2):91–110, 2004. (Zitiert auf Seite 67.)
- [LP98] LIU, F. und R. PICARD: *Finding periodicity in space and time*. In: *Proceedings of the 6th International Conference on Computer Vision*, Seiten 376–383, 1998. (Zitiert auf Seite 80.)
- [LWP07] LINDL, R., L. WALCHSHÄUSL und V. PARASCHOUDIS: *Autonomous Braking for Collision Mitigation Purposes by Means of Multi Sensor Perception*. In: *Proceedings of the 6th European Congress and Exhibition on Intelligent Transport Systems and Services*, Aalborg, Denmark, June 2007. IEEE. (Zitiert auf Seite 15.)
- [Mac86] MACKINLAY, J.: *Automating the design of graphical presentations of relational information*. ACM Transactions on Graphics, 5(2):110–141, 1986.
- [MACH96] MATTHEWS, N.D., P.E. AN, D. CHARNLEY und C.J. HARRIS: *Vehicle detection and recognition in greyscale imagery*. Control Engineering Practice, 4(4):473–479, 1996. (Zitiert auf Seite 34.)
- [MC93] MORI, H. und N.M. CHARKARI: *Shadow and rhythm as sign patterns of obstacle detection*. In: *Proceedings of the IEEE International Symposium on Industrial Electronics*, Seiten 271–277, 1993. (Zitiert auf Seite 34.)
- [McC86] MCCONNELL, R.K.: *Method of and apparatus for pattern recognition*, January 1986. US Patent 4,567,610. (Zitiert auf Seite 66.)

- [MDB87] MCCORMICK, B.H., T.A. DEFANTI und M.D. BROWN: *Visualization in scientific computing - A synopsis*. IEEE Computer Graphics & Applications, 7(7):61–70, 1987.
- [MDG⁺05] MERTZ, C., D. DUGGINS, J. GOWDY, J. KOZAR, R. MACLACHLAN, A. STEINFELD, A. SUPPE, T. CHUCK, THORPE und C. WANG: *Collision Warning and Sensor Data Processing in Urban Areas*. In: *Proceedings of the 5th international conference on ITS telecommunications*, Seiten 73–78, June 2005. (Zitiert auf Seite 15.)
- [MDK⁺97] MARTIN, A., G. DODDINGTON, T. KAMM, M. ORDOWSKI und M. PRZYBOCKI: *The DET Curve in Assessment of Detection Task Performance*. Fifth European Conference on Speech Communication and Technology, 1997.
- [MS05] MAURER, M. und C. STILLER: *Fahrerassistenzsysteme mit maschineller Wahrnehmung*. Springer, 2005. (Zitiert auf Seite 11.)
- [MSP06] MOHAMMED, H.S., N. STEPENOSKY und R. POLIKAR: *An Ensemble Technique to Handle Missing Data from Sensors*. In: *Proceedings of the 2006 IEEE Sensors Applications Symposium*, Seiten 101–105, 2006. (Zitiert auf Seite 101.)
- [Naa04] NAAB, K.: *Sensorik- und Signalverarbeitungsarchitekturen für Fahrerassistenz und Aktive Sicherheit*. In: *Tagung Aktive Sicherheit durch Fahrerassistenz*, Garching, Germany, März 2004. (Zitiert auf den Seiten 12 und 13.)
- [ND02] NANDA, H. und L. DAVIS: *Probabilistic template based pedestrian detection in infrared videos*. In: *Proceedings of the 2002 IEEE Intelligent Vehicles Symposium*, Band 1, 2002. (Zitiert auf Seite 63.)
- [Nev05] NEVE, A.: *Fußgängererkennung in FIR und Video Sequenzen mit Hilfe entfernungsgebender Sensorik*. Diplomarbeit, Technische Universität München, 2005.
- [NMTS05] NGUYEN, V., A. MARTINELLI, N. TOMATIS und R. SIEGWART: *A comparison of line extraction algorithms using 2D laser rangefinder for indoor mobile robotics*. In: *Proceedings of the IEEE International Conference on Intelligent Robots and Systems (IROS 2005)*, Seiten 1929–1934, 2005. (Zitiert auf Seite 37.)
- [NO87] NIELSON, G. und D. OLSEN: *Direct manipulation techniques for 3D objects using 2D locator devices*. In: *Proceedings of the 1986 Workshop on Interactive 3D graphics (SI3D '86)*, Seiten 175–182, New York, NY, USA, 1987. ACM.
