RUHR
UNIVERSITÄT
BOCHUM

RUB

Erkennung der Querungsintention von Fußgängern für das automatisierte Fahren im städtischen Umfeld

Dissertation zur Erlangung des Grades eines
Doktor-Ingenieurs
der Fakultät für Elektrotechnik und Informationstechnik
an der Ruhr-Universität Bochum

Friederike Schneemann
geboren in Düsseldorf

2018

1. Gutachter	2. Gutachter
Prof. Dr. Gregor Schöner	PD Dr. Rolf Würtz

Tag der mündlichen Prüfung: 21. August 2018

Audi-Dissertationsreihe, Band 134

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

1. Aufl. - Göttingen: Cuvillier, 2018
Zugl.: Bochum, Univ., Diss., 2018

Nonnenstieg 8, 37075 Göttingen
Telefon: 0551-54724-0
Telefax: 0551-54724-21
www.cuvillier.de

1. Auflage, 2018
Gedruckt auf umweltfreundlichem, säurefreiem Papier aus nachhaltiger Forstwirtschaft.

ISBN 978-3-7369-9916-9
eISBN 978-3-7369-8916-0

Danksagung

Die vorliegende Dissertation entstand während meiner Tätigkeit als Doktorand der AUDI AG unter wissenschaftlicher Betreuung am Institut für Neuroinformatik der Ruhr Universität Bochum. In dieser Zeit haben mich viele Menschen begleitet und bei der Erstellung dieser Arbeit unterstützt. Dafür möchte ich mich aufrichtig bedanken.

Mein besonderer Dank gilt Prof. Dr. Gregor Schöner für die bereitwillige Übernahme der Erstbetreuung, obwohl die Richtung dieser Arbeit anfänglich noch unklar war. Seine konstruktiven Hinweise haben meine Blickweise auf das Thema „Intention“ entschieden erweitert und die erreichte Interdisziplinarität dieser Arbeit maßgeblich geprägt. Bei PD Dr. Rolf Würtz möchte ich mich herzlich für die Übernahme der Zweitbetreuung bedanken.

Weiterer Dank gilt meinen Kollegen bei der Audi Electronics Venture GmbH; allen vorweg meinem Betreuer Dr. Patrick Heinemann, der mir mit intensiven Diskussionen geholfen hat, den Untersuchungsschwerpunkt dieser Arbeit zu finden, mir in schwierigen Phasen Mut zugesprochen und sich bis zur Fertigstellung stets Zeit für mich genommen hat. Zudem danke ich meinen Vorgesetzten Dr. Miklós Kiss und Andreas Reich für die erhaltene Unterstützung und Rückendeckung sowie für die ermöglichten Freiräume zur Erstellung der Arbeit. Vielen Dank auch an meine Kollegen der AEV-31 und an das „Doktorandenbüro“ für die freundschaftliche Arbeitsumgebung, insbesondere an Dr. Harald Altinger für den technischen Support und an Christoph Sippl für das bereitwillige Teilen seiner Workstation. Ebenso bedanke ich mich bei meinen Studenten: Sascha Rosbach für die Einblicke in das Deep Learning und das Aufzeigen der Herausforderungen bei der Erkennung von Fußgängerorientierungen, Irene Gohl für das außerordentlich hohe Engagement bei der Analyse der Fahrer-Fußgänger Interaktion und Florian Engel für die Markierung jedes noch so kleinen Fußgängerkopfes. Meinen Kollegen Dr. Ingo Totzke, Leonie Gauer, Amelie Stephan und Dr. Isis Mennig möchte ich für das geduldige Beantworten aller meiner „Psychologie“-Fragen danken. Ein weiterer Dank gilt den Kollegen am Electronics Research Laboratory, die meinen Aufenthalt in Kalifornien zu einer lehrreichen und unvergesslichen Zeit gemacht haben.

Ein herzlicher Dank gilt zudem Dr. Frederik Diederichs vom Fraunhofer IAO: Mit der frühzeitigen Zurverfügungstellung seiner Doktorarbeit sowie seiner Bereitschaft zu einem anregenden fachlichen Austausch über das Jordan-Modell hat er maßgeblich zu den Ergebnissen dieser Arbeit beigetragen. Ebenso bedanke ich mich bei allen Teilnehmern meiner Beobachterstudie, die mit der mühsamen Annotation der Videodaten einen wichtigen Beitrag zu den Ergebnissen geliefert haben.

Abschließend möchte ich mich bei meiner Familie und meinen Freunden für die Geduld und den Rückhalt bedanken. Ein besonders großer Dank gilt hier meinen Eltern für die mir im Leben gebotenen Möglichkeiten und die stets erfahrene vielseitige Unterstützung. Zudem bedanke ich mich bei meinem Vater für die unermüdliche Verbesserung meines Schreibstils und das Setzen von Semikolons, Doppelpunkten und Kommas an die richtigen Stellen. Zu guter Letzt gilt mein Dank meinem Freund Michael für sein Verständnis gegenüber den unzählige Stunden, die er hinter der Arbeit zurückstecken musste.

Kurzfassung

Die vorliegende Dissertation befasst sich mit der Erkennung der Querungsintention von Fußgängern, um das „Situationsbewusstsein“ zukünftig automatisiert fahrender Fahrzeuge im städtischen Umfeld zu verbessern. Auf Basis einer ausführlichen Analyse bestehender Definitionen und Modelle zur menschlichen Intention wird zunächst der Begriff „Fußgängerintention“ eindeutig definiert und präzise abgegrenzt von im bisherigen Stand der Technik synonym verwendeten Begriffen. Weiter wird nach ausführlicher Analyse des beobachtbaren Querungsverhaltens von Fußgängern ein bereits bekanntes Modell zur Erkennung von Fahrerintentionen an die Erkennung der Querungsintention von Fußgängern angepasst – mit dem Ergebnis, dass die im Modell beschriebenen Verhaltensweisen querungswilliger Fußgänger über kontext- und posenbasierte Informationen beobachtet werden können. Diese Erkenntnis bildet die Basis für den Entwurf eines neuen Erkennungssystems, bei dem merkmalsbasierte Methoden des maschinellen Lernens unter Verwendung der Support Vector Regression eingesetzt werden. Da aus dem Stand der Technik keine geeignete Beschreibungsform für das kontextuelle Bewegungsverhalten eines Fußgängers bekannt ist, wird ein eigenes Verfahren entwickelt, das die Fußgängerbewegung relativ zu relevanten Szenenelementen abbildet. Zur Beschreibung posenbasierter Informationen sind bereits mehrere theoretisch geeignete Merkmale bekannt. Die nicht direkte Beobachtbarkeit der Intention stellt die Bestimmung der Referenz, die für das Training und die Evaluation des neuen Ansatzes benötigt wird, schließlich vor neue Herausforderungen. Diesen wird in der vorliegenden Arbeit mit der Entwicklung einer beobachterbasierten Videoannotationsmethode begegnet, mit Hilfe derer das Urteil menschlicher Beobachter als Referenz verwendet werden kann. Die auf Basis von – im deutschen Straßenverkehr aufgenommenen – Videodaten durchgeführte Evaluation zeigt schließlich, dass die Erkennung der Querungsintention von Fußgängern mit dem entwickelten Ansatz möglich ist. Verbesserungsbedarf besteht vor allem noch hinsichtlich der Erkennung positiv ausgeprägter Querungsintentionen sowie bei der Vorhersage des genauen Unsicherheitswerts. Für zukünftige Entwicklungsprojekte wird eine Reduktion der mit einem einzelnen Prädiktor zu erfassenden Situationskomplexität empfohlen, die durch den Einsatz mehrerer, spezifisch trainierter Situationsexperten in Kombination mit einer Situationsklassifikation erreicht werden kann.

Inhaltsverzeichnis

Kapitel 1

Einleitung

Automatisiertes Fahren hat das Potential, die Mobilität der Zukunft äußerst sicher, komfortabel und effizient zu gestalten. Erste prototypische Systeme übernehmen bereits in speziellen Situationen, wie etwa auf der Autobahn oder im Parkhaus ohne Mischverkehr, die Fahraufgabe komplett. Diese Systeme sind bisher jedoch auf Situationen in stark strukturierter Umgebung und mit begrenzten Umgebungsvariablen beschränkt. Zukünftig sollen auch im städtischen Umfeld automatisierte Fahrfunktionen realisiert werden.

1.1 Motivation

Im städtischen Umfeld besteht im Vergleich zu den bereits beherrschbaren Szenarien eine deutlich komplexere Verkehrssituation. Diese ist geprägt durch eine Vielzahl an Umgebungsvariablen und ein hohes Interaktionslevel zwischen den Verkehrsteilnehmern. Entgegen heutiger technischer Systeme, ist der menschliche Fahrer sehr gut in der Lage, sich auch in solchen komplexen Situationen seiner Umgebung zutreffend bewusst zu sein und, vor dem Hintergrund des eigenen Ziels, Entscheidungen für ein sicheres und effizientes Handeln zu treffen (Anthony, 2016). Ein Ansatz, um zukünftig auch in komplexen Situationen wie dem städtischen Umfeld automatisiert fahren zu können, ist somit die Übertragung der kognitiven Fähigkeiten des Menschen zum Situationsbewusstsein auf ein Fahrzeugsystem. In Kombination mit einem angemessenen

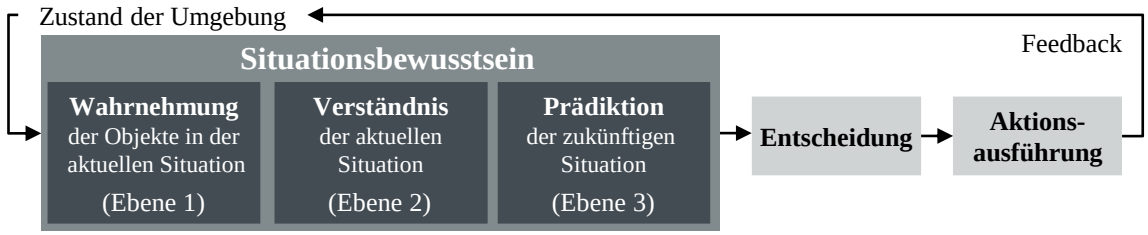


Abbildung 1.1: Vereinfachte Darstellung des Modells zum Situationsbewusstsein von Endsley (1995).

Interaktionskonzept ermöglicht ein entsprechend entwickeltes technisches Situationsbewusstsein zukünftigen automatisierten Fahrzeugen die Generierung menschenähnlichen Verhaltens, das konform den Erwartungen aller menschlichen Teilnehmer ist. Hierdurch wird ein hinreichend konfliktfreier Ablauf des Straßenverkehrs sowie ein hohes Komfortlevel der Passagiere gewährleistet (Schneemann und Gohl, 2016).

Entsprechend dem Modell von Endsley (1995) muss das automatisierte Fahrzeug zur Bildung eines Situationsbewusstseins drei Ebenen durchlaufen (s. Abb. 1.1). Die erste Ebene beschreibt die Wahrnehmung aller Objekte und Elemente der aktuellen Situation innerhalb einer dynamischen Umgebung. Auf der zweiten Ebene wird das Verständnis für die aktuelle Situation gewonnen, indem die wahrgenommenen Objekte miteinander in Bezug gesetzt und mit angeeignetem Wissen verknüpft werden. Ebene drei beschreibt schließlich die Prädiktion der zukünftigen Situation auf Basis der in den vorherigen Ebenen gewonnen Informationen sowie des Wissens über die mögliche Dynamik der identifizierten Elemente. Das so gebildete Situationsbewusstsein bildet die Grundlage für die anschließende Entscheidungsfindung bezüglich des eigenen Handelns, welche schließlich zur Ausführung entsprechender Aktionen führt. Diese Aktionen nehmen wiederum Einfluss auf den Zustand der dynamischen Umgebung.

Vor dem Hintergrund der Entwicklung aktiver Komfort- und Sicherheitsfunktionen im Bereich der Fahrerassistenzsysteme (FAS) gibt es bezüglich der Abbildung des menschlichen Situationsbewusstseins bereits Erfolge bei der Detektion anderer Objekte mittels fahrzeugeigener Sensorik (Ebene 1), sowie bei der kurzzeitigen Prädiktion der Objektposition (Ebene 3) (s. Abb. 1.2, obere Zeile). Auf Basis dieser Prädiktion werden gemeinhin über die Time-to-Collision (TTC) (Lee, 1976) potentielle Kol-

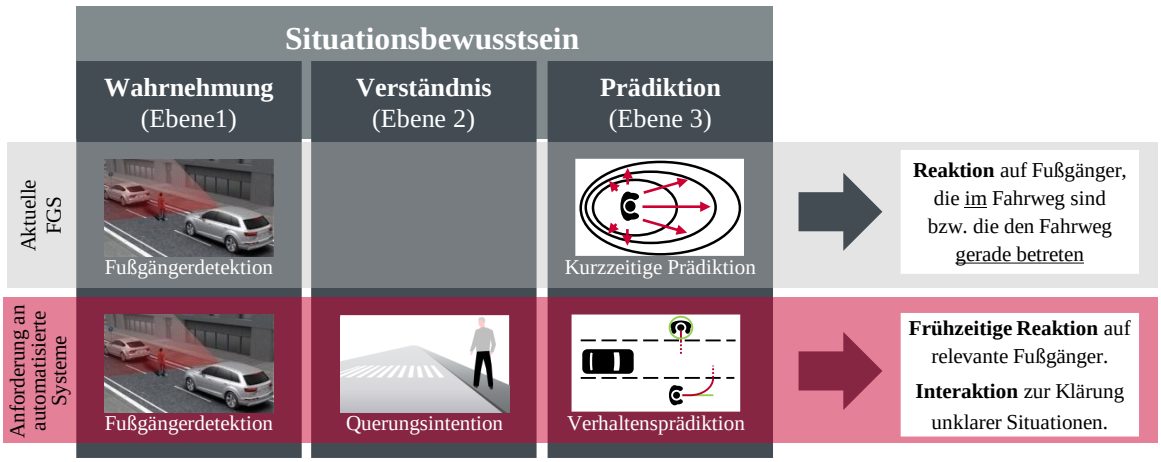


Abbildung 1.2: Vergleich des Situationsbewusstseins aktueller Fahrerassistenzsysteme (FAS) mit den Anforderungen an zukünftige automatisierte Systemen am Beispiel aktueller Fußgängerschutzsysteme (FGS).

lisionsrisiken mit den anderen Verkehrsteilnehmern erkannt, um daraufhin geeignete Maßnahmen zur Kollisionsvermeidung, wie eine Fahrerwarnung oder eine automatische Notbremsung, einzuleiten (Winner et al., 2015). Die bisher zur Prädiktion der Objektpositionen verwendeten Modelle basieren auf den physikalischen Zusammenhängen zwischen den kinematischen Größen Position, Geschwindigkeit und Beschleunigung (Scherf und Zecha, 2009). Dadurch können die zukünftigen Zustände der detektierten Objekte nur durch eine, auf der theoretischen Dynamik der Verkehrsteilnehmer basierenden, Extrapolation des aktuellen Bewegungszustands prädiziert werden. Dieses resultiert in kurzen Vorhersagehorizonten, da die Prädiktionsunsicherheit mit steigenden Zeithorizonten schnell über ein vertretbares Niveau ansteigt (Rehder et al., 2015). Der Handlungsspielraum aktueller FAS ist deswegen auf ein rein reaktives Verhalten begrenzt, was für ein automatisiertes Fahrzeug nicht ausreichend ist. Eine allein auf physikalischen Dynamikparametern basierende Prädiktion genügt daher nicht, um für das automatisierte Fahren in hochdynamischen Bereichen, wie dem städtischen Umfeld, ein akzeptables Fahrverhalten zu generieren.

Abhilfe kann eine auf der Ebene 2 einzuordnende Erkennung der Intention der anderen Verkehrsteilnehmer schaffen (s. Abb. 1.2, untere Zeile). Die Intention wird in der philosophischen Handlungstheorie als Ursache menschlicher Handlung betrachtet

(s. Abschn. 2.1). Ihre Erkennung ermöglicht somit die Prädiktion der zukünftigen Situation (Ebene 3) anhand der zu erwartenden Handlung und den dazu durchzuführenden Aktionen. Zudem kann bei der Prädiktion die wechselseitige Beeinflussung des Handelns der Verkehrsteilnehmer berücksichtigt werden. Im Vergleich zur Verwendung rein physikalischer Bewegungsmodelle können so deutlich längere Prädiktionshorizonte erreicht und der Einfluss des eigenen geplanten Verhaltens auf die anderen Verkehrsteilnehmer einbezogen werden.

Im Gegensatz zu den in bisherigen Fahrerassistenzsystemen verwendeten kinematischen Größen ist die Intention jedoch kein quantitativer Parameter, der direkt über eine Sensormessung erfasst werden kann. Dieses stellt die Entwicklung eines technischen Systems zur Intentionserkennung vor neue Herausforderungen: Die Intention muss im Rahmen einer Operationalisierung zunächst beobachtbar und messbar gemacht werden. Dazu bedarf es eines Modells, das die Erkennung der Intention auf Basis beobachtbarer Größen zulässt. Zudem fordert die nicht direkte Messbarkeit der Intention neue Ansätze zur Referenzbestimmung. Eine besondere Herausforderung bei der Operationalisierung bilden schließlich die Unsicherheiten, mit denen Hypothesen zur Intention der anderen Verkehrsteilnehmer vermutlich behaftet sind. Denn selbst ein menschlicher Fahrer ist sich bei der Einschätzung des Gefahrenpotentials anderer Verkehrsteilnehmer nicht immer sicher (Tsimhoni et al., 2008) und sagt deren zukünftiges Verhalten mitunter falsch vorher (Schmidt et al., 2008; Vollrath et al., 2006). Dennoch ist ein aufmerksamer und erfahrener Fahrer eher selten in kritische Situationen mit anderen Verkehrsteilnehmern involviert (OECD/ITF, 2015), da er sein Fahrverhalten an die Unsicherheit der Situation anpasst (Schneemann und Gohl, 2016).

Im Vergleich zur Vorhersage des Bewegungsverhalten anderer Fahrzeuge oder Radfahrer bildet die Verhaltensvorhersage von Fußgängern eine besondere Herausforderung. Fußgänger können ihre Bewegungszustände sehr dynamisch ändern (Rehder et al., 2015), wodurch die auf Basis kinematischer Parameter erreichbaren Prädiktionshorizonte auf einige Zehntelsekunden begrenzt sind (Kloeden et al., 2014). Aktuelle Fußgängerschutzsysteme (FGS) sind daher nur in der Lage auf Fußgänger zu reagieren, die den Verkehrsraum des Fahrzeugs bereits betreten haben oder gerade dabei sind, dieses zu tun. Daher verspricht vor allem bei Fußgängern die Erkennung der Intention

eine deutliche Verbesserung der Prädiktionsgüte und bildet somit einen wesentlichen Baustein bei der technischen Abbildung des menschlichen Situationsbewusstseins. An diesem Punkt setzt die vorliegende Arbeit an.

Im Hinblick auf das automatisierte Fahren sind für das Fahrzeug nur Verhaltensweisen des Fußgängers relevant, die eine Anpassung des eigenen Fahrverhaltens fordern. Da die Verkehrsräume von Fußgängern und Fahrzeugen in der Regel getrennt sind[1], kann die Erkennung der Fußgängerintention daher auf die Erkennung seiner Absicht, den Fahrstreifen queren zu wollen, beschränkt werden. Diese Absicht wird im Folgenden als Querungsintention (QI) bezeichnet. Eine Klassifikation der Fußgänger in solche mit und solche ohne Querungsintention ermöglicht einem automatisierten Fahrzeug somit die frühzeitige Identifikation der Fußgänger, deren zukünftige intendierte Handlungen für die eigene Verhaltensplanung relevant sind. Hierdurch ist das automatisierte Fahrzeug nicht nur in der Lage frühzeitig auf querungswillige Fußgänger zu reagieren; es kann zudem mit diesen in Interaktion treten, um beispielsweise unklare Situationen zu lösen, indem es durch die eigenen Verhaltensweisen, das Verhalten des querungswilligen Fußgängers positiv beeinflusst (Schneemann und Gohl, 2016). Im Vergleich zu einer reinen Aktionserkennung, wie dem Betreten der Straße durch den Fußgänger (s. Abschn. 2.3), bietet diese Intentionserkennung somit einen deutlichen Mehrwert für die Gestaltung des Verhaltens zukünftiger automatisierter Fahrzeuge.

[1] Ausnahmen bilden verkehrsberuhigte Bereiche und Shared Spaces (Bad architects group, 2012), bei denen die gesamte Verkehrsfläche von Fahrzeugen und Fußgängern gleichermaßen genutzt werden darf.

1.2 Zielsetzung

Ziel dieser Arbeit ist die Entwicklung eines technischen Systems zur Erkennung der Querungsintention von Fußgängern. Die vorliegende Arbeit leistet damit einen Beitrag zur Übertragung des menschlichen Situationsbewusstseins auf ein technisches System und bildet eine Basis für das automatisierte Fahren im städtischen Umfeld. Im Rahmen der Entwicklung des Systems werden zunächst bestehende Modelle zur menschlichen Intention vorgestellt und auf Ihre Anwendbarkeit in einem technischen System zur Intentionserkennung analysiert. Der aus dieser Analyse basierende Systementwurf wird auf bereits vorhandenem Wissen über das beobachtbare Verhalten von Fußgängern zurückgreifen, dessen Ausprägung im Rahmen der Implementierung anhand von Beispielen erlernt wird. Hierzu werden Methoden des maschinellen Lernens angewendet, über die die menschliche Einschätzung einer Situation möglichst genau nachgebildet wird. Das zu entwickelnde System muss dabei auch die bei der Erkennung der Querungsintention potentiell bestehenden Unsicherheiten in der Beobachtung abbilden können.

Damit automatisierte Fahrzeuge unabhängig vom Ausbau infrastrukturseitiger Sensorik zukünftig in allen Bereichen des städtischen Umfelds agieren können, soll das zu entwickelnde System mit fahrzeugeigener Sensorik arbeiten. Vor dem Hintergrund kultureller Einflüsse auf das Fußgängerverhalten, beschränkt sich das im Rahmen dieser Dissertation entwickelte System auf die Anwendung im deutschen Verkehrsraum. Die für das System notwendige Bestimmung physikalisch messbarer Größen, wie die Position und Körperorientierung der Fußgänger, sowie die Position von Fahrstreifenbegrenzungen werden in dieser Arbeit als gelöstes Problem betrachtet.

1.3 Aufbau der Arbeit

Die vorliegende Arbeit gliedert ist in die folgenden Themenbereiche gegliedert: Kapitel 2 gibt einen Überblick über den Hintergrund der Fußgängerintentionserkennung mit Definitionen und Modellen zur menschlichen Intention, Studien zu beobachtbarem Fußgängerverhalten und bereits bekannten Ansätzen zur Erkennung der Fußgängerintention. Kapitel 3 erläutert den zum Verständnis des eigenen Ansatzes er-

forderlichen Hintergrund aus dem Bereich des maschinellen Lernens. Kapitel 4 befasst sich mit der in dieser Arbeit entwickelten, beobachterbasierten Referenzmethode und geht auf die Reliabilität dieser ein. In Kapitel 5 wird das Konzept des zur Erkennung der Querungsintention von Fußgängern entwickelten, eigenen Ansatzes vorgestellt und auf Details der Implementierung eingegangen. Kapitel 6 stellt den zur Evaluation verwendeten Datensatz vor und zeigt die Leistungsfähigkeit des entwickelten Systems, indem die durchgeführte Evaluation des neuen Ansatzes sowie die erreichten Ergebnisse dargestellt und im Detail diskutiert werden. Kapitel 7 fasst die Ergebnisse schließlich zusammen, zieht Schlussfolgerungen und gibt einen Ausblick für zukünftige Arbeiten.

Kapitel 2

Hintergrund der Fußgängerintentionserkennung

Die Erkennung von Fußgängerintentionen ist ein junges, aufstrebendes Forschungsfeld im Bereich der Verhaltenserkennung und -prädiktion. Dieses Kapitel beschäftigt sich mit den Hintergründen und dem Stand der Technik zu diesem Thema. In Abschnitt 2.1 werden hierzu allgemein der Begriff „Intention“ erläutert und bestehende Modelle aus dem Bereich der Handlungstheorie vorgestellt. In Abschnitt 2.2 werden anschließend Erkenntnisse über das Verhalten von Fußgängern beschrieben, die für den Entwurf eines technischen Systems zur Intentionserkennung relevant sind. Abschnitt 2.3 beschäftigt sich mit bisher bekannten Systemen zur Fußgängerverhaltenserkennung und -prädiktion und stellt in diesem Rahmen das bisherige Verständnis des Begriffs „Fußgängerintention“ vor. Abschnitt 2.4 geht anschließend auf bekannte Referenzmethoden zur Bestimmung der Leistung von intentionserkennenden Systemen ein. Da sich die Bewertung des Vorgestellten auf das gesamte Forschungsfeld stützt, erfolgt diese unter dem Ziel, bestehende Forschungsdefizite in Bezug auf die Entwicklung eines Systems zur Erkennung der Querungsintention von Fußgängern aufzudecken, gesamtheitlich in Abschnitt 2.5.

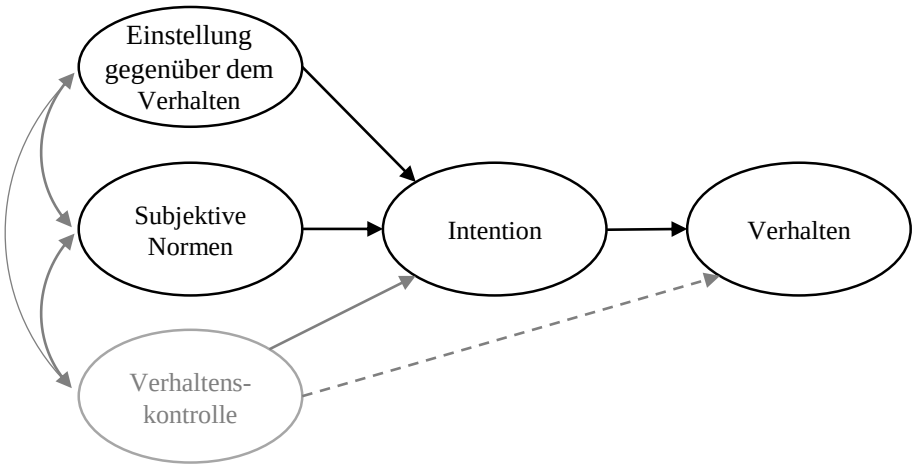


Abbildung 2.1: Die Theorie des überlegten Handelns (schwarz) von Fishbein und Ajzen (1975) und die Erweiterung zur Theorie des geplanten Verhaltens (grau) von Ajzen (1985).

2.1 Definitionen und Modelle der Intention

Der Begriff „Intention“ findet seinen Wortursprung im lateinischen *intendere* (= sein Streben/seine Aufmerksamkeit auf etwas richten) und beschreibt die Absicht eine bestimmte Handlung durchzuführen bzw. ein bestimmtes Ziel zu erreichen (Puca, 2014). Die Intention ist als ein kognitiver, nicht von außen beobachtbarer Zustand oder Prozess zu verstehen (Diederichs, 2017) und wird in der philosophischen Handlungstheorie als Ursache menschlicher Handlung betrachtet (Anscombe, 1957; Davidson, 1963).

2.1.1 Handlungstheoretische Modelle der Intention

Die Kausalität zwischen Intention und Handeln findet sich in der häufig zur Verhaltensvorhersage verwendeten **Theorie des überlegten Handelns** (*Theory of Reasoned Action*) von Fishbein und Ajzen (1975) wieder. Der Theorie zufolge werden Handlungen direkt von Intentionen gesteuert, wobei die Intentionen wiederum aus einer positiven Einstellung gegenüber der Handlung, sowie einer als subjektive Norm bezeichneten, positiv wahrgenommenen Einstellung relevanter anderer Personen bezüglich der Ausführung des Verhaltens entstehen (s. Abb. 2.1). Die Erweiterung der

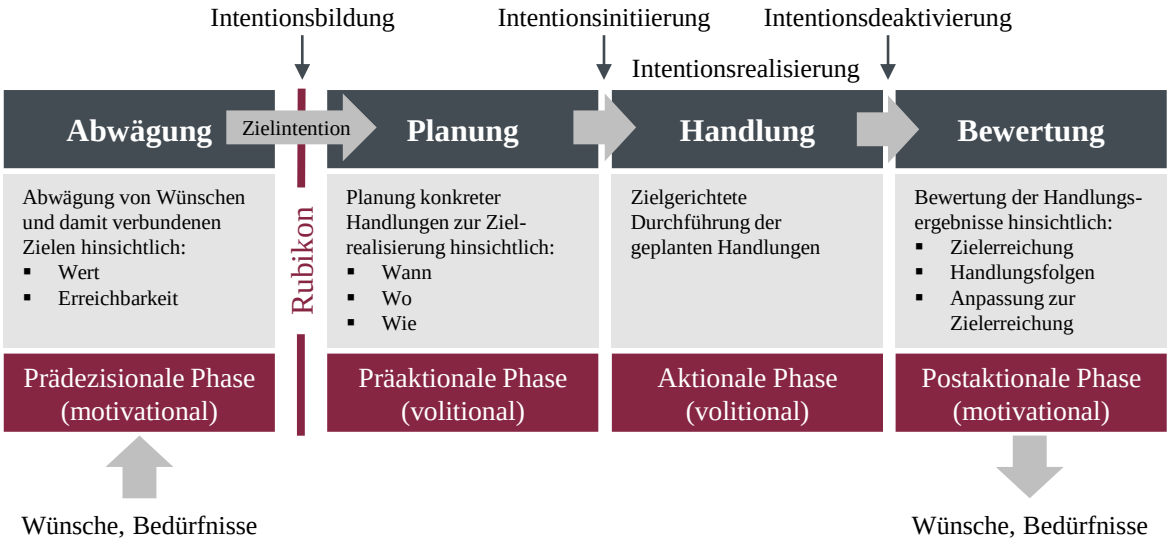


Abbildung 2.2: Das Rubikon-Modell der Handlungsphasen von Heckhausen und Gollwitzer (1987). Darstellung nach (Diederichs, 2017).

Theorie zur **Theorie des geplanten Verhaltens** (*Theory of Planned Behavior*) von Ajzen (1985) führt zusätzlich die subjektiv wahrgenommene Verhaltenskontrolle ein. Diese beschreibt die Überzeugung einer Person, über genug Fähigkeiten und Ressourcen zur Ausführung des geplanten Verhaltens zu verfügen. Die wahrgenommene Verhaltenskontrolle kann die Intention beeinflussen, aber auch das Verhalten selbst. Über die Intention kann somit immer nur der Versuch einer bestimmten Verhaltensausführung vorhergesagt werden, nicht unbedingt die tatsächliche Ausführung. Denn die Verhaltensrealisierung ist von der Person gegebenenfalls nicht kontrollierbar.

Das aus der Motivationstheorie stammende **Rubikon-Modell der Handlungsphasen** von Heckhausen und Gollwitzer (1987) unterstreicht diese Trennung zwischen Intentionsbildung und -realisierung. Das Modell postuliert das Verfolgen eines Ziels über intendierte Handlungen als einen vierphasigen Prozess (s. Abb. 2.2). Die erste und letzte Phase (Abwägung und Bewertung) beschreiben dabei die beiden motivationalen Phasen, die der Zielwahl und Zielsetzung dienen, wohingegen die zweite und dritte Phase (Planung und Handlung) als sogenannte volitionale Phasen die Realisierung dieser Ziele beschreiben. Das Rubikon-Modell unterscheidet damit deutlich zwischen

der Wahl von Handlungszielen einerseits und der Realisierung dieser Ziele andererseits, wobei die beiden Prozesse funktional verknüpft sind (Achtziger und Gollwitzer, 2006).

Im Detail beginnt jede intendierte Handlung mit einer Phase des Abwägens verschiedener Wünsche und damit verbundener Ziele bezüglich ihres Wertes und ihrer Erreichbarkeit. Wird ein Wunsch zur Realisierung ausgewählt, muss dieser zunächst in ein konkretes Ziel umgewandelt werden. Diese Umwandlung wird als das Überschreiten des Rubikons bezeichnet: Analog zu Julius Caesar, der mit dem Überschreiten des Rubikons 49 v. Chr. eine Kriegserklärung implizierte und sich damit unwiderruflich auf einen Bürgerkrieg einließ, wird mit der Umwandlung des Wunsches in ein Ziel eine verpflichtende Zielintention gebildet. Mit der Intentionsbildung erfolgt der Eintritt in die präaktionale Planungsphase, in der konkrete Handlungen zur Realisierung geplant werden. Da viele Ziele nicht unmittelbar nach der Intentionsbildung umgesetzt werden können, müssen Pläne entwickelt werden, die bestimmen wann, wo und auf welche Art und Weise die zielförderlichen Handlungen durchgeführt werden sollen. Liegen die zur Ausführung als günstig definierten Bedingungen vor, beginnt mit der Handlungsinitiierung der Übergang in die aktionale Handlungsphase. In dieser Phase wird versucht, die geplanten Handlungen tatsächlich zu realisieren und zu einem erfolgreichen Ende zu bringen. Schließlich erfolgt in der postaktionalen Bewertungsphase eine Bewertung der Ergebnisse der durchgeführten Handlungen. Bei dieser wird geprüft, inwieweit das ursprünglich gesetzte Ziel tatsächlich erreicht wurde und welche Handlungen ggf. noch auszuführen sind, um den Handlungsverlauf zum gewünschten Abschluss zu bringen (Achtziger und Gollwitzer, 2006).

2.1.2 Beobachtungsbasierte Modelle der Intention

Die vorgestellten Definitionen und Modelle legen nahe, dass sich Intentionen in Handlungen und Verhaltensweisen zeigen. Die obigen, in der Handlungspsychologie etablierten Modelle eignen sich besonders dazu, die kognitive Bildung von Intention und deren Realisierung in Handlungen formal zu beschreiben. Zur Anwendung der Modelle für die Intentionserkennung fehlt jedoch der Bezug zu beobachtbaren Verhaltensweisen, über die auf die Intention zurückgeschlossen werden kann (Diederichs, 2017).

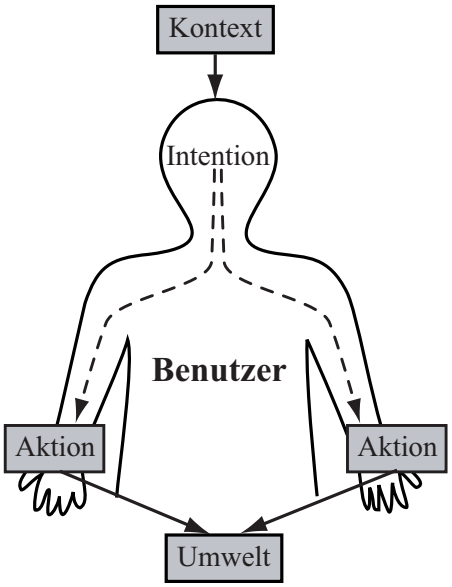


Abbildung 2.3: Menschmodell zur Intentionserkennung von Schrempf (2008).

Ein einfaches Modell, das an diesem Punkt ansetzt, ist das in Abbildung 2.3 gezeigte **Menschmodell zur Intentionserkennung** nach Schrempf (2008). Das Modell beschreibt die Intention als versteckten Zustand im Kopf des Menschen, der von einem relevanten Teil der Umwelt, dem Kontext, beeinflusst wird. Die Intention wird auch hier als Ursache menschlicher Handlung verstanden, die wiederum durch Folgen beobachtbarer Aktionen beschrieben wird. Die Ausführung von Aktionen beeinflusst die ebenfalls beobachtbare Umwelt. Dadurch kann auch bei schwer zu erkennenden Aktionen von der Beobachtung der Umwelt ein Rückschluss auf die ausgeführte Aktion und damit auf die zugrundeliegende Intention gezogen werden.

Diederichs (2017) geht noch einen Schritt weiter und löst die im obigen Menschmodell sehr allgemein gehaltene Beschreibung der menschlichen Handlung detailliert auf. Hierzu überträgt Diederichs das Rubikon-Modell auf die beobachtbaren Verhaltensweisen einer Intention und entwickelt unter Verwendung empirischer Daten das in Abbildung 2.4 gezeigte **Jordan-Modell** zur Erkennung von Fahrerintentionen. Im Jordan-Modell entsteht die Intention aus dem Zusammenspiel zwischen Voraussetzungen, die Handlungen erst ermöglichen und persönlichen Reiz-Reaktionsmustern, die bereits in der Vergangenheit beobachtet wurden und dadurch die Wahrscheinlichkeit

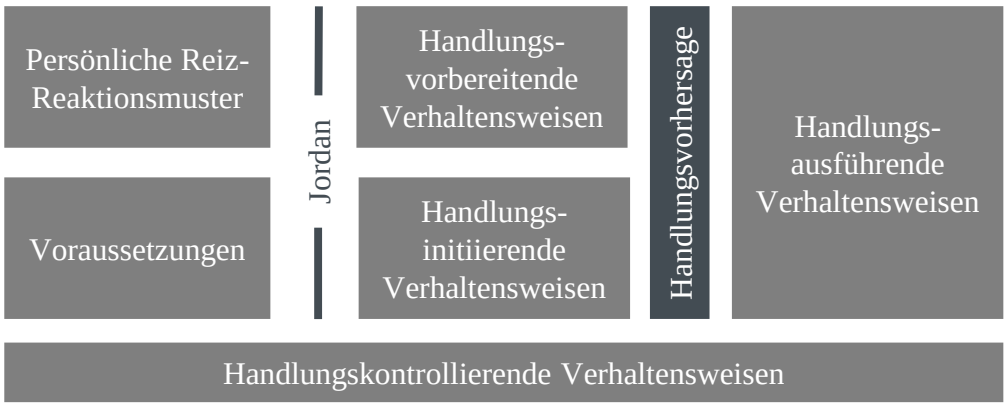


Abbildung 2.4: Das Jordan-Modell der Fahrmanöverintention von Diederichs (2017).

erhöhen, unter gleichen Voraussetzungen erneut diese Handlung auszuführen. Obwohl diese beiden Intentionsauslöser beobachtbare Größen sind, können an dieser Stelle noch keine Intentionen erkannt oder sogar genaue Ausführungszeitpunkte intendierter Handlungen vorhergesagt werden. Denn die Voraussetzungen für eine entsprechende Handlung müssen nicht nur bestehen, sondern sie müssen auch vom Handelnden wahrgenommen werden, was wiederum kein beobachtbarer Prozess ist. Ob und wann tatsächlich eine Intention gebildet wurde und damit der Jordan[1] tatsächlich überschritten wird, ist erst messbar, wenn beobachtbare Verhaltensweisen begonnen werden. Hier unterscheidet das Modell zwischen handlungsvorbereitenden und handlungsinitiierenden Verhaltensweisen, die in einem dynamischen Umfeld zeitgleich oder in vermischter zeitlicher Abfolge auftreten können. Handlungsvorbereitende Verhaltensweisen dienen der Überprüfung der Voraussetzungen zum Handeln, um zu planen wann und wie die Intention realisiert werden kann. Hierzu können im Konkreten beispielsweise gerichtete Blicke und Kopfbewegungen zählen. Handlungsinitiierende Verhaltensweisen bezeichnen hingegen Handlungen, die das bereits geplante Handeln einleiten, wie zum Beispiel das Einnehmen einer handlungsangemessenen Körperhaltung oder im Fall einer Fahrerintention zum Fahrstreifenwechsel, das Setzen des Blinkers. Es folgt der Übergang in die konkrete Handlungsausführung. An diesem Übergang wird die zeitlich

[1] Analog zum Rubikon hat der Jordan eine historische Bedeutung als der Fluss, den das Volk Israel nach dem Auszug aus Ägypten überschreiten musste, um jenseits des Flusses eine neue Heimat zu finden (Diederichs, 2017).

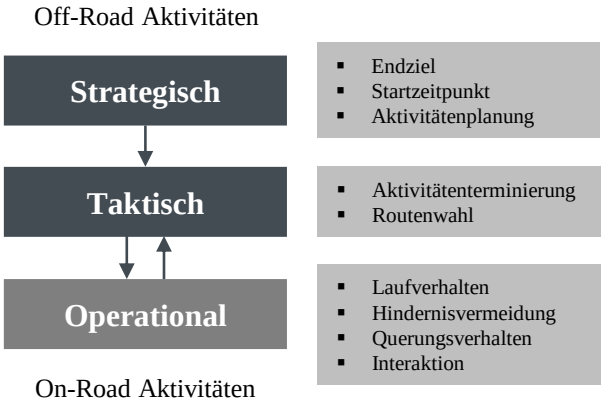


Abbildung 2.5: Hierarchisches Drei-Ebenen-Modell des Fußgängerverhaltens von Hoogendoorn und Bovy (2004). Darstellung nach (Papadimitriou et al., 2009).

noch unspezifische Intention zu einem unmittelbar bevorstehenden Handlungsbeginn, wodurch eine konkrete Handlungsvorhersage auf Basis der Intention an dieser Stelle erfolgen kann. Nach Überschreitung dieser Grenze beginnen die handlungsausführenden Verhaltensweisen. Parallel zu jeder Phase können handlungskontrollierende Verhaltensweisen beobachtet werden, d.h. im Gegensatz zum Rubikon-Modell erfolgt eine Bewertung nicht erst nach Abschluss der Intentionsrealisierung, sondern kontinuierlich über die gesamte Dauer des Prozesses (Diederichs, 2017).

2.2 Beobachtbares Fußgängerverhalten

Das Verhalten von Fußgängern kann nach Hoogendoorn und Bovy (2004) hierarchisch in drei Ebenen eingeteilt werden (s. Abb. 2.5). Die oberste Ebene beschreibt das strategische Verhalten des Fußgängers. Hierzu zählen die Bestimmung des Endziels, des Startzeitpunkts, sowie die Planung der durchzuführenden Aktivitäten. Diese oberste Ebene bezieht sich auf das Entscheidungsverhalten des Fußgängers vor seinem Fußweg und somit vor der Teilnahme am Straßenverkehr (*Off-Road* Aktivitäten). Die mittlere Ebene beschreibt das taktische Fußgängerverhalten. Dieses umfasst die Terminierung der geplanten Aktivitäten, sowie die davon abhängige Routenwahl. Die der taktischen Ebene zuzuordnenden Verhaltensentscheidungen können sowohl vor, als auch während

der Teilnahme am Straßenverkehr (*On-Road* Aktivitäten) gefällt werden. Beispielsweise beeinflusst das Wissen über das Straßennetz die Routenwahl des Fußgängers bereits vor seinem Fußweg, die während des Fußwegs auftretenden Bedingungen können seine Wahl jedoch ändern. Die untere Ebene beschreibt schließlich das operationale Verhalten des Fußgängers bei der Ausführung der Gehaufgabe. Diese Ebene umfasst somit ausschließlich On-Road Aktivitäten. Hierzu zählt das Gehverhalten an sich, die Hindernisvermeidung, das Querungsverhalten des Fußgängers, sowie die Interaktion mit anderen Verkehrsteilnehmern (Papadimitriou et al., 2009). Für die Entwicklung des in dieser Arbeit angestrebten Systems sind vor allem Verhaltensweisen von Fußgängern relevant, die vor der Durchführung einer Querung beobachtbar sind, da diese die Grundlage für die Erkennung der Querungsintention eines Fußgängers bilden. Aus diesem Grund beschränkt sich die folgende Darstellung des beobachtbaren Fußgängerverhaltens auf Verhaltensweisen der operationalen Ebene, mit einem Fokus auf dem Querungsverhalten von Fußgängern.

2.2.1 Theoretische Fußgängerdynamik

Das allgemeinen Gehverhalten auf der operationalen Ebene und die theoretische Dynamik von Fußgängern sind empirisch bereits gut erforscht. Es existieren zahlreiche Studien zur durchschnittlichen Gehgeschwindigkeit des Fußgängers in Abhängigkeit verschiedener Faktoren wie Alter, Geschlecht, Bewegungsart (Eberhardt und Himbert, 1977; Petersen, 2003; Yanqing et al., 2014), Verkehrszweck (Weidmann, 1992) und den Witterungsverhältnissen (Li und Fernie, 2010), sowie zur durchschnittlichen Querungsgeschwindigkeit in Abhängigkeit der Straßengestaltung (Chandra et al., 2013) oder infrastrukturseitiger Querungshilfen (Geruschat et al., 2003; Galanis und Nikolaos, 2012). Im Mittel kann eine Durchschnittsgeschwindigkeit von 1,34 m/s ($\hat{=}$ 4,82 km/h) festgestellt werden (Weidmann, 1992), wobei jedoch eine starke Streuung zwischen 0,7 m/s und 1,8 m/s zu beobachten ist (Schnabel und Lohse, 2011). Die Untergrenze bilden hierbei Kinder in Begleitung von Erwachsenen, sowie alte und in der Mobilität eingeschränkte Personen. Die Obergrenze des Spektrums besetzen Jugendliche und junge Erwachsene. In den meisten Altersgruppen queren Frauen langsamer als Männer

und Individuen schneller als Personen in Gruppen (Hagen et al., 2010). Diese Werte spiegeln sich auch in den deutschen Regelwerken zur Gestaltung von Straßenverkehrsanlagen wieder. Das *Handbuch für die Bemessung von Straßenverkehrsanlagen* (FGSV, 2015) geht bei den standardisierten Berechnungsverfahren beispielsweise auch von einer durchschnittlichen Fußgängergeschwindigkeit von 1,34 m/s aus und die *Richtlinien für Lichtsignalanlagen* (FGSV, 2010) empfiehlt für Fußgängerampeln eine Räumungsgeschwindigkeit von 1,2 m/s mit Variationen von 1,0 m/s bis höchstens 1,5 m/s.

Neben der Betrachtung der Gehgeschwindigkeit wurde zur Bestimmung der theoretischen Dynamik eines Fußgängers auch seine mögliche Beschleunigung untersucht. Laut Tiemann (2012) liegt diese in Abhängigkeit vom Ausgangszustand (stehen, gehen, laufen, rennen) zwischen 1,6 m/s^2 und 3 m/s^2. Auch hier zeigt sich eine hohe Varianz innerhalb der ermittelten Werte. Zudem konnte beobachtet werden, dass Fußgänger bereits beim ersten Schritt auf ihre durchschnittliche Gehgeschwindigkeit beschleunigen (Breniere und Do, 1986).

2.2.2 Querungsverhalten

Das Querungsverhalten von Fußgängern umfasst im engeren Sinne sowohl die Querung an sich, als auch deren Vorbereitung (s. Abb. 2.6). Die Querungsvorbereitung kann weiter in eine konzeptionelle Phase unterteilt werden, die der Planung und Auswahl einer bestimmten Querungsstelle dient, sowie in die unmittelbare Querungsvorbereitung, die der Sicherung der Querung dient (Hagen et al., 2010). Die eigentliche Querung wird im Folgenden weiter in den Querungsbeginn, sowie die Querungsdurchführung unterteilt. Noch bevor ein Fußgänger die Querung durch das Betreten der Straße beginnt, lässt vor allem das bei der Querungsvorbereitung beobachtbare Sicherungsverhalten Rückschlüsse auf die Querungsintention des Fußgänger zu. Aus diesem Grund ist dieser Teil des beobachtbaren Fußgängerverhaltens für das im Rahmen dieser Arbeit zu entwickelnde System von besonderer Bedeutung und wird daher im Folgenden fokussiert behandelt.

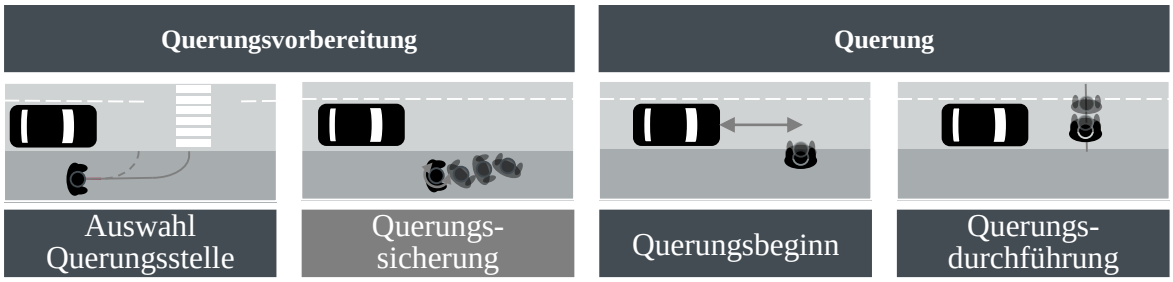


Abbildung 2.6: Differenzierung bei der Beschreibung des Querungsverhaltens von Fußgängern.

Auswahl der Querungsstelle

Die Auswahl einer Querungsstelle ist zunächst von der auf der taktischen Ebene erfolgten Routenwahl abhängig, wobei die Routenwahl bereits von den vorhandenen Querungsmöglichkeiten beeinflusst ist (s. Abschn. 2.2). Bei der finalen Wahl einer konkreten Querungsstelle spielen vor allem Faktoren der Fußwegbeziehung selbst, wie die Weglänge und die Lage von Start- und Zielpunkt (Hagen et al., 2010), sowie das Vorhandensein und die Attraktivität von Querungsalternativen eine Rolle. Die Attraktivität einer Querungsstelle ist hierbei von dem Verkehrsaufkommen, der Sicherheit, dem zur Querung benötigten Zeitbedarf, sowie der Vorhersagbarkeit der sich entwickelnden Verkehrssituation abhängig (Chu et al., 2004; Montel et al., 2013). Zudem haben persönliche Faktoren wie Alter, Geschlecht, Risikotoleranz und Gesundheit einen großen Einfluss auf die Wahl des konkreten Querungsortes (Hagen et al., 2010).

Unabhängig von persönlichen Faktoren weisen Querungshilfen wie Fußgängerampeln, Zebrastreifen oder Fußgängerinseln für alle Fußgänger eine hohe Attraktivität auf. Bei Studien in der Schweiz (Schweizer et al., 2009) und in den USA (Nee und Hallenbeck, 2003) nutzten über 96 % der beobachteten Fußgänger die angebotenen Querungshilfen. Auch für den deutschen Verkehrsraum konnte Roehder (2011) eine Konzentration von querenden Fußgängern an Fußgängerüberwegen beobachten. Zudem konnte Roehder zeigen, dass die genaue Eintrittsstellen in die Straße einer am Zebrastreifen zentrierte Normalverteilung entspricht. Bei der Nutzung von Querungshilfen ist jedoch eine große Umwegempfindlichkeit seitens der Fußgänger zu erkennen (Hagen et al., 2010). Querungsangebote außerhalb kürzester Gehwegverbindungen werden sehr viel seltener

genutzt als solche in direkter Gehwegrelation (Sisiopiku und Akin, 2003). Dieses gilt jedoch weniger für mobilitätseingeschränkte Personen. Diese bevorzugen das Queren an gesicherten Querungsstellen und nehmen dafür auch längere Umwege in Kauf als gesunde Personen (Schilde, 2007).

Auch ein menschlicher Beobachter scheint entsprechende Erfahrungen über die Nutzung von Querungshilfen bei seiner Einschätzung bezüglich der Querungsintention eines Fußgängers mit einzubeziehen. So zeigt die Studie von Schmidt und Färber (2009) einen Einfluss der Präsenz eines Fußgängerüberwegs auf die Einschätzung der Intention eines Fußgängers.

Auf freier Strecke ist die Auswahl der konkreten Stelle zur Querung neben den oben beschrieben Faktoren der Fußwegbeziehung, maßgeblich von der Möglichkeit zur adäquaten Querungssicherung, wie etwa ausreichender Sichtverhältnisse, beeinflusst (Hagen et al., 2010). Zudem hat auch die Gestaltung der Umgebung, vor allem der Gehwege einen deutlichen Einfluss auf die Auswahl der Querungsstelle (Montel et al., 2013). Ist das Gehen auf dem Gehweg nicht ausreichend gut möglich, queren Fußgänger auch an unsicheren oder ungünstigen Stellen, um ihren Weg auf einem akzeptablen Gehweg fortzusetzen. Doch vor allem bei der Querung auf freier Strecke bleibt neben den vielen bekannten Faktoren stets ein bedeutsamer Anteil an Zufall, der bei der Wahl der Querungsstelle zwischen den bestehenden Alternativen situativ entscheidet (Hagen et al., 2010).

Sicherungsverhalten

Das als allgemeingültig anzusehende Ziel des Fußgängers bei der Querung ist die sichere Überquerung der Straße bei Unversehrtheit von Leib und Leben. Aus diesem Grund sichern Fußgänger die eigentliche Querung vor Querungsbeginn durch ein bewusst ausgeführtes Sicherungsverhalten, als auch durch unbewusste Verhaltensweisen ab. Die Sicherung dient dabei der Wahrnehmung der aktuellen Situation zur Abschätzung, ob und wann die Straße sicher überquert werden kann. Die bei der Sicherung aufgenommenen Informationen können sowohl visuell, als auch akustisch gespeist werden. Der akustische Anteil des Sicherungsverhaltens ist jedoch grundsätzlich nicht beobachtbar

(Hagen et al., 2010) und daher für die Entwicklung des in dieser Arbeit angestrebten Systems nicht relevant.

Die genaue Ausprägung des gezeigten Sicherungsverhaltens ist stark situationsabhängig. Die für den Fußgänger komplexeste Situation stellt die Querung auf freier Strecke dar. Hier muss der Fußgänger zunächst den subjektiven Querungsaufwand abschätzen und anschließend über die Wahrnehmung des momentanen Fahrzeugstroms verfügbare Lücken auf ihre Eignung zur sicheren Querung bewerten (Hagen et al., 2010). Die zur Verkehrserziehung beispielsweise im Green Cross Code (RoSPA, 2014) vorgegebenen Verhaltensrichtlinien zur Straßenquerung sehen in solchen Situationen eine visuelle Querungssicherung mit mindestens drei Blicken vor: einen ersten Blick nach links, einen Zweiten nach rechts und einen dritten Blick wieder nach links (bei Rechtsverkehr). Nach Schoon (2006) benötigt ein Fußgänger zur Ausführung einer solchen visuellen Querungsvorbereitung zwischen 2,4 s und 3,2 s. Dass die meisten Erwachsenen ein den Empfehlungen entsprechendes Blickverhalten zur Querungssicherung durchführen, zeigt die Studie von Wilson und Grayson (1980). In dieser konnten bei erwachsenen Fußgänger durchschnittlich 2,4 bis 3,0 Kopfbewegungen vor dem Beginn der Querung beobachtet werden. Kinder zeigen mit 1,2 bis 2,9 Kopfbewegungen hingegen nur ein reduziertes visuelles Absicherungsverhalten (Grayson, 1975). Das bestätigt auch die Eye-Tracking Studie von Egan et al. (2008), nach der erwachsene Fußgänger ein aufgabenabhängiges Blickverhalten mit vermehrter Fixation nach links und rechts aufweisen. Kinder zeigten sich hingegen von für die Querung irrelevanten Informationen, wie Gebäuden und Bäumen leicht abgelenkt, was zu einem weniger eindeutigen Blickverhalten führt.

Die obigen Angaben zur Absicherung mit drei Blicken gelten für die Querung aus dem Stand. Die Untersuchungen von Kloeden et al. (2014) zeigen zusätzlich, dass Fußgänger sich bereits vor dem Erreichen der Bordsteinkante, noch während des Gehens, visuell absichern und sich die Kopfpositionsdaten eines querungswilligen Fußgängers schon 3,8 s vor dem Abbiegen in Richtung Bordstein signifikant von denen eines Fußgängers ohne Querungsintention unterscheiden. Hamaoka et al. (2013) konnten zudem beobachten, dass bei einer Querung aus der Bewegung die Kopfdrehfrequenz gleichmäßig ansteigt, je näher der Fußgänger an die gewählte Querungsstelle kommt.

Die visuelle Sicherung mit einer aufgabenbezogenen Blickführung ist zudem das am bewusstesten ausgeführte querungsvorbereitende Verhalten. Das zeigen die Ergebnisse verschiedener Fußgängerbefragungen. Sowohl die von Sullman et al. (2011) befragte Gruppe jugendlicher Fußgänger, als auch die Gruppen nicht blinder Personen in (Hagen et al., 2010) nannten am häufigsten „Blicke nach beiden Seiten der Fahrbahn“ als sicherheitsbezogenes Querungsverhalten.

Unterbewusst weist ein Fußgänger bei der Querungsvorbereitung jedoch noch mehr typische Verhaltensweisen auf, als nur die Kopf- und Blickbewegungen. Wie die von Schmidt et al. (2008) durchgeführte Beobachterstudie zeigt, bezieht ein menschlicher Beobachter zur Abschätzung der Fußgängerintention die gesamte Körpersprache des Fußgängers, sowie Eigenschaften der Bewegungsdynamik mit ein. Hinzu kommen Kontextinformationen, wie die Bewegungsrichtung in Bezug zum Straßenverlauf, sowie der Abstand zur Straße. Die Studie zeigt weiterhin, dass Dynamik- und Kontextinformationen alleine, ohne Wissen über die Körpersprache, zu deutlich schlechteren Vorhersagen der Fußgängerintention führen.

In Situationen, in denen der Fußgänger durch einen Fußgängerüberweg oder eine grün geschaltete Fußgängerampel explizit Vorrang bei der Querung hat, weicht das beobachtbare Sicherungsverhalten von dem bei einer freien Querung gezeigten Verhalten ab. Ein Fußgänger muss durch sein Vorrangrecht in solchen Situationen nicht aktiv nach einer Lücke im Fahrzeugstrom suchen. Daher zeigt sich in dem Blickverhalten des Fußgängers eher eine Rückversicherung, ob das heranfahrende Fahrzeug anhält, sowie der Versuch, Blickkontakt mit dem Fahrer aufzunehmen (Sullman et al., 2011; Schneemann und Gohl, 2016). Bei Querungen an Fußgängerinseln sind die Kopfbewegungen eines Fußgängers hingegen ähnlich zu denen bei Querungen auf freier Strecke. Obwohl sich ein Fußgänger bei einer Fußgängerinsel zunächst nur gegen den Verkehr von einer Seite absichern muss, schauen 89 % der Fußgänger vor dem Überqueren des ersten Fahrstreifens in beide Richtungen (Schweizer et al., 2009). Allgemein kann bei der Querung an Stellen mit Querungshilfen bereits bei der Bewegung zum Bordstein eine Fixation des Blicks auf Elemente der Querungshilfe, wie die Zebrastreifenlinien (Geruschat et al., 2003) oder das Grünsignal bei Ampelanlagen (Hatfield und Murphy, 2007), beobachtet werden. Vor allem bei der Querung an Fußgängerüberwegen

mit Zebrastreifen wird Fußgängern die Verwendung von Gesten zur Signalisierung ihres Querungswunsches empfohlen (Limbourg, 2011). Schweizer et al. (2009) zeigen jedoch, dass die Anzeige einer Querungsintention hauptsächlich durch indirekte Kommunikation stattfindet. So konnten Schweizer et al. in nur 0,4 % der untersuchten Fußgänger-Fahrer-Interaktionen Handzeichen seitens des Fußgängers beobachten, die der Anzeige seiner Querungsintention und dem Stoppen des sich nähernden Fahrzeugs dienten.

Querungsbeginn

Schätzt ein Fußgänger die aktuelle Verkehrssituation auf Basis der durchgeführten Sicherung als für eine Querung ausreichend sicher ein, beginnt die eigentliche Querung mit dem Betreten der Straße. Bei der freien Querung schätzt der Fußgänger zur Bewertung der aktuellen Situation ab, ob eine gegebene Lücke im Fahrzeugstrom zur Querung ausreicht. In zahlreichen Studien konnte bereits gezeigt werden, dass die Lückenakzeptanz sowohl von subjektiven Parametern des Fußgängers, wie Alter und Geschlecht (Oxley et al., 2005), sowie einem Führerscheinbesitz (Holland und Hill, 2007) abhängig ist, als auch von situativen Parametern, wie der Laufrichtung (Schmidt und Färber, 2009), der Gruppendynamik (Dipietro und King, 1970), der Geschwindigkeit (Schmidt und Färber, 2009) und der Größe der sich nähernden Fahrzeuge (Kotte und Pütz, 2016), sowie von der bereits abgewarteten Zeit (Hamed, 2001). In der Regel wird die Größe einer Lücke in Form der Time-to-Collision (TTC) zu dem sich nähernden Fahrzeug operationalisiert. Die kritischen Grenzen bei der Akzeptanz von Verkehrslücken liegen bei einer TTC von 2–3 s bzw. 7–8 s (Das et al., 2005; Schmidt und Färber, 2009). Bei Lücken mit einer TTC unter 2–3 s beginnt niemanden eine Querung; Lücken mit einer TTC über 7–8 s werden von jedem als sicher akzeptiert.

Neben der reinen Größe einer Lücke kann beobachtet werden, dass die Ausführung der Querung stark von der Interaktion mit dem sich nähernden Fahrzeug beeinflusst wird. So zeigt die Studie von Kadali und Vedagiri (2013), dass die durchschnittliche TTC bei der Akzeptanz von Verkehrslücken um 30 % sinkt, wenn der Fahrer durch verlangsamen seines Fahrzeugs anzeigt, dem Fußgänger Vorrang zu gewähren. Entsprechende Ein-

flussnahmen des Annäherungsverhaltens des Fahrzeugs auf die Querungsentscheidung des Fußgängers konnten auch von Schneemann und Gohl (2016) beobachtet werden.

An durch Ampelanlagen geregelten Querungsstellen beginnt die Querung in der Regel mit dem Umschalten des Lichtsignals auf Grün. Allgemein erzeugt das Rotsignal einer Ampel hohe Compliance. Bei der in Griechenland durchgeführten Studie von Galanis und Nikolaos (2012) konnte beispielsweise beobachtet werden, dass 85 % der Fußgänger auf Grün warten, um mit der Querung zu beginnen. Rotsünder treten nach der Studie am häufigsten bei geringem oder langsamen Verkehrsaufkommen auf. Zudem konnte eine erhöhte Quote an Rotsündern bei Ampelanlagen in Einbahnstraßen und bei Ampelanlagen mit Mittelinseln beobachtet werden (Cambon de Lavalette et al., 2009). Neben den infrastrukturseitigen Einflussfaktoren hat die Gruppendynamik einen bedeutsamen Einfluss auf das Warteverhalten an Ampeln. Bei Studien in Finnland (Himanen und Kulmala, 1988) sowie in Italien (Rosenbloom, 2009) konnte ein deutlicher Anstieg der Wahrscheinlichkeit, dass ein Fußgänger über Rot geht beobachtet werden, wenn ein anderer Fußgänger bereits bei Rot gegangen ist.

Verhalten während der Querung

Nach der deutschen Straßenverkehrs-Ordnung (StVO) haben Fußgänger „Fahrbahnen unter Beachtung des Fahrzeugverkehrs zügig auf dem kürzesten Weg quer zur Fahrtrichtung zu überschreiten“ (§ 25, Abs. 3 StVO). Diese geforderte rechtwinklige Querung der Fahrbahn findet in der Praxis vor allem an Zebrastreifen (Schweizer et al., 2009) und Fußgängerinseln (Nee und Hallenbeck, 2003) statt. Bei freier Querung führt die Umwegempfindlichkeit der Fußgänger oft zu einem eher diagonalen Laufweg (Nee und Hallenbeck, 2003). Dass eine Querung auf freier Strecke trotz Absicherung prinzipiell als ein Risiko empfunden wird, zeigt sich an einer erhöhten Gehgeschwindigkeit im Vergleich zur normalen Durchschnittsgeschwindigkeit beim Gehen (Jain et al., 2014) oder beim Queren an Fußgängerampeln (Galanis und Nikolaos, 2012) sowie Mittelinseln (Geruschat et al., 2003). Vor allem bei Kindern ist zu beobachten, dass jedes vierte bis fünfte Kind über die Straße rennt (Schweizer et al., 2009; Li et al., 2013), was ein hohes Unfallrisiko darstellt.

Neben dem Sicherungsverhalten vor dem Querungsbeginn, sichert sich ein Fußgänger auch während der Querung weiter ab. So haben Wilson und Grayson (1980) beispielsweise während dem Überqueren der Straße 3,2 bis 3,9 Kopfdrehungen seitens des Fußgängers beobachtet. Auch die von Sullman et al. (2011) befragten Jugendlichen gaben an, während der Querung sich weiter umzuschauen und auf den Verkehr zu hören, bis sie auf der anderen Seite angekommen sind.

Allgemein kann beobachtet werden, dass ein Fußgänger, der eine Querung einmal begonnen hat, diese konsequent durchzieht und nicht zurück geht. In den von Hagen et al. (2010) durchgeführten Interviews gaben die Fußgänger an, dass sie konsequent weiter die Straße überqueren, auch wenn sich bei der Querung Unsicherheiten über die Situation ergeben. Nur wenige beschreiben, dass sie bei Unklarheiten die Querung abbrechen und zurückgehen.

2.3 Erkennung und Vorhersage von Fußgängerverhalten

Die Fußgängerintentionserkennung ist ein Teilgebiet der operationalen Verhaltenserkennung und eng mit dem Gebiet der Verhaltensvorhersage verknüpft. Nachdem empirische Forschungen gezeigt haben, dass die theoretische Dynamik eines Fußgängers eine langfristige Vorhersage der Fußgängerposition allein auf Basis physikalischer Dynamikparameter unmöglich macht (s. Abschn. 2.2.1), beschäftigt sich die Wissenschaft seit etwa fünf Jahren zunehmend mit der Erkennung und Vorhersage komplexerer Verhaltensmuster. In diesem Rahmen wird auch der Begriff „Fußgängerintentionserkennung“ (*Pedestrian Intention Recognition*) immer wieder verwendet. Dabei variiert das Begriffsverständnis jedoch stark, so dass auch Ansätze zur Trajektorienprädiktion (*Trajectory/Path Prediction*) oder Aktionserkennung (*Action/Behavior Recognition*) oft als Intentionserkennung bezeichnet werden. Daher muss zur Erarbeitung des Stand der Technik der Fußgängerintentionserkennung das gesamte Feld der Verhaltenserkennung und Verhaltensvorhersage betrachtet werden und eine Differenzierung der Begrifflichkeiten der einzelnen Teilaspekte erfolgen.

2.3.1 Trajektorienprädiktion

Die Trajektorie eines Fußgängers beschreibt den zeitabhängigen Verlauf seiner Position. Somit wird bei der Trajektorienprädiktion die Position des Fußgängers für jeden Zeitpunkt innerhalb eines relevanten zukünftigen Zeithorizonts vorhergesagt (Bonnin et al., 2014b). Abhängig von der Länge des Vorhersagehorizonts kann zwischen einer Kurzzeit-Trajektorienprädiktion (*Short-Term Path Prediction*) und einer Langzeit-Trajektorienprädiktion (*Long-Term Path Prediction*) unterschieden werden (Rehder et al., 2015). Zudem kann die Langzeit-Trajektorienprädiktion, in Abhängigkeit von der verwendeten Methodik, weiter in eine pfadbasierte und eine zielgerichtete Prädiktion unterteilt werden.

Kurzzeit-Trajektorienprädiktion

Bei der Kurzzeit-Trajektorienprädiktion wird die bisherige Bewegung des Fußgängers in der Regel über rekursive Schätzverfahren modellbasiert extrapoliert (Rehder et al., 2015). Als Standardverfahren kommen hier vor allem der Kalman-Filter (Kalman, 1960; Bar-Shalom et al., 2001) sowie seine Variationen zum Einsatz (Schneider und Gavrila, 2013). Die Länge der vertretbaren Vorhersagehorizonte ist somit explizit von der Dynamik der zu prädizierenden Größe abhängig. Dadurch ist die Kurzzeit-Trajektorienprädiktion bei den theoretisch hochdynamischen Fußgängern auf einen Vorhersagehorizont von 0,5 s bis maximal 2 s begrenzt (Schneider und Gavrila, 2013; Kloeden et al., 2014). Selbst bei der Verwendung eines situationsspezifischen Dynamikmodells, das beispielsweise wie in (Goldhammer et al., 2013) auf einen rechtwinklig querenden Fußgänger angepasst ist, liegt der Prädiktionsfehler für einen Zeithorizont von 2,4 s bereits bei 26 cm. Die Kurzzeit-Trajektorienprädiktion bildet aktuell noch die Grundlage der heutigen, reaktiv auf die TTC reagierenden, Fußgängerschutzsysteme (Tiemann, 2012). Neue Ansätze der Forschung verbessern die Prädiktionsgüte durch die Verwendung zustandspezifischer Bewegungsmodelle in Kombination mit der Erkennung einer Bewegungszustandsänderung. Solche hybriden Ansätze werden in dieser Arbeit der Kategorie der Aktionserkennung zugeordnet und im Abschnitt 2.3.2 vorgestellt.

Pfadbasierte Langzeit-Trajektorienprädiktion

Ansätze zur Langzeit-Trajektorienprädiktion basieren auf der Beobachtung, dass Fußgänger zielabhängig, typischen Pfaden folgen (vgl. Abschn. 2.2.2). Bei der pfadbasierten Trajektorienprädiktion werden diese Pfade über die Beobachtung eines Szenarios, unter Zuhilfenahme eines Fußgängertrackingsystems, a priori bestimmt. Zur Prädiktion werden schließlich die bekannten Pfade mit der bisher beobachteten Trajektorie eines Fußgängers verglichen, um anschließend seine zukünftige Position anhand des ähnlichsten typischen Pfads vorherzusagen (Bonnin et al., 2014b). Für die Trajekorienprädiktion können so Vorhersagehorizonte von mehreren Sekunden erreicht werden (Rehder et al., 2015). Dieses Verfahren findet bisher vor allem Anwendung in Bereichen mit stationären Videokameras (Chen und Yung, 2009; Yi et al., 2015); eine Anwendung mit fahrzeugeigener Sensorik ist jedoch auch möglich.

Dieses wird in (Roehder, 2011) gezeigt. Das vorgestellte System basiert auf einer a priori Beobachtung von Verkehrsstellen. Bei dieser werden die Bewegungen der Fußgänger aus der Vogelperspektive aufgenommen und als zusammenhängende Bewegungspfade abgelegt. Nach der Beobachtungsphase werden ähnliche Bewegungsabläufe verschiedener Fußgänger zu einem charakteristischen Pfadsegment gruppiert und durch einen Polygonzug modelliert. Die Stützstellen des Polygonzugs beinhalten dabei sowohl die örtliche Standardabweichung der gruppierten Bewegungspfade, als auch die mittlere Geschwindigkeit der Fußgänger mit zugehöriger Standardabweichung. Roehder wendet dieses Verfahren an drei verschiedenen Verkehrsstellen mit jeweils einem Fußgängerüberweg an. Pro Szenario ergaben sich 6 bis 12 charakteristische Pfadsegmente. In der Anwendungsphase wird schließlich unter Verwendung eines Maximum Likelihood Classifier (MLC) jedem, über die Fahrzeugsensorik detektierten Fußgänger das charakteristischste Pfadsegmente zugeordnet. Hierdurch können 1,4 s bevor ein querungswilliger Fußgänger die Straßenkante erreichen würde, bis zu 80 % der Fußgänger richtig querenden bzw. nicht querenden Pfadsegmenten zugeordnet werden.

Zielgerichtete Langzeit-Trajektorienprädiktion

Die zielgerichtete Langzeit-Trajektorienprädiktion basiert auf der Annahme, dass Fußgänger sich aus der Absicht heraus bewegen, ein bestimmtes örtliches Ziel zu erreichen, und dass der zu diesem Ziel geplante Pfad von der Umgebung beeinflusst ist (Rehder et al., 2015). Bei der zielgerichteten Trajektorienprädiktion wird daher der Einfluss einzelner Umgebungsfaktoren sowie die Positionierung der möglichen Fußgängerziele a priori bestimmt. Da das konkrete Ziel des einzelnen Fußgängers bei der Prädiktion jedoch nicht bekannt und auch nicht direkt beobachtbar ist, wird es in der Regel in Form einer Wahrscheinlichkeitsverteilung über alle möglichen Ziele geschätzt. Die Verteilung wird dabei zu jedem Zeitschritt über einen Vergleich der bisher beobachteten Trajektorie mit den, zu den Zielen modellbasiert geplanten Trajektorien aktualisiert. Im Vergleich zur pfadbasierten Trajektorienprädiktion ist die Generalisierungsfähigkeit der zielgerichteten Trajektorienprädiktion deutlich höher, da der Einfluss einzelner Szenenelemente nicht nur implizit in den beobachteten Pfaden inbegriffen ist, sondern explizit modelliert wird und somit leichter auf andere Szenarios übertragbar ist (Kitani et al., 2012).

Kitani et al. (2012) wenden ihren zielgerichteten Ansatz zur Prädiktion von Fußgängerpositionen auf ein über stationäre Kameras beobachtetes Parkplatzszenario an. Kitani et al. gehen in diesem Szenario von drei möglichen Intentionen des Fußgängers aus: *„zum Auto gehen“*, *„vom Auto weggehen“* und *„durchlaufen“*. Dazu werden jeder Intention örtliche Ziele auf einer diskretisierten Karte zugeordnet. Die Vorhersage der Fußgängertrajektorie wird schließlich über einen Markow-Entscheidungsprozess (Markov Decision Process (MDP)) umgesetzt, bei dem die einzelnen Zellen der Karte die Zustände des MDP darstellen. Um den Einfluss von Umgebungselementen wie Bürgersteigen, Grasflächen, geparkten Autos und Gebäuden auf den Fußgängerpfad zu berücksichtigen, werden die Ergebnisse eines Detektionsalgorithmus zur Erkennung dieser Elemente in die Kostenfunktion des MDP eingebunden. Kitani et al. können die theoretische Generalisierungsfähigkeit der zielgerichteten Ansätze bestätigen, indem sie zeigen, dass ein auf einem Szenario trainiertes MDP auch bei der Übertragung auf neue Szenarien vergleichbar gute Ergebnisse bei der Trajektorienprädiktion liefert.

Abbildung 2.7: Ergebnisse verschiedener zielgerichteter Ansätze zur Langzeit-Trajektorienprädiktion. Von links nach rechts: (Kitani et al., 2012), (Karasev und Soatto, 2016), (Rehder et al., 2015).

Während der Ansatz in (Kitani et al., 2012) auf den Daten stationärer Kameras basiert, stellen Karasev und Soatto (2016) ein zielgerichtetes Verfahren vor, das auf Basis fahrzeugeigener Sensorik und Kartendaten arbeitet. Auch in diesem Ansatz wird das Bewegungsverhalten von Fußgängern über ein MDP modelliert, bei dem der Zustand eines Fußgängers jedoch über die Position, Orientierung und Geschwindigkeit des Fußgängers beschrieben wird, sowie über sein von außen nicht beobachtbares Ziel. Die möglichen Ziele eines Fußgängers werden dabei an die Ränder des jeweils betrachteten Kartenausschnitts, sowie an die Eingänge von Gebäuden gesetzt. Den Einfluss der Umgebung auf die Fußgängerbewegung integrieren Karasev und Soatto über die Belohnungsfunktion des MDP. Hierzu wird jedem Punkt in der Umgebung ein semantisches Label wie beispielsweise *„Bürgersteig“*, *„Straße“*, *„Gebäude“* oder *„Gras “* zugewiesen, das wiederum mit einem kategorieabhängigen Belohnungswert verknüpft ist. Zudem zeigen Karasev und Soatto auf, wie auch der Einfluss dynamischer Umgebungselemente berücksichtigt werden kann. So integrieren sie die dynamische Zustandsänderung einer Ampel beispielsweise über eine adaptierte Version des Hidden Markov Model (HMM) in die Strategie π des MDP. Um zu berücksichtigen, dass Fußgänger nicht ausschließlich rational handeln und somit nicht immer dem „optimalen“ Pfad folgen, wird schließlich die Verwendung einer stochastischen Strategie vorgeschlagen, über die multiple Hypothesen über das Fußgängerverhalten betrachtet werden können. Die Evaluation des Systems zeigt, dass es bei einem Vorhersagehorizont von beispielsweise 10 s mit einem Prädiktionsfehler von 4 m die Trajektorie eines Fußgängers in verschiedenen innerstädtischen Szeanrios deutlich genauer vorhersagen kann, als klassische Ansätze, die

keine semantischen oder dynamischen Umgebungsinformationen verwenden. Zudem wird gezeigt, dass die Verwendung der Körperorientierung des Fußgängers zu einer deutlich schnelleren Erkennung der intendierten Ziele des Fußgängers führt und damit der Prädiktionsfehler frühzeitig reduziert wird.

Die Verwendung von MDPs ist im allgemeinen ein beliebter Ansatz, um das zielgerichtete Bewegungsverhalten von Fußgängern und die dabei herrschende Unsicherheit über das intendierte Ziel eines Fußgängers zu modellieren. Bandyopadhyay et al. (2013) gehen noch einen Schritt weiter und stellen einen Mixed Observable Markov Decision Process (MOMDP) vor, der die zielgerichtete Trajektorienprädiktion mit der Bewegungsplanung eines Fahrzeugs verknüpft. Dadurch kann im Fahrzeug die Unsicherheit über das Ziel des Fußgängers direkt bei der eigenen Bewegungsplanung berücksichtigt werden. Durch die Aufteilung des Zustandsraumes in voll beobachtbare Variablen (Fußgängerposition, Fahrzeugposition und -geschwindigkeit) und nicht beobachtbare Variablen (Fußgängerziel) reduzieren Bandyopadhyay et al. die Berechnungskomplexität des MOMDP, so dass die Strategie online in einem Fahrzeug ausgeführt werden kann. Trotzdem bleibt das Problem aller MDP-basierten Ansätze, dass das betrachtete Szenario a priori auf eine feste Anzahl an Zuständen begrenzt werden muss.