- [Nor94] NORM, D.: *Din 70 000 Fahrzeugdynamik und Fahrverhalten (Vehicle dynamics and handling)*, 1994. (Zitiert auf Seite 27.)
- [NWB04] NEWMAN, J., M. WAGNER und M. BAUER: *Ubiquitous Tracking for Augmented Reality*. In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR'04)*, Arlington, VA, USA, November 2004. (Zitiert auf Seite 28.)
- [Oxf08] OXFORD TECHNICAL SOLUTIONS: *RT3003*. <http://www.oxts.co.uk/>, 2008.
- [PB88] PHILLIPS, C. und N. BADLER: *JACK: a toolkit for manipulating articulated figures*. In: *Proceedings of the 1st annual ACM SIGGRAPH symposium on User Interface Software (UIST '88)*, Seiten 221–229, New York, NY, USA, 1988. ACM.
- [PHBK06] PUSTKA, D., M. HUBER, M. BAUER und G. KLINKER: *Spatial Relationship Patterns: Elements of Reusable Tracking and Calibration Systems*. In: *Proceedings of*

- the IEEE International Symposium on Mixed and Augmented Reality (ISMAR'06)*, October 2006. (Zitiert auf Seite 28.)
- [Pla00] PLATT, J.C.: *Probabilities for SV Machines*. Advances in Large-Margin Classifiers, 2000. (Zitiert auf Seite 94.)
- [Pol06] POLIKAR, R.: *Ensemble Based Systems in Decision Making*. IEEE Circuits and Systems Magazine, 6(3):21–45, 2006. (Zitiert auf den Seiten 90, 93 und 95.)
- [PP95] PARODI, P. und G. PICCIOLI: *A feature-based recognition scheme for traffic scenes*. In: *Proceedings of the 1995 IEEE Intelligent Vehicles Symposium*, Seiten 229–234, 1995. (Zitiert auf Seite 64.)
- [PP00] PAPAGEORGIOU, C. und T. POGGIO: *A Trainable System for Object Detection*. International Journal of Computer Vision, 38(1):15–33, 2000. (Zitiert auf den Seiten 34 und 64.)
- [RBW05] REGENBRECHT, H., G. BARATOFF und W. WILKE: *Augmented reality projects in the automotive and aerospace industries*. IEEE Computer Graphics and Applications, 25(6):48–56, 2005.
- [RC02] ROGALSKI, A. und K. CHRZANOWSKI: *Infrared devices and techniques*. Optoelectronics Review, 10(2):111–36, 2002. (Zitiert auf Seite 17.)
- [RDB01] ROLLAND, J.P., L.D. DAVIS und Y. BAILLOT: *A Survey of Tracking Technology for Virtual Environments*. Fundamentals of Wearable Computers and Augmented Reality, 2001.
- [RG00] RUTA, D. und B. GABRYS: *An overview of classifier fusion methods*. Computing and Information Systems, 7(1):1–10, 2000. (Zitiert auf Seite 89.)
- [RG05] RASSHOFER, R.H. und K. GRESSER: *Automotive Radar and Lidar Systems for Next Generation Driver Assistance Functions*. Advances in Radio Science, 3(10):205–209, 2005. (Zitiert auf Seite 17.)
- [RH95] ROBINETT, W. und R. HOLLOWAY: *The visual display transformation for virtual reality*. Presence: Teleoperators and Virtual Environments, 4(1):1–23, 1995. (Zitiert auf Seite 28.)
- [RJ03] ROSS, A. und A. JAIN: *Information fusion in biometrics*. Pattern Recognition Letters, 24(13):2115–2125, 2003. (Zitiert auf Seite 89.)
- [Rüp04] RÜPING, S.: *A Simple Method For Estimating Conditional Probabilities For SVMs*. LWA, 2004. (Zitiert auf Seite 94.)
- [RPL07] RUSER, H. und F. PUENTE LEÓN: *Informationsfusion - Eine Übersicht*. Technisches Messen, 74(3):93–102, 2007. (Zitiert auf den Seiten 21 und 22.)
- [RR91] ROBINETT, W. und J.P. ROLLAND: *Computational model for the stereoscopic optics of a head-mounted display*. Proceedings of SPIE, 1457:140, 1991.
- [RZW⁺05] RAN, Y., Q. ZHENG, I. WEISS, LS DAVIS, W. ABD-ALMAGEED und L. ZHAO: *Pedestrian Classification from Moving Platforms Using Cyclic Motion Pattern*. In: *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Band 2, 2005. (Zitiert auf Seite 80.)