Der in (Rehder et al., 2015) vorgestellte Ansatz ist durch seinen modularen Charakter deutlich flexibler. Zudem arbeitet das System auch mit fahrzeugeigener Sensorik und kommt ohne Kartendaten aus. Die einzelnen Module des Ansatzes beschreiben den Einfluss unterschiedlicher Faktoren auf den Fußgängerpfad in Form von Wahrscheinlichkeitsverteilungen über die zukünftige Fußgängerposition. Konkret betrachten Rehder et al. den Einfluss der Fußgängerdynamik, der Zielausrichtung sowie der Umgebung. Das System kann aber leicht um weitere Faktoren erweitert werden. Die Wahrscheinlichkeitsverteilungen werden dabei jeweils als Belegungsgitter diskretisiert. Da das Ziel eines Fußgängers nicht bekannt ist, wird dieses als eine latente Variable behandelt und die Schätzung der Verteilung der Ziele über einen Partikelfilter realisiert. Hierdurch können auch bei diesem Ansatz multiple Hypothesen über das Fußgängerverhalten betrachtet werden. Die in (Rehder und Kloeden, 2015) ausführlicher beschriebene Evaluation des Verfahrens zeigt, dass der vorgestellte Ansatz vor allem bei längeren Vorhersagehorizonten einem rein auf Dynamik-Parametern basie-

renden Kalman-Filter deutlich überlegen ist. Zudem stellen Rehder et al. fest, dass die Berücksichtigung der Umgebungsinformationen einen deutlich geringeren Einfluss auf die Prädiktionsgüte hat, als der zielgerichtete Anteil der Fußgängerbewegung. Brouwer et al. (2015) zeigen, dass dieser Prädiktionsansatz auch in Kombination mit der posenbasierten Aktionserkennung von Kloeden et al. (2014) (s. Abschn. 2.3.2, S. 34) zur Generierung von Fahrerwarnungen verwendet werden kann.

2.3.2 Aktionserkennung

Für heutige, reaktiv reagierende Fußgängerschutzsysteme ist ausschließlich die Frage relevant, ob ein Fußgänger mit dem sich nähernden Fahrzeug kollidieren wird oder nicht. Aus diesem Grund wird das Problem der Fußgängerintentionserkennung bisher meistens als Aktionserkennungsproblem verstanden, bei dem die Frage „Geht er oder geht er nicht?“ beantwortet wird. Bei der Aktionserkennung wird somit zwischen verschiedenen Aktions- oder Verhaltenskategorien wie *„gehen“* und *„stehen“* beziehungsweise *„anhalten“* unterschieden. Im Gegensatz zur Trajektorienprädiktion (s. Abschn. 2.3.1) wird bei einer Aktionserkennung nicht die genaue Position des Fußgängers vorhergesagt (Bonnin et al., 2014b). Es existieren jedoch einige hybride Ansätze, die durch die Integration einer Kurzeit-Trajektorienprädiktion auch bei der Aktionserkennung die Fußgängerposition als Ausgangsgröße bestimmen. Diese Ansätze werden in dieser Arbeit ebenfalls als aktionserkennende Ansätze betrachtet. Abhängig von der zur Aktionserkennung verwendeten Informationsbasis können die bisherigen Arbeiten in dynamikbasierte, posenbasierte und kontextbasierte Ansätze unterteilt werden.

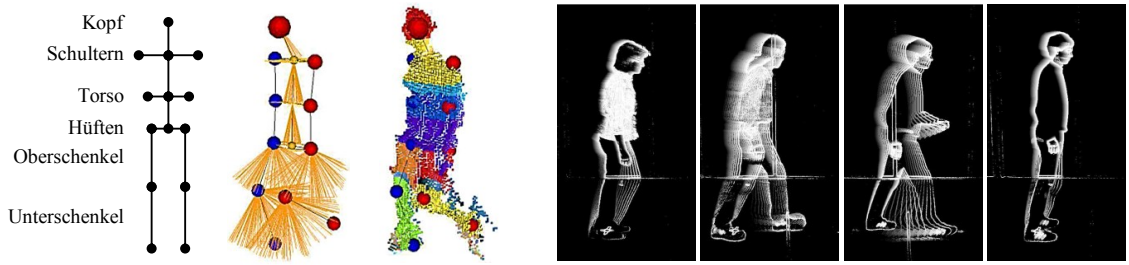
Dynamikbasierte Aktionserkennung

Die dynamikbasierte Aktionserkennung tritt ausschließlich in Kombination mit einer Kurzzeit-Trajektorienprädiktion auf, um bei dieser zustandsspezifische Bewegungsmodelle zu verwenden. Die Aktionserkennung dient dabei der Erkennung der Zustandsänderung und erfolgt in diesem Fall ausschließlich über die beobachtete Fußgängerdynamik. Wakim et al. (2004) unterscheiden beispielsweise zwischen den vier Bewegungszuständen *„stehen“*, *„gehen“*, *„joggen“* und *„rennen“*. Die vier Zustände sowie ihre Ver-

bindungen werden in Form einer Markov-Kette erster Ordnung modelliert. Hierzu wird jedem Zustand auf Basis von Bewegungsdaten aus der Literatur eine Wahrscheinlichkeitsverteilung über die Geschwindigkeit eines Fußgängers, sowie über die Änderung seiner Orientierung zugewiesen. Die Erkennung einer Zustandsänderung ist dabei von dem aktuellen Zustand und der aktuell beobachteten Geschwindigkeit des Fußgängers abhängig.

Posenbasierte Aktionserkennung

Die posenbasierte Aktionserkennung basiert auf aktionsabhängigen Veränderungen in der Pose des Fußgängers. Zur Erfassung dieser Daten stellen Keller et al. (2011) das **Histograms of Orientation Motion (HOM)** vor. Zur Berechnung des HOM wird zunächst der optische Fluss (Wedel et al., 2008) zwischen zwei aufeinanderfolgenden Fußgängerdetektionen bestimmt und bezüglich der Tiefe, sowie der durchschnittlichen Geschwindigkeit normalisiert. Keller et al. verwenden hierzu die Tiefen- und Bilddaten eines im Fahrzeug integrierten Stereokamerasystems. Anschließend werden die Orientierungshistogramme des oberen und unteren Flussfelds berechnet, wodurch das HOM die Oberkörper- und Beinbewegung eines Fußgängers implizit beschreibt. Keller et al. kombinieren das HOM schließlich mit Positionsdaten des Fußgängers, um über eine probabilistische Trajektorienzuordnung zu unterscheiden, ob ein sich lateral auf eine Fahrbahnkante zu bewegender Fußgänger weiter läuft oder stehen bleibt. Mit dem **Lateral Scene Flow Features (LSFF)** stellen Keller und Gavrila (2014) eine weitere Technik zur Beschreibung der Körperbewegung des Fußgängers vor. Ebenfalls auf dem optischen Fluss basierend, werden bei diesem Merkmal zunächst die lateralen Geschwindigkeiten aller Fußgängerpixel berechnet und diese anschließend zu einem Merkmalsvektor zusammengefasst. Über ein trainiertes Gaussian Process Dynamical Model (GPDM) (Wang et al., 2008) wird dieser Merkmalsvektor schließlich in einen niedrigdimensionalen Raum projiziert und zur Positionsprädiktion verwendet. Der in (Keller und Gavrila, 2014) durchgeführte Vergleich der beiden Ansätze mit zwei Varianten des Kalman-Filters zeigt, dass durch die zusätzliche Verwendung der posenbasierten Bewegungsinformationen eine deutliche Verbesserung der kurzzei-



(a) (b)

Abbildung 2.8: Zur posenbasierten Aktionserkennung verwendete (a) 3D Joints von Quintero et al. (2014) und (b) Motion History Images von Köhler et al. (2012).

tigen Pfadvorhersage von 10–50 cm erreicht wird. Ein menschlicher Beobachter kann nach (Keller et al., 2011) die ausgeführte Fußgängeraktion jedoch immer noch deutlich früher korrekt einschätzen.

Dasselbe Aktionserkennungsproblem wird auch von Quintero et al. (2014) betrachtet. Quintero et al. schlagen zur Beschreibung der Körperhaltung des Fußgängers die Verwendung von **3D Joints** vor, die die Position charakteristischer Körperteile und Gelenke beschreiben (s. Abb. 2.8a). Diese können über ein hierarchisches Segmentierungsverfahren in den Tiefendaten einer Stereokamera gefunden werden. Auch hier werden über ein GPDM anschließend die aktionstypischen Positionen der Joints und deren Verschiebungsvektoren erlernt und das trainierte Modell schließlich zur Positionsprädiktion angewendet. Der Ansatz von Quintero et al. erreicht ähnliche Ergebnisse wie die zwei von Keller et al. vorgeschlagenen Methoden.

Hariyono und Jo (2015) erweitern den Ansatz von Quintero et al., indem sie zusätzlich zu der Position und den Verschiebungsvektoren der Joints auch die Geschwindigkeit der Verschiebung berücksichtigen. Mit dem erweiterten Merkmalsvektor und einem naiven Bayes-Klassifikator (Bishop, 2006) können Hariyono und Jo die vier Aktionen *„losgehen“*, *„gehen“*, *„anhalten“* und *„vorbeugen“* zu 98 % korrekt klassifizieren.

Köhler et al. (2012) stellen zur posenbasierten Aktionserkennung die **Motion Contour Histograms of Oriented Gradients (MCHOG)** vor. Die MCHOG basieren auf lokalen Richtungsänderungen der Fußgängerkontur und erfassen somit Aspekte der Körpersprache, wie ein Vorbeugen des Oberkörpers oder eine Bewegung der Beine

explizit. Zur Erstellung der MCHOG wird zunächst eine Sequenz von binären Fußgängerkonturbildern $\Psi(I(x,y,t))$ erstellt. Hierzu wird für jeden Fußgängerbildbereich über eine Hochpassfilterung ein Gradientenbild erstellt, das anschließend über ein statisches Hintergrundmodell auf den Vordergrund begrenzt und schließlich über einen kleinen, empirisch bestimmten Schwellwert binarisiert wird. Abschließend werden die Fußgängerkonturbilder mehrerer Zeitschritte zu einem Motion History Image (MHI) $H_\tau(x,y,t)$ überlagert:

$$H_\tau(x,y,t) = \begin{cases} \tau & \text{wenn } \Psi(I(x,y,t)) \neq 0 \quad (2.1a) \\ max(0, H_\tau(x,y,t-1) - 1) & \text{sonst} \quad (2.1b) \end{cases}$$

mit τ als Zerfallsvariable, die die Länge der betrachteten Historie bestimmt. Köhler et al. setzen $\tau = 10$ ($\hat{=}0{,}2\,s$ bei 50 fps), da nach der durchgeführten Evaluation die Betrachtung einer längeren Historie zu keiner nennenswerten Ergebnisverbesserung führt. Abbildung 2.8b zeigt beispielhafte MHIs verschiedener Fußgänger.

Um die im MHI erfassten lokalen Bewegungen zu beschreiben, verwenden Köhler et al. anschließend eine Adaption der aus dem Bereich der bildbasierten Fußgängerdetektion bekannten Histograms of Oriented Gradients (HOG) (s. Abschn. 3.4.1). Im Gegensatz zu der von Dalal und Triggs (2005) beschriebenen, originalen Implementierung verzichten Köhler et al. jedoch auf die abschließende Normalisierung der Zellen über überlappende Blöcke. Diese Blocknormalisierung führt bei natürlichen Bildern zu einer gesteigerten Invarianz gegenüber Beleuchtungsänderungen. Im MCHOG sollen die lokalen Änderungen benachbarter Zellen jedoch zur Beschreibung der lokalen Körperbewegungen erhalten bleiben. Köhler et al. zeigen, dass eine trainierte Support Vector Machine (SVM) (s. Abschn. 3.2) auf Basis der MCHOG zuverlässig unterscheiden kann, ob ein an der Fahrbahnkante stehender Fußgänger weiter stehen bleibt oder anfängt los zu laufen. Zudem zeigt eine Analyse des Einflusses der einzelnen Körperteile, dass die Initialbewegung der Beine der wichtigste Indikator bei der Entscheidung ist, ob ein Fußgänger losläuft oder nicht. Als zweitwichtigster Indikator folgt die Neigung des Oberkörpers. Zur Validierung werden Bilddaten stationärer, hochauflösender Videokameras verwendet. In (Köhler et al., 2013) wird die Anwendbarkeit dieser Aktionserkennung zur Auslösung eines ausweichenden Fußgängerschutzsystems gezeigt. Das MHI wird hierbei auf eine Größe von 72×128 px skaliert und die MCHOG an-

schließend mit einer Zellengröße von 8×8 px und 12 Histogrammbins bestimmt. Um den Einfluss von Hintergrundbewegungen auf das MHI zu reduzieren, greifen Köhler et al. hier auf die zur Verfügung stehenden Tiefendaten zurück, statt dem in (Köhler et al., 2012) verwendeten statischen Hintergrundbild.

Auch Furuhashi und Yamada (2011) verwenden ein auf den HOG basierendes Merkmal zur Aktionserkennung. Im Gegensatz zu den obigen Arbeiten unterscheiden Furuhashi und Yamada jedoch zwischen vier möglichen Aktionen des Fußgängers: a) Querung der Straße, b) Geradeaus laufen ohne Querung, c) Versuchen zu Queren, Querungsversuch aber abbrechen und vor der Bordsteinkante zum Stehen kommen, d) Versuchen zu Queren, Querungsversuch aber abbrechen und weiter geradeaus laufen. Zur Aktionserkennung bestimmen Furuhashi und Yamada zunächst für jeden detektierten Fußgänger die HOG, mit einer Fenstergröße von 30×60 px, 5×5 px großen Zellen, 15×15 px großen Blöcken sowie einem Histogramm mit 9 Bins. Zur Reduktion der Modellkomplexität wird dieser 3.240 elementige Merkmalsvektor anschließend über eine Hauptkomponentenanalyse (Principal Component Analysis (PCA)) auf einen $d = 10$ dimensionalen Unterraum projiziert und über ein Feature Selection Verfahren weiter auf die $d' = 6$ relevantesten Elemente begrenzt. Abschließend wird die Variationsvielfalt der dimensionsreduzierten HOG über eine, auf dem k-means Clusterverfahren (Bishop, 2006) basierende, Vektorquantisierung mit $k = 12$ weiter reduziert. Dieses **Posture Appearance Feature (PAF)** wird anschließend für eine Sequenz von n Fußgängerdetektionen, die einen Abstand von l Frames zueinander haben, berechnet. Die Wahrscheinlichkeit, dass der so bestimmte Vektor eine der vier betrachteten Aktionen beschreibt, wird schließlich über eine angepasste Version des Bayes-Theorem bestimmt. Die durchgeführte Evaluation zeigt, dass die Betrachtung einer Historie ($n \geq 2$) zur Aktionserkennung deutlich effizienter ist, als die Betrachtung der aktuellen Einzelpose ($n = 1$). Die besten Ergebnisse erzielen Furuhashi und Yamada mit der Betrachtung von $n = 2$ Einzelposen, die im Abstand von $l = 7$ Frames gesampelt werden. Dadurch wird eine Sequenz von $(n-1) * l + 1 = 8$ Frames betrachtet, was bei einer Bildwiederholrate von 30 fps einer zeitlichen Distanz von $0{,}2\overline{6}$ s entspricht.

Motiviert durch die Beobachtung, dass ein querungswilliger Fußgänger bei der Erkennung eines potenziell gefährlichen Fahrzeugs sein Sicherungsverhalten unterbricht und

das Fahrzeug fixiert, fokussieren sich Kloeden et al. (2014) auf die Kopforientierung des Fußgängers, um zu entscheiden, ob dieser auf die Straße tritt oder vorher stehen bleibt. Zur Klassifikation wird ein Mixture of Gaussians Hidden Markov Model (MoG-HMM) verwendet, dessen bestimmender Merkmalsvektor aus dem **Blickwinkel** des Fußgängers auf das Fahrzeug, sowie der **Quergeschwindigkeit** des Fußgängers und seiner Distanz zum Fahrschlauch besteht. Kloeden et al. zeigen, dass unter Verwendung eines Referenzsystems, welches die Kopforientierung alle 100 ms mit einer Genauigkeit von einem Grad genau bestimmt, alle 16 Testsequenzen korrekt als *„Stehenbleiben“* oder *„Überqueren“* klassifiziert wurden. Eine durch additives Rauschen simulierte Ungenauigkeit des Sensors führt mit $\sigma = 40°$ zu einer Genauigkeit von 75 %. Dabei ist jedoch zu beachten, dass die Testsequenzen unter Verwendung von eingewiesenen Probanden aufgenommen wurden und diese abhängig von der Probandeneinweisung ein ausgeprägteres Sicherungsverhalten als natürliche Personen zeigen.

Kontextbasierte Aktionserkennung

Die kontextbasierten Aktionserkennung basiert auf der Relation zwischen der Fußgängerbewegung und der kontextuellen Querungssituation, die sowohl über den Abstand des Fußgängers zu Szenenelementen, wie Zebrastreifen und Fahrbahnkanten, als auch über die Annäherung eines Fahrzeugs an die Querungsstelle beschrieben wird.

Bonnin et al. (2014b) stellen zur Modellierung der Kontextinformationen ein merkmalsbasiertes, multimodales System vor. Dieses besteht aus einem generischen innerstädtischen Modell, welches die üblichen Verhaltensmuster von Fußgängern bei freier Querung abbildet, sowie aus einem speziell auf die Querungssituation an Zebrastreifen angepassten Modell. Beide Modelle verwenden neun bis zwölf Merkmale, um die Bewegung des Fußgängers in Relation zur Fahrbahnkante, zum sich nähernden Fahrzeug, sowie gegebenenfalls zum Zebrastreifen zu beschreiben. Die Merkmalsliste umfasst dabei die laterale Distanz zu verschiedenen charakteristischen Punkten, wie dem Bordstein oder dem Kollisionspunkt, an dem der Fußgänger und das Fahrzeug aufeinander treffen würden, ebenso wie die Zeit die der Fußgänger braucht, um diese Punkte zu erreichen. Zudem werden verschiedene Orientierungswerte, wie der Winkel

zwischen dem Fußgänger und der Straße, oder dem Fahrzeug berücksichtigt. Zur Aktionsklassifikation verwenden Bonnin et al. ein Single Layer Perceptron (SLP) (Bishop, 2006), eine vereinfachte Form der künstlichen neuronalen Netze. Die durchgeführte Evaluation zeigt, dass durch die Verwendung kontextspezifischer Modelle, alle Querungen 2,59 s vor Querungsbeginn mit einer True Positive Rate (TPR) von 62 % fehlerfrei (False Positive Rate (FPR) von 0 %) vorhergesagt werden. Im Vergleich dazu konnten unter der Verwendung eines generischen Modells alle Querungen nur 0,72 s vor Querungsbeginn zu 31 % fehlerfrei erkannt werden.

Kooij et al. (2014) verwenden Kontextinformationen, um in Situationen mit einem lateral querenden Fußgänger zu unterscheiden, ob der Fußgänger vor dem sich nähernden Fahrzeug auf die Straße tritt oder vorher anhält. Die Kontextinformationen werden als latente Variablen eines Dynamic Bayesian Network (DBN) modelliert, über die in Form eines Switching Linear Dynamical System (SLDS) entschieden wird, welches Bewegungsmodell (gehen vs. stehen) bei der Prädiktion der Fußgängerposition Anwendung findet. Als Kontextinformationen verwenden Kooij et al. die zeitliche Distanz des Fußgängers zur Fahrbahnkante, die Kritikalität der aktuellen Situation in Form der Distanz zwischen dem Fußgänger und dem Fahrzeug, sowie das Situationsbewusstsein des Fußgängers, bei dem über die Orientierung seines Kopfes abgeschätzt wird, ob er das sich nähernde Fahrzeug gesehen hat oder nicht. Kooij et al. können zeigen, dass die Verwendung der Kontextinformationen im Vergleich zu reinen Positionsdaten zu einer deutlich verbesserten Positionsprädiktion führt (0,39 m bei einem Prädiktionshorizont von 1 s).

2.3.3 Intentionserkennung

Die Intention eines Fußgängers, definiert als seine prinzipielle Absicht eine bestimmte Handlung durchzuführen, unabhängig vom Zeitpunkt der Handlungsausführung, ist ein neues Verständnis des Begriffs *„Fußgängerintention“*. Nach Wissen der Autorin existiert aktuell nur die Arbeit von Voelz et al. (2015), die sich mit der Erkennung solcher prinzipiellen Handlungsabsichten bei Fußgängern beschäftigt. Voelz et al. verwenden in ihrer Arbeit an einem Zebrastreifen mit einem stationären LiDAR-System

(Velodyne, 2016) aufgezeichnete Fußgänger- und Fahrzeugtrajektorien, um über eine Merkmalsselektion die Merkmale zu ermitteln, welche relevant bei der Vorhersage sind, ob ein Fußgänger an dem Zebrastreifen quert oder nicht. Voelz et al. untersuchen dazu 15 verschiedene Merkmale. Vier der Merkmale beziehen sich auf die richtungsabhängige Geschwindigkeit des Fußgängers. Zudem wird mit den minimalen Abständen des Fußgängers zum Bordstein und zum Zebrastreifen die Relation der Fußgängerposition zu den für das Kontextverständnis relevanten Szenenelementen berücksichtigt. Die restlichen neun Merkmale beziehen sich auf den möglichen Einfluss eines sich dem Zebrastreifen nähernden Fahrzeugs auf die Fußgängerbewegung. Hier wird sowohl die richtungsabhängige Geschwindigkeit und die relative Position des Fahrzeugs zum Zebrastreifen, als auch die relative Geschwindigkeit und der Abstand zwischen Fahrzeug und Fußgänger berücksichtigt. Um auch die Historie der aktuellen Situation abzubilden, werden die jeweils letzten fünf Ausprägungen eines Merkmals zu einem Merkmalsvektor mit insgesamt 75 Elementen kombiniert. Über die, auf trainierten SVMs basierende Recursive Feature Elimination (RFE) (Guyon et al., 2002) zeigen Voelz et al. schließlich, dass die laterale Geschwindigkeit des Fußgängers in Kombination mit dem Abstand des Fußgängers zum Bordstein sowie zum Zebrastreifen die höchste Genauigkeit bei der Bestimmung der Intention des Fußgängers, die Straße an dem Zebrastreifen zu queren, erreicht. Erwartungsgemäß haben die dem Fahrzeug zugeordneten Merkmale wenig Einfluss auf die prinzipielle Querungsabsicht. Mit dem auf die besten 10 Merkmale begrenzten Merkmalsvektor können Voelz et al. beispielsweise 3 m vor dem Zebrastreifen 95 % der querenden Fußgänger und fast 100 % der nicht querenden Fußgänger korrekt klassifizieren. Dass die fahrzeugbezogenen Merkmale zwar nicht für die Erkennung der Querungsintention relevant sind, wohl aber bei der Prädiktion des Zeitpunkts der Querungsausführung, zeigt sich in (Voelz et al., 2016). Hier verwenden Voelz et al. die gesamte Merkmalsliste, um für die querungswilligen Fußgänger über Quantile Regression Forests (Meinshausen, 2006) eine Time-to-Cross zu schätzen.

2.4 Referenzmethoden

Die nicht direkte Beobachtbarkeit und damit auch Messbarkeit der Intention birgt eine Herausforderung bei der Bestimmung einer für intentionserkennende Systeme geeigneten Referenzmethode. Wie in Abschnitt 2.4.1 erläutert wird, verwenden alle bisher bekannten Systeme zur Fußgängerverhaltenserkennung und -vorhersage das im späteren Verlauf tatsächlich gezeigte Verhalten des Fußgängers als Referenz und werden im Folgenden daher als Ground-Truth-basierte Referenzmethoden bezeichnet. Da diese Methoden jedoch Schwächen bei der Anwendung für intentionserkennende Systeme haben, wird in Abschnitt 2.4.2 auf das Konzept und die Auswertungsmöglichkeiten beobachterbasierter Referenzmethoden eingegangen. Diese Verfahren finden im Bereich der Fußgängerverhaltenserkennung aktuell noch selten Einsatz; sie gelten bei der Erhebung psychologischer Merkmale aber als anerkannte, empirische Messmethode (Wirtz und Caspar, 2002).

2.4.1 Ground-Truth-basierte Referenzmethoden

Die Ground-Truth-basierten Referenzmethoden setzen das im späteren Verlauf tatsächlich gezeigte Verhalten des Fußgängers als Referenz für die vorigen Zeitschritte. Bei der Trajektorienprädiktion (s. Abschn. 2.3.1) wird im Rahmen der Evaluation in der Regel der euklidische Abstand der vorhergesagten Fußgängerposition zur Ground-Truth bestimmt und als Leistungsmaß in Abhängigkeit des Prädiktionshorizonts angegeben (vgl. Schneider und Gavrila, 2013). Wird bei der pfadbasierten Langzeit-Trajektorienprädiktion zwischen Pfaden unterschieden, die mit verschiedenen Aktionen wie beispielsweise dem Queren und Nicht-Queren verknüpft sind, dann wird zudem als Leistungsmaß die Häufigkeit angegeben, mit der einem Fußgänger ein Pfad der richtigen Aktion zugeordnet wird (vgl. Roehder, 2011).

Auch bei der Aktionserkennung (s. Abschn. 2.3.2) wird die wirkliche Durchführung der Aktion als Referenz für die Erkennungsleistung herangezogen. Der Startzeitpunkt der Aktion wird hierbei in der Regel ereignisbasiert durch einen menschlichen Beobachter markiert. So wird der Zeitpunkt des Querungsbeginns beispielsweise in (Köhler et al., 2012) über die initiale Bewegung des Fußes vor dem Betreten der Straße

beschrieben. Ist die Aktionserkennung mit einer Kurzzeit-Trajektorienprädiktion verknüpft, wird als Leistungsmaß ebenfalls der metrische Prädiktionsfehler des Ansatzes verwendet (vgl. Keller et al., 2011). In den anderen Fällen, in denen die Aktionserkennung als ein Klassifikationsproblem betrachtet wird, werden die standardmäßig zur Beurteilung von Klassifikatoren verwendeten Gütemaße (s. Abschn. 3.3.2), wie die TPR oder FPR, zur Beurteilung herangezogen (vgl. Hariyono und Jo, 2015). Um nicht nur die Erkennungs-, sondern auch die Vorhersageleistung dieser Ansätze anzuführen, wird die Klassifikationsleistung oft in Abhängigkeit der Sekunden vor dem Aktionsbeginn angegeben (vgl. Bonnin et al., 2014b). Auch die bisher einzige Arbeit, die die Querungsintention eines Fußgängers als seine prinzipielle Querungsabsicht versteht (s. Voelz et al., 2015 in Abschn. 2.3.3), zieht als Referenz eine im späteren Verlauf tatsächlich beobachtete Querung des Fußgängers heran. Als Leistungsmaß wird hier die Klassifikatorgüte in Abhängigkeit des Abstandes des Fußgängers zum Zebrastreifen angegeben.

Die Verwendung des im späteren Verlauf tatsächlich gezeigten Verhaltens des Fußgängers als Referenz hat bei der Aktions- sowie bei der Intentionserkennung den Nachteil, dass es zwar eine Vergleichbarkeit der Ansätze untereinander zulässt, jedoch kann die Frage, ab wann eine Aktion oder Intentionen prinzipiell erkennbar ist, so nicht beantwortet werden. Zudem werden bei der Intentionserkennung so keine Intentionen berücksichtigt, die letztlich nicht zu Handlungen führen (Diederichs, 2017).

2.4.2 Beobachterbasierte Referenzmethoden

Bei beobachterbasierten Referenzmethoden wird der Mensch als Experte betrachtet und statt dem tatsächlich gezeigten Verhalten das Urteil eines oder mehrerer menschlicher Beobachter[2] als Referenz verwendet. Vor allem in Disziplinen wie der empirischen Sozialwissenschaft und der Psychologie haben sich beobachterbasierte Methoden zur Erfassung von Merkmalen, die nicht direkt gemessen oder beim Merkmalsträger erfragt werden können, etabliert (Wirtz und Caspar, 2002). Im Speziellen ermöglicht

[2]In der Literatur werden synonym für den Begriff „Beobachter“ auch die Begriffe „Rater“, „Codierer“ oder „Beurteiler“ verwendet.

die Methode der strukturierten, quantitativen Beobachtung eine Erhebung numerischer Beobachtungsdaten über das Verhalten anderer Personen. In Kombination mit einer Ratingskala kann so beispielsweise der Ausprägungsgrad eines vorab definierten Merkmals gemessen werden. Voraussetzung hierfür ist ein zielgerichtetes und systematisches Vorgehen mit einem standardisierten Ablauf der Beobachtung. Methoden, die diese Anforderung erfüllen können, sind beobachterbasierte Videoannotationen, bei denen einem oder mehreren geschulten Beobachtern dieselben Videoaufzeichnungen unterschiedlicher Objekte oder Situationen gezeigt werden. Abhängig davon, ob die Beurteilung separater Einzelereignisse oder längerer Verhaltensströme angestrebt ist, werden die Videos ereignisbasiert oder in einem festen Zeitintervall zur Abgabe des Beobachterurteils pausiert (Döring und Bortz, 2016a).

Im Bereich der Fußgängerverhaltenserkennung sind nur wenige Arbeiten bekannt, die eine solche beobachterbasierte Videoannotation einsetzen. Zum einen ist hier die viel zitierte Human Factors Studie von Schmidt et al. (2008) zu nennen, in der mehreren Beobachtern Videos realer Verkehrsszenen mit querenden und nicht querenden Fußgängern gezeigt wurden. Die Videos wurden an verschiedenen Zeitpunkten vor dem Stoppen beziehungsweise Queren automatisch pausiert und der Beobachter aufgefordert, auf einer 7-stufigen Skala zu beurteilen, ob der Fußgänger im nächsten Moment die Straße queren wird oder nicht. Durch die Maskierung einzelner Parameter in den Videos, wie etwa dem Kopfbereich des Fußgängers oder dem Verkehr, konnten Schmidt et al. die Parameter ermitteln, auf die zur Vorhersage des Fußgängerverhaltens aus Sicht eines menschlichen Beobachters nicht verzichten werden kann (vgl. Abschn. 2.2.2).

Tsimhoni et al. (2008) setzen eine beobachterbasierte Videoannotation ein, um den Einfluss fünf unabhängiger Variablen, wie die Laufrichtung des Fußgängers, seinen Abstand zur Fahrbahnkante oder die Geschwindigkeit des sich nähernden Fahrzeugs, auf das für einen Fußgänger empfundene Gefahrenpotential hin zu untersuchen. Die Beobachter mussten hierzu nach der Beobachtung eines 3–15 sekündigen Videoclips auf einer Skala von 0 bis 100 beurteilen, wie sehr ein Fahrer den Fußgänger beobachten muss, um eine Kollision mit diesem zu vermeiden.

Keller et al. (2011) sind schließlich die einzigen, die die Ergebnisse einer beobachterbasierten Videoannotation direkt bei der Evaluation eines fußgängerverhaltenerken-

nenden Systems einsetzen. Analog zu Schmidt et al. (2008) beurteilten die von Keller et al. befragten Beobachter, ob der Fußgänger am Fahrbahnrand stehen bleibt oder queren wird und gaben ihre Unsicherheit darüber auf einer kontinuierlichen Skala von 0 bis 1 an. Die durchschnittlichen Unsicherheitswerte verwenden Keller et al. schließlich, um die Leistung ihres aktionserkennenden Systems in Kontext zur menschlichen Leistungsfähigkeit zu setzen. Als Ground-Truth wird jedoch weiterhin die im späteren Verlauf tatsächlich durchgeführte Aktion verwendet, und eine genaue Auswertung der Beobachterurteile erfolgt nicht.

Die Zuverlässigkeit und Genauigkeit beobachterbasierter Methoden kann empirisch anhand der Beobachterübereinstimmung und der Beobachter- bzw. Interraterreliabilität überprüft werden. Ein vollständiger Überblick über Methoden zur Bestimmung der Übereinstimmung und der Reliabilität zwischen Beobachtern ist in (Wirtz und Caspar, 2002) zu finden. Die folgende Darstellung der für diese Arbeit relevantesten Punkte basiert, wenn nicht anders angegeben, auf dieser Quelle.

Beobachterübereinstimmung

Die Beobachterübereinstimmung erfasst, inwiefern verschiedene Beobachter die verschiedenen Objekte jeweils exakt gleich bewerten. Damit kann die Beobachterübereinstimmung bereits für mindestens nominalskalierte[3] Daten bestimmt werden. Bei mehr als zwei Beobachtern lässt sich die Übereinstimmung sowohl für alle Beobachter gemeinsam, als auch in Form eines Mittelwerts über alle möglichen Beobachterpaare berechnen. Da im ersten Fall nur von einer Übereinstimmung ausgegangen wird, wenn alle Beobachter dasselbe Urteil für ein Objekt abgeben, empfehlen Wirtz und Caspar (2002) stets die gemittelte, paarweise Übereinstimmung zu verwenden, da ansonsten das Urteil einzelner Beobachter ein zu starkes Gewicht bekommt. Aus diesem Grund wird im Folgenden die Berechnung der Übereinstimmungsmaße nur am Beispiel von zwei Beobachtern dargestellt.

[3]Mit einer Nominalskala kann unterschieden werden, ob oder welche Ausprägung eines Merkmals vorliegt (z.B. Raucher oder Nichtraucher). Entgegen Ordinal-, Intervall- oder Verhältnisskalen erfolgt bei Nominalskalen keine Einordnung in eine Rangfolge (Döring und Bortz, 2016b).

Tabelle 2.1: Übereinstimmungsmatrix für zwei Beobachter (Wirtz und Caspar, 2002).

		Beobachter 2				
		c_1	c_2	...	c_s	$\sum$
Beobachter 1	c_1	$\mathbf{n_{11}}$	n_{12}	...	n_{1s}	$n_{1\cdot}$
	c_2	n_{21}	$\mathbf{n_{22}}$	...	n_{2s}	$n_{2\cdot}$
	...	...	...	...	...	...
	c_s	n_{s1}	n_{s2}	...	$\mathbf{n_{ss}}$	$n_{s\cdot}$
	$\sum$	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot s}$	N

Übereinstimmungsmatrix Die Grundlage der Übereinstimmungsberechnung bildet die Übereinstimmungsmatrix, die die erfassten Beobachterurteile in einer standardisierten Form darstellt. Tabelle 2.1 zeigt die Übereinstimmungsmatrix für $k = 2$ Beobachter, die N Objekte mittels einem s-stufigen Kategoriesystem beurteilt haben. Die Matrixelemente n_{ij} geben die absoluten Häufigkeiten an, mit der Beobachter 1 Objekte zur Kategorie i und Beobachter 2 dieselben Objekte zur Kategorie j zuordnet. Die Elemente n_{ii} der Hauptdiagonalen entsprechen somit den beobachteten Übereinstimmungen in der jeweiligen Kategorie. Anstatt den absoluten Häufigkeiten n_{ij}, kann eine Übereinstimmungsmatrix auch die relativen Häufigkeiten $h_{ij} = n_{ij}/N$ angeben. Über die Randhäufigkeiten der Zeilen $n_{i\cdot}$ und die der Spalten $n_{\cdot j}$, lässt sich zudem eine Matrix mit den bei Zufall erwarteten Übereinstimmungen $e_{ij} = n_{\cdot j} \cdot n_{i\cdot}/N$ bilden.

Übereinstimmungsmaße Das einfachste Maß zur Beschreibung der absoluten Übereinstimmung zwischen Beobachtern ist die **prozentuale Übereinstimmung ($P\ddot{U}$)**. Diese beschreibt den prozentualen Anteil der Fälle, in denen die Beobachter das gleiche Urteil abgeben. Es gilt

$$P\ddot{U} = \sum_{i=1}^{s} \frac{n_{ii}}{N} \cdot 100\,\% = \sum_{i=1}^{s} h_{ii} \cdot 100\,\%. \tag{2.2}$$

Da die PÜ nicht gegenüber dem Zufall bereinigt ist, ist ihr jedoch nicht zu entnehmen, in welchem Maß der gemessene Wert größer ist, als bei rein zufälligem Beurteilungsverhalten. Somit überschätzt die PÜ die wahre Übereinstimmung zwischen den

Beobachtern grundsätzlich. Sie sollte daher stets mit der bei Zufall erwarteten Übereinstimmung $P\ddot{U}_{Zufall}$ angegeben

$$P\ddot{U}_{Zufall} = \sum_{i=j=1}^{s} \frac{e_{ij}}{N} \cdot 100\,\% \tag{2.3}$$

und der Unterschied zwischen der $P\ddot{U}$ und $P\ddot{U}_{Zufall}$ auf Signifikanz geprüft werden.

Alternativ können zufallskorrigierte Übereinstimmungsmaße, wie das **Cohens κ** oder das **Scotts π**, verwendet werden. Beide Gütemaße berücksichtigen das Verhältnis der beobachteten Übereinstimmung P_o zu der bei Zufall erwarteten Übereinstimmung P_e und liefern somit eine standardisierte Maßzahl mit einem Wertebereich von $[-1,\ +1]$. Sowohl Cohens κ, als auch Scotts π basieren auf der prozentualen Übereinstimmung mit $P_o = P\ddot{U}/100\,\%$. Der Unterschied zwischen den beiden Maßen liegt lediglich in der Definition der Zufallserwartung. So gilt bei Cohens κ

$$P_{e,\kappa} = \frac{1}{N^2} \cdot \sum_{i=j=1}^{s} n_{\cdot j} \cdot n_{i\cdot} = \sum_{i=j=1}^{s} h_{\cdot j} \cdot h_{i\cdot} \tag{2.4}$$

und bei Scotts π hingegen

$$P_{e,\pi} = \frac{1}{N^2} \cdot \sum_{i=1}^{s} \left(\frac{n_{\cdot j} + n_{i\cdot}}{2} \right)^2 = \sum_{i=j=1}^{s} \left(\frac{h_{\cdot j} \cdot h_{i\cdot}}{2} \right)^2. \tag{2.5}$$

Die Gütemaße berechnen sich schließlich über

$$\kappa,\pi = \frac{P_o - P_e}{1 - P_e}. \tag{2.6}$$

Cohens κ und Scotts π haben nur dann eine unterschiedliche Ausprägung, wenn die Randsummenverteilung beider Beobachter unterschiedlich ist. Da Scotts π den Effekt unterschiedlicher Grundwahrscheinlichkeiten stärker gewichtet, ist es das strengere Übereinstimmungsmaß. Trotzdem ist Cohens κ das in der Literatur am häufigsten verwendete Maß zu Bestimmung der Beobachterübereinstimmung.

Welche κ- und π-Werte als ausreichend oder gut zu beurteilen sind, ist abhängig von den zu beurteilenden Objekten und Merkmalen. Nach Wirtz und Caspar (2002) kann beispielsweise für ein schwer zu erfassendes Merkmal 0,5 ein zufriedenstellender Wert sein, für ein einfaches Merkmal 0,8 hingegen ein zu niedriger Wert. In der Literatur werden allgemein als Faustregel Werte größer 0,4 als akzeptable, Werte größer

0,6 als gute und Werte größer 0,75 als sehr gute Übereinstimmung bewertet (Fleiss et al., 1973). Altman (1991) gibt vergleichbare Grenzwerte an. Maßgebend sollte jedoch sein, welche Werte in dem jeweiligen Forschungsgebiet gefunden wurden und ob alternative Testverfahren existieren, mit denen die beobachteten Merkmale zuverlässiger erfasst werden können (Wirtz und Caspar, 2002). Bezüglich der Intention von Fußgängern sind aktuell noch keine Werte bekannt, die einem Vergleich dienlich sind. Daher können höchstens Werte aus verwandten, ähnlich komplexen Forschungsgebieten herangezogen werden. Bei der Fahrmanöverintention wurden beispielsweise bereits beobachterbasierte Referenzmethoden eingesetzt (Diederichs und Pöhler, 2014). Abhängig von dem betrachteten Fahrmanöver und dem zu identifizierenden Verhalten konnten hierbei κ-Werte von $\kappa = 0{,}30$ bis $\kappa = 1{,}0$, mit einem Mittelwert von $\hat{\kappa} = 0{,}62$ erreicht werden. Im Speziellen erreichen Verhaltenskategorien, die dem Verhalten eines Fußgängers am ähnlichsten sind, wie beispielsweise ein manöverspezifisches Blickverhalten oder eine Bewegung des Torsos, Übereinstimmungswerte von $\kappa = 0{,}50$ beziehungsweise $\kappa = 0{,}87$.

Beobachterreliabilität

Im Gegensatz zur Beobachterübereinstimmung wird bei der Beobachterreliabilität keine exakte Übereinstimmung der Urteile gefordert. Stattdessen wird die, für mindestens ordinalskalierten Daten oft sinnvollere, Ähnlichkeit der relativen Lage der Beobachterurteile zum Mittelwert der untersuchten Stichprobe untersucht. Eine Beobachtung ist also dann reliabel, wenn jeder Beobachter unterschiedliche Objekte unterschiedlich beurteilt und verschiedene Beobachter bei demselben Objekt zu ähnlichen Urteilen kommen. Für eine hohe Reliabilität müssen folglich die Unterschiede der Urteile für ein Objekt im Verhältnis zu den Unterschieden der Mittelwerte zwischen den verschiedenen Objekten möglichst klein sein.

Reliabilitätsmaße Für intervallskalierte Daten ist die **Intraklassenkorrelation (ICC)** das geeignetste Reliabilitätsmaß. Im Gegensatz zu den oben beschriebenen Übereinstimmungsmaßen kann die ICC direkt für beliebig viele Beobachter bestimmt werden. Abhängig davon, ob bei der weiteren Verwendung der Beobachterurteile das

Urteil eines einzelnen Beobachters oder der Mittelwert k verschiedener Beobachter verwendet wird, sollten für die Reliabilitätsbestimmung entweder die einzelnen Rohwerte der Urteile oder deren Mittelwert verwendet werden. Zudem muss entschieden werden, ob im Rahmen einer unjustierten Auswertung (ICC_{unjust}) die Mittelwertunterschiede zwischen den Beobachtern zulasten der Reliabilitätsschätzung verrechnet werden, oder eine justierte Auswertung (ICC_{just}) vorgenommen wird, bei der die Mittelwerte der Beobachter verschieden sein dürfen. Hierbei ist zu beachten, dass es sich bei der ICC_{unjust} nur dann um ein Reliabilitätsmaß handelt, wenn die Beobachter zufällig ausgewählt sind und diese somit eine repräsentative Stichprobe aller Beobachter bilden, für die die Reliabilitätsaussage gültig sein soll. Die ICC_{just} gilt hingegen nur bei fest eingesetzten Beobachtern als Reliabilitätsmaß und kann somit nicht auf andere Beobachter außerhalb der Stichprobe verallgemeinert werden. Andernfalls sind die ICC-Werte nur als Korrelationsmaß zu interpretieren. Die Frage nach der Justierung der Mittelwerte ist somit unmittelbar an die Verallgemeinerbarkeit der Ergebnisse hinsichtlich anderer Beobachter gebunden.

Vorteilhaft ist schließlich, wenn alle Objekte von derselben Beobachtergruppe beurteilt werden, da dieses die Berechnung der ICC über ein zweifaktorielles Modell zulässt. Dadurch kann die Merkmalsvarianz exakter geschätzt werden, als mit einem einfaktoriellen Modell, da sich Ratereffekte für alle beurteilten Objekte gleich auswirken. Wie in Wirtz und Caspar (2002) erläutert wird, kann die ICC jedoch nur über eine zweifaktorielle Varianzanalyse berechnet werden, wenn angenommen werden kann, dass keine systematischen Interaktionseffekte zwischen Objekten und Beobachtern vorliegen. Zur Überprüfung dieser Voraussetzung empfehlen Wirtz und Caspar die Anwendung des Additivitätstest von Tukey (1949).

Für die Berechnung der ICC_{unjust} im zweifaktoriellen Modell gilt schließlich

$$\text{ICC}_{unjust} = \frac{MS_{obj} - MS_{err}}{MS_{obj} + (k-1) \cdot MS_{err} + \frac{k}{N} \cdot (MS_{rat} - MS_{err})} \tag{2.7}$$

wobei MS_{obj} die Varianz zwischen den Objekten, MS_{rat} die Varianz zwischen den Beobachtern und MS_{err} die Restvarianz beschreibt.

Für die ICC_{just} gilt analog

$$\mathrm{ICC}_{just} = \frac{MS_{obj} - MS_{err}}{MS_{obj} + (k-1) \cdot MS_{err}}. \tag{2.8}$$

Die Berechnung der Varianzschätzungen MS über eine zweifaktorielle Varianzanalyse ist ausführlich in Anhang A.1 beschrieben.

Zur Überprüfung der Signifikanz der Abweichung der ICC wird die Prüfgröße

$$F_0 = \frac{MS_{obj}}{MS_{err}} \tag{2.9}$$

mit den kritischen Werten der F-Verteilung verglichen ($df_{Zähler} = (N-1)$ und $df_{Nenner} = (k-1) \cdot (N-1)$). Die Berechnung der Konfidenzintervalle ist Anhang A.2 zu entnehmen.

Der Wertebereich der ICC ist definitionsgemäß auf $[0, 1]$ beschränkt. Allgemein wird in der Literatur ein ICC von mindestens 0,7 als Indiz für gute Reliabilität angesehen. Auch hier ist die Ausprägung des Koeffizienten aber immer in Abhängigkeit des zu messenden Merkmals zu bewerten.

Reliabilität der Mittelwerte mehrerer Beobachter Eine Möglichkeit die Reliabilität der Daten zu erhöhen, ist die Verwendung der Mittelwerte mehrerer Beobachter, anstatt nur die Urteile eines Beobachters als Informationsgrundlage zu verwenden. Liegen die Urteile mehrerer Beobachter bereits vor, kann die unjustierte Reliabilitätsschätzung des Mittelwerts von k Beobachtern über

$$ICC_{unjust,\,MW} = \frac{MS_{obj} - MS_{err}}{MS_{obj} + \frac{1}{N} \cdot (MS_{rat} - MS_{err})} \tag{2.10}$$

empirisch bestimmt werden. Für die justierte Reliabilitätsschätzung gilt entsprechend:

$$ICC_{just,\,MW} = \frac{MS_{obj} - MS_{err}}{MS_{obj}} \tag{2.11}$$

Ist mindestens die Reliabilitätsschätzung der Einzelwerte (ICC_1) bekannt, lässt sich die theoretisch durch die Verwendung der Mittelwerte von k Beobachtern erreichbare Zuverlässigkeit über die Spearman-Brown-Formel

$$ICC_{(k)} = \frac{k \cdot ICC_1}{1 + (k-1) \cdot ICC_1} \tag{2.12}$$

schätzen.

2.5 Diskussion und Bewertung

Die in Abschnitt 2.1 vorgestellten Definitionen und Modelle der Intention bestätigen die zur Motivation dieser Arbeit führende Annahme (vgl. Abschn. 1.1), dass die Intention als Ursache menschlicher Handlung zu verstehen ist, und eine Intentionserkennung somit zu einer verbesserten Prädiktion der zukünftigen Situation führt. Dass besonders bei Fußgängern die Erkennung der Intention zur langfristigen Vorhersage des Verhaltens notwendig ist, wird von den in Abschnitt 2.2.1 aufgeführten Erkenntnissen über die theoretische Fußgängerdynamik bestätigt. Die großen Varianzen in den empirischen Ergebnissen zeigen hierbei, dass eine rein auf physikalischen Dynamikparametern basierende Prädiktion nicht zielführend sein kann, um die Position eines Fußgängers langfristig vorherzusagen.

Die vorgestellten Definitionen und Modelle zur Intention bestätigen jedoch auch, dass die Intention eine nicht direkt beobachtbare Größe ist und für ein intentionserkennendes Systems daher zunächst operationalisiert werden muss. Während die in Abschnitt 2.1.1 aufgeführten handlungstheoretischen Modelle hierbei nahelegen, dass sich Intentionen in Handlungen und Verhaltensweisen zeigen, bilden die in Abschnitt 2.1.2 vorgestellten beobachtungsbasierten Modelle den, für die Intentionserkennung nötigen, Bezug zu beobachtbaren Verhaltensweisen ab. Vor allem das Jordan-Modell von Diederichs (2017) bestätigt hierbei, dass eine Handlungsvorhersage bereits vor der Beobachtung handlungsausführender Verhaltensweisen möglich ist, da bereits über die Beobachtung von handlungsvorbereitenden und handlungsinitiierenden Verhaltensweisen auf die Bildung einer Intention geschlossen werden kann. Das Jordan-Modell wurde jedoch für die Fahrerintentionserkennung entwickelt und bisher ausschließlich für diesen Anwendungsfall validiert. Wie die folgende Analyse der in Abschnitt 2.2.2 beschriebenen Erkenntnisse über das Querungsverhalten von Fußgängern zeigt, kann das Modell aber auch auf die Querungsintention von Fußgängern angewendet werden. Abbildung 2.9 gibt einen zusammenfassenden Überblick über die Validierungsergebnisse.

Wie in Abschnitt 2.1.2 erläutert, entsteht im Jordan-Modell die Intention aus dem Zusammenspiel zwischen Voraussetzungen, die Handlungen erst ermöglichen und diese

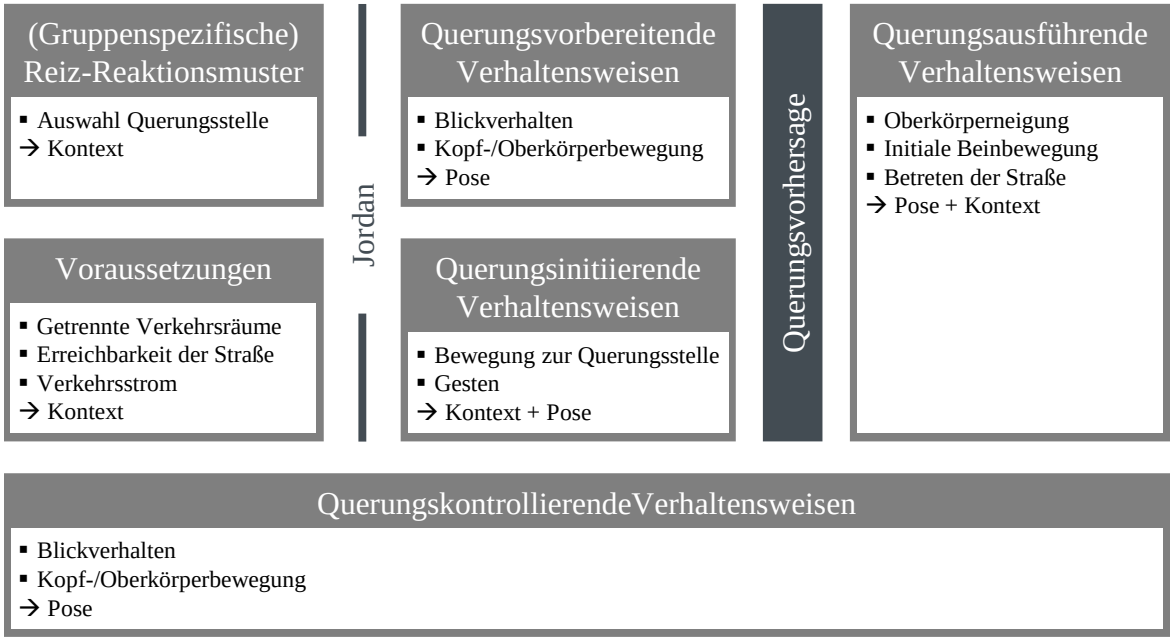


Abbildung 2.9: Die Validierung des Jordan-Modells für die Querungsintention von Fußgängern.

gegebenenfalls stimulieren sowie persönlichen Reiz-Reaktionsmustern, die bereits in der Vergangenheit beobachtet werden konnten. Auch im Anwendungsfall der Fußgängerintentionserkennung kann ein Fußgänger eine Querungsintention nur dann bilden, wenn die **Voraussetzungen** zur Ausführung der Straßenquerung überhaupt gegeben sind: Die Verkehrsräume zwischen Fahrzeug- und Fußverkehr müssen getrennt sein, die Straße, als Verkehrsraum des Fahrzeugverkehrs, muss durch den Fußgänger prinzipiell erreichbar, also nicht durch bauliche Maßnahmen wie Mauern oder Zäune abgesperrt sein und der Verkehrsstrom lässt prinzipiell eine sichere Querung zu.

Da ein System zur Fußgängerintentionserkennung in den seltensten Fällen mehrfach auf denselben Fußgänger trifft, ist die Verwendung der im Jordan-Modell beschriebenen persönlichen Reiz-Reaktionsmuster bei dem hier betrachteten Anwendungsfall nicht zielführend. Aus den empirischen Studien zur Auswahl der Querungsstelle (Abschn. 2.2.2, S. 18 f.) sind jedoch **Reiz-Reaktionsmuster** bekannt, die für mindestens eine bestimmte Teilgruppe von Fußgängern gelten. So weisen beispielsweise angebotene Querungshilfen und Möglichkeiten zur adäquaten Querungssicherung eine hohe

Attraktivität bei der Wahl der konkreten Querungsstelle auf. Solche kontextbasierten Szeneninformationen können somit einen Beitrag bei der Erkennung der Querungsintention von Fußgängern leisten. Dieser Kontextbezug ist konform mit dem Menschmodell zur Intentionserkennung von Schrempf (2008) (vgl. Abschn. 2.1.2, S. 13).

Trotz der vielen bekannten Faktoren bezüglich der Auswahl der Querungsstelle, bleibt neben der nicht beobachtbaren Routenwahl des Fußgängers immer ein bedeutsamer Anteil an Zufall, der situativ über die Ausprägung einer Querungsintention entscheidet. Somit gilt auch im Anwendungsfall der Fußgängerintentionserkennung: Ob und wann der „Jordan“ tatsächlich überschritten ist und damit eine Querungsintention seitens des Fußgängers feststeht, ist erst messbar, wenn beobachtbare Verhaltensweisen beginnen.

Analog zu der Unterteilung von Diederichs (2017) können auch bei querungswilligen Fußgängern querungsvorbereitende, querungsinitiierende und querungskontrollierende Verhaltensweisen beobachtet werden. Wie die vorgestellten empirischen Studien zum Sicherungsverhalten von Fußgängern zeigen (Abschn. 2.2.2, S. 19 ff.), überprüft ein querungswilliger Fußgänger die Voraussetzungen für die geplante Handlung in der Regel durch ein bewusst ausgeführtes Sicherungsverhalten. Die beobachtbaren Verhaltensweisen zeigen sich dabei vor allem in einem aufgabenabhängigen Blickverhalten und einer daraus resultierenden typischen Kopf- und Oberkörperbewegung. Abhängig davon, ob die Blicke dabei explorativ, ohne spezielles Ziel oder wiederholt und zielgerichtet auf bereits vorher als relevant identifizierte Ziele erfolgen, ist dieses Verhalten den **querungsvorbereitenden** oder den **querungskontrollierenden Verhaltensweisen** zuzuordnen. Da die Unterscheidung der zwei Phasen und damit die diskrete Trennbarkeit des beobachteten Verhaltens in der Praxis nicht immer möglich ist, können diese zwei Phasen auch auf eine reduziert werden. Da eine Trennung aus kognitiver Sicht jedoch sinnvoll ist, empfiehlt sich diese auch im Anwendungsfall der Fußgängerintention beizubehalten.

Als **querungsinitiierende Verhaltensweisen** kann bei querungswilligen Fußgängern die Bewegung des Fußgängers zur gewählten Querungsstelle und damit eine Ausrichtung seiner Position und Orientierung an die Fahrbahnkante beobachtet werden.

In wenigen Fällen kommt ein aktives Anzeigen der geplanten Querung durch Gesten hinzu.

Die beobachtbaren querungsvorbereitenden, -kontrollierenden und -initiierenden Verhaltensweisen basieren somit sowohl auf der, die Körperhaltung sowie die Körper- und Kopfbewegung umfassende Pose des Fußgängers, als auch auf kontextuellen Informationen, wie der Position und Orientierung des Fußgängers relativ zur Fahrbahnkante und zu Fußgängerüberwegen. Alle drei Verhaltensweisen können dabei, wie im Jordan-Modell dargestellt, auch bei Fußgängern zeitgleich oder in vermischter zeitlicher Abfolge auftreten. Dieses wird durch die Beobachtung von Kloeden et al. (2014) (vgl. Abschn. 2.2.2, S. 20) gestützt, dass sich Fußgänger bereits vor dem Erreichen der Bordsteinkante, noch während des Gehens, visuell absichern.

Über eine Erkennung der beschriebenen Verhaltensweisen kann also eine bevorstehende Querung der Straße durch den Fußgänger vorhergesagt werden. Wie der Teilabschnitt zum Querungsbeginn zeigt (Abschn. 2.2.2, S. 22 f.), hängt der genaue Zeitpunkt der Querungsausführung jedoch von der Verkehrssituation und dem Verhalten des sich nähernden Fahrzeugs ab. Dieses bestätigt die in der Motivation zu dieser Arbeit angeführten Annahme, dass zukünftige automatisierte Fahrzeuge, durch ein entsprechend gestaltetes Annäherungsverhalten, den Fußgänger bezüglich seiner Querungsausführung beeinflussen könne; wobei als Grundvoraussetzung dafür zunächst die Querungsintention des Fußgängers erkannt werden muss.

Die konkrete Querungsausführung wird schließlich durch das Betreten der Straße durch den Fußgänger initiiert. Wie die Analyse von Köhler et al. (2012) zeigt (vgl. Abschn. 2.3.2, S. 33), können jedoch bereits vor dem ersten Schritt auf die Straße **querungsausführende Verhaltensweisen**, wie eine Neigung des Oberkörpers oder eine Initialbewegung der Beine, beobachtet werden. Die Studien über das Fußgängerverhalten während der Querung (Abschn. 2.2.2, S. 23 f.) zeigen zudem, dass ein Fußgänger eine einmal begonnene Querung in der Regel konsequent durchzieht, sich dabei aber weiterhin visuell absichert. Diese Beobachtung unterstreicht damit die Darstellung des Jordan-Modells, das handlungskontrollierende Verhaltensweisen kontinuierlich über die gesamte Dauer der Handlung auftreten.

Die Erkenntnisse über die bei querungswilligen Fußgängern beobachtbaren Verhaltensweise zusammenfassend, scheinen **kontextbasierte** und **posenbasierte Informationen** gut geeignet zu sein, um die Basis für ein technisches System, das die Querungsintention von Fußgängern erkennt, zu bilden. Wie in Abschnitt 2.3 gezeigt, bilden diese zwei Kategorien auch die Informationsbasis bei bisherigen Ansätzen zur Erkennung und Vorhersage von Fußgängerverhalten.

Neben der Beschreibung dieser Informationen muss das zu entwickelnde intentionserkennende System auch in der Lage sein, die Ausprägung der Posen- und Kontextdaten eines querungswilligen Fußgängers von denen eines Fußgängers ohne Querungsintention zu unterscheiden. Anstatt die Grenzen explizit zu modellieren, zeigt der Stand der Technik, dass es empfehlenswert ist, diese anhand von Beispielen zu erlernen. Hierzu kann auf Methoden des **maschinellen Lernens** zurückgegriffen werden (s. Kap. 3).

Zum Lernen der intentionstypischen Ausprägungen der Daten sowie zur Bewertung der Leistungsfähigkeit des zu entwickelnden Systems muss für die Beispieldaten jedoch zunächst eine **Referenz** bestimmt werden. Wie in Abschnitt 2.4.1 erläutert, hat die beim aktuellen Stand der Technik übliche Verwendung des im späteren Verlauf tatsächlich gezeigten Verhaltens des Fußgängers als Referenz bei der Intentionserkennung den Nachteil, dass keine Intentionen berücksichtigt werden, die letztlich nicht zu Handlungen führen. Zudem kann nicht garantiert werden, dass solche Ground-Truth-basierten Referenzen das Situationsbewusstsein eines menschlichen Fahrers widerspiegeln, da das in Zukunft gezeigte Verhalten des Fußgängers nicht der aktuellen Einschätzung eines menschlichen Beobachters entsprechen muss. Daher scheint die in Abschnitt 2.4.2 vorgestellte Methode der beobachterbasierten Videoannotation zur Referenzbildung für intentionserkennende Systeme geeignet zu sein. Die Zuverlässigkeit und Genauigkeit der Methode kann über die Übereinstimmungs- und Reliabilitätsmessung überprüft werden. Der Prozess der Referenzbildung mittels einer beobachterbasierten Videoannotation für das in dieser Arbeit entwickelte System wird in Kapitel 4 beschrieben.

Durch das Heranziehen menschlicher Beobachter ist der Prozess zur Referenzbildung jedoch sehr aufwendig und zeitintensiv. Daher ist die in dieser Arbeit zur Verfügung stehende Datenmenge begrenzt. Das schränkt die für diese Arbeit geeigneten Methoden des maschinellen Lernens auf **merkmalsbasierte Methoden** ein, bei denen die

zu bewertenden Objekte in Form manuell gestalteter Merkmale oder Eigenschaften repräsentiert werden. Denn bei den alternativ zu manuell gestalteten Merkmalen einsetzbaren Methoden aus dem Bereich des **Deep Learning** ist der Bedarf an Trainingsbeispielen deutlich höher, da die tiefen neuronalen Netze die zur Unterscheidung der Daten relevanten Merkmale während des Trainingsprozesses selbst erlernen (Goodfellow et al., 2016). Zudem bildet die Gestaltung der Eingangsdaten für die tiefen neuronalen Netze eine Herausforderung, da jeder Fußgänger im Bild einzeln bezüglich der Ausprägung einer Querungsintention bewertet werden muss, eine klassischerweise eingesetzte Begrenzung des Eingangsbilds auf die Fußgänger Bounding Box jedoch dazu führt, dass die kontextuellen Informationen zum Fußgänger dem neuronalen Netz nicht zur Verfügung stehen. Das nachfolgende Kapitel 3 beschränkt sich bei der Darstellung des Hintergrunds zum maschinellen Lernen daher auf den für diese Arbeit relevanten Teil der merkmalsbasierten Methoden.

Die bisher aus dem Stand der Technik bekannten Ansätze zur Erkennung des Fußgängerverhaltens setzen ebenfalls überwiegend auf merkmalsbasierte Ansätze. Daher scheinen die aus dem Stand der Technik bekannten Ansätze, in einer angepassten Form, prinzipiell auch für die Erkennung der Querungsintention von Fußgängern geeignet zu sein. Gleichwohl sich bisher nur ein Verfahren, nach dem in dieser Arbeit gesetzten Begriffsverständnis, ebenfalls mit der Erkennung der Fußgängerintention beschäftigt (vgl. Abschn. 2.3.3).

Vor allem die bei der **posenbasierten** Aktionserkennung (Abschn. 2.3.2, S. 31 ff.) verwendeten Merkmale zur Beschreibung der **Körperbewegung** des Fußgängers versprechen eine gute Möglichkeit, die für querungswillige Fußgänger typische Körperbewegung abzubilden. Hierzu zählen die Histograms of Orientation Motion (HOM) von Keller et al. (2011), die Lateral Scene Flow Features (LSFF) von Keller und Gavrila (2014) und die Motion Contour Histograms of Oriented Gradients (MCHOG) von Köhler et al. (2012). Im Gegensatz zu den Systemen von Keller und Gavrila stehen bei dem in dieser Arbeit zu entwickelnden System jedoch keine Tiefendaten zur Verfügung. Daher können hier nur die MCHOG zum Einsatz kommen. Anders als bei Köhler et al. soll das zu entwickelnde System jedoch auf Daten nichtstationärer Kameras basieren.

Daher muss geprüft werden, ob die MCHOG dazu geeignet sind, die Körperbewegung eines Fußgängers, trotz der veränderten Datenbasis, abzubilden.

Durch die Verwendung des Motion History Image (MHI) bilden die MCHOG zwar die Körperbewegung des Fußgängers explizit ab, nicht aber die **Körperhaltung** eines still stehenden Fußgängers. Diese kann jedoch gerade bei am Fahrbahnrand stehenden Fußgängern zur Differenzierung von Fußgängern mit und ohne Querungsintention entscheidend sein. Diese Anforderung wird von dem auf der Verkettung zeitlich versetzter HOG basierenden PAF von Furuhashi und Yamada (2011) erfüllt. Ob eine solche implizite Erfassung der Historie jedoch ausreichend ist, um die relevante Körperbewegung eines Fußgängers zu beschreiben, muss erst evaluiert werden.

Zudem ist fraglich, ob sowohl die MCHOG als auch das PAF die für die Querungsintentionserkennung wichtige **Kopfbewegung** des Fußgängers ausreichend genug beschreiben. Denn keines der beiden Merkmale weist dieser eine, im Vergleich zur restlichen Körperbewegung, erhöhte Aufmerksamkeit zu. Abhilfe können hier gegebenenfalls Merkmale wie das Local Binary Pattern (LBP) (s. Abschn. 3.4.2) schaffen, das bereits erfolgreich zur Erkennung der Kopforientierung von Fußgängern eingesetzt wurde.

Zusammenfassend bedarf es zur posenbasierten Erkennung der Querungsintention von Fußgängern einer Evaluation der zwei aus dem Bereich der posenbasierten Aktionserkennung bekannten Merkmale, die auf Basis von Mono-Videodaten grundsätzlich generiert werden können. Bestätigen sich die vermuteten Schwächen bezüglich der expliziten Abbildung der Kopf- und Körperbewegung sowie der statischen Körperhaltung, sind entsprechende Ableitungen für die Entwicklung eines neuen Merkmals, das all diese als relevant identifizierten Informationen abbildet, zu treffen. Die in diesem Rahmen betrachteten Merkmale und angepassten Merkmalskombinationen werden in Abschnitt 5.3 vorgestellt.

Zur Beschreibung der querungsinitiierenden Verhaltensweisen, wie die Bewegung des Fußgängers zur Querungsstelle, scheinen **kontextbasierte** Merkmale, wie die von Kooij et al. (2014) oder Bonnin et al. (2014b) zur Aktionserkennung (Abschn. 2.3.2, S. 35 ff.) sowie die von Voelz et al. (2015) zur Intentionserkennung (Abschn. 2.3.3) verwendeten Merkmale, geeignet zu sein. Alle drei Ansätze modellieren den kontextuellen Bezug zwischen Fußgängern und Szenenelementen über die laterale und/oder

temporale Distanz des Fußgängers zu den Szenenelementen. Die Historie der kontextuellen Fußgängerbewegung muss hierbei jedoch explizit modelliert werden, beispielsweise über eine Aneinanderreihung der vergangenen Werte. Das hat den Nachteil, dass die Größe des Merkmalsvektors mit steigender Historie wächst bzw. der betrachtbare Zeithorizont durch die, vor dem Klassifikatortraining zu definierende Größe des Merkmalsvektors beschränkt ist.

Durch die Verwendung von Punktdistanzen beinhalten die Merkmale zudem keine weiteren Informationen über die Szenenelemente, wie beispielsweise die Krümmung der Straße oder die Breite des Zebrastreifens. Diese Eigenschaften können aber zu einem situativ unterschiedlichen Bewegungsverhalten seitens des Fußgängers führen und sind daher wichtig für die Erkennung der Querungsintention von Fußgängern. Es fehlt somit an einer allgemeinen Beschreibungsform, um das kontextuelle Bewegungsverhalten eines Fußgängers abbilden und darüber seine Querungsintention erkennen zu können. Wie konzeptionell aus der zielgerichteten Langzeit-Trajektorienprädiktion (Abschn. 2.3.1, S. 27 ff.) bekannt, sollte der Einfluss einzelner Szenenelemente dabei getrennt voneinander beschrieben werden, da nur so eine hohe Generalisierungsfähigkeit und Übertragbarkeit auf neue Szenarios gewährleistet wird. Das auf Basis dieser Bewertung neu entwickelte, kontextbasierte Merkmal zur Erkennung der Querungsintention wird in Abschnitt 5.2 vorgestellt.

Kapitel 3

Hintergrund des maschinellen Lernens

Maschinelles Lernen ist eine Methode der künstlichen Intelligenz, bei der ein System anhand repräsentativer Beispieldaten eine allgemeine Konzeptbeschreibung erlernt, um diese anschließend zur Beurteilung neuer Daten heranzuziehen. Dieses Kapitel stellt den für das Verständnis des eigenen Ansatzes notwendigen Hintergrund dieses Themenkomplexes vor. Entsprechend der Diskussion in Abschnitt 2.5 beschränkt sich diese Arbeit dabei auf merkmalsbasierte Methoden, bei denen die zu bewertenden Objekte in Form manuell gestalteter Merkmale repräsentiert werden.

Abschnitt 3.1 führt mit einem kurzen Überblick über die verschiedenen Arten des maschinellen Lernens in den Themenkomplex ein. In Abschnitt 3.2 folgt eine detailliertere Vorstellung des im eigenen Ansatz verwendeten Lernverfahrens, der Support Vector Machine. Der Fokus liegt hierbei auf der Vorstellung des grundsätzlichen Konzepts, sowie auf den beim Lernen mit unausgewogenen Daten entstehenden Herausforderungen. Abschnitt 3.3 stellt anschließend das Vorgehen zur Beurteilung der erlernten Modelle vor. Auch hier liegt ein besonderes Augenmerk auf der Beurteilung bei unausgewogenen Daten. In Abschnitt 3.4 werden visuelle Deskriptoren vorgestellt, die im Bereich Computer Vision erfolgreich zur Repräsentation von Bilddaten verwendet werden. Abschließend erfolgt in Abschnitt 3.5 eine Bewertung des Vorgestellten, um

ein geeignetes Vorgehen zur Entwicklung eines trainierten Systems zur Erkennung der Querungsintention von Fußgängern abzuleiten.

3.1 Arten des maschinellen Lernens

Die Methoden des maschinellen Lernens werden, abhängig davon, ob während der Trainingsphase die Zielwerte bekannt sind oder nicht, in zwei Hauptkategorien unterteilt: überwachtes Lernen (*supervised Learning*) und unüberwachtes Lernen (*unsupervised Learning*).

Im Folgenden werden die für diese Arbeit relevanten Arten des maschinellen Lernens vorgestellt. Für eine umfassendere Darstellungen wird auf (Murphy, 2012) verwiesen. Die folgende Darstellung basiert, wenn nicht anders angegeben, auf dieser Quelle.

3.1.1 Überwachtes Lernen

Beim überwachten Lernen werden Problemstellungen betrachtet, bei denen die als Labels oder Ground Truth bezeichneten Zielwerte der Beispieldaten bekannt sind. Der Trainingsdatensatz besteht mit

$$D = \{\boldsymbol{x}_i, y_i\}_{i=1}^{N} \tag{3.1}$$

somit aus N beispielhaften Eingangsinstanzen $\boldsymbol{x}_i$, die jeweils mit einem Label y_i verknüpft sind. Jede Eingangsinstanz beschreibt dabei problemspezifische Eigenschaften oder Merkmale eines Objekts in Form eines d-dimensionalen Merkmalsvektors $\boldsymbol{x}_i = [x_1, x_2, \cdots, x_d]$.

Das Ziel des Lernprozesses ist die Bildung eines Modells $y = f(x)$, das den Zusammenhang zwischen den Eingangsdaten und den Zielwerten generalisiert abbildet. Abhängig davon, ob der Zielwert ein kategorischer oder ein numerischer Wert ist, werden zur Modellbildung Algorithmen der **Klassifikation** oder der **Regression** eingesetzt. Das gewonnene Modell kann schließlich auf unbekannte Daten zur Prädiktion der Zielwerte $\hat{y}$ angewendet werden.

Klassifikation

Bei der Klassifikation wird der Merkmalsvektor $\boldsymbol{x}_i$ auf einen diskreten Ausgabewert y_i abgebildet, beispielsweise mit $y_i \in \{0, \ldots, M-1\}$, wobei M die Anzahl der Klassen beschreibt. Ist $M = 2$, wird von einem binären Klassifikationsproblem mit negativen ($y_i = -1$) und positiven ($y_i = 1$) Daten gesprochen. $M > 2$ beschreibt ein Multiklassen-Problem.

Abhängig vom verwendeten Klassifikationsalgorithmus kann nicht nur eine diskrete Klassenzugehörigkeit mit $\hat{y}_i \in \{0, \ldots, M-1\}$, sondern auch die Wahrscheinlichkeit $p(\hat{y}_i = M_j | \boldsymbol{x}_i)$, dass das durch $\boldsymbol{x}_i$ repräsentierte Objekt zur Klasse M_j gehört, prädiziert werden.

Zudem existieren Methoden im Bereich des Learning on Probabilistic Labels (LoPL), die mit unsicheren Labels während des Trainingsprozesses umgehen können (Nguyen et al., 2014; Peng et al., 2014). Methoden des LoPL zielen somit auf Probleme, bei denen die genaue Klassenzugehörigkeit eines Trainingsbeispiels nicht bekannt ist. Der Trainingsdatensatz beinhaltet hierzu mit

$$D = \{\boldsymbol{x}_i, y_i, b_i\}_{i=1}^{N} \tag{3.2}$$

ein zusätzliches weiches Label b_i mit $\{b_i \in \mathbb{R} | 0 \leq b_i \leq 1\}$, dass die Wahrscheinlichkeit angibt, mit der die durch $\boldsymbol{x}_i$ repräsentierte Instanz zu der durch y_i angegebenen Klasse gehört.

Regression

Im Gegensatz zur Klassifikation ist der Zielwert bei der Regression numerisch mit $y_i \in \mathbb{R}$. Somit wird bei der Regression keine Trennebene zwischen den Klassen modelliert, sondern der kontinuierliche Verlauf des in den Daten vorhandenen Musters abgebildet. Abbildung 3.1 skizziert den Unterschied zwischen der Klassifikation und der Regression.

Wird ein Regressionsmodell auf probabilistische Eingangsdaten angepasst, können die Ausgabewerte des Modells außerhalb des Wertebereichs $[0, 1]$ fallen und wären damit Inkonsistenz mit dem Wertebereich von Wahrscheinlichkeiten. Alternativ können die

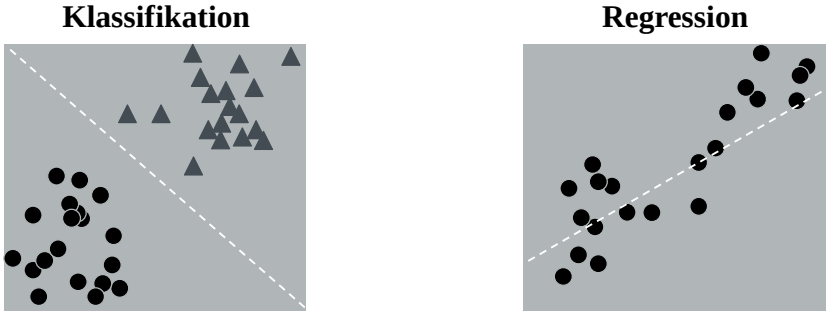


Abbildung 3.1: Schematischer Unterschied zwischen der Klassifikation und der Regression. Darstellung nach (Rossant, 2014).

Eingangsdaten vor der Anpassung des Regressionsmodells mit der, im Intervall $[0, 1]$ monotonen Logit Funktion L, mit

$$L(y_i) = ln\frac{y_i}{1 - y_i} \tag{3.3}$$

transformiert werden. Das Regressionsmodell wird dann über den Trainingsdatensatz

$$D = \{\boldsymbol{x}_i, L(y_i)\}_{i=1}^{N} \tag{3.4}$$

bestimmt. Über die Inverse der Logit Funktion

$$L^{-1}(\hat{y}) = \frac{e^{\hat{y}}}{e^{\hat{y}} + 1} \tag{3.5}$$

lässt sich der Ausgabewert des Modells schließlich zurück in den probabilistischen Eingangsraum transformieren (Nguyen et al., 2014).