- [SBM02] SUN, Z., G. BEBIS und R. MILLER: *On-road vehicle detection using Gabor filters and support vector machines*. In: *Proceedings of the 14th International Conference on Digital Signal Processing*, Band 2, 2002. (Zitiert auf Seite 64.)

-
- [SBM05] SUN, Z., G. BEBIS und R. MILLER: *On-road vehicle detection using evolutionary Gabor filter optimization*. IEEE Transactions on Intelligent Transportation Systems, 6(2):125–137, 2005. (Zitiert auf Seite 64.)
- [SBM06] SUN, Z., G. BEBIS und R. MILLER: *On-road vehicle detection: a review*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(5):694–711, 2006. (Zitiert auf den Seiten 33 und 63.)
- [SBW99] STEINBERG, A.N., C. BOWMAN und F.E. WHITE: *Revisions to the JDL data fusion process model*. In: *Proceedings of Sensor Fusion: Architectures, Algorithms, and Applications III (SPIE)*, Band 3719, Seiten 430–441, 1999. (Zitiert auf Seite 21.)
- [SCFW04] SCHEUNERT, U., H. CRAMER, B. FARDI und G. WANIELIK: *Multi sensor based tracking of pedestrians: a survey of suitable movement models*. In: *Proceedings of the 2004 IEEE Intelligent Vehicles Symposium*, Seiten 774–778, 2004. (Zitiert auf Seite 34.)
- [Sch90] SCHAPIRE, R.E.: *The strength of weak learnability*. Machine Learning, 5(2):197–227, 1990. (Zitiert auf Seite 93.)
- [Sch04] SCHNEBERGER, M.: *Infrared Optical Tracking Systems Mathematical and Operation Principles*. Technischer Bericht, A.R.T. GmbH, 2004.
- [SG06] SEECK, A. und T.M. GASSER: *Klassifizierung und Würdigung der rechtlichen Rahmenbedingungen im Zusammenhang mit der Einführung moderner FAS*. In: *Tagung Aktive Sicherheit durch Fahrerassistenz*, Garching, Germany, April 2006. (Zitiert auf Seite 14.)
- [SGRB05] SUARD, F., V. GUIGUE, A. RAKOTOMAMONJY und A. BENSRAIR: *Pedestrian detection using stereo-vision and graph kernels*. In: *Proceedings of the 2005 IEEE Intelligent Vehicles Symposium*, Seiten 267–272, 2005. (Zitiert auf Seite 64.)
- [Shn96] SHNEIDERMAN, B.: *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In: *IEEE Visual Languages*, Seiten 336–343, College Park, Maryland 20742, U.S.A., 1996.
- [Shn03] SHNEIDERMAN, B.: *Why not make interfaces better than 3D reality?* Computer Graphics and Applications, IEEE, 23(6):12–15, 2003.
- [Sho92] SHOEMAKE, K.: *ARCBALL: A User Interface for Specifying Three-Dimensional Orientation Using a Mouse*. Graphics Interface, 92:151–156, 1992.
- [SK00] SCHNEIDERMAN, H. und T. KANADE: *A Statistical Model for 3D Object Detection Applied to Faces and Cars*. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Band 1, Seiten 1746–1751, 2000. (Zitiert auf Seite 64.)
- [SKK⁺97] SIADAT, A., A. KASKE, S. KLAUSMANN, M. DUFAUT und R. HUSSON: *An Optimized Segmentation Method for a 2D Laser-Scanner Applied to Mobile Robot Navigation*. In: *Proceedings of the 3rd IFAC Symposium on Intelligent Components and Instruments for Control Applications*, 1997. (Zitiert auf Seite 37.)
- [Sko90] SKOLNIK, M.I.: *Radar Handbook*. McGraw-Hill Professional, 1990. (Zitiert auf Seite 77.)
- [SM00] SCHUMAN, H. und W. MÜLLER: *Visualisierung*. Springer Verlag, 2000.

- [SMBD02] SUN, Z., R. MILLER, G. BEBIS und D. DiMEO: *A real-time precrash vehicle detection system*. In: *Proceedings of the 6th IEEE Workshop on Applications of Computer Vision (WACV 2002)*, Seiten 171–176, 2002. (Zitiert auf Seite 34.)
- [SMP05] SVOBODA, T., D. MARTINEC und T. PAJDLA: *A Convenient Multicamera Self-Calibration for Virtual Environments*. *Presence: Teleoperators and Virtual Environments*, 14(4):407–422, 2005.