Überanpassung und Unteranpassung

Sowohl bei der Klassifikation, als auch bei der Regression kann es bei der Modellbildung zu zwei Arten von Fehlern kommen: zur Überanpassung (*Overfitting*) oder zur Unteranpassung (*Underfitting*) (s. Abb. 3.2). Bei einer Überanpassung werden auch sehr kleine Variationen innerhalb der Eingangsdaten modelliert. Die im Trainingsdatensatz vorgegebenen Beziehungen zwischen den Eingangsinstanzen und den Labeln werden hierbei eher auswendig gelernt, als dass das dahinter liegende Muster erkannt

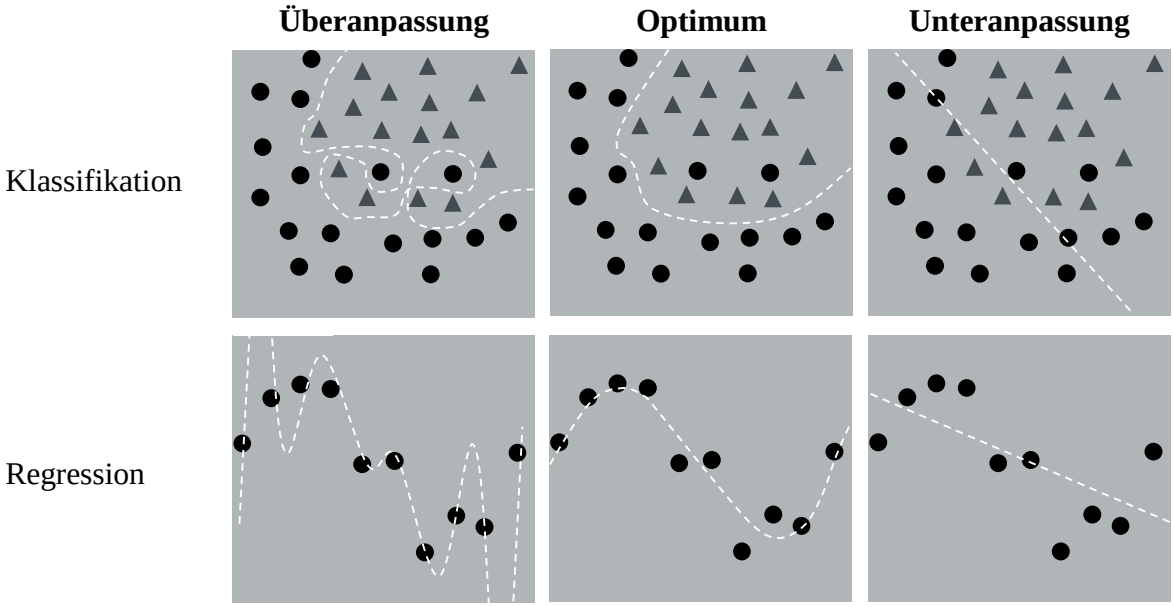


Abbildung 3.2: Beispiele für überangepasste, optimale und unterangepasste Klassifikations- sowie Regressionsmodelle. Darstellung nach (Lavrenko und Goddard, 2014).

wird. Dadurch besitzen überangepasste Modelle eine schlechte Generalisierungsfähigkeit und können unbekannte Daten nur unzureichend vorhersagen. Eine Überanpassung entsteht vor allem bei der Anwendung übermäßig komplexer Modelle oder wenn die Dimension d der Eingangsdaten im Verhältnis zur Anzahl der Trainingsbeispiele N zu groß ist. Überangepasste Modelle können über die in Abschnitt 3.3.1 vorgestellte Kreuzvalidierung erkannt werden.

Von einer Unteranpassung wird gesprochen, wenn ein Modell das in den Daten enthaltene Trennmuster nicht abbilden kann. Dies ist beispielsweise der Fall, wenn ein lineares Modell auf nicht lineare Daten angepasst wird. Im Gegensatz zur Überanpassung können unterangepasste Modelle den Ausgabewert selbst bei den Trainingsdaten nur unzureichend vorhersagen.

Das Problem, beide Fehlerquellen (Überanpassung und Unteranpassung) gleichzeitig zu minimieren, wird als **Bias-Varianz-Dilemma** bezeichnet. Der Bias beschreibt dabei den, durch das gewählte Modell bedingten, systematischen Fehler. Er gibt somit an, wie präzise ein Modell in Bezug auf die Trainingsdaten ist und kann durch eine

Erhöhung der Modellkomplexität verkleinert werden. Die Varianz beschreibt, wie sensitiv das Modell bezüglich kleiner Änderungen in den Trainingsdaten ist. Ein robustes Modell sollte nicht übermäßig sensitiv bezüglich kleiner Änderungen sein. Das Dilemma besteht darin, dass Bias- und Varianzfehler im Konflikt stehen und stets nur ein Kompromiss möglich ist. Einfache Modelle sind weniger präzise, dafür aber robuster (→ Unteranpassung). Komplexere Modelle sind präziser, reagieren aber auch sensibler auf kleine Änderungen (→ Überanpassung) (Rossant, 2014).

3.1.2 Unüberwachtes Lernen

Unüberwachtes Lernen wird eingesetzt, wenn die Zielwerte der Beispieldaten nicht bekannt sind. Die Lernaufgabe besteht hierbei in der Erkennung von Mustern innerhalb der Beispieldaten. Dieses kann zur automatischen Segmentierung der Daten oder zur Reduktion der Datendimension herangezogen werden.

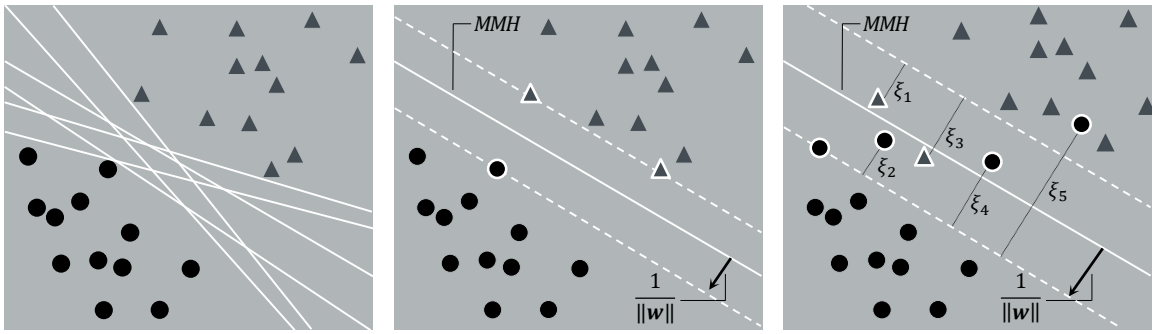
3.2 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) sind Methoden des überwachten Lernens, die sowohl zur Klassifikation als auch zur Regression eingesetzt werden können. In ihrer heutigen Form wurde die SVM von Cortes und Vapnik (1995) vorgestellt und zählt bis heute zu den beliebtesten Methoden im Bereich des Merkmalsbasierten Lernens. Im Folgenden wir das zur Klassifikation und Regression eingesetzte Konzept der SVM erklärt. Die Darstellung basiert, wenn nicht anders angegeben, auf (James et al., 2013).

3.2.1 SVMs zur Klassifikation

SVMs basieren auf der Idee, die durch ihre Merkmalsvektoren $\boldsymbol{x}_i$ repräsentierten Objekte mittels einer linearen Hyperebene

$$\langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b = 0 \tag{3.6}$$



(a) (b) (c)

Abbildung 3.3: Die Hyperebene einer SVM. (a) Sind die Daten linear trennbar, existieren beliebig viele Trennebenen. (b) Die Maximum Margin Hyperebene (MMH) verspricht die beste Generalisierungsfähigkeit. Ihre Lage wird durch die Support Vektoren (weiß umrandete Objekte) definiert. (c) Bei einem *Soft Margin* ist es den Trainingsbeispielen erlaubt, auf der falschen Seite der Margin-Grenze ($\xi_1, \xi_2 > 0$) oder auf der falschen Seite der Hyperebene ($\xi_3, \xi_4, \xi_5 > 1$) zu liegen. Darstellung nach (Hastie et al., 2009).

in zwei Klassen zu unterteilen. Die Hyperebene wird dabei durch den Normalenvektor $\boldsymbol{w}$ und den Bias b beschrieben. Mit $y_i \in \{-1, 1\}$ kann die Klassenzugehörigkeit eines Objekts somit über die Entscheidungsfunktion

$$\hat{y}_i = sgn\left(\langle \boldsymbol{w}, \boldsymbol{x_i} \rangle\right) + b \tag{3.7}$$

bestimmt werden. Die Entscheidungsfunktion sagt dabei aus, ob der Merkmalsvektor $\boldsymbol{x_i}$ oberhalb oder unterhalb der Hyperebene liegt und damit zur positiven oder negativen Klasse gehört. Die Ebenenparameter $\boldsymbol{w}$ und b werden während des Lernvorgangs auf Basis des Trainingssatzes bestimmt.

Maximum Margin Hyperebene

Sind die Trainingsdaten linear trennbar, gibt es jedoch unendlich viele Hyperebenen, die die beiden Klassen voneinander trennen (s. Abb. 3.3a). Die beste Generalisierungsfähigkeit verspricht dabei die Hyperebene, die einen maximalen Abstand $\frac{1}{\|\boldsymbol{w}\|}$ zu den am nächsten gelegenen Trainingsbeispielen hat (s. Abb. 3.3b). Diese Ebene wird als

MMH bezeichnet und kann durch die Lösung des folgenden Optimierungsproblems gefunden werden:

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 \\ \text{s.t. } &y_i\left(\langle \boldsymbol{w}, \boldsymbol{x_i}\rangle + b\right) \geq 1 \quad \forall\, i = 1,\ldots,N. \end{aligned} \tag{3.8}$$

Dabei ist die Lage der MMH bereits durch die Trainingsbeispiele definiert, die direkt auf dem Margin liegen (s. Abb. 3.3b). Diese Stützvektoren werden als **Support Vektoren** bezeichnet.

Soft Margin Hyperebene

Die bisher angenommene perfekte Trennbarkeit der Trainingsbeispiele liegt in der Praxis nur selten vor. In solchen Fällen können die Daten durch einen *Soft Margin* getrennt werden. Bei diesem ist es den Trainingsbeispielen erlaubt, auf der falschen Seite der Margin-Grenze oder der falschen Seite der Hyperebene zu liegen. Um die Nebenbedingung der Gleichung 3.8 hier hingehend verletzen zu können, werden Schlupfvariablen ξ_i eingeführt, die die Lage jedes Trainingsbeispiels angeben (s. Abb. 3.3c):

$\xi_i = 0 : \boldsymbol{x_i}$ liegt auf der korrekten Seite der Margin-Grenze

$\xi_i > 0 : \boldsymbol{x_i}$ liegt auf der falschen Seite der Margin-Grenze

$\xi_i > 1 : \boldsymbol{x_i}$ liegt auf der falschen Seite der Hyperebene

Das zu lösende Optimierungsproblem ändert sich damit zu:

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\xi_i \\ \text{s.t. } &y_i\left(\langle \boldsymbol{w}, \boldsymbol{x_i}\rangle + b\right) \geq 1 - \xi_i \\ &\xi_i \geq 0, \qquad\qquad \forall\, i = 1,\ldots,N. \end{aligned} \tag{3.9}$$

Der in dieser Arbeit verwendeten Implementierung von (Pedregosa et al., 2011) entsprechend, ist C hierbei ein positiver Modellparameter, über den die Bestrafung bei einer Verletzung des Margin gewichtet wird. C beeinflusst somit die Größe des Margin,

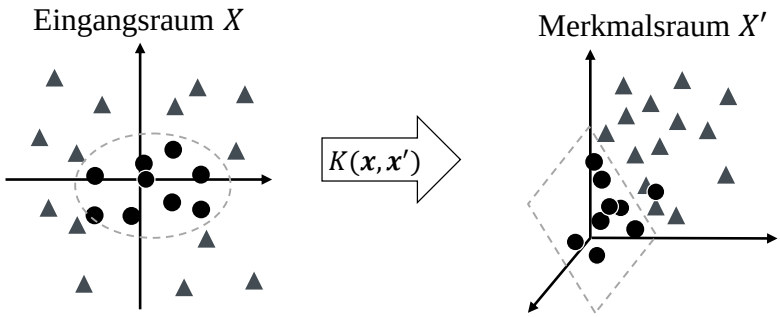


Abbildung 3.4: Der Kernel Trick bei SVMs. Nicht-linear trennbare Daten werden in einen höherdimensionalen Raum projiziert, in dem eine lineare Trennung einfacher ist. Darstellung nach (Thornton, 2008).

die wiederum Einfluss auf die Anzahl der Support Vektoren hat: je kleiner C, desto größer ist der Margin und desto mehr Support Vektoren existieren.[1]

Zur Lösung des Optimierungsproblems werden die Nebenbedingungen mithilfe von Lagrange-Multiplikatoren in das Problem integriert. Details zur Lösung können (Wang, 2010) entnommen werden.

Es kann schließlich gezeigt werden, dass eine lineare SVM nur auf Basis der Skalarprodukte der Trainingsbeispiele mit

$$f(\boldsymbol{x}) = b + \sum_{i=1}^{N} \alpha_i \langle \boldsymbol{x}, \boldsymbol{x_i} \rangle, \tag{3.10}$$

repräsentiert werden kann, wobei der Koeffizient α_i für alle Trainingssamples, die keine Support Vektoren sind, gleich Null ist.

Kernel Trick

Liegen komplexere Datenstrukturen vor, reicht eine Trennung der Daten mittels einer linearen Trennebene oft nicht aus. Die SVM bedient sich daher des Kernel Tricks. Bei diesem werden die Eingangsdaten durch die Anwendung einer nicht-linearen Kernel-Funktion $K(\boldsymbol{x}, \boldsymbol{x}')$ in einen höherdimensionalen Raum projiziert, in dem eine lineare

[1] In anderen Darstellungen (z.B. in (James et al., 2013)), wird der Modellparameter C nicht als Gewichtungsfaktor, sondern mit $\sum_{i=1}^{N} \xi_i \leq C$ als maximale Summe der Schlupfvariablen definiert. Der Zusammenhang zwischen C und der Anzahl der Support Vektoren ist dann reziprok.

Trennung der Klassen einfacher ist (s. Abb. 3.4). Hierdurch werden Entscheidungsgrenzen ermöglicht, die nicht-linearen Trennungen im Eingaberaum entsprechen. $K(\boldsymbol{x}, \boldsymbol{x}')$ ersetzt dabei das Skalarprodukt in Gleichung 3.10:

$$f(\boldsymbol{x}) = b + \sum_{i=1}^{N} \alpha_i K(\boldsymbol{x}, \boldsymbol{x_i}), \tag{3.11}$$

Ein häufig verwendeter Kernel ist der Gaußsche Kernel, der auf der radialen Basisfunktion (RBF) basiert:

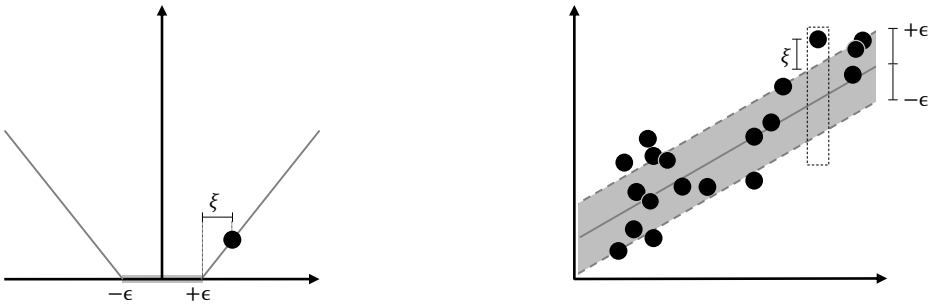
$$K_{RBF}(\boldsymbol{x}, \boldsymbol{x'}) = exp(-\gamma \|\boldsymbol{x} - \boldsymbol{x'}\|^2). \tag{3.12}$$

Der Skalierungsfaktor γ definiert dabei den Wirkungsbereich der Support Vektoren: je größer γ desto kleiner ist der lokale Wirkungsbereich. Hierdurch haben weniger Trainingsbeispiele einen Einfluss auf den Ausgabewert $\hat{y}$, was die Gefahr einer Überanpassung des Modells an die Trainingsdaten steigert. Ein zu kleines γ führt hingegen zu einer Unteranpassung des Modells.

Da die Parameter C und γ stark interagieren, wird die für die betrachteten Daten beste Parameterkombination in der Regel über ein *Grid Search*-Verfahren mit logarithmisch oder exponentiell steigenden Werten im Rahmen einer Kreuzvalidierung (s. Abschn. 3.3.1) gefunden (Murphy, 2012).

Probabilistische Klassifikation

Konzeptbedingt kann eine SVM keine direkte probabilistische Aussage $p(\hat{y}_i = M_j|\boldsymbol{x_i})$ über die Zugehörigkeit des durch $\boldsymbol{x_i}$ repräsentierten Objekts zur Klasse M_j treffen. Da der Abstand des Merkmalsvektors $\boldsymbol{x_i}$ zur Hyperebene als Indikator für die Konfidenz der Klassifikation interpretiert werden kann, können jedoch Methoden wie das **Platt Scaling** (Platt, 1999) zum Generieren probabilistischer Aussagen angewendet werden. Beim Platt Scaling wird eine Logistische Regression (Rohrlack, 2009) an den Ausgangswert eines Klassifikators angepasst, um eine Wahrscheinlichkeitsverteilung über die möglichen Klassen zu erhalten. Das Platt Scaling führt jedoch vor allem bei großen Trainingsdatensätzen zu einer deutlichen Erhöhung des Rechenaufwands während des Trainings (Pedregosa et al., 2011).



(a) (b)

Abbildung 3.5: SVMs zur Regression. (a) Die bei der SVR verwendete ϵ-insensitive Kostenfunktion mit einem beispielhaften Datenpunkt (b) Fehlprädiktionen innerhalb des ϵ-Schlauchs werden nicht bestraft, außerhalb des ϵ-Schlauchs ist der Bestrafungswert linear zum Abstand ξ des Datenpunkts zum ϵ-Schlauch. Darstellung nach (Murphy, 2012).

3.2.2 SVMs zur Regression

Das für die Klassifikation entwickelte Konzept der SVMs kann auch für Regressionsprobleme angewendet werden. Die von Drucker et al. (1996) hierzu vorgestellte Version der SVM wird auch als **Support Vector Regression (SVR)** bezeichnet.

Die SVR basiert auf der epsilon-insensitiven Kostenfunktion, die über

$$L_\epsilon(y_i,\hat{y_i}) = \begin{cases} 0 & \text{wenn } |y_i - \hat{y}_i| < \epsilon \quad (3.13a) \\ |y_i - \hat{y}_i| - \epsilon & \text{sonst} \quad (3.13b) \end{cases}$$

definiert ist (s. Abb. 3.5a). Die Kostenfunktion ignoriert somit alle Fehler $|y_i - \hat{y}_i|$, die kleiner als ein vorgegebene Wert ϵ sind. Vergleichbar mit dem Soft Margin bei der Klassifikation, ergibt sich hierdurch um die Prädiktion ein ϵ-Schlauch, innerhalb dessen Fehlprädiktionen nicht bestraft werden (s. Abb. 3.5b).

Unter Betrachtung des linearen Regressionsmodells

$$\hat{y}(\boldsymbol{x_i}) = \langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b \quad (3.14)$$

ergibt sich die zu minimierende Zielfunktion damit zu:

$$\min_{\boldsymbol{w},b_0} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N} L_\epsilon(y_i, \hat{y}_i). \quad (3.15)$$

Analog zur Klassifikation werden zur Lösung auch hier Schlupfvariablen (ξ_i^+ und ξ_i^-) eingeführt, die die Abweichung einer Instanz vom ϵ-Schlauch angeben. Das beim Training der SVR zu lösende Optimierungsproblem kann damit über

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum_{i=1}^{N}\left(\xi_i^+ + \xi_i^-\right) \\ \text{s.t. } & y_i - \hat{y}(\boldsymbol{x_i}) \leq \epsilon + \xi_i^+ \\ & \hat{y}(\boldsymbol{x_i}) - y_i \leq \epsilon + \xi_i^- \\ & \xi_i^+, \xi_i^- \geq 0, \qquad\qquad \forall\, i = 1,\ldots,N. \end{aligned} \tag{3.16}$$

beschrieben und analog zu dem Vorgehen bei der Klassifikation gelöst werden.

Auch bei der SVR kann für nicht linear trennbare Daten eine Kernel-Funktionen $K(\boldsymbol{x}, \boldsymbol{x}')$ eingesetzt werden. Die Prädiktion erfolgt dann über

$$\hat{y}(\boldsymbol{x_i}) = b + \sum_{i}^{N} \alpha_i K(\boldsymbol{x_i}, \boldsymbol{x_i}') \tag{3.17}$$

mit $\alpha_i > 0$ für alle Support Vektoren.

3.2.3 SVMs bei unausgewogenen Daten

Bei vielen Problemstellungen beinhaltet der Trainingsdatensatz unterschiedlich viele Trainingsbeispiele für die einzelnen Klasse. Eine solche ungleiche Klassenverteilung wird als *between-class imbalance* bezeichnet. Abhängig vom Verhältnis der häufig und der selten auftretenden Klasse, wird der Datensatz als marginal unausgewogen (2 : 1), mäßig unausgewogen (10 : 1) oder extrem unausgewogen (1000 : 1) betrachtet. Zudem existieren auch Problemstellungen, in dem einzelne Fälle innerhalb einer Klasse deutlich seltener vorkommen, als andere Fälle. Diese Unausgewogenheit wird als *within-class imbalance* bezeichnet. In beiden Situationen ist es sehr schwer, die seltenen Fälle zu lernen (He und Ma, 2013).

Wie auch bei vielen anderen maschinellen Lerntechniken, führt ein unausgewogener Datensatz bei den SVMs zu einer Verzerrung der Ergebnisse in Richtung der häufig auftretenden Fälle. Bei den SVMs ist dieses auf die bei der Suche nach der optimalen Hyperebene angestrebte Minimierung der Falschklassifikationskosten zurückzuführen

(vgl. Gl. 3.9). Da für alle Fehlklassifikationen, unabhängig von ihrer Klassenzugehörigkeit, derselbe Strafwert angewendet wird, ist die Minimierung der Falschklassifikationskosten mit einer Minimierung der absoluten Anzahl an Falschklassifikationen gleichzusetzen. Bei unausgewogenen Daten führt dies folglich dazu, dass die Hyperebene zu Gunsten der dichter verteilten, häufiger auftretenden Klasse verschoben wird. Dies kann im Extremfall dazu führen, dass eine SVM alle Beispiele stets der am häufigsten auftretenden Klasse zuweist (Akbani et al., 2004).

Um einen solchen Bias zu vermeiden, werden meistens externe Methoden zur Datenvorverarbeitung oder interne Methoden zur Modifikation des Lernalgorithmus eingesetzt. Zu den externen Methoden zählen die Sampling-Verfahren, bei denen durch ein Überabtasten der seltenen Beispiele (*Oversampling*) oder ein Unterabtasten der häufigen Beispiele (*Undersampling*) der Datensatz vor dem Trainingsprozess künstlich ausgeglichen wird. Bei den internen Methoden werden häufig die einzelnen Trainingsbeispiele mit Hilfe unterschiedlicher Fehlerkosten gewichtet. Im Fall einer Falschklassifikation wird dem Trainingsbeispiel ein Kostenwert zugewiesen, der von der Auftrittshäufigkeit der Klasse abhängig ist. Bei einem binären Klassifikationsproblem, bei dem deutlich mehr negative als positive Beispiele vorhanden sind, wird das Optimierungsproblem aus Gleichung 3.9 somit zu

$$\begin{aligned} &\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|^2 + C^+ \sum_{i|y_i=+1}^{N} \xi_i + C^- \sum_{i|y_i=-1}^{N} \xi_i \\ \text{s.t. } & y_i \left(\langle \boldsymbol{w}, \boldsymbol{x_i} \rangle + b\right) \geq 1 - \xi_i \\ & \xi_i \geq 0, \qquad\qquad \forall\, i = 1, \ldots, N. \end{aligned} \tag{3.18}$$

modifiziert, wobei C^+ und C^- jeweils die Fehlerkosten der positiven beziehungsweise negativen Beispiele beinhalten, mit $C^+ > C^-$. Für eine ausführlichere Darstellung der beim Lernen mit unausgewogenen Daten eingesetzten Methoden wird auf (He und Ma, 2013) verwiesen.

3.2.4 Vor- und Nachteile der SVMs

Ein großer Vorteil der SVMs ist, dass ihre Entscheidungsfunktion nur auf den Support Vektoren und nicht auf dem gesamten Trainingsdatensatz basiert. Dadurch kann

eine Klassifikation beziehungsweise Vorhersage sehr schnell erfolgen und der Speicherbedarf des Modells ist sehr gering. Zudem weisen SVMs im Allgemeinen eine hohe Generalisierungsfähigkeit auf, haben einen soliden mathematischen Hintergrund, sind gut anwendbar auf reale Probleme und arbeiten auch in hoch-dimensionalen Eingangsräumen effektiv. Durch die Möglichkeit, verschiedene Kernel-Funktionen zu verwenden, lassen sich SVMs außerdem sehr flexibel auf verschiedenste Datenstrukturen anwenden. Nachteile der SVM sind, dass der Lernvorgang recht zeitaufwendig ist und eine SVM beim Hinzufügen zusätzlicher Trainingsdaten neu angelernt werden muss. Zudem ist die Leistung einer SVM stark abhängig von der Wahl des C-Parameters, sowie der Wahl des Kernels und seiner Parameter. Die optimale Kombination muss im Rahmen eines aufwendigen Evaluierungsverfahrens empirisch ermittelt werden. Schließlich können SVMs durch ihren nicht-probabilistischen Charakter nur indirekt Aussagen über die Qualität einer Prädiktion treffen.

3.3 Beurteilung maschineller Lernverfahren

In diesem Abschnitt wird erläutert, wie die Leistung eines trainierten Modells beurteilt werden kann. Abschnitt 3.3.1 stellt die hierzu eingesetzte Technik der Kreuzvalidierung vor. Abschnitt 3.3.2 und Abschnitt 3.3.3 gehen anschließend jeweils auf die bei der Klassifikation und bei der Regression verwendeten Beurteilungsmetriken ein.

3.3.1 Kreuzvalidierung

Um die Leistung maschineller Lernverfahren beurteilen zu können, muss die Evaluation mit Daten durchgeführt werden, die nicht während des Trainingsprozesses verwendet wurden. Andernfalls kann nicht überprüft werden, ob eine Überanpassung des Modells an die Trainingsdaten vorliegt (vgl. Abschn. 3.1.1, S. 58). Hierzu werden in der Regel die vorhandenen Trainingsbeispiele in einen Trainingsdatensatz $D_{Training}$ und ein Testdatensatz D_{Test}, mit

$$D_{Test} \cap D_{Training} = \emptyset \tag{3.19}$$

aufgeteilt.

Um die Generalisierungsfähigkeit des Modells zu evaluieren und trotzdem nicht zu viele Daten für das Training des Modells als Testdaten zu verlieren, kann eine k-fold Kreuzvalidierung durchgeführt werden. Hierbei werden die vorhandenen Beispieldaten in k Teildatensätze unterteilt, von denen jeweils $k-1$ Datensätze zum Training und ein Datensatz zum Evaluieren des trainierten Modells verwendet werden. Der Durchschnitt aller k Teilergebnisse wird schließlich als Gesamtergebnis angegeben.

Je größer k, desto mehr Beispiele können für das Training des Modells herangezogen werden. Allerdings steigt damit auch der Trainingsaufwand, da das Modell k-mal mit den verschiedenen Trainingsdatensätze trainiert werden muss. Typischerweise wird eine Kreuzvalidierung mit $k = 5$ oder $k = 10$ durchgeführt.

Bei der Aufteilung der Beispieldaten in die k Teildatensätze sind gegebenenfalls bestehende Abhängigkeiten zwischen den einzelnen Trainingsbeispielen zu berücksichtigten. So sollten Daten, die von demselben Objekt stammen, stets dem gleichen Datensatz zugewiesen werden, da das Modell sonst mit Beispielen von Objekten evaluiert wird, von denen bereits Daten im Trainingsprozess verwendet wurden. Diese Berücksichtigung von Abhängigkeiten zwischen den Beispieldaten wird als **Group *k*-fold** bezeichnet (Pedregosa et al., 2011).

Zudem ist sicherzustellen, dass jeder der Teildatensätze eine repräsentative Teilmenge der gesamten Beispieldaten darstellt. Dies ist insbesondere bei unausgewogenen Daten von Bedeutung, da ansonsten das Risiko besteht, dass bestimmte Fälle nur im Trainings- oder nur im Testdatensatz vertreten sind. Eine k-fold Kreuzvalidierung, bei der die Teildatensätze annähernd gleich verteilt sind, wird als **Stratified *k*-fold** bezeichnet (Japkowicz, 2013).

3.3.2 Beurteilungsmetriken der Klassifikation

Im Allgemeinen werden die im Kontext der Klassifikation verwendeten Beurteilungsmetriken in drei Kategorien eingeteilt: Schwellwert Metriken, Rang Metriken und probabilistische Metriken (Japkowicz, 2013). Da die probabilistischen Metriken auch zur Beurteilung von Regressionsmodellen eingesetzt werden, werden diese in Abschnitt 3.3.3 vorgestellt.

Tabelle 3.1: Konfusionsmatrix eines binären Klassifikators.

		Tatsächliche Klasse y		
		1	0	$\sum$
Prädizierte Klasse $\hat{y}$	1	True Positive (TP)	False Positive (FP)	Predicted Positive (PP)
	0	False Negative (FN)	True Negative (TN)	Predicted Negative (PN)
	$\sum$	Real Positive (RP)	Real Negative (RN)	

Alle im folgenden Abschnitt vorgestellten Metriken basieren auf dem Konzept der Konfusionsmatrix (s. Tab. 3.1). Die Konfusionsmatrix gibt für jede Klasse an, wie viele der Testbeispiele korrekt als Teil dieser Klasse prädiziert wurden und wie viele der Testbeispiele jeweils falschen Klassen zugeordnet wurden. Im Fall eines binären Klassifikators gibt die Konfusionsmatrix die Anzahl der richtig positiv (True Positive, TP), der falsch positiv (False Positive, FP), der falsch negativ (False Negative, FN) und der richtig negativ (True Negative, TN) prädizierten Beispieldaten an (Powers, 2007).

Schwellwert Metriken

Schwellwert Metriken basieren auf einem Grenzwert, bei dem alle Beispiele, deren Prädiktion über dem Wert liegen, als positiv und der Rest als negativ klassifiziert wird. Schwellwert Metriken berücksichtigen somit nicht, wie nah die Prädiktion an dem Grenzwert liegt, sondern nur ob sie den Wert über- oder unterschreitet (Jeni et al., 2013). Schwellwert Metriken können weiter in Metriken mit Ein-Klassen Fokus und Metriken mit Multi-Klassen Fokus unterteilt werden (Japkowicz, 2013)

Ein-Klassen Fokus Schwellwert Metriken mit einem Ein-Klassen Fokus bestimmen die Leistung eines Klassifikators ausschließlich unter der Berücksichtigung der positiven oder der negativen Klasse. Die meisten Metriken in dieser Kategorie basieren auf der

als Recall (RC) bezeichneten Trefferquote und der als Precision (PR) bezeichneten Genauigkeit:

$$Recall\ (RC) = \frac{TP}{TP + FN} \tag{3.20}$$

$$Precision\ (PR) = \frac{TP}{TP + FP}. \tag{3.21}$$

Dieses Metrikenpaar hat seinen Ursprung im Bereich des Information Retrieval (Manning et al., 2008), bei dem der Anteil als relevant klassifizierter Daten zusammen mit der Menge tatsächlich relevanter Daten von Interesse ist. Der Fokus liegt hier somit einzig auf der positiven Klasse.

Um die Leistung mehrerer Modelle vergleichen zu können, ist es oft nötig, die Leistung eines Klassifikators über einen einzelnen Wert abzubilden. Hierzu wird in der Regel das als F_1-Maß bekannte, harmonische Mittel aus Recall und Precision bestimmt. Mit $\alpha > 0$ lässt die allgemeinere Form

$$F_\alpha = \frac{(1 + \alpha) \cdot PR \cdot RC}{\alpha \cdot PR + RC} \tag{3.22}$$

eine dem betrachteten Problem angepasste Gewichtung der zwei Metriken zu.

Durch die Fokussierung auf nur eine Klasse werden die Aussagekraft obiger Metriken durch eine ungleiche Klassenverteilung nicht negativ beeinflusst. Jedoch werden auch keinerlei Informationen über die Leistung des Klassifikators bezüglich der anderen, in diesem Fall der negativen Klasse angegeben (Japkowicz, 2013).

Multi-Klassen Fokus Schwellwert Metriken mit einem Multi-Klassen Fokus berücksichtigen die Leistung des Klassifikators über alle im Datensatz existierenden Klassen (Japkowicz, 2013). Zu diesen Metriken gehört die Accuracy (AC):

$$Accuracy\ (AC) = \frac{TP + TN}{TP + FP + TN + FN} \tag{3.23}$$

Diese gibt die Wahrscheinlichkeit an, dass ein zufällig ausgewähltes Beispiel korrekt klassifiziert wird. Die Accuracy ist eine der verbreitetsten Beurteilungsmetriken; im Fall einer ungleichen Klassenverteilung verschleiert die Accuracy jedoch ein Overfitting des Klassifikators auf die häufiger auftretende Klasse (Jeni et al., 2013). Im Fall

einer *between-class imbalance* von 100 : 1 würde ein Klassifikator, der beispielsweise alle Samples ausschließlich der häufig auftretenden Klasse zuordnet, trotzdem eine Accuracy von 99 % erreichen.

Um ein Overfitting auf eine der zwei Klassen zu erkennen, empfiehlt sich die Verwendung von Metriken, die auf der einzeln bestimmten True Positive Rate (TPR) und True Negative Rate (TNR) mit

$$\textit{True Positive Rate (TPR)} = \frac{TP}{TP + FN} \tag{3.24}$$

$$\textit{True Negative Rate (TNR)} = \frac{TN}{TN + FP}. \tag{3.25}$$

basieren. Da beide Metriken den Anteil der richtig klassifizierten Beispiele separat für eine Klasse angeben, beeinträchtigt eine ungleiche Klassenverteilung die Aussagefähigkeit der einzelnen Werte nicht (Japkowicz, 2013). Somit ist auch das, analog zum F_1-Maß bestimmte, harmonische Mittel (HM) der TPR und TNR mit

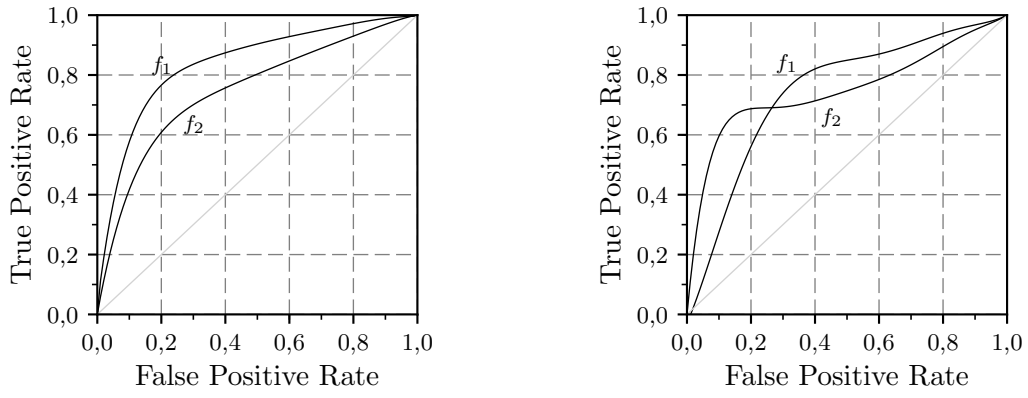
$$\textit{Harmonische Mittel (HM)} = \frac{2 * \mathrm{TPR} * \mathrm{TNR}}{\mathrm{TPR} + \mathrm{TNR}} \tag{3.26}$$

invariant gegenüber einer *between-class imbalance*.

Ranking Methoden und Metriken

Bei Klassifikatoren, die eine probabilistische Klassenzugehörigkeit ausgeben, werden die bewerteten Samples in der Regel der wahrscheinlichsten Klasse zugewiesen. Abhängig von der problemspezifischen Gewichtung der FP- zu den FN-Fehlprädiktionen, kann der Schwellwert für die Klassenzuweisung jedoch beliebig im Bereich [0,1] verschoben werden. Während die oben vorgestellten Schwellwert Metriken die Leistung eines Klassifikators somit nur an einem bestimmten Arbeitspunkt betrachten, wird bei den Ranking Methoden der gesamte Arbeitsbereich des Klassifikators berücksichtigt (Japkowicz, 2013). Ranking Metriken können somit als Zusammenfassung der Klassifikatorleistung über alle möglichen Schwellwerte betrachtet werden (Jeni et al., 2013).

Die bekannteste Ranking Methode ist die **Receiver Operation Characteristic (ROC)-Kurve**, die die TPR als Funktion der FPR darstellt (s. Abb. 3.6). Die ROC-Kurve erlaubt eine schnelle qualitative Bewertung: Jede ROC-Kurve geht stets durch

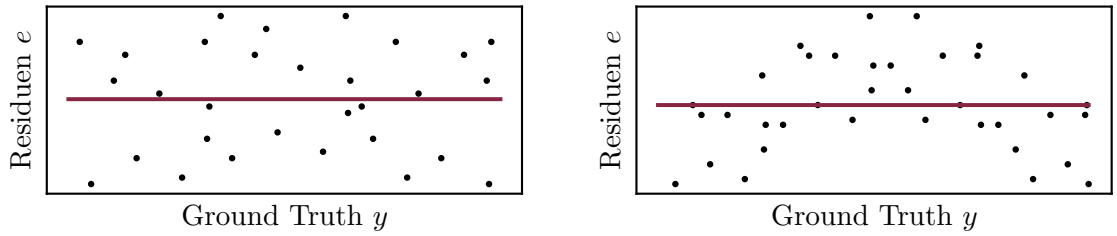


(a) (b)

Abbildung 3.6: ROC-Kurven für zwei beispielhafte Klassifikatoren. (a) f_1 dominiert eindeutig über f_2 (b) Keiner der Klassifikatoren ist dem anderen eindeutig überlegen. Abbildung nach (Japkowicz, 2013).

die Punkte (0,0) und $(1, 1)$, an denen der Klassifikator jeweils die eine Klasse komplett richtig und die andere komplett falsch klassifiziert. Ein perfekter Klassifikator wird durch einen Punkt in der oberen linken Ecke beschrieben und ein Klassifikator dessen Güte einem zufälligen Raten entspricht, weist eine ROC-Kurve nahe der Diagonalen auf. ROC-Kurven unterhalb der Diagonalen zeigen, dass eine negative Korrelation zwischen den Daten und der Klassifikatorausgabe vorliegt (Powers, 2007).

Die Fläche unter der ROC-Kurve (**Area Under the ROC-Curve (AUROC)**) dient schließlich als quantitative Bewertungsmetrik. Dabei gibt der Wert der AUROC die Wahrscheinlichkeit an, dass ein zufälliges positives Testbeispiel höher eingestuft wird, als ein zufälliges negatives Testbeispiel. Die Area Under the ROC-Curve (AUROC) lässt einen Vergleich mehrerer Klassifikatoren auch dann zu, wenn keine der ROC-Kurven deutlich dominiert. Da die ROC Analyse bei Problemstellungen mit *between-class imbalances* keine Modelle bevorteilt, die gut auf der häufig vertretenen Klasse arbeiten, nicht aber auf der seltenen Klasse, ist die AUROC eine De-facto-Standardmetrik bei der Evaluierung unausgewogener Daten (Japkowicz, 2013).



(a) (b)

Abbildung 3.7: Beispielhafte Residuendiagramme. (a) Horizontales Muster. (b) Gekrümmtes Muster. Abbildung nach (OriginLab Corporation, 2012).

3.3.3 Beurteilungsmetriken der Regression

Die Güte eines Regressionsmodells wird auf Basis der Abweichung e_i des prädizierten Werts $\hat{y}_i$ vom wahren Wert y_i bestimmt. Es gilt:

$$e_i = y_i - \hat{y}_i. \tag{3.27}$$

Dieser durch das Modell nicht prädizierbare Fehlerterm wird auch als Residuum bezeichnet.

Residuendiagramme

Residuen sollten zufällig um Null verteilt sein und kein systematisches Muster aufweisen. Zur Prüfung dieser Annahme können Residuendiagramme verschiedener Art eingesetzt werden. Residuendiagramme sind Streudiagramme, die den Fehler e_i in der Regel als Funktion des Zielwerts y_i abbilden. Das Residuendiagramm besteht somit aus den Punkten $(y_1, e_1), \cdots (y_N, e_N)$.

Zeigt die grafische Residuenanalyse beispielsweise, dass die Streuung der Punkte um die Nulllinie konstant ist (s. Abb. 3.7a), haben alle Fehlerterme die gleiche Varianz, was als Voraussetzung für viele Standardtests der Statistik gilt. Weist ein Residuendiagramm hingegen ein von der Horizontalen abweichendes Muster auf (s. Abb. 3.7b), ist dies ein Hinweis dafür, dass ein systematischer Fehler vorliegt und beispielsweise die Komplexität des Modells für die betrachteten Daten nicht ausreicht (Engel, 2010).

Hinweise zur Auswertung von Mustern in Residuendiagrammen können (OriginLab Corporation, 2012; Minitab Inc., 2016) entnommen werden.

Alternativ eignen sich zur visuellen Überprüfung der Normalverteilung auch Histogramme der Residuen oder Streudiagramme, die den prädizierten Wert $\hat{y}_i$ als Funktion des Zielwerts y_i abbilden.

Fehlermetriken

Auf Basis des Residuum lassen sich eine Vielzahl von Fehlermetriken bestimmen, die die Prädiktionsfähigkeit des Regressionmodells quantitativ repräsentieren. Da die Differenz zwischen dem prädizierten und dem wahren Wert auch negativ sein kann, wird zur Bestimmung der Gesamtgröße des Fehlers über alle betrachteten Testsamples in der Regel der Betrag oder das Quadrat des einzelnen Fehlers e_i herangezogen. Die darauf basierenden Fehlermetriken messen schließlich die Prädiktionsleistung eines Regressionsmodells über die mittleren Abweichung der Vorhersage von den wahren Werten. Niedrige Fehlerwerte bedeuten, dass das Modell bei der Vorhersage genauer ist. Eine Gesamtfehler-Metrik von 0 bedeutet, dass das Modell perfekt auf die Daten passt.

Häufig verwendete Fehlermaße sind der mittlere absolute Fehler (Mean Absolute Error (MAE)) und die Wurzel der mittleren quadratischen Abweichung (Root Mean Squared Error (RMSE)):

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| = \frac{1}{N}\sum_{i=1}^{N}|e_i| \tag{3.28}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}e_i^2}. \tag{3.29}$$

Sowohl der MAE als auch der RMSE werden in derselben Einheit gemessen, wie die zu prädizierende Variable. Beide Fehlermaße sind damit skalenabhängig und daher nur zum Vergleich von Modellen geeignet, deren Variablen derselbe Maßstab zugrunde liegt (Hyndman und Athanasopouls, 2013).

Die Größenordnung des MAE und des RMSE sind in der Regel ähnlich zueinander. Der RMSE gewichtet größere Fehler jedoch stärker und ist somit sensibler gegenüber Ausreißern. Bis zu welcher Größe ein Fehlerwert als gut zu beurteilen ist, ist abhängig

von der betrachteten Problemstellung und der geforderten Prädiktionsgüte. Daher kann kein fester Schwellwert festgelegt werden, ab dem ein Regressionsmodell als gut oder schlecht zu bewerten ist.

Darüber hinaus kann unter der Annahme normalverteilter Residuen, über den RMSE die Unsicherheit der Prädiktion angegeben werden: Die Grenzen des 95 % Konfidenzintervalls einer Prädiktion ergeben sich über $\pm 2RMSE$.

Zudem ist es vorteilhaft, dass beide Fehlermaße eine unterschiedlichen Gewichtungen für verschiedene Samples erlauben, was bei der Bewertung von Problemstellungen mit unausgewogenen Daten relevant ist. Die Gleichung 3.29 ändert sich für den gewichteten RMSE dann zu

$$RMSE_w = \sqrt{\frac{1}{N}\sum_{i=1}^{N} w_i(y_i - \hat{y}_i)^2} \tag{3.30}$$

mit

$$\sum_{i=1}^{N} w_i = 1. \tag{3.31}$$

3.4 Visuelle Deskriptoren für Merkmalsbasiertes Lernen

Im Bereich Computer Vision existiert bereits eine Vielzahl visueller Deskriptoren, die objekttypische Eigenschaften anhand der Pixelwerten der Bilddaten domänenspezifisch extrahieren. Im Folgenden werden die zwei Deskriptoren vorgestellt, die sich im Bereich der Fußgänger- und Kopfdetektion (Dollár et al., 2012; Rehder et al., 2014) sowie der Erkennung von Fußgängeraktionen (vgl. Abschn. 2.3.2) bereits bewährt haben und daher auch im Rahmen dieser Arbeit einen Einsatz finden.

3.4.1 Histograms of Oriented Gradients (HOG)

Histograms of Oriented Gradients (HOG) sind normalisierte Histogramme, die die Häufigkeitsverteilung der Gradientenorientierung eines Bildbereichs abbilden. Erstmals von Dalal und Triggs (2005) vorgestellt, basiert der HOG-Deskriptor auf der Idee, dass Objekte über ihre Kantenrichtungen, genauer über die Verteilung ihrer lokalen Intensitätsgradienten, beschreibbar sind.

Wie Abbildung 3.8 zeigt, werden zur Bestimmung des HOG-Deskriptors für ein Grauwertbild zunächst die Gradienten des Eingangsbilds, über eine Filterung in x- sowie in y-Richtung mit dem Filterkern $[-1,0,1]$, berechnet. Anschließend werden für jeden Bildpunkt der Winkel und die Amplituden der Gradienten bestimmt: Mit g_x als ein beliebiger Punkt im in x-Richtung gefilterten Gradientenbild und g_y als entsprechender Punkt im in y-Richtung gefilterten Gradientenbild, wird der Winkel α über

$$\alpha = arctan\left(\frac{g_y}{g_x}\right) \tag{3.32}$$

und die Amplitude A über

$$A = \sqrt{g_x^2 + g_y^2} \tag{3.33}$$

berechnet.

Der $M \times N$ px große Bildbereich wird anschließend in $M_c \times N_c$ px große Zellen aufgeteilt und für jede dieser Zellen wird ein Gradientenhistogramm mit b Bins berechnet. Hierbei werden die Amplitudenwerte A entsprechend ihrem zugehörigen Winkel in die Bins eingeordnet. Abhängig von der zu beschreibenden Objektklasse decken die Bins dabei einen Wertebereich von $[0° - 180°)$ oder von $[0° - 360°)$ ab.

Der zellenweisen Berechnung der Gradientenhistogramme folgt eine Normalisierung über $M_b \times N_b$ px große Blöcke, die sich jeweils um p px überlappen. Zur Blocknormalisierung können verschiedene Normalisierungsmethoden zum Einsatz kommen. Die am häufigsten vertretene sind die $L2$-Norm, $L2$-Hys, $L1$-Norm oder $L1$-Sqrt (Sonka et al., 2013). Die über die Blöcke normalisierten Histogrammwerte werden schließlich zu einem d-elementigen Merkmalsvektor mit

$$d = \frac{M - (M_b - p)}{p} * \frac{N - (N_b - p)}{p} * \frac{M_b * N_b}{M_c * N_c} * b \tag{3.34}$$

verkettet.

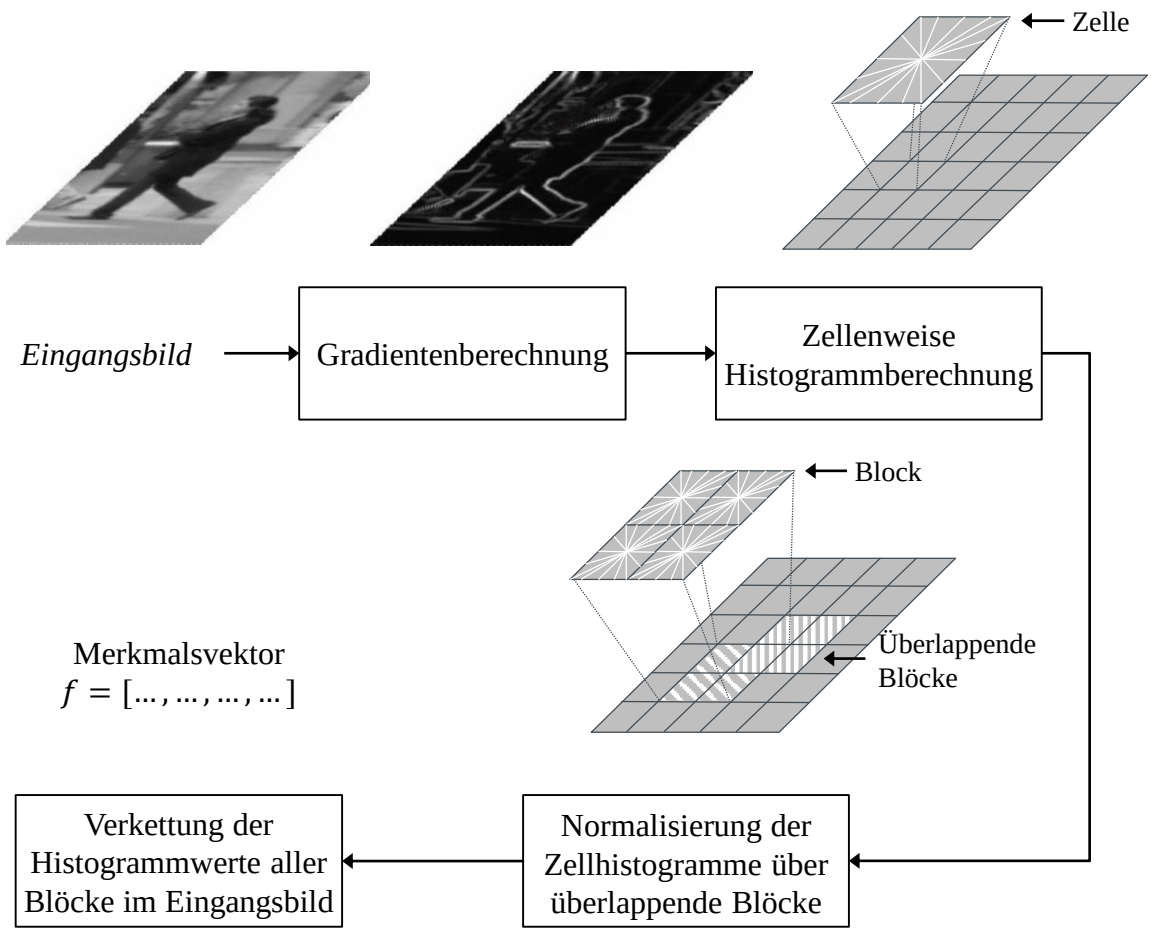


Abbildung 3.8: Zur Erstellung eines HOG-Deskriptors zu durchlaufende Schritte bei Verwendung eines Grauwertbilds als Eingangsdaten.

3.4.2 Local Binary Pattern (LBP)

Das von Pietikainen und Maenpaa (2002) erstmalig vorgestellte Local Binary Pattern (LBP) ist ein Texturoperator, der die räumliche Struktur eines lokalen Bildbereichs durch Codierung der Unterschiede zwischen dem Grauwert des zentralen Pixels und denen seiner Nachbarn beschreibt. Indem bei dem Vergleich der Pixelwerte nur das Vorzeichen ausgewertet wird, kann so ein binäres Muster für jeden Bildbereich gebildet werden. Dieses wird schließlich in einen Dezimalwert gewandelt und in einem LBP-Bild gespeichert.

Wie Abbildung 3.9 zeigt, wird zur Berechnung des LBP der Wert eines beliebigen Pixels x_c mit denen seiner p Nachbarpixel $\{x_{r,p,n}\}_{n=0}^{p-1}$, die auf einem Kreis mit dem Radius r gleichmäßig um x_c verteilt sind, verglichen:

$$LBP_{r,p}(x_c) = \sum_{n=0}^{p-1} s(x_{r,p,n} - x_c) * 2^n, \tag{3.35}$$

mit

$$s(x) = \begin{cases} 1 & \text{wenn } x \geq 0 \quad (3.36a) \\ 0 & \text{sonst.} \quad (3.36b) \end{cases}$$

Die Grauwerte von Nachbarn $x_{r,p,n}$, die nicht genau in den Mittelpunkt eines Pixels fallen, werden über Interpolation bestimmt (Liu et al., 2016).

Analog zum HOG kann die Textur des Eingangsbildes schließlich über die Häufigkeitsverteilung der 2^p LBP Muster beschrieben werden. Hierzu wird das Bild ebenfalls in $M_c \times N_c$ px große Zellen aufgeteilt, um pro Zelle ein Histogramm der LBP-Werte zu erstellen. Eine Normalisierung über Blöcke findet in der Regel nicht statt.

Im Fall einer $p = 8$ großen Nachbarschaft kann das LBP theoretisch 256 Werte annehmen, was zu Zellenhistogrammen mit $b = 256$ Bins und einem entsprechend hochdimensionalen Merkmalsvektor führt. Wie Pietikainen und Maenpaa (2002) beobachtet haben, treten bei natürlichen Bildern jedoch überwiegend Binärcodes mit maximal zwei Wechseln zwischen 0 und 1 auf. Diese Bitmuster werden als Uniform Pattern bezeichnet und stellen, wie Abbildung 3.10 illustriert, Merkmale wie Flächen, Linien oder Ecken dar. Bei einem 8-Bit Binärcode existieren 58 Uniform Patterns, deren Häufigkeit im LBP-Histogramm jeweils mit einem eigenen Bin erfasst wird.

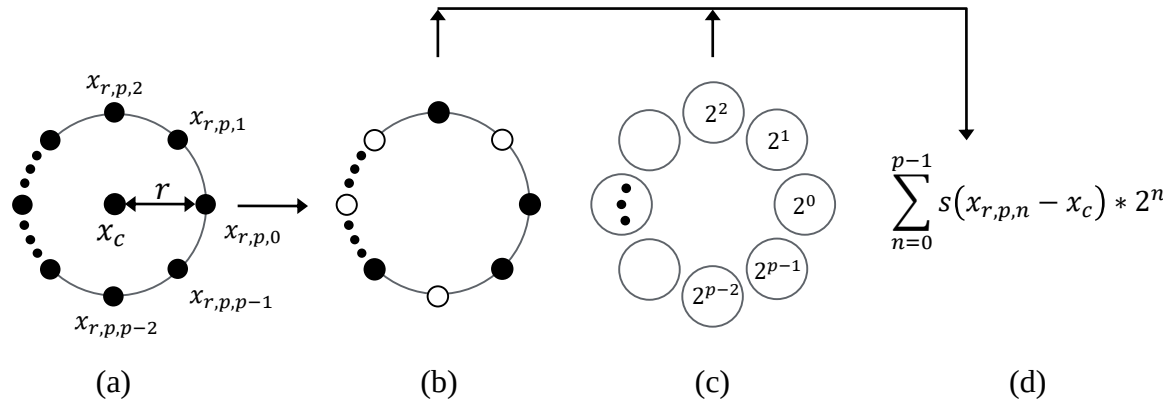


Abbildung 3.9: Prozess zur Erstellung eines LBP. (a) Typischerweise betrachtete Nachbarschaft: zentraler Pixel x_c und seine p gleichmäßig auf einem Kreis mit dem Radius r verteilten Nachbarn. (b) Binäres Muster. (c) Gewichte. (d) Dezimalwert. Abbildung nach (Liu et al., 2016).

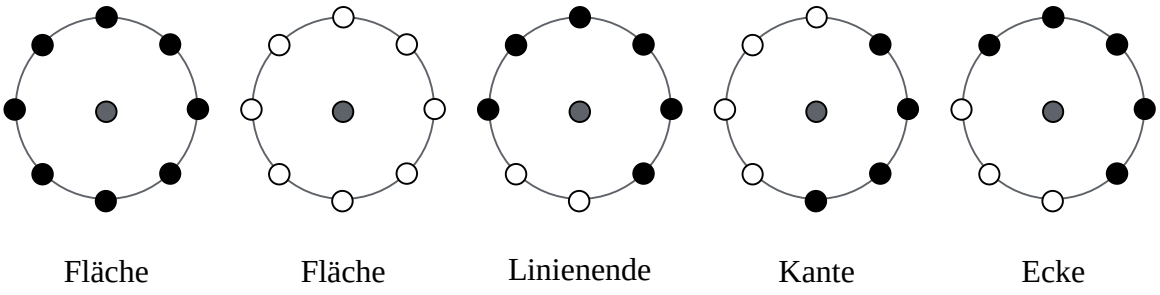


Abbildung 3.10: Beispiele der 58 Uniform Patterns. Abbildung nach (Guennouni et al., 2015).

Alle weiteren Non-Uniform Patterns werden einheitlich dem 59. Histogramm-Bin zugewiesen. Die Verwendung von Uniform Patterns erlaubt somit eine Reduktion der Zellenhistogrammgröße von 256 auf 59 Werten.

LBPs finden Ihren Ursprung in der Texturklassifikation; auf Grund ihrer Invarianz gegenüber monotonen Grauwertänderungen werden sie aber auch im Bereich der Gesichtserkennung (Ahonen et al., 2006) sowie zur Kopfdetektion (Rehder et al., 2014) mit Erfolg eingesetzt.

3.5 Diskussion und Bewertung

Durch die in Kapitel 4 vorgestellte Referenzmethode sind die Zielwerte der in dieser Arbeit verwendeten Beispieldaten bekannt. Daher können bei der Entwicklung des eigenen Ansatzes Methoden des **überwachten Lernens** eingesetzt werden. Entsprechend der Diskussion in Abschnitt 2.5 beschränkt sich diese Arbeit dabei auf **merkmalsbasierte Methoden**, bei denen die zu bewertenden Objekte in Form manuell gestalteter Merkmale oder Eigenschaften repräsentiert werden. Wie Abschnitt 3.4, sowie die in Abschnitt 2.5 bereits bewerteten posenbasierten Verfahren zur Aktionserkennung zeigen, existieren im Bereich Computer Vision bereits eine Vielzahl visueller Deskriptoren, von denen sich einige auch schon im Bereich der Fußgänger- und Kopfdetektion sowie bei der Erkennung von Fußgängeraktionen bewährt haben.

Im Speziellen hat sich zur Detektion von Fußgängern vor allem das **Histograms of Oriented Gradients** (HOG) durchgesetzt, da dieses die kantenbasierten Strukturen aufrecht stehender Personen besonders markant abbildet (s. Abschn. 3.4.1). Wie in Abschnitt 2.5 diskutiert, lässt sich jedoch vermuten, dass die auf dem HOG basierenden Merkmale der posenbasierten Aktionserkennung die für die Erkennung der Querungsintention relevante Ausrichtung des Kopfes nicht explizit genug abbildet. Da monotone Grauwertverläufe, die charakteristisch für den Kopfbereich eines Fußgängers sind, deutlich besser von dem in Abschnitt 3.4.2 vorgestellten **Local Binary Pattern** (LBP) beschrieben werden, verspricht ein Deskriptor, der den Körperbereich des Fußgängers über ein HOG und den Kopfbereich über ein LBP abbildet, gut geeignet zu sein, die für die Querungsintention relevante, statische Pose eines Fußgängers zu beschreiben. Kombiniert mit den ebenfalls in Abschnitt 2.5 diskutierten Methoden der posenbasierten Aktionserkennung, kann zudem die über mehrere Frames verlaufende Körperbewegungen des Fußgänger abgebildet werden. Ob eine solche Merkmalskombination die, entsprechend der Bewertung aus Abschnitt 2.5, zu evaluierenden Schwächen bisheriger zur posenbasierten Aktionserkennung eingesetzten Merkmale überwindet, ist ebenfalls im Rahmen einer Evaluierung festzustellen.

Wie die in Abschnitt 4.2 präsentierten Ergebnisse der beobachterbasierten Referenzbildung zeigen, ist die Querungsintention eines Fußgängers eine mit Unsicher-

heiten behaftete Größe, die theoretisch eine kontinuierliche Ausprägung zwischen 0 und 1 hat. Somit ist die Erkennung der Querungsintention eines Fußgängers als ein **Regressionsproblem** zu betrachten. Abhängig von dem Anwendungsfall, in dem ein technisches System zur Erkennung der Querungsintention eingesetzt werden soll, kann aber auch nur die binäre Entscheidung über die Ausprägung einer Querungsintention – ohne die Abbildung der Unsicherheit – von Interesse sein. Dies ist beispielsweise der Fall bei warnenden Assistenzfunktionen, die ausschließlich vor querungswilligen Fußgängern warnen, die vom Fahrer übersehen wurden. Für solche Anwendungsfälle ist auch der Einsatz von **Klassifikationsmethoden** möglich.

Diese Arbeit fokussiert sich jedoch auf die Übertragung des menschlichen Situationsbewusstseins auf ein technisches System für den Anwendungsfalls des automatisierten Fahrens. Da ein automatisiertes Fahrzeug – analog zum menschlichen Fahrer – sein eigenes Handeln an bestehende Unsicherheiten anpassen sollte, ist im Rahmen des eigenen Ansatzes ein **Regressionsmodell** zu entwickeln, das die Querungsintention eines Fußgängers als kontinuierlichen Wert zwischen 0 und 1 bestimmt. Über die Anwendung eines Schwellwerts kann das zu entwickelnde Modell auch zur Generierung binärer Ergebnisse verwendet werden. Diese Ergebnisse sollten daher auch mit angegeben und ausgewertet werden, der Fokus der Modellentwicklung und -optimierung liegt jedoch auf der Bestimmung der Querungsintention eines Fußgängers unter Berücksichtigung der bestehenden Unsicherheiten.

Eine Methodik des überwachten Lernens, die sowohl zur Anwendung bei Klassifikations- als auch bei Regressionsproblemen geeignet ist, ist die **Support Vector Machine** (SVM) bzw. die **Support Vector Regression** (SVR). Diese Methodik überzeugt durch eine Vielzahl von Vorteilen (s. Abschn. 3.2.4) und wurde im Bereich der posenbasierten Aktionserkennung bereits erfolgreich in Kombination mit den für die Erkennung der Querungsintention als geeignet bewerteten Merkmale eingesetzt. Die SVR ist daher eine vielversprechende Methode, um den Zusammenhang zwischen den Merkmalsvektoren und der mit Unsicherheiten behafteten Ausprägung einer Querungsintention eines Fußgängers zu erlernen. Als Kernelfunktion empfiehlt sich dabei die Radiale Basisfunktion (RBF), da diese im allgemeinen flexibler ist als lineare oder polynomiale Kernels. Zudem zeigt die in (Schneemann und Heinemann,

2016) durchgeführte Analyse, das ein RBF-Kernel gut geeignet ist, um die über den kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ beschriebenen Daten binär zu trennen.

Bezüglich des **Training** der SVR ist jedoch zu beachten, dass die in dieser Arbeit verwendeten Daten unausgewogen sind und sowohl eine *between-class imbalance* als auch eine *within-class imbalance* vorliegt (s. Abschn. 6.1). Um während des Trainingsprozesses ein Overfitting auf die häufig vorkommenden Samples zu vermeiden, muss daher eine geeignete Funktion zur Gewichtung der Samples gefunden werden.

Die vorliegende Unausgewogenheit der Daten muss zudem bei der **Beurteilung** der Leistung des trainierten Modells berücksichtigt werden. Dies gilt sowohl für die Kreuzvalidierung, als auch für die zur Beurteilung verwendeten Metriken. Erschwerend kommt hinzu, dass bei den verwendeten Daten sowohl eine zeitliche, als auch eine situative Korrelation zwischen den einzelnen Samples vorliegt. Dabei rührt die zeitliche Korrelation aus dem Tracking eines Fußgängers über mehrere Frames und die situative Korrelation entsteht durch Situationen, in denen mehrere Fußgänger gleichzeitig auftreten, beispielsweise wenn mehrere Fußgänger an derselben Ampel warten. Somit bedarf es bei der Kreuzvalidierung einer Kombination aus der Stratified- und der Group k-fold-Methode. Dies stellt sicher, dass auch unter Berücksichtigung der bestehenden Abhängigkeiten jeder Teildatensatz eine repräsentative Teilmenge der gesamten Beispieldaten darstellt (s. Abschn. 3.3.1).

Als **Beurteilungsmetrik** eignet sich der gewichtete RMSE ($RMSE_w$, s. Gl. 3.30) auf Grund der expliziten Möglichkeit, für verschiedene Samples unterschiedliche Gewichte zu berücksichtigen. Als Gewichtungsfunktion kann hier dieselbe Funktion verwendet werden, die für die Gewichtung der Trainingssamples gefunden werden muss, um ein Overfitting des Modells zu vermeiden.

Zur Bewertung der **Klassifikationsgüte** des Modells ist die in der Literatur verbreitetste Accuracy-Metrik auf Grund der unausgewogenen Daten nicht geeignet (s. Gl. 3.23). Und auch wenn das F_1-Maß eine der beliebtesten Metriken zur Beurteilung unausgewogener Daten ist, führt der Ein-Klassen Fokus zur einer Sensibilisierung auf die True Positive (TP), wodurch diese gegenüber den True Negative (TN) stärker gewichtet sind. Dies ist zwar zielführend bei Information Retrieval Anwendungen, bei der binären Erkennung der Querungsintention eines Fußgängers ist es jedoch abhängig

von der Ausprägung des Systems. Hier muss je nach Anwendungsfall entschieden werden, ob eine nicht erkannte Querungsintention höher zu gewichten ist, als ein falsch positiver Alarm. Daher sollte zur allgemeinen Evaluation der Klassifikationsleistung eine Metrik eingesetzt werden, die die TP und die TN gleichermaßen im Fokus hat. Dies gilt für das Harmonische Mittel (HM) aus TPR und TNR (s. Gl. 3.26).

Kapitel 4

Referenzbildung durch beobachterbasierte Videoannotation

Wie die Analyse des Stands der Technik zeigt (s. Abschn. 2.5), bedarf es zur Entwicklung eines Systems zur Erkennung der Querungsintention von Fußgängern einer neuen Referenzmethode. Die in Bereichen außerhalb der Fußgängerverhaltenserkennung bereits etablierten, beobachterbasierten Methoden (s. Abschn. 2.4.2) scheinen hier besonders geeignet zu sein. Daher wird im Folgenden eine beobachterbasierte Videoannotation zur Referenzbildung vorgeschlagen. Die eingesetzte Methode wird in Abschnitt 4.1 ausführlich beschrieben. In Abschnitt 4.2 werden anschließend die Ergebnisse der durchgeführten Übereinstimmungs- und Reliabilitätsprüfung vorgestellt und in Abschnitt 4.3 vor dem Hintergrund des in dieser Arbeit zu entwickelnden Systems bewertet.

4.1 Methode

Bei der in dieser Arbeit eingesetzten, beobachterbasierten Videoannotation, werden die im Entwicklungsprozess als Trainings- und Testdaten eingesetzten Videosequenzen (s. Abschn. 4.1.1) zunächst von einem Experten beurteilt. Um die Reliabilität

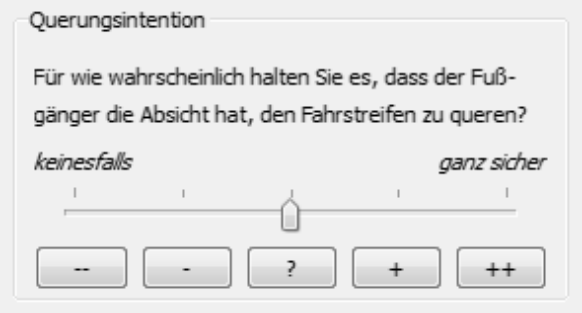


Abbildung 4.1: Zur Beurteilung der Querungsintention eingesetzte Ratingskala.

dieser Methode zu überprüfen, wird ein Teil der Daten anschließend von unabhängigen Beobachtern ein weiteres Mal beurteilt und eine Prüfung der Übereinstimmung und Reliabilität der Beobachterurteile durchgeführt. Die Videoannotation wird dabei wie im Folgenden beschrieben durchgeführt und gilt sowohl für die Beurteilung durch den Experten, als auch für die Beurteilung durch die Beobachter.

Jede der zu bewertenden Videosequenzen beinhaltet mindestens einen Fußgänger. Zur Annotation werden die Videos mit der halben Aufnahmegeschwindigkeit abgespielt und eventbasiert, bei jedem neu erscheinenden Fußgänger pausiert. Anschließend wird der Beobachter aufgefordert, anhand einer fünfstufigen, intervallskalierten Ratingskala zu beurteilen, für wie wahrscheinlich er es hält, dass der Fußgänger die Absicht hat, den Fahrstreifen zu queren (s. Abb. 4.1). Um sicherzustellen, dass die Abstände zwischen den Stufen als äquidistant aufgefasst werden, werden die Ausprägungen der fünf Stufen, basierend auf den Empfehlungen von Rohrmann (1978), wie folgt formuliert:

– –	Der Fußgänger hat *keinesfalls* die Absicht den Fahrstreifen zu queren
–	Der Fußgänger hat *wahrscheinlich nicht* die Absicht den Fahrstreifen zu queren
?	Der Fußgänger hat *vielleicht* die Absicht den Fahrstreifen zu queren
+	Der Fußgänger hat *ziemlich wahrscheinlich* die Absicht den Fahrstreifen zu queren
+ +	Der Fußgänger hat *ganz sicher* die Absicht den Fahrstreifen zu queren

Um den Fahrstreifen, auf den sich die Querungsintention des Fußgängers bezieht, eindeutig zu beschreiben, wird stets die rechte und linke Begrenzung des Ego-Fahrstreifens im Video angezeigt.

Abbildung 4.2: Das PCI_Labeltool: rechts ist der Videobereich mit der zu bewertenden Fußgängerdetektion und dem eingezeichneten Ego-Fahrstreifen zu sehen. Links befindet sich ein Informations-, Steuer- und Bewertungsbereich (von oben nach unten).

Neben der Beurteilung der Querungsintention jedes neuen Fußgängers, werden die Beobachter zeitbasiert aufgefordert zu überprüfen, ob ihr zuletzt abgegebenes Urteil weiterhin zutrifft. Dazu wird das Video im Fall einer aktiven Fußgängerdetektion in einem festen Intervall von 18 Frames ($\hat{=}$ 2 s) pausiert und dem Beobachter sein zuletzt abgegebenes Urteil angezeigt. Dieses kann der Beobachter bestätigen oder durch die Abgabe eines neuen Urteils ändern. Ein geändertes Urteil gilt immer ab dem Änderungszeitpunkt. Ein Revidieren vorheriger Urteile ist nicht erlaubt.

Die technischen Umsetzung der Videoannotation erfolgt über das neu entwickelte und selbst implementierte PCI_Labeltool – eine auf der C++-Klassenbibiliothek QT basierende und in das Framework ADTF integrierte grafische Benutzeroberfläche (GUI). Abbildung 4.2 zeigt die Oberfläche des Tools während der Beurteilung eines Fußgängers.

4.1.1 Datenbasis

Wie in Abschnitt 6.1 beschrieben, besteht der im Rahmen dieser Arbeit erstellte Datensatz aus 175 Videosequenzen, die mit einem serienmäßigen Mono-Frontkamera-System bei Fahrten im deutschen Straßenverkehr aufgenommen wurden und insgesamt 639 Fußgänger beinhalten.

Jeder der 639 Fußgänger wird zunächst von dem Experten bezüglich seiner Querungsintention beurteilt. Da der Annotationsprozess sehr aufwendig ist und weit über 20 Stunden in Anspruch nimmt, wird für die anschließende Reliabilitätsprüfung nur ein Teil der gesamten Datenbasis verwendet. Anstatt eine zufällige Auswahl zu treffen, werden für die Reliabilitätsmessung alle Videosequenzen ausgewählt, die nach dem Urteil des Experten mindestens einen Fußgänger mit einer nicht eindeutigen (Stufe –, ? oder +) oder mit einer sich ändernden Querungsintention enthalten. Diese Situationen sind bei der Bewertung der Reliabilität des Videoannotationsverfahrens von besonderem Interesse, da zu erwarten ist, dass die Beobachterübereinstimmung in diesen nicht-eindeutigen Situationen weniger hoch ist, als in eindeutigen Situationen.

Die so bestimmte Teilmenge der Videodaten bildet mit 41 Sequenzen 23,4 % der gesamten Datenbasis ab. Die ausgewählten Sequenzen beinhalten 278 Fußgänger ($\hat{=}$ 44 % aller Fußgänger), von denen nach dem Urteil des Experten 30 Fußgänger eindeutig eine Querungsintention haben ($\hat{=}$ 37 % aller Fußgänger mit QI) und 179 Fußgänger keinesfalls die Absicht haben, die Straße zu queren (36 % aller Fußgänger ohne QI). Den restlichen 69 Fußgängern weist der Experte somit entweder kein eindeutiges oder ein sich über die Beobachtungszeit änderndes Urteil zu. Insgesamt stehen bei der Übereinstimmungs- und Reliabilitätsprüfung von jedem Beobachter $N = 27.361$ Samples zur Verfügung.

4.1.2 Stichprobe

Die oben beschriebene Videoannotation wurde zunächst von der als Experte fungierenden Autorin dieser Arbeit (b_0) durchgeführt. Anschließend wurden die Urteile von weiteren $k = 6$ unabhängigen Beobachtern (b_1–b_6) über die gleiche Methodik erfasst. Voraussetzung für die Rekrutierung als unabhängiger Beobachter ist eine langjähri-

Tabelle 4.1: Demografische Daten der Beobachter.

	Alter	Geschlecht	Führerscheinbesitz in Jahren	Kilometerleistung im letzten Jahr	Fahrzeugnutzung pro Woche
Experte b_0	29	w	11	$\leq$ 20.000	täglich
Beobachter b_1	62	w	40	$\leq$ 10.000	3–5
Beobachter b_2	63	m	45	$\leq$ 15.000	täglich
Beobachter b_3	28	m	10	$\leq$ 5.000	3–5
Beobachter b_4	33	m	15	$\leq$ 20.000	täglich
Beobachter b_5	26	m	9	$\leq$ 15.000	3–5
Beobachter b_6	22	m	5	$\leq$ 5.000	$<$ 1

ge Erfahrung als aktiver Teilnehmer im deutschen Straßenverkehr, der Besitz eines gültigen PKW-Führerscheins sowie die Unkenntnis des in der Studie gezeigten Videomaterials. Vor der Durchführung der Studie wurden die unabhängigen Beobachter mit der in Anhang B.1 angefügten Anleitung bezüglich des richtigen Verständnis des Begriffs „Querungsintention“ sowie der Handhabung des PCI_Labeltools geschult. Wenn nicht anders angegeben, wird im Folgenden die gesamte Gruppe aus Experte und unabhängigen Beobachtern als Beobachter bezeichnet.

Im Anschluss an die Videoannotation wurden über den in Anhang B.2 gezeigten Fragebogen demografische Daten der Beobachter erhoben. Wie Tabelle 4.1 zeigt, sind die Beobachter zwischen 22 und 63 Jahren alt (MW: 38, SD: 16). Fünf der Beobachter sind männlich, zwei sind weiblich. Die Beobachter haben seit 5 bis 45 Jahren (MW: 19, SD: 15) einen Führerschein. Fast alle Beobachter fahren mindestens 3–5 mal pro Woche mit dem Auto, drei sogar täglich. Nur Einer fährt seltener als einmal pro Woche mit dem Auto.

Neben den demografischen Daten wurde über den Fragebogen auch der subjektiv eingeschätzte Fahrstil der Beobachter anhand der fünfstufigen Skala von Kobiela (2011) erhoben. Wie die Ergebnisse in Abbildung 4.3 zeigen, schätzen sich die Beobachter als durchschnittliche bis eher schnelle, offensive und risikobereite Fahrer ein.

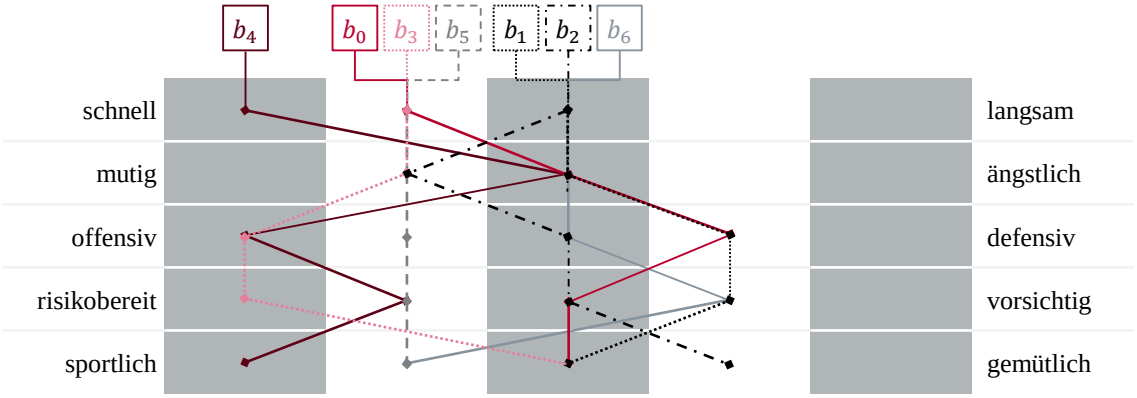


Abbildung 4.3: Subjektiv eingeschätzter Fahrstil der Beobachter.

4.2 Ergebnisse

Im Folgenden werden die Ergebnisse zur Evaluierung der Beobachterübereinstimmung und Beobachterreliabilität vorgestellt. Die Ergebnisse wurden entsprechend den in Abschnitt 2.4.2 beschriebenen Methoden bestimmt.

4.2.1 Verteilung der Beobachterurteile

Tabelle 4.2 gibt einen allgemeinen Überblick über die Verteilung der Urteile der einzelnen Beobachter bezüglich der fünf Stufen der Ratingskala. Wie die Ergebnisse zeigen, weisen die Urteile aller Beobachter, mit Ausnahme derer von Beobachter b_6, eine ähnliche Verteilung auf. Demnach beinhalten die Videodaten deutlich mehr Fußgänger, die den Fahrstreifen keinesfalls queren wollen (65 %), als solche, die ganz sicher queren wollen (19 %). Zudem zeigen die Ergebnisse, dass sich die Beobachter bei den meisten Samples (84 %) über die Ausprägung der Querungsintention des Fußgängers eindeutig sicher sind (Stufe $--$ und $++$) und nur in 5 % der Fälle keine Aussage über die Querungsintention treffen können (Stufe ?). Einzig die Urteile von Beobachter b_6 weisen eine sich stark unterscheidende Verteilung auf. Mit insgesamt 61 % weist Beobachter b_6 den Fußgängern zwar im Vergleich zu den anderen Beobachtern mit einer ähnlich hohen Häufigkeit tendenziell keine Querungsintention zu, er ist sich dabei aber nur in 6 % der Fälle absolut sicher, dass der Fußgänger keinesfalls die Absicht hat, den Fahrstreifen zu queren. Auch die Stufe ? wird mit 14 % von Beobachter b_6 häufiger ge-

Tabelle 4.2: Verteilung der Beobachterurteile.