- [Spe00] SPENCE, R.: *Information Visualization*. Addison Wesley, December 2000.
- [SRBB06] SUARD, F., A. RAKOTOMAMONJY, A. BENSRAHAIR und A. BROGGI: *Pedestrian Detection using Infrared images and Histograms of Oriented Gradients*. *Machine Learning*, 9(5):12, 2006. (Zitiert auf Seite 64.)
- [ST88] SAMET, H. und M. TAMMINEN: *Efficient component labeling of images of arbitrary dimension represented by linear bintrees*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):579–586, 1988. (Zitiert auf Seite 35.)
- [Sta06] STATISTISCHES BUNDESAMT: *Verkehrsunfälle - Fachserie 8 Reihe 7 - 2006*. Technischer Bericht 8, Statistisches Bundesamt, 2006. Reihe 7. (Zitiert auf den Seiten 2 und 14.)
- [STP07] SAAR-TSECHANSKY, M. und F. PROVOST: *Handling Missing Values when Applying Classification Models*. *The Journal of Machine Learning Research*, 8:1623–1657, 2007. (Zitiert auf Seite 101.)
- [Svo03] SVOBODA, T.: *Quick guide to multi-camera self-calibration*. Technischer Bericht, Technical Report BiWi-TR-263, Computer Vision Lab, Swiss Federal Institute of Technology, August 2003.
- [SWND05] SHREINER, D., M. WOO, J. NEIDER und T. DAVIS: *OpenGL(R) Programming Guide: The Official Guide to Learning OpenGL(R), Version 2 (5th Edition) (OpenGL)*. Addison-Wesley Professional, 2005.
- [SZN+06] SHAO, X., H. ZHAO, K. NAKAMURA, R. SHIBASAKI, R. ZHANG und Z. LIU: *Analyzing Pedestrians' Walking Patterns Using Single-Row Laser Range Scanners*. In: *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (ICSMC'06)*, Band 2, 2006. (Zitiert auf Seite 80.)
- [Tho05] THOMA, S.: *Fusion von Infrarotvideo- und Laserscanner-Daten zur Fußgängererkennung*. Diplomarbeit, Technische Universität München, Fakultät für Informatik, Dezember 2005. (Zitiert auf den Seiten 36 und 80.)
- [THWN02] TSUJI, T., H. HATTORI, M. WATANABE und N. NAGAOKA: *Development of night-vision system*. *IEEE Transactions on Intelligent Transportation Systems*, 3(3):203–209, 2002. (Zitiert auf Seite 34.)
- [TK04] TAYLOR, G. und L. KLEEMAN: *A Multiple Hypothesis Walking Person Tracker with Switched Dynamic Model*. In: *Proceedings of the Australasian Conference on Robotics and Automation (ACRA)*, 2004. (Zitiert auf Seite 80.)
- [TN00] TUCERYAN, M. und N. NAVAB: *Single point active alignment method (SPAAM) for optical see-through HMD calibration for AR*. In: *Proceedings of the IEEE and ACM International Symposium on Augmented Reality*, Seiten 149–158, 2000.
- [TRC01] TAN, D.S., G.G. ROBERTSON und M. CZERWINSKI: *Exploring 3D navigation: combining speed-coupled flying with orbiting*. In: *Proceedings of the SIGCHI confe-*

- rence on Human factors in computing systems, Seiten 418–425. ACM Press New York, NY, USA, 2001.
- [TvS98] TZOMAKAS, C. und W. VON SEELEN: *Vehicle Detection in Traffic Scenes Using Shadows*. INI, 1998. (Zitiert auf Seite 34.)
- [Vap00] VAPNIK, V.N.: *The Nature of Statistical Learning Theory*. Springer, 2000.
- [Var97] VARSHNEY, P.K.: *Multisensor data fusion*. Electronics & Communication Engineering Journal, 9(6):245–253, December 1997. (Zitiert auf den Seiten 21 und 23.)
- [vdHvW01] HEIJDEN, R. VAN DER und K. VAN WEES: *Introducing Advanced Driver Assistance Systems: Some Legal Issues*. European Journal for Transport and Infrastructure Research, 1(3):309–326, 2001. (Zitiert auf Seite 14.)
- [Vic01] VICCI, L.: *Quaternions and Rotations in 3-Space: The Algebra and its Geometric Interpretation*. TR01-014, Department of Computer Science, UNC, Chapel Hill, NC, 2001. (Zitiert auf Seite 30.)