	– –	–	?	+	++
Experte b_0	0,69	0,04	0,02	0,05	0,20
Beobachter b_1	0,72	0,02	0,04	0,05	0,17
Beobachter b_2	0,62	0,09	0,06	0,06	0,17
Beobachter b_3	0,54	0,14	0,11	0,05	0,16
Beobachter b_4	0,61	0,09	0,04	0,05	0,21
Beobachter b_5	0,69	0,03	0,03	0,05	0,20
Beobachter b_6	0,06	0,55	0,14	0,09	0,16
MW	0,56	0,14	0,06	0,06	0,18
$\text{MW}_{\setminus b_6}$	0,65	0,07	0,05	0,05	0,18

wählt, als von den anderen Beobachtern. Auf Grund dieser Abweichungen werden im Folgenden die Übereinstimmungs- und Reliabilitätswerte sowohl für alle Beobachter, als auch für die Gruppe b_0 bis b_5 angegeben, also ohne b_6 ($\setminus b_6$).

4.2.2 Beobachterübereinstimmung

Wie in Abschnitt 2.4.2 beschrieben, wird bei der Beobachterübereinstimmung erfasst, inwiefern die verschiedenen Beobachter die verschiedenen Samples jeweils exakt gleich bewertet haben. Tabelle 4.3 zeigt die nach Formel 2.2 bestimmten, paarweisen prozentualen Übereinstimmungen (PÜ) aller 21 Beobachterpaare. Die über alle Beobachterpaare gemittelte Übereinstimmung liegt bei $P\ddot{U} = 61\,\%$. Ohne die Urteile von Beobachter b_6 erhöht sich diese auf $P\ddot{U}_{\setminus b_6} = 74\,\%$. Mit $P\ddot{U}_{Zufall} = 37\,\%$ und $P\ddot{U}_{Zufall,\setminus b_6} = 46\,\%$ liegen beide PÜ-Werte deutlich über der bei Zufall erwarteten Übereinstimmung.

Die überzufällige Übereinstimmung der Beobachterurteile wird durch die nach den Formeln 2.4 - 2.6 bestimmten zufallskorrigierten Übereinstimmungsmaße Cohens κ und Scotts π bestätigt. Wie Tabelle 4.4 und Tabelle 4.5 zeigen, ergeben sich über die Mittelung aller paarweise bestimmten Maße Werte von $\kappa = 0{,}43$ und $\kappa_{\setminus b_6} = 0{,}52$ sowie $\pi = 0{,}38$ und $\pi_{\setminus b_6} = 0{,}52$.

Tabelle 4.3: Paarweise Prozentuale Übereinstimmung ($P\ddot{U}$).

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	$P\ddot{U}$	$P\ddot{U}_{\setminus b_6}$
b_0	1,00	0,81	0,77	0,66	0,78	0,86	0,26	0,69	0,78
b_1		1,00	0,75	0,63	0,76	0,81	0,23	0,66	0,75
b_2			1,00	0,63	0,73	0,76	0,28	0,65	0,73
b_3				1,00	0,66	0,70	0,34	0,60	0,66
b_4					1,00	0,79	0,29	0,69	0,74
b_5						1,00	0,25	0,69	0,78
b_6							1,00	0,27	–
								0,61	**0,74**

Da sich die Werte von κ und π unterscheiden, die Ausprägungen von $\kappa_{\setminus b_6}$ und $\pi\setminus b_6$ hingegen gleich sind, unterstreichen diese Ergebnisse die in Abschnitt 4.2.1 bereits beobachteten Gemeinsamkeiten und Unterschiede bezüglich der Randsummenverteilungen der einzelnen Beobachter. Die Randsummenverteilungen der Beobachter b_0 bis b_5 sind sehr ähnlich zueinander, wodurch $\kappa_{\setminus b_6}$ und $\pi\setminus b_6$ ähnliche bis gleiche Werte annehmen.

4.2.3 Beobachterreliabilität

Für die Eignungsprüfung der beobachterbasierten Videoannotationsmethode ist nicht nur das Maß an exakten Übereinstimmungen zwischen den Beobachtern relevant, sondern auch die Ähnlichkeit der relativen Lage der Beobachterurteile. Daher wird zusätzlich zu der Beobachterübereinstimmung auch die Beobachterreliabilität bestimmt. Da die zur Beurteilung verwendete Ratingskala intervallskaliert ist, alle Fußgänger von allen Beobachtern beurteilt wurden und die Gruppe der in dieser Studie eingesetzten Beobachter eine zufällige Stichprobe deutscher PKW-Fahrer repräsentiert, ist die über ein zweifaktorielles Modell berechnete ICC_{unjust} das geeignete Reliabilitätsmaß. Die zur Bestimmung der ICC_{unjust} benötigten Größen werden analog der Beschreibung in Anhang A.1 berechnet. Die Tabellen 4.6 und 4.7 zeigen die Ergebnisse dieser zwei-

Tabelle 4.4: Zufallskorrigiertes Übereinstimmungsmaß Cohens κ.

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	κ	$\kappa_{\setminus b_6}$
b_0	1,00	0,59	0,57	0,43	0,58	0,71	0,17	0,51	0,58
b_1		1,00	0,51	0,36	0,53	0,58	0,15	0,45	0,51
b_2			1,00	0,39	0,53	0,54	0,18	0,46	0,51
b_3				1,00	0,45	0,50	0,22	0,39	0,42
b_4					1,00	0,60	0,18	0,51	0,54
b_5						1,00	0,17	0,69	0,78
b_6							1,00	0,18	–
								0,43	**0,52**

Tabelle 4.5: Zufallskorrigiertes Übereinstimmungsmaß Scotts π.

	b_0	b_1	b_2	b_3	b_4	b_5	b_6	π	$\pi_{\setminus b_6}$
b_0	1,00	0,59	0,57	0,42	0,58	0,71	−0,02	0,48	0,57
b_1		1,00	0,51	0,34	0,52	0,58	−0,06	0,41	0,51
b_2			1,00	0,39	0,53	0,54	0,03	0,43	0,51
b_3				1,00	0,44	0,49	0,11	0,36	0,42
b_4					1,00	0,59	0,03	0,48	0,53
b_5						1,00	−0,03	0,48	0,58
b_6							1,00	0,01	–
								0,38	**0,52**

Tabelle 4.6: Ergebnisse der zweifaktoriellen Varianzanalyse für alle Beobachter.

Varianzquelle	QS	df	MS	F	p
Objekte (obj)	23.340	27.360	0,85	25,70	$0{,}000 < \alpha$
Beobachter (rat)	752	6	125,40	3.778,06	$0{,}000 < \alpha$
Rest (err)	5.449	164.160	0,03		

Tabelle 4.7: Ergebnisse der zweifaktoriellen Varianzanalyse ohne Beobachter b_6

Varianzquelle	$QS_{\setminus b_6}$	$df_{\setminus b_6}$	$MS_{\setminus b_6}$	F	$p_{\setminus b_6}$
Objekte (obj)	21.477	27.360	0,79	22,24	$0{,}000 < \alpha$
Beobachter (rat)	66	5	13,21	374,28	$0{,}000 < \alpha$
Rest (err)	4.829	136.800	0,04		

faktoriellen Varianzanalyse. Unter Verwendung der Formel 2.7 von Seite 45 berechnet sich die unjustierte Intraklassenkorrelation schließlich zu $\text{ICC}_{unjust} = 0{,}76$.

Da für die über Formel 2.9 bestimmte Prüfgröße $F_0 = 25{,}70 > F_{krit(0{,}95,\,27.360,\,164.160)} = 1{,}02$ gilt, kann angenommen werden, dass die Abweichung überzufällig ist. Das Konfidenzintervall der ICC_{unjust} ist $P\,(0{,}73 < \rho < 0{,}78) = 0{,}95$.

Ohne die Urteile des Beobachters b_6 erhöht sich die unjustierte Intraklassenkorrelation zu $\text{ICC}_{unjust,\setminus b_6} = 0{,}78$. Im justierten Modell führt der Ausschluss der Urteile des Beobachters b_6 hingegen zu keiner unterschiedlichen Koeffizientenausprägung. Es gilt $\text{ICC}_{just} = \text{ICC}_{just,\setminus b_6} = 0{,}78$. Auf Grund der zufällig ausgewählten Beobachter ist die ICC_{just} jedoch nur als Korrelationsmaß zu interpretieren.

Werden statt den Urteilen eines einzelnen Beobachters die Mittelwerte aller $k = 7$ Beobachter zur Einschätzung der Merkmalsausprägung der beurteilten Objekte verwendet, erhöht sich die Reliabilität nach Formel 2.10 auf $ICC_{unjust,\,MW} = 0{,}96$. Die Spearman-Brown-Schätzung (Formel 2.12) sagt denselben Wert vorher. Abbildung 4.4 zeigt die theoretisch geschätzte Reliabilität der Mittelwerte in Abhängigkeit der Anzahl der Beobachter.

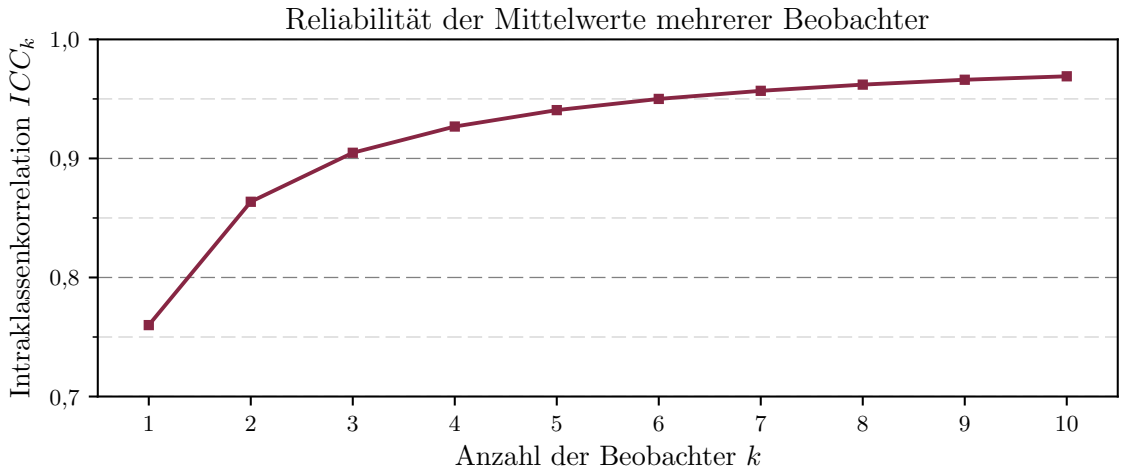


Abbildung 4.4: Nach der Spearman-Brown-Formel geschätzte Reliabilität der Mittelwerte in Abhängigkeit der Anzahl der Beobachter.

4.3 Diskussion und Bewertung

Wie oben beschrieben, wurde die zur Referenzbildung entwickelte, beobachterbasierte Videoannotation von einem Experten auf dem gesamten Datensatz sowie von sechs weiteren unabhängigen Beobachtern auf einem besonders relevanten Teil des Datensatzes durchgeführt.

Die **Verteilung der Beobachterurteile** zeigt, dass alle Beobachter alle fünf Stufen der Ratingskala zur Beurteilung der Querungsintention der Fußgänger verwendet haben. Dies bestätigt die in der Zielsetzung dieser Arbeit getätigte Annahme (s. Abschn. 1.2), dass bei der Erkennung der Querungsintention Unsicherheiten in der Beobachtung bestehen. Ein System, das auf die Abbildung des menschlichen Situationsbewusstseins zielt, muss somit in der Lage sein, auch die bestehenden Unsicherheiten in der Beobachtung abbilden zu können.

Die Verteilung der Beobachterurteile zeigt jedoch auch, dass sich trotz der stets gleich durchgeführten Beobachterschulung, das Verständnis über bestehende Unsicherheiten bei einzelnen Beobachtern unterscheidet. So zeigt Beobachter b_6 eine starke Tendenz zu den Kategorien, die unsicherheitsbehaftete Beobachtungen beschreiben. Beispielsweise gibt Beobachter b_6 in nur 6 % der Fälle an, dass der gezeigte Fußgänger *keinesfalls*

die Absicht hat, den Fahrstreifen zu queren. Der Mittelwert der restlichen Beobachter liegt in dieser Kategorie bei 65 %. Beobachter b_6 hat somit in vielen Situationen, die für die anderen Beobachter eindeutig sind, einen Restzweifel. Dies kann darauf zurückzuführen sein, dass Beobachter b_6 als jüngster Teilnehmer der Studie, die wenigsten Fahrerfahrungen hat und auch der einzige Teilnehmer ist, der seltener als einmal pro Woche ein Fahrzeug nutzt (s. Tabelle 4.1). Um diese Hypothese zu verifizieren, müssten jedoch weitere Studien durchgeführt werden, bei denen die Fahrerfahrung der Beobachter systematisch variiert wird. Es ist somit zu vermuten, dass die Angaben von Beobachter b_6 nicht dem Situationsbewusstsein eines erfahrenen, sicheren Fahrers entspricht. Daher sollten die Urteile nicht als Grundlage für das in dieser Arbeit zu entwickelnde System herangezogen werden.

Trotz der unterschiedlichen Randsummenverteilungen von Beobachter b_6 zu den anderen Beobachtern, liegt die **prozentuale Übereinstimmung** aller Beobachterurteile deutlich über der bei Zufall erwarteten Übereinstimmung. Dies spricht dafür, dass die Ausprägung der Querungsintention von Fußgängern kein subjektiv empfundenes Maß ist und die vorgestellte Methode prinzipiell dazu geeignet ist, diese auch zu messen.

Laut den in der Literatur angegebenen Grenzwerten (s. Abschn. 2.4.2, S. 43) zeigen auch die **zufallskorrigierten κ- und π-Werte** eine akzeptabel bis gute Übereinstimmung der Beobachterurteile. Der $\kappa_{\setminus b_6}$-Wert liegt mit 0,52 jedoch unter den aus der Literatur bekannten Werten zur beobachterbasierten Beurteilung von Manöverintentionen ($\hat{\kappa} = 0{,}62$). Hier ist jedoch zu beachten, dass die Beobachter bei Diederichs und Pöhler (2014) das Auftreten kleinerer, vordefinierter Verhaltenskategorien oder Bewegungseinheiten bewerten sollten, was eine weniger komplexe Aufgabe darstellt, als die Bewertung einer nicht direkt beobachtbaren Intention.

Aus der Übereinstimmung der Werte für $\kappa_{\setminus b_6}$ und $\pi_{\setminus b_6}$ lässt sich zudem ableiten, dass eine verringerte Koeffizientenausprägung ausschließlich auf den Effekt mangelnder Konsistenz zurück zu führen ist. Dass unter Berücksichtigung der Urteile von Beobachter b_6 hingegen $\pi < \kappa$ gilt, bestätigt, dass sich die Randsummenverteilung von Beobachter b_6 deutlich von der der anderen Beobachter unterscheidet. Die Ausprägung der Koeffizienten ist hier somit zusätzlich von dem Effekt unterschiedlicher Grundwahrscheinlichkeiten beeinflusst.

Die Ergebnisse zur **Beobachterübereinstimmung** zusammenfassend, können mit der zur Referenzbildung entwickelten Methode gute Ergebnisse erreicht werden, selbst wenn zur Abbildung eines genauen Unsicherheitslevels der Querungsintention eines Fußgängers eine exakte Übereinstimmung der Urteile gefordert wird.

Noch bessere Ergebnisse werden erreicht, wenn nur die Ähnlichkeit der relativen Lage der Beobachterurteile gefordert ist. Eine **Inter-Rater-Reliabilität** (ICC_{unjust}) von 0,76 bestätigt die Reliabilität der eingesetzten Videoannotationsmethode. Da bei der ICC_{unjust} die Mittelwertsunterschiede zwischen den Beobachtern zulasten der Reliabilitätsschätzung verrechnet werden, unterstreicht die Steigerung der $ICC_{unjust,\setminus b_6}$ auf 0,78 das unterschiedliche Mittelwertverständnis von Beobachter b_6 zu den anderen Beobachtern. Da die um den Effekt unterschiedlicher Mittelwerte bereinigte ICC_{just} auch unter Ausschluss der Urteile des Beobachters b_6 konstant bei 0,78 bleibt, zeigt aber auch, dass Beobachter b_6 in Relation zum jeweils eigenen Mittelwert die gleichen Objekte zuverlässig als „besser“ oder „schlechter“ einschätzt.

Die gute Inter-Rater-Reliabilität zwischen den unterschiedlichen Beobachtern bestätigt zudem, dass die Querungsintention von Fußgängern eine objektive Größe ist und mit der vorgestellten, beobachterbasierten Referenzmethode erfasst werden kann. Dabei kann die Verwendung der Urteile eines einzelnen Beobachters im Allgemeinen bereits als reliable Methode betrachtet werden. Die Gefahr, dass, wie bei Beobachter b_6, das Situationsbewusstsein eines zufällig ausgewählten Beobachters deutlich von dem der Allgemeinheit abweicht, kann dabei mit den oben empfohlenen Maßnahmen zur Auswahl und Schulung der Beobachter reduziert werden.

Da die **Verwendung der Mittelwerte** von nur zwei Beobachtern bereits zu einer theoretischen Steigerung der Reliabilität um 10 % auf 0,86 führt (s. Abb. 4.4), wird jedoch empfohlen, die Daten stets von mindestens zwei Beobachtern bewerten zu lassen und schließlich den Mittelwert der Urteile als Referenz zu verwenden. Der mit der Anzahl der Beobachter steigende Aufwand ist im Einzelfall gegen die gesteigerte Reliabilität abzuwägen.

Zusammenfassend empfiehlt es sich in dieser Arbeit, für den von allen Beobachtern bewerteten Teildatensatz die Mittelwerte aller Beobachter ohne die Urteile des Beobachters b_6 als Referenz für das zu entwickelnde System zur Erkennung der Que-

rungsintention von Fußgängern zu verwenden. Auf Grund der guten Beobachterübereinstimmung und -reliabilität ist aber auch davon auszugehen, dass bei den Daten, die ausschließlich von dem Experten beurteilt wurden, die Expertenurteile repräsentativ für das Situationsbewusstsein eines erfahrenen, sicheren Fahrers sind und sich diese somit ebenfalls als Referenz eignen.

Kapitel 5

Algorithmus zur Erkennung der Querungsintention

Auf Basis der Bewertungsergebnisse der in den vorangegangenen Kapiteln vorgestellten Inhalte wird ein neuer Ansatz zur Erkennung der Querungsintention von Fußgängern entwickelt. Der folgende Abschnitt 5.1 gibt einen Überblick über das Konzept des eigenen Ansatzes und geht auf die erforderlichen Eingangsdaten ein. Anschließend werden in den Abschnitt 5.2 bis Abschnitt 5.4 die einzelnen Komponenten des neuen Verfahrens ausführlich vorgestellt.

5.1 Überblick

Entsprechend der in Abschnitt 2.5 durchgeführten Bewertung, beruht der neue Ansatz zur Erkennung der Querungsintention von Fußgängern auf kontextbasierten sowie posenbasierten Informationen (s. Abb. 5.1). Diese Informationen werden, wie in der Zielsetzung dieser Arbeit definiert (s. Abschn. 1.2), auf Basis eines Mono-Frontkamera-Systems mit integrierter Fußgänger- und Fahrstreifenerkennung erzeugt (s. Abschn. 6.1).

Da aus dem Stand der Technik keine geeignete Beschreibungsform bekannt ist, die das kontextuelle Bewegungsverhalten eines Fußgängers generalisiert abbildet, wird ein eigenes Verfahren entwickelt, das die Fußgängerbewegung relativ zu relevanten Sze-

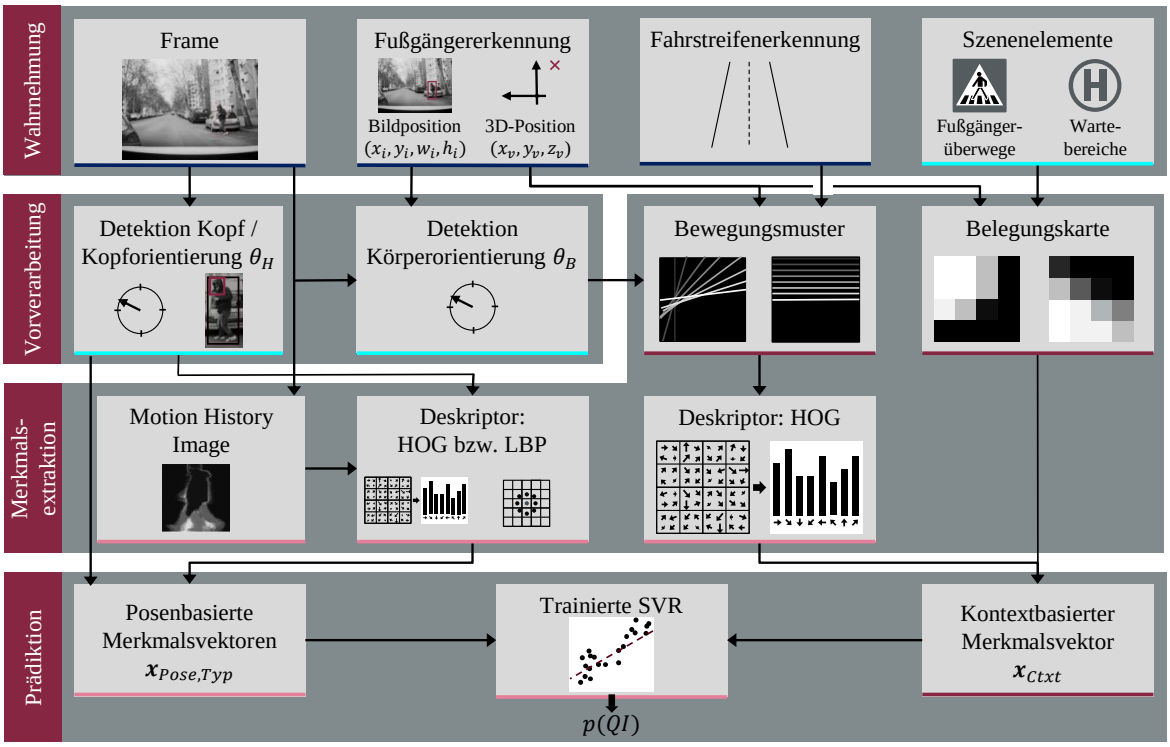


Abbildung 5.1: Neues Verfahren zur Erkennung der Querungsintention von Fußgängern. Die blau unterstrichenen Module werden von dem verwendeten Mono-Frontkamera-System bereitgestellt. Zudem wird angenommen, dass ein solches System zukünftig auch die cyan unterstrichenen Module bereitstellt, die in dieser Arbeit durch die Verwendung von Labels simuliert werden. Die rot unterstrichenen Module werden im Rahmen dieser Arbeit selbst entwickelt und implementiert, während die pink unterstrichenen Module Reimplementierungen bekannter Verfahren aus dem Stand der Technik entsprechen, die an das in dieser Arbeit betrachtete Problem und die verwendete Datenbasis angepasst sind.

nenelementen in Form eines Merkmalsvektors $\boldsymbol{x}_{Ctxt}$ beschreibt. Das neue Verfahren benötigt als Eingangsgrößen die (x_v, y_v, z_v)-Position sowie die Körperorientierung θ_B des Fußgängers relativ zum Fahrzeugkoordinatensystem (s. Anhang C.1). Zudem muss der Verlauf der Fahrstreifen und der Fahrbahnbegrenzung, sowie die Position von Infrastrukturelementen wie Fußgängerüberwege und Bushaltestellen bekannt sein. Das Vorgehen zur Bestimmung von $\boldsymbol{x}_{Ctxt}$ wird in Abschnitt 5.2 im Detail vorgestellt.

Zur Abbildung posenbasierter Informationen sind mit den MCHOG, dem PAF und den HOG bereits drei theoretisch geeignete Merkmale aus dem Stand der Technik bekannt. Daher wird im Rahmen dieser Arbeit evaluiert, inwieweit diese Merkmale geeignet sind, den kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ mit posenbasierten Informationen anzureichern. Zudem wird evaluiert, inwieweit eine explizite Beschreibung der Kopfpose zu einer Steigerung der Prädiktionsgüte führt.

Die betrachteten posenbasierten Merkmale basieren auf den von dem Kamerasystem gelieferten Bilddaten, sowie der in Form einer Bounding Box (x_i, y_i, w_i, h_i) angegebenen Bildposition des Fußgängers. Zudem wird die Bildposition des Fußgängerkopfes (x_h, y_h, w_h, h_h) sowie die Kopforientierung θ_H verwendet. Die in dieser Arbeit betrachteten posenbasierten Merkmalsvektoren $\boldsymbol{x}_{Pose,Typ}$ werden in Abschnitt 5.3 vorgestellt.

Entsprechend den Ergebnissen der in Abschnitt 3.5 durchgeführten Bewertung bekannter maschineller Lerntechniken, wird im neuen Ansatz zur Bestimmung des Zusammenhangs zwischen den Merkmalsvektoren und der Querungsintention eines Fußgängers eine SVR eingesetzt. Details des Trainings- und Anwendungsprozesses werden in Abschnitt 5.4 vorgestellt.

Die Implementierung des neuen Ansatzes erfolgte in dem auf der Programmiersprache C++ basierenden Automotive Data and Time-Triggered Framework (ADTF) (Elektrobit, 2017). Die im Folgenden beschriebenen Komponenten wurden dabei als einzelne ADTF-Filter umgesetzt. Für die Bildverarbeitungsanteile wurde die OpenCV-Bibiliothek (OpenCV, 2017) eingesetzt.

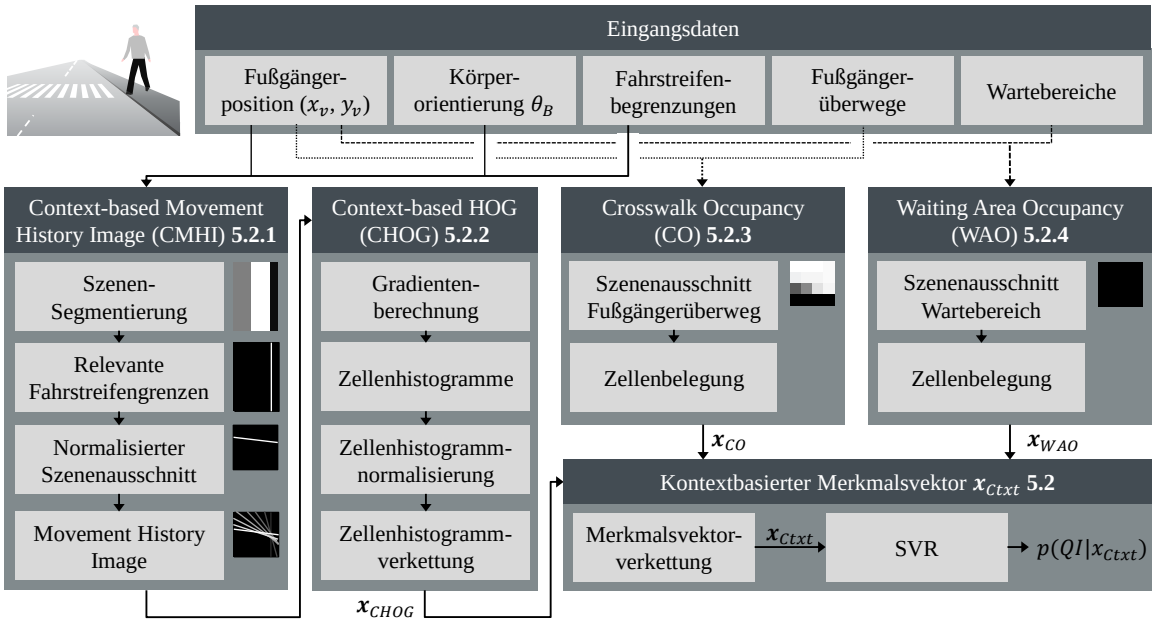


Abbildung 5.2: Neues Verfahren zur kontextbasierten Erkennung der Querungsintention von Fußgängern.

5.2 Kontextbasierte Erkennung der Querungsintention

Das neue Verfahren zur kontextbasierten Erkennung der Querungsintention eines Fußgängers basiert auf einer generischen Repräsentation der Fußgängerbewegung relativ zu Szenenelementen, die relevant für die Bewertung der Querungsintention des Fußgängers scheinen. Abbildung 5.2 gibt einen Überblick über das neue Verfahren.

Das neue Verfahren erstellt einen Merkmalsvektor $\boldsymbol{x}_{Ctxt}$, der aus drei Elementen besteht. Zum einen wird ein Merkmalsvektor $\boldsymbol{x}_{CHOG}$ erstellt, der das Bewegungsmuster des Fußgängers in Relation zu relevanten Fahrstreifen- oder Fahrbahnbegrenzungen beschreibt. Hierzu wird ein Bild erzeugt, das die kontext-basierte Bewegungshistorie des Fußgängers abbildet. Die Erstellung dieses Context-based Movement History Image (CMHI) wird in Abschnitt 5.2.1 beschrieben. Auf Basis des CMHI werden anschließend normalisierte Histogramme der Kantenorientierungen (Context-based Histograms of Oriented Gradients (CHOG)) bestimmt, die in Kombination mit drei

zusätzlichen situationsbezogenen Merkmalen zum Merkmalsvektor $\boldsymbol{x}_{CHOG}$ verkettet werden (s. Abschn. 5.2.2).

Zu dem $\boldsymbol{x}_{CHOG}$ werden zwei weitere Merkmalsvektoren $\boldsymbol{x}_{CO}$ (s. Abschn. 5.2.3) und $\boldsymbol{x}_{WAO}$ (s. Abschn. 5.2.4) erstellt, die die räumliche Anordnung von Szenenelementen wie Fußgängerüberwege aller Art und Wartebereiche beschreiben. Hierzu wird bestimmt, welche Bereiche um den Fußgänger mit einem Fußgängerüberweg (Crosswalk Occupancy (CO)) oder einem Wartebereich (Waiting Area Occupancy (WAO)) belegt sind. Der finale kontextbasierte Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ wird schließlich aus der Verkettung der drei Vektoren mit

$$\boldsymbol{x}_{Ctxt} = [\boldsymbol{x}_{CHOG}, \boldsymbol{x}_{CO}, \boldsymbol{x}_{WAO}] \tag{5.1}$$

gebildet.

5.2.1 Context-based Movement History Image (CMHI)

Wie in Abschnitt 2.5 diskutiert, kann als querungsinitiierende Verhaltensweise eine Bewegung des Fußgängers zur gewählten Querungsstelle und damit eine Ausrichtung seiner Position und Orientierung an die Fahrbahnkante beobachtet werden. Um dieses abzubilden, wird das Context-based Movement History Image (CMHI) vorgestellt, dass die Bewegung des Fußgängers in Relation zu relevanten Fahrstreifen- oder Fahrbahnbegrenzungen beschreibt.

Szenen-Segmentierung

Zur Entscheidung, welche Fahrstreifen- oder Fahrbahnbegrenzungen relevant sind, um auf die Querungsintention eines Fußgängers zu schließen, ist zunächst ein Szenenverständnis zu generieren. Wie in Abbildung 5.3 dargestellt, wird die aktuelle Szene hierzu in vier Zonen eingeteilt:

1. die Ego-Zone Z_{ego}: definiert durch den Bereich des eigenen Fahrstreifens
2. die Straßen-Zone Z_{str}: bestehend aus allen Fahrstreifen, außer dem eigenen Fahrstreifen

Abbildung 5.3: Beispielhafte Segmentierung einer Szene mit einer Ego-Zone (Z_{ego}, Pink), einer Straßen-Zone (Z_{str}, Grau), einer Gemischten-Zone (Z_{mix}, Cyan) und zwei Gehweg-Zonen (Z_{swk}, Blau). Das weiße Rechteck im rechten Bild repräsentiert die Kontur des an der Position des Fußgängers (rotes Kreuz) zentrierten, relevanten Szenenausschnitts. Die rote gestrichelte Linie markiert die relevante Fahrstreifengrenzen.

3. die Gemischte-Zone Z_{mix}: bestehend aus Bereichen, die zwischen der befahrenen Fahrstreifen und dem Gehweg liegen und sowohl vom Fahrzeug-, als auch vom Fußverkehr genutzt werden (beispielsweise Bushaltebuchten oder Parkbuchten)

4. die Gehweg-Zone Z_{swk}: definiert als Bereich der nur für den Fußverkehr zugelassen ist und von Z_{str} oder Z_{mix} im Regelfall durch einen Bordstein abgegrenzt ist.

Zur Segmentierung der Szene wird die als Grundebene bezeichnete Fläche um das Fahrzeug als ein $M_s \times N_s$ px großes Szenenbild mit einer Auflösung von r_s px/m repräsentiert und jeder Zone ein definierter Pixelwert zugewiesen (s. Abb. 5.2). Die Zone Z_{ped}, in der sich ein Fußgänger aktuell befindet, lässt sich somit über eine Transformation der Fußgängerposition in Fahrzeugkoordinaten (x_v, y_v, z_v) zu Szenenbildkoordianten (x_s, y_s) mit

$$x_s = round(M_s * 0{,}5 - y_v * r_s) \tag{5.2}$$

$$y_s = N_s - round(x_v * r_s)) \tag{5.3}$$

und der anschließenden Bestimmung des Pixelwerts an der Stelle (x_s, y_s) ermitteln.

Relevante Fahrstreifengrenzen

Die oben beschriebenen Zonen werden durch die vom Fahrstreifenerkennungssystem detektierten Fahrstreifen begrenzt (s. Abschn. 6.1.2). Die Fahrstreifengrenzen E_{ped}, die relevant sind, um die auf eine Querungsintention hinweisende Bewegung des Fußgängers zum Egofahrstreifen zu beschrieben, sind abhängig von der aktuellen Zone des Fußgängers Z_{ped}, der Seite S_{ped} auf der sich der Fußgänger gerade befindet, mit

$$S_{ped} = \begin{cases} links & \text{wenn } y_v > 0\,\text{m} \quad (5.4) \\ rechts & \text{sonst,} \quad (5.5) \end{cases}$$

sowie von der aktuellen Zonenanordnung. Gilt beispielsweise $S_{ped} = rechts$ und $Z_{ped} = Z_{swk}$, und grenzt eine Gemischte-Zone an die Gehweg-Zone Z_{ped}, dann ist eine Bewegung des Fußgängers in Richtung der Grenze zwischen Z_{mix} und Z_{swk} bereits ein Indikator für die Intention des Fußgängers, die an Z_{mix} grenzende Ego-Zone Z_{ego} queren zu wollen und nicht erst die Bewegung des Fußgängers in Richtung der Grenze zwischen Z_{mix} und Z_{ego}. Die Bestimmung der relevanten Fahrstreifengrenzen ist als Lookup-Tabelle umgesetzt. Abbildung 5.4 gibt einen Überblick über die relevanten Fahrstreifengrenzen in Situationen mit maximal einem Fahrstreifen pro Fahrtrichtung.

Normalisierter Szenenausschnitt

Um für jeden Fußgänger eine normalisierte Beschreibung der aktuellen Situation zu erhalten, wird ein $M_p \times N_p$ px großer Ausschnitt $I(t)$ des Szenenbildes erstellt, der die relevanten Fahrstreifengrenzen E_{ped} zeigt und an der Position des Fußgängers zentriert ist. Zur Normalisierung wird der Szenenausschnitt um die gegebene Körperorientierung des Fußgängers θ_B gedreht und horizontal gespiegelt, falls $S_{ped} = links$ und $Z_{ped} \neq Z_{ego}$. Somit wird eine Situation, in der ein Fußgänger einer geraden Straße zugewandt ist, durch einen normalisierten Szenenausschnitt mit einer horizontalen Linie abgebildet. Eine Situation mit einem Fußgänger, der parallel zu einer geraden Straße ausgerichtet ist, resultiert in einem normalisierten Szenenausschnitt mit einer vertikalen Linie. Die Seite der Linie ist dabei abhängig davon, ob der Fußgänger dem Fahrzeug zugewandt

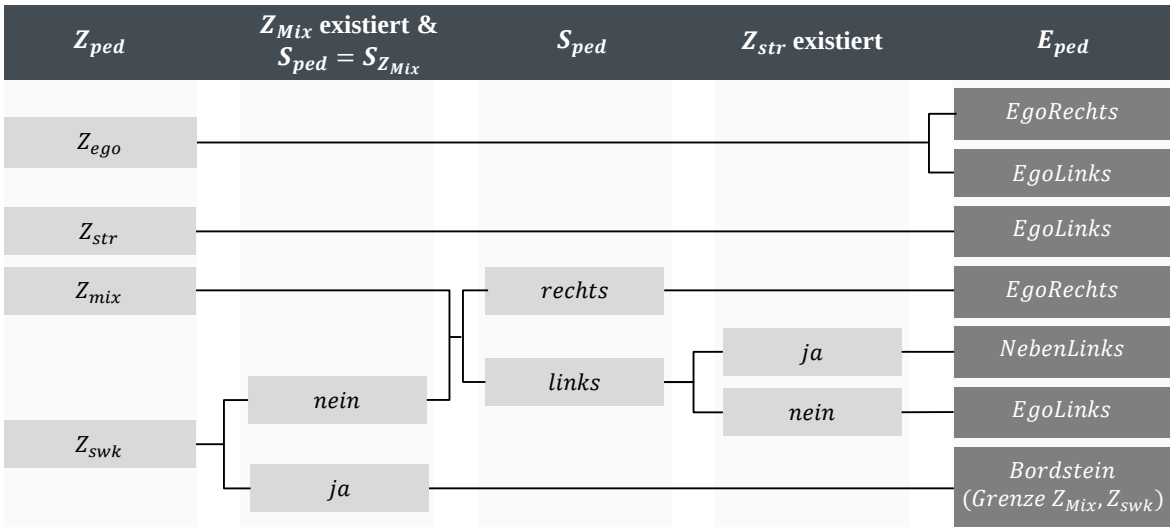


Abbildung 5.4: Relevante Fahrstreifengrenzen abhängig von der Fußgängerposition und der Zonenanordnung für Situationen mit maximal einem Fahrstreifen pro Fahrtrichtung.

ist (Linie ist auf der rechten Seite) oder mit dem Rücken zum Fahrzeug steht (Linie ist auf der linken Seite). Durch die Zentrierung des Szenenausschnitts an der Fußgängerposition gilt: Je näher sich die Linie in der Mitte des normalisierten Szenenausschnitts befindet, desto näher ist der Fußgänger an der relevanten Fahrbahnkante.