- [VJ01] VIOLA, P. und M.J. JONES: *Rapid object detection using a boosted cascade of simple features*. In: *Proceedings of the CVPR*, Band 1, Seiten 511–518, 2001. (Zitiert auf Seite 64.)
- [VJS05] VIOLA, P., M.J. JONES und D. SNOW: *Detecting Pedestrians Using Patterns of Motion and Appearance*. International Journal of Computer Vision, 63(2):153–161, 2005. (Zitiert auf Seite 64.)
- [VM02] VALENTINI, G. und F. MASULLI: *Ensembles of Learning Machines*. Neural Nets: 13th Italian Workshop on Neural Nets, Wirm Vietri 2002, Vietri Sul Mare, Italy, May 30-June 1, 2002. Revised Papers, 2002. (Zitiert auf Seite 89.)
- [Vog07] VOGEL, K.: *High-Accuracy Reference Data Acquisition for Evaluation of Active Safety Systems by means of a RTK-GNSS-Surveying System*. In: *6th European Congress and Exhibition on Intelligent Transport Systems and Services (ITS)*, June 2007.
- [WCH92] WENG, J., P. COHEN und M. HERNIOU: *Camera Calibration with Distortion Models and Accuracy Evaluation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(10):965–980, 1992.
- [WD05] WENDER, S. und K.C.J. DIETMAYER: *Statistical Approaches for Vehicle Environment Classification at Intersections with a Laserscanner*. In: *Proceedings of the Intelligent Transport Systems (ITS)*, 2005. (Zitiert auf Seite 90.)
- [WD07] WENDER, S. und K.C.J. DIETMAYER: *A Feature Level Fusion Approach for Object Classification*. In: *Proceedings of the 2007 IEEE Intelligent Vehicles Symposium*, Istanbul, June 2007. (Zitiert auf Seite 90.)
- [WDQ05] WEST, D., S. DELLANA und J. QIAN: *Neural network ensemble strategies for financial decision applications*. Computers and Operations Research, 32(10):2543–2559, 2005. (Zitiert auf Seite 89.)
- [Wei03] WEISSTEIN, E.W.: *Least Squares Fitting–Perpendicular Offsets*. Online: <http://mathworld.wolfram.com/LeastSquaresFittingPerpendicularOffsets.html> last accessed, 18(06), 2003. (Zitiert auf Seite 37.)
- [WGS⁺04] WISSELMANN, D., K. GRESSER, H. SPANNHEIMER, K. BENGLER und A. HUESMANN: *ConnectedDrive–ein methodischer Ansatz für die Entwicklung zukünftiger*

- Fahrerassistenzsysteme*. In: *Tagung Aktive Sicherheit durch Fahrerassistenz*, 2004. (Zitiert auf Seite 19.)
- [Whi96] WHITTLE, M.W.: *Gait analysis*. Butterworth-Heinemann Boston, 1996. (Zitiert auf Seite 80.)
- [WO90] WARE, C. und S. OSBORNE: *Exploration and virtual camera control in virtual three dimensional environments*. In: *Proceedings of the 1990 symposium on Interactive 3D graphics*, Seiten 175–183, New York, NY, USA, 1990. ACM.
- [WWFD06] WENDER, S., T. WEISS, K.C. FÜRSTENBERG und K.C.J. DIETMAYER: *Feature Level Fusion for Object Classification*. PReVENT Fusion Forum e-Journal, 1:31–36, 2006. (Zitiert auf Seite 90.)
- [XCQW06] XU, Y.W., X.B. CAO, H. QIAO und F.Y. WANG: *A Cascaded Classifier for Pedestrian Detection*. In: *Proceedings of the 2006 IEEE Intelligent Vehicles Symposium*, Seiten 336–343, 2006. (Zitiert auf Seite 64.)
- [XLF05] XU, F., X. LIU und K. FUJIMURA: *Pedestrian Detection and Tracking With Night Vision*. IEEE Transactions on Intelligent Transportation Systems, 6(1):63, 2005. (Zitiert auf Seite 34.)
- [ZBvS92] ZIELKE, T., M. BRAUCKMANN und W. VON SEELEN: *Intensity and Edge-Based Symmetry Detection Applied to Car-Following*. In: *Proceedings of the European Conference on Computer Vision*, Seiten 865–873. Springer, 1992. (Zitiert auf Seite 34.)
- [Zha99] ZHANG, Z.: *Flexible camera calibration by viewing a plane from unknown orientations*. In: *International Conference on Computer Vision (ICCV'99)*, Seiten 666–673, Sep. 1999.