Movement History Image

Um die Bewegung des Fußgängers über mehrere Zeitschritte zu beschreiben, wird ähnlich dem MHI von Köhler et al. (2012) (s. Abschn. 2.3.2, S. 119) der jeweils aktuelle normalisierte Szenenausschnitt $I(t)$ mit einer Sequenz vorangegangener normalisierter Szenenausschnitte $\Psi(I(t-1))$ zu einem Movement History Image $H(t)$ über

$$H_{x,y}(t) = \begin{cases} \tau & \text{wenn } I_{x,y}(t) \neq 0 \quad (5.6) \\ \max(0, H_{x,y}(t-1) - \delta) & \text{sonst} \quad (5.7) \end{cases}$$

zusammengefügt. Analog zur Gleichung 5.10 entspricht τ einer Zerfallsvariable, die die Länge der betrachteten Historie bestimmt. δ bestimmt die Stärke der Abschwächung der vorangegangenen Szenenausschnitte.

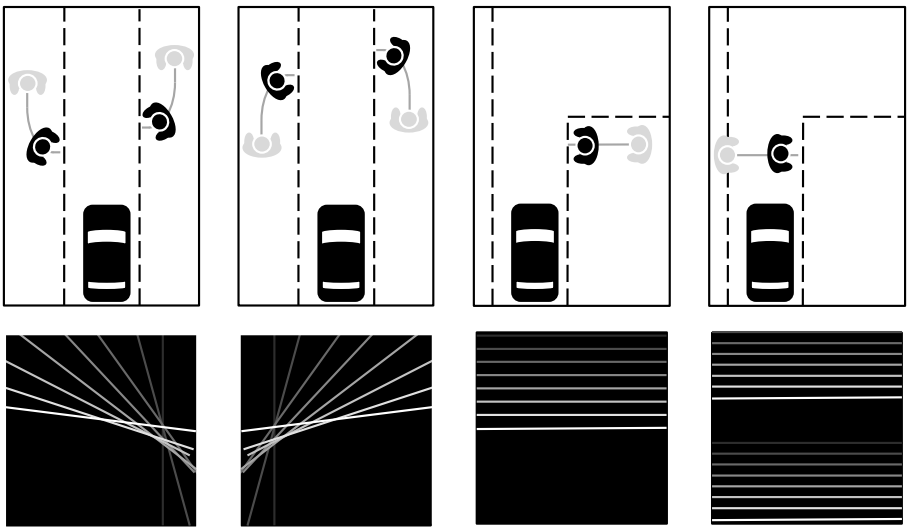


Abbildung 5.5: Schematische Darstellung des CMHI für verschiedene Fußgängerbewegungen (der Fußgänger startet an der grauen Position und bewegt sich entlang der Linie bis zur schwarzen Position).

Wird der Abstand zwischen zwei Zeitschritte konstant gehalten, beschreibt das CMHI über die Distanz zweier zeitlich aufeinanderfolgenden Fahrstreifengrenzen implizit die Geschwindigkeit der Bewegung eines Fußgängers. Dieses gilt nicht bei Fußgängern, die sich parallel zu einer geraden Straße bewegen. Abbildung 5.5 zeigt die CMHIs typischer Fußgängerbewegungen. Wie zu sehen ist, repräsentiert das CMHI die verschiedenen Situationen über charakteristische Bildmuster.

5.2.2 Context-based Histograms of Oriented Gradients (CHOG)

Zur Erfassung der im CMHI abgebildeten Bewegungsinformationen werden normalisierte Histogramme der Kantenorientierungen des CMHI bestimmt. Hierbei wird analog zum Vorgehen von Köhler et al. (2012) (s. Abschn. 2.3.2, S. 33) eine adaptierte Version des originalen HOG-Ansatzes verwendet. Die adaptierte Version verzichtet nach der Normalisierung der Zellenhistogramme auf eine weitere Normalisierung der Zellen über Blöcke, damit lokale Unterschiede zwischen benachbarten Zellen erhal-

ten bleiben. Die restliche Bestimmung des HOG entspricht dem in Abschnitt 3.4.1 beschriebenen Vorgehen.

Das im CMHI abgebildete Bewegungsmuster ist nur im Zusammenhang mit der aktuellen Zone des Fußgängers aussagekräftig. Daher wird den zu einem Vektor verketteten Zellenhistogrammen die aktuelle Zone des Fußgängers Z_{ped} in Form eines Integerwertes vorangestellt. Da diese Kombination aus Z_{ped} und dem CMHI nach der in (Schneemann und Heinemann, 2016) durchgeführten Evaluation jedoch nicht ausreicht, um einen mit einer wechselnden Querungsintention verknüpften Zonenwechsel des Fußgängers (z.B. von Z_{ego} zu Z_{sdw}) zu erlernen, wird zudem die Zone des Fußgängers im vorangegangenen Zeitschritt $t-1$ dem Vektor hinzugefügt. Schließlich wird die Anzahl der Zeitschritte $(t-t_0)$, seitdem der Fußgänger erstmalig detektiert wurde, als zusätzliches Merkmal ergänzt. Hierdurch sind stets Informationen über die Länge der Beobachtung vorhanden, auch in Situationen, in denen das CMHI diese nicht beinhaltet, da der Fußgänger einen konstanten Abstand und eine konstante Orientierung zu E_{ped} hat. Dies gilt zum Beispiel für stehende Fußgänger oder Fußgänger, die sich parallel zu einer geraden Fahrstreifengrenze bewegen. Die Kombination der normalisierten Zellenhistogramme mit den drei zusätzlichen Merkmalen bildet schließlich den Merkmalsvektor $\boldsymbol{x}_{CHOG}$.

5.2.3 Crosswalk Occupancy (CO)

Um die Präsenz von Fußgängerüberwegen wie Fußgängerampeln, Zebrastreifen oder Fußgängerinseln zu beschreiben, wird der Merkmalsvektor $\boldsymbol{x}_{CO}$ bestimmt. Dieser bildet die Position eines Fußgängerüberwegs relativ zur aktuellen Position und Orientierung des Fußgängers zum aktuellen Zeitpunkt t ab. Eine Historie wird beim $\boldsymbol{x}_{CO}$ nicht betrachtet, da die Begrenzung eines Fußgängerüberwegs stets mit der Begrenzung der Straße korrespondiert, und die Bewegung eines Fußgängers zu einem Überweg somit bereits im CMHI abgebildet ist.

Zur Erstellung des CO-Merkmalsvektors wird analog zum CMHI ein ebenfalls $M_p \times N_p$ px großer Szenenausschnitt erstellt, der in diesem Fall die durch einen Fußgängerüberweg belegte Fläche aus der Vogelperspektive beschreibt, mit der aktuellen

Fußgängerposition als Zentrum. Abhängig vom Typ des Überwegs, wird allen Pixeln, die zum Fußgängerüberweg gehören, ein definierter Wert zugewiesen. Anschließend wird analog zum CMHI der Szenenausschnitt um die Körperorientierung θ_B gedreht und horizontal gespiegelt, falls $S_{ped} = links$ und $Z_{ped} \neq Z_{ego}$.

Anschließend wird der Szenenausschnitt in r_o Zellen pro Meter aufgeteilt und pro Zelle bestimmt, wie viel Prozent der Pixel durch einen Fußgängerüberweg belegt sind. Für den finalen Merkmalsvektor $\boldsymbol{x}_{CO}$ werden schließlich alle Zellenbelegungswerte miteinander verkettet und der Überwegtyp in Form eines Integerwertes vorangestellt.

5.2.4 Waiting Area Occupancy (WAO)

Die Präsenz von Wartebereichen, wie Bus- oder Straßenbahnhaltestellen, kann ein Indikator dafür sein, dass ein Fußgänger trotz einer Ausrichtung seiner Position und Orientierung in Richtung Straße, keine Querungsintention hat. Daher wird die räumliche Anordnung solcher Szenenelemente relativ zur Fußgängerposition in dem Merkmalsvektor $\boldsymbol{x}_{WAO}$ abgebildet. Dieser wird analog zum $\boldsymbol{x}_{CO}$ gebildet, mit der einzigen Ausnahme, dass kein Integerwert für den Wartebereichtyp angehängt wird.

5.3 Posenbasierte Erweiterung zur Erkennung der Querungsintention

Um den kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ mit posenbasierten Informationen anzureichern, werden mehrere aus dem Stand der Technik bekannten Merkmale implementiert und evaluiert. Ziel ist es, ein Merkmal zu finden, dass die für die Erkennung der Querungsintention relevanten Elemente der Körper- und Kopfpose hinreichend beschreibt.

5.3.1 Betrachtete Merkmale

Wie in Abschnitt 2.5 diskutiert, werden zwei der im Bereich der posenbasierten Aktionserkennung bereits eingesetzten Merkmale als für diese Arbeit prinzipiell geeignet bewertet. Hierbei handelt es sich um die **Motion Contour Histograms of Oriented Gradients (MCHOG)** von Köhler et al. (2012) und das **Posture Appearance Feature (PAF)** von Furuhashi und Yamada (2011).

Während das MCHOG über ein MHI die Körperbewegung des Fußgängers explizit beschreibt, kombiniert das PAF die, über eine dimensionsreduzierte Variante der HOG beschriebene, aktuelle Körperhaltung des Fußgängers mit zeitlich zurückliegenden Körperhaltungen, um so eine implizite Beschreibung der Körperbewegung des Fußgängers zu erhalten. Für Details zu den zwei Merkmalen wird auf die Beschreibung in Abschnitt 2.3.2 verwiesen. Der über das MCHOG erstellte Merkmalsvektor wird im Folgenden als $\boldsymbol{x}_{Pose,MCHOG}$ bezeichnet und der des PAF als $\boldsymbol{x}_{Pose,PAF}$.

Zudem werden die nicht dimensionsreduzierten, aber nur die Körperhaltung und nicht die -bewegung beschreibenden, klassischen **Histograms of Oriented Gradients (HOG)** von Dalal und Triggs (2005) evaluiert. Diese werden im Folgenden als $\boldsymbol{x}_{Pose,HOG}$ bezeichnet. Details zu der Erstellung sind in Abschnitt 3.4.1 zu finden.

Da, wie in in Abschnitt 2.5 diskutiert, alle drei bisher genannten Merkmale Schwächen bezüglich der expliziten Abbildung der für die Erkennung der Querungsintention wichtigen Kopfpose aufweisen, werden zudem zwei Merkmale betrachtet, die sich aus-

schließlich auf den Kopfbereich beziehungsweise die Kopforientierung des Fußgängers beziehen.

Um zunächst evaluieren zu können, welchen Beitrag die Kopforientierung und Kopfbewegung des Fußgängers bei der Erkennung der Querungsintention leisten kann, wird als weiteres posenbasiertes Merkmal der absolute Wert der Kopforientierung θ_H in Kombination mit der Differenz θ_Δ zur Körperorientierung θ_B mit

$$\theta_\Delta = \begin{cases} \theta_B - \theta_H - 360° & \text{wenn } \theta_B - \theta_H > 180° \quad (5.8a) \\ \theta_B - \theta_H + 360° & \text{wenn } \theta_B - \theta_H < -180° \quad (5.8b) \\ \theta_B - \theta_H & \text{sonst} \quad (5.8c) \end{cases}$$

betrachtet. Um damit auch die Bewegung des Kopfes zu erfassen, werden n aufeinander folgende θ_Δ-Werte in einem Abstand von l Frames gesampelt und miteinander kombiniert. Der im Folgenden als **Orientation Feature (OF)** bezeichnete Merkmalsvektor $\boldsymbol{x}_{Pose,OF_{(n,l)}}$ ergibt sich somit aus:

$$\boldsymbol{x}_{Pose,OF_(n,l)} = [(\theta_{H,t_0}, \theta_{\Delta,t_0}), \ldots, (\theta_{H,t_{n*l}}, \theta_{\Delta,t_{n*l}})] \quad (5.9)$$

und betrachtet eine Historie von $t = (n-1) * l + 1$ Frames.

Um schließlich die Leistungsfähigkeit visueller Deskriptoren bei der Abbildung der Kopfpose zu evaluieren, wird als letztes Merkmal das in Abschnitt 3.4.2 beschriebene und im Stand der Technik zur Detektion der Kopforientierung zum Beispiel bei Rehder et al. (2014) eingesetzte **Local Binary Pattern (LBP)** betrachtet. Das LBP wird dabei nur für den von einem System zur Kopfdetektion ausgegebenen Bildbereich bestimmt. Analog zum PAF wird auch hier über die Verkettung von n zeitlich um l Frames versetzter LBPs versucht, die Bewegung des Fußgängerkopfes implizit zu beschreiben. Der so erstellte Merkmalsvektor wird im Folgenden als $\boldsymbol{x}_{Pose,LBP}$ bezeichnet.

5.3.2 Implementierung der Merkmalsextraktion

Abbildung 5.6 zeigt das zur Extraktion der fünf betrachteten posenbasierten Merkmale implementierte Verfahren. Zunächst ist für alle Merkmale außer dem OF eine Vorverarbeitung der von dem Fußgänger- bzw. Kopfdetektionssystem ausgegebenen

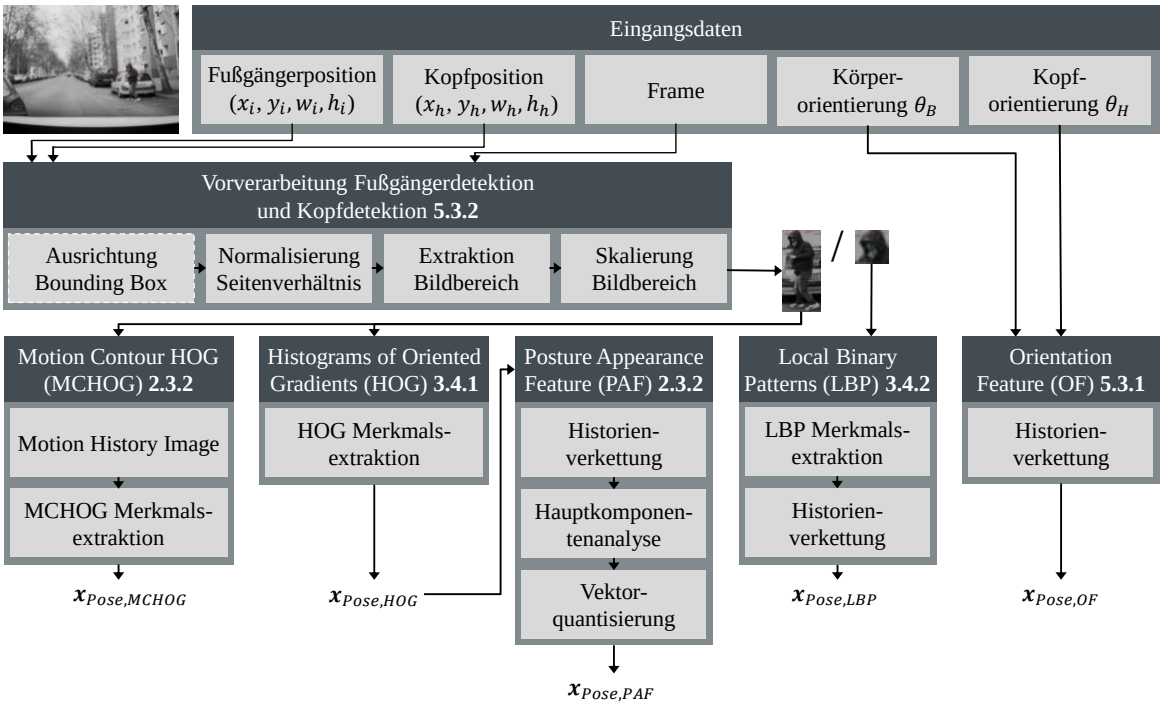


Abbildung 5.6: Verfahren zur Extraktion der betrachteten posenbasierten Merkmale.

Bounding Box nötig. Während der Implementierung der Merkmalsextraktionsverfahren musste festgestellt werden, dass das in dieser Arbeit verwendete Fußgängerdetektionssystem ein instabiles Tracking der Bounding Box über aufeinanderfolgende Frames aufweist und sich die detektierte Bildposition (x_i, y_i, w_i, h_i) daher oft sprunghaft ändert (s. Abb. 5.7). Das ist vor allem für die bildbasierte Beschreibung der Bewegung des Fußgängers problematisch, da beim Vergleich der extrahierten Bildbereiche verschiedener Zeitschritte die durch die sprunghaften Bounding Box Positionen erzeugten Bildunterschiede deutlich stärker sind, als die durch die Bewegung des Fußgängers erzeugten Änderungen. Dadurch bildet beispielsweise ein MHI eher die Bewegung der Bounding Box ab, als die des Fußgängers (s. Abb. 5.7, links).

Aus diesem Grund werden in dem ersten Vorverarbeitungsschritt die Bildpositionen der Fußgängerdetektionen an die, in dieser Arbeit deutlich stabiler zur Verfügung stehenden, Detektionen der Fußgängerköpfe (x_h, y_h, w_h, h_h) ausgerichtet, um diese so, auch über mehrere Frames, zu stabilisieren (s. Abb. 5.7, rechts). Hierzu wird zunächst die obere Kante der Fußgänger Bounding Box an der oberen Kante des Kopfes ausge-

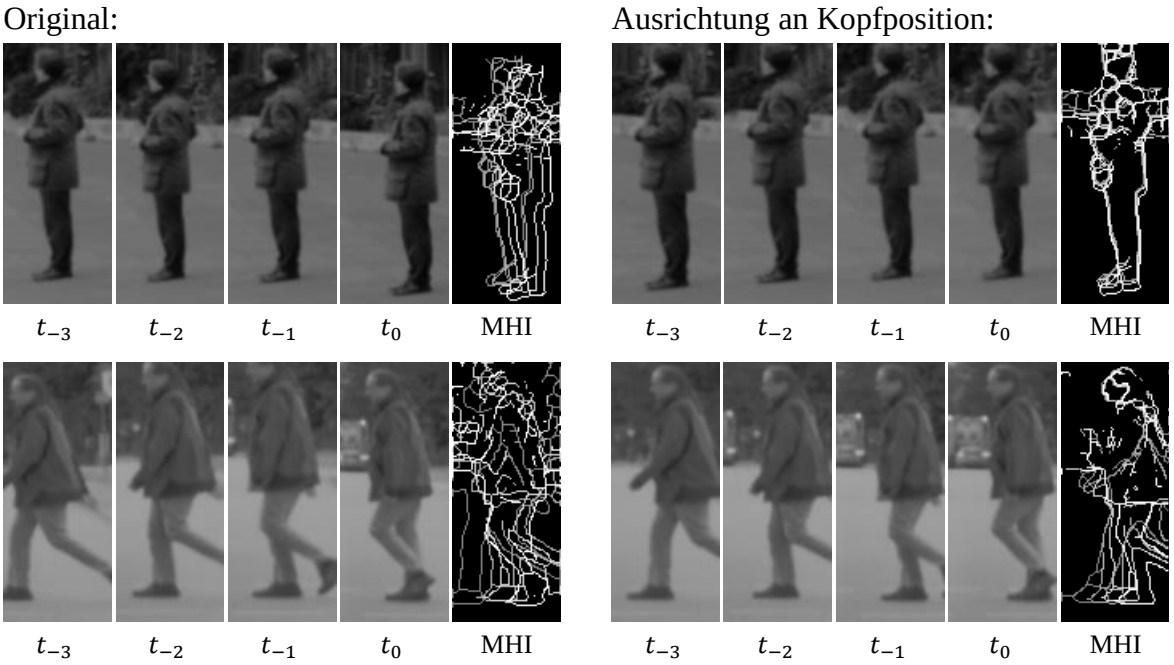


Abbildung 5.7: Sprunghafte Änderungen der Bounding Box Positionen führen zu dominierenden Bewegungen im MHI (links). Eine Ausrichtung über die Kopfposition führt zu einem verbesserten MHI (rechts).

richtet ($y_i = y_h$). Anschließend wird die horizontale Position der Fußgänger Bounding Box an den horizontalen Mittelpunkt des Kopfes mit $x_c = x_h + w_h/2$ angeglichen.

Während diese Ausrichtung nur für die Fußgänger Bounding Box durchgeführt wird und ausschließlich für Merkmale relevant ist, die eine Bewegungshistorie betrachten, sind die nachfolgenden drei Vorverarbeitungsschritte sowohl für die Fußgänger- als auch für die Kopfdetektionen notwendig. Da alle in dieser Arbeit betrachteten Merkmale Bilddaten mit einer festen Fenstergröße voraussetzen, wird zunächst jede Bounding Box so vergrößert, dass ihr Seitenverhältnis mit dem der vorgegebenen Fenstergröße übereinstimmt. Anschließend wird der durch die normalisierte Bounding Box begrenzte Bildbereich aus dem aktuellen Frame extrahiert und auf die vorgegebene Fenstergröße skaliert.

Die anschließende Merkmalsextraktion wird entsprechend der, in den jeweils referenzierten Abschnitten beschriebenen, Originalimplementierung durchgeführt, mit den im Folgenden beschriebenen Anpassungen an die vorhandene Datenbasis.

In dieser Arbeit werden die Videodaten eines nichtstationären Mono-Kamerasystems verwendet. Daher können bei den MCHOG zur Bestimmung des MHI die zum Hintergrund gehörenden Pixel weder auf Basis eines statischen Hintergrundbilds, wie in (Köhler et al., 2012), noch über Tiefendaten, wie in Köhler et al. (2013), gefiltert werden. Um die Störungen der Hintergrundbereiche auf das MHI trotzdem möglichst gering zu halten, wird die in Abbildung 5.8 gezeigte, angepasste Prozesskette zur Erstellung des MHI, verwendet.

Bei dieser wird nach der Skalierung und Gradientenberechnung zunächst der Hintergrundbereich, der sich seitlich zum detektierten Kopfbereich befindet, gefiltert. Anschließend wird das Gradientenbild binarisiert und mit Morphologischen Operatoren (Dougherty und Lotufo, 2003) gefiltert, um kleine, in der Regel dem Hintergrund zugehörige Bereiche, zu entfernen sowie kleine Löcher innerhalb der Konturen zu schließen. Abschließend werden die Konturen mit dem *Thinning*-Algorithmus von Zhang und Suen (1984) auf eine Liniestärke von 1 px skelettiert und damit das MHI des vorangegangenen Zeitschritts aktualisiert.

Wie in Abschnitt 6.1.1 angegeben, liegt die durchschnittliche Bounding Box Größe eines Fußgängers in dem in dieser Arbeit verwendeten Datensatz bei 52×139 px.

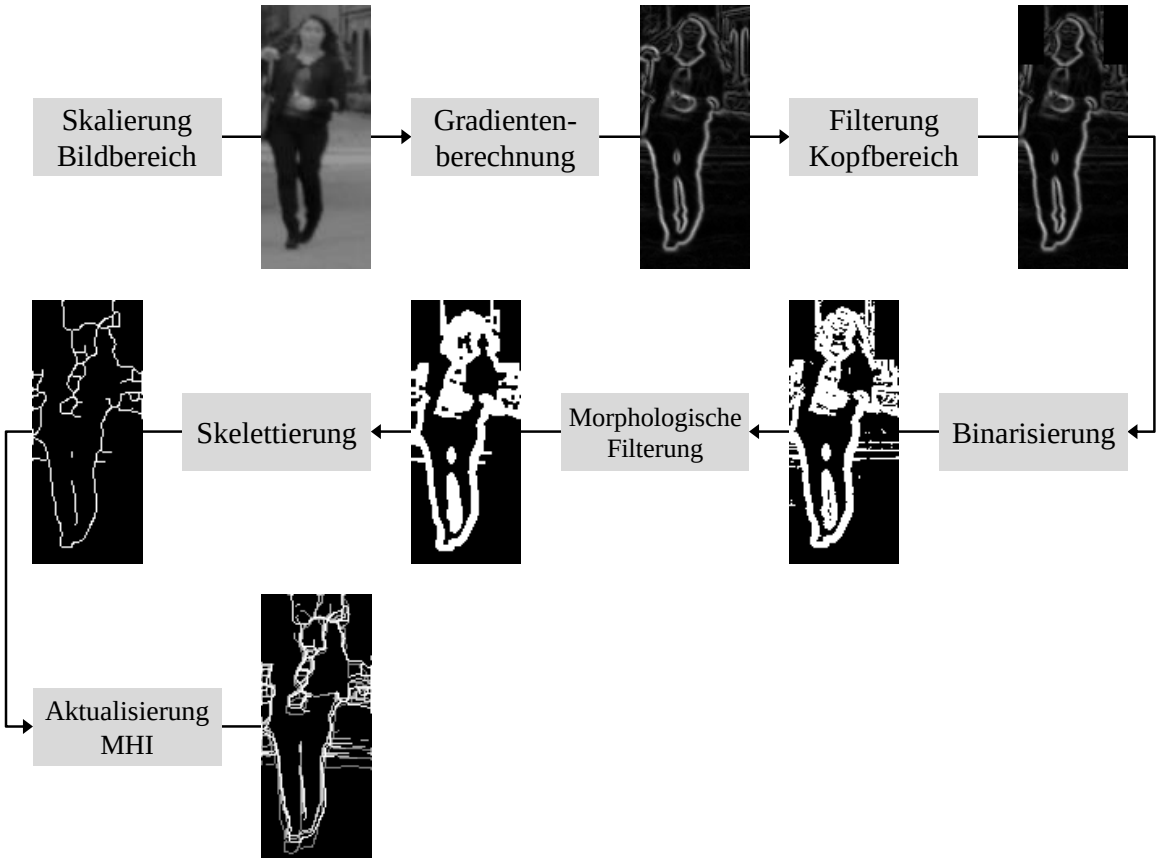


Abbildung 5.8: An die vorhandene Datenbasis angepasster Erstellungsprozess des MHI.

Tabelle 5.1: Parametrisierung des MCHOG-, PAF- und HOG-Merkmalsvektor.

Merkmal	Fenstergröße	Zellengröße	Blockgröße	Bins	Historie
$\boldsymbol{x}_{Pose,MCHOG}$	$48 \times 136\,\text{px}$	$8 \times 8\,\text{px}$	–	12	$\tau = 4$
$\boldsymbol{x}_{Pose,PAF}$	$50 \times 140\,\text{px}$	$10 \times 10\,\text{px}$	$30 \times 30\,\text{px}$	9	$n = 2, l = 4$
$\boldsymbol{x}_{Pose,HOG}$	"	"	"	"	–

Sie weicht damit von der von Köhler et al. (2013) für die MCHOG sowie von der von Furuhashi und Yamada (2011) für das PAF verwendeten Fenstergröße ab. Um die Zellenhistogramme der MCHOG und des, auf dem klassischen HOG basierenden PAF, möglichst ähnlich zur Originalimplementierung zu gestalten, wird die in Tabelle 5.1 gezeigte Parametrisierung verwendet. Zudem werden die, die Historie beeinflussenden Parameter an die vorhandene Bildwiederholrate von 18 fps (s. Abschn. 6.1) angepasst. Dadurch ergibt sich für die MCHOG eine Historie von 0,2 s und für das PAF eine von 0,28 s.

Das OF findet kein direktes Vorbild im Stand der Technik. Daher werden die Historienparameter n und l im Rahmen der durchgeführten Evaluation variiert. Bei den betrachteten Parameterkombinationen wird sich sowohl an dem PAF orientiert ($n = 2, l = 4$), als auch an dem in (Kloeden et al., 2014) durchgeführten Sampling der Kopforientierung in 100 ms-Schritten sowie an der vom kontextbasierten Merkmalsvektor betrachteten Historie von 2,8 s (s. Abschn.6.4.1, S. 147 f.).

Der Merkmalsvektor $\boldsymbol{x}_{Pose,LBP}$ wird folglich nur für die als am besten bewerteten Historienparameter evaluiert. Die LBP-Histogramme werden dabei, entsprechend der Standardimplementierung (s. Abschn. 3.4), mit $p = 8$ Nachbarn, einem Radius von $r = 1$ und einer Zellengröße von $8 \times 8\,\text{px}$ bestimmt. Entsprechend der durchschnittlichen Größe der Fußgängerköpfe im verwendeten Datensatz (s. Tab. 6.3, S. 124) werden die Kopfdetektionen hierzu auf eine Größe von $24 \times 24\,\text{px}$ skaliert.

Diese über die beschriebenen Prozesse generierten Merkmalsvektoren $\boldsymbol{x}_{Pose,MCHOG}$, $\boldsymbol{x}_{Pose,PAF}$, $\boldsymbol{x}_{Pose,HOG}$, $\boldsymbol{x}_{Pose,OF_{(n,l)}}$ und $\boldsymbol{x}_{Pose,LBP}$ werden schließlich einzeln sowie in Kombination mit dem kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ zum Training und zur Anwendung der SVR herangezogen (s. Abschn. 5.4).

5.4 Training und Anwendung der SVR

Das Training und die Evaluation der SVR erfolgt in der Programmiersprache Python unter Verwendung der scikit-learn Bibliothek (Pedregosa et al., 2011). Als Kernel wird, wie in Abschnitt 3.5 empfohlen, ein RBF-Kernel verwendet.

5.4.1 Training der SVR

Für das Training der SVR ist, über das Abspielen der Videosequenzen im ADTF, für jeden Fußgänger eine csv-Datei zu erstellen, die den Merkmalsvektor $\boldsymbol{x}$ und das Label y für jeden Zeitschritt beinhaltet. Um bei der, im Rahmen der Evaluation durchgeführten, Aufteilung der Daten in Trainings- und Testdaten bestehende Abhängigkeiten zwischen den Samples beachten zu können, wird dabei die Zuordnung der Merkmalsvektoren zu jedem Fußgänger sowie die Zuordnung jedes Fußgängers zur Videosequenz bewahrt. Die csv-Dateien bilden die Grundlage des in Python realisierten Trainingsprozesses, dessen algorithmischer Ablauf in Abbildung 5.9 gezeigt wird.

Da die SVR gegenüber der Skalierung der Eingangsdaten invariant ist, werden die Trainingsdaten zunächst so skaliert, dass der Mittelwert ihrer Amplituden $\mu = 0$ und ihre Varianz $\sigma = 1$ entspricht. Die dazu verwendeten Skalierungsparameter μ_{scale} und σ_{scale} werden dabei für die Anwendungsphase gespeichert.

Zudem werden die Labels y vor dem Training mit der Logit Funktion $L(y_i)$ aus Gleichung 3.3 transformiert. Gemäß der Empfehlung in Abschnitt 3.1.1 wird so verhindert, dass die Ausgabewerte der trainierten SVR außerhalb des probabilistischen Wertebereichs von $[0, 1]$ fallen.

Wie in Abschnitt 6.1 gezeigt wird, besteht bei dem in dieser Arbeit verwendeten Datensatz sowohl eine *between-class imbalance* als auch eine *within-class imbalance* (s. Abschn. 3.2.3). Daher gilt es, eine Überanpassung der SVR auf die häufig auftretenden Fälle zu vermeiden (Fußgänger mit eindeutig keiner Querungsintention, die auf dem Gehweg parallel zur Fahrbahn laufen). Hierzu wird jedem Sample während des Trainingsprozesses ein Gewicht zugewiesen. Das Gewicht ist antiproportional zur Häufigkeit der durch das Sample repräsentierten Situation. Zur Clusterung der theoretisch unendlich vielen Situationen wird den Samples abhängig von ihrem mit fünf Stufen

Eingaben

1 $D_{Train} = [(x_1, y_1), \ldots, (x_N, y_N)]$ // *Satz mit Trainingsbeispielen und Labels,* $y_n \in [0, 1]$

2 $S_{Id} = [s_{Id_1}, \ldots, s_{Id_N}]$ // *Situationskennungen jedes Datensamples*

3 C, γ // *SVR Parameter*

Algorithmus

4 `Bestimme` μ_{scale} `und` σ_{scale} // *Mittelwert und Varianz der Trainingsdaten*

5 `for` i `:= 1 to` N `do` // *Für jedes Trainingsbeispiel*

6 x_s `:=` $(x_i - \mu_{scale}) \cdot \sigma_{scale}$ // *Skaliere Trainingsbeispiel*

7 y_s `:= ln(`y_i `/ (1 -` y_i`))` // *Transformiere Label*

8 w_s `:=` N`/sum(`S_{Id} `==` S_{Id}`[`i`])` // *Bestimmte Gewicht des Trainingsbeispiels*

9 `trainSVR`$_{C,\gamma}$`(mit allen` x_s, y_s, w_s`)` // *Training*

Rückgabe

10 $\mu_{scale}, \sigma_{scale}$ // *Skalierungsparameter*

11 $\alpha = [\alpha_1, \ldots, \alpha_i]$ // *Support Vektoren der trainierten SVR*

Abbildung 5.9: Algorithmischer Ablauf zum Training der SVR.

Tabelle 5.2: Zusammensetzung der Situationskennung S_{Id}.

Ziffer	Wertebereich	Beschreibung
1.	$[0, 4]$	Querungsintention (quantisiert mit 5 Stufen)
2.	$[0, 2]$	Präsenz von Szenenelementen
3.	$[0, 3]$	Zone des Fußgängers zum aktuellen Zeitpunkt t
4.	$[0, 3]$	Zone des Fußgängers zum Zeitpunkt $t - 1$
5.	$[0, 7]$	Körperorientierung relativ zur Straße θ_{Str} (quantisiert mit 8 Stufen)

diskretisierten Label y, der Präsenz von Fußgängerüberwegen oder Wartebereichen, der aktuellen und der vorangegangenen Zone des Fußgängers Z_{ped} sowie der der in acht Stufen diskretisierten Körperorientierung des Fußgängers relativ zur Straße θ_{Str} mit:

$$\theta_{Str} = \begin{cases} \theta_B + 180° & \text{wenn } S_{ped} = rechts \wedge \theta \neq 0° \wedge \theta \neq 180° \quad (5.10a) \\ \theta_B & \text{sonst} \quad (5.10b) \end{cases}$$

eine eindeutige Situationskennung S_{Id} zugewiesen (s. Tabelle 5.2). Anschließend wird der proportionale Anteil aller Samples mit derselben S_{Id} in Bezug auf den gesamten Datensatz bestimmt und der Kehrwert als Gewichtungsfaktor w_s für alle Samples mit dieser Situationskennung verwendet.

Das Training der SVR verläuft schließlich gemäß der Beschreibung in Abschnitt 3.2.2. Nach Abschluss des Trainings werden die bestimmten Support Vektoren $\boldsymbol{\alpha}$ für die Anwendungsphase gespeichert.

5.4.2 Anwendung der SVR

Abbildung 5.10 zeigt den algorithmischen Ablauf zur Anwendung des trainierten Modells. In der Anwendungsphase wird der zu bewertende Merkmalsvektor $\boldsymbol{x_i}$ zunächst über die aus der Trainingsphase bekannten Parameter μ_{scale} und σ_{scale} skaliert. Anschließend ist unter Verwendung der gespeicherten Support Vektoren $\boldsymbol{\alpha}$ und Gleichung 3.16 der Ausgangswert $\hat{y}_s$ zu bestimmen. Dieser wird schließlich über die in Gleichung 3.5 angegebene Inverse der Logit Funktion $L^{-1}(\hat{y}_s)$ in den probabilistischen Eingangsraum rücktransformiert, um mit $\hat{y}$ die prädizierte Wahrscheinlichkeit über die Ausprägung einer Querungsintention des Fußgängers zu erhalten.

Eingaben

1 $D_{Test} = [(x_1, y_1), \ldots, (x_N, y_N)]$ // *Satz mit Testbeispielen und Labels,* $y_n \in [0, 1]$

2 $\mu_{scale}, \sigma_{scale}$ // *Skalierungsparameter*

3 $\alpha = [\alpha_1, \ldots, \alpha_i]$ // *Support Vektoren der trainierten SVR*

4 C, γ // *SVR Parameter*

Algorithmus

5 `for` i `:= 1 to` N `do` // *Für jedes Testbeispiel*

6 x_s `:=` $(x_i - \mu_{scale}) \cdot \sigma_{scale}$ // *Skaliere Testbeispiel*

7 $\hat{y}_s$ `:= applySVR`$_{C,\gamma,\alpha}(x_s)$ // *Prädiktion*

8 $\hat{y}_i$ `:=` $e^{\hat{y}_s} / (\hat{y}_s + 1)$ // *Transformiere Prädiktion*

Rückgabe

9 $\hat{y} = [\hat{y}_1, \ldots, \hat{y}_N]$ // *Prädizierte Zielwerte*

Abbildung 5.10: Algorithmischer Ablauf zur Anwendung der trainierten SVR.

Kapitel 6

Evaluation

Zur Bestimmung der für die Erkennung der Querungsintention eines Fußgängers am besten geeigneten Kombination aus Merkmalsvektor und SVR wird eine ausführliche Evaluation durchgeführt. In den folgenden Abschnitten werden die hierzu verwendete Datenbasis vorgestellt, die zur Evaluation angewendete Methodik beschrieben sowie die erreichten Ergebnisse vorgestellt und diskutiert.

6.1 Datenbasis

Zum Training der SVR und zur Evaluation des Verfahrens wurde ein Datensatz erstellt, das aus 175 Videosequenzen mit einer Gesamtlänge von 40,1 Minuten besteht. Die Videosequenzen wurden mit einer serienmäßigen 1 MP-Fahrzeugkamera (1280×971 px, 18 fps, FOV = 42°) auf Fahrten im deutschen Straßenverkehr aufgenommen. 102 Sequenzen wurden dabei auf zufällig ausgewählten Strecken im Innenstadtbereich verschiedener deutscher Städte aufgenommen. Diese Sequenzen bilden somit ein repräsentatives Bild der typischerweise im deutschen Straßenverkehr auftretenden Situationen ab, die das in dieser Arbeit zu entwickelnde System beherrschen muss. Da diese Daten verhältnismäßig wenige Fußgänger beinhalten, die an Stellen ohne Querungshilfen queren, wurden die restlichen 73 Sequenzen speziell für diese Arbeit in Bereichen der Stadt Ingolstadt aufgenommen, in denen typischerweise viele Fußgänger ohne bauliche Hilfsmaßnahmen die Straße queren. Diese Sequenzen beinhalten damit viele Fußgänger, die

Abbildung 6.1: Beispielhafte Frames des in dieser Arbeit verwendeten Datensatzes.

ihre Querungsintention hauptsächlich durch ihr Absicherungsverhalten anzeigen. Auf Grund des geringen Öffnungswinkels der Kamera, bildet der Datensatz nur Interaktionen mit Fußgängern bei geradeaus Fahrten ohne Abbiegemanöver ab. Abbildung 6.1 zeigt eine Auswahl der verwendeten Videodaten.

6.1.1 Fußgängererkennung

Das zur Aufzeichnung verwendete Kamerasystem verfügt über ein integriertes Fußgängerdetektionssystem. Dieses gibt die Fußgängerposition sowohl in 2D-Bildkoordinaten (x_i, y_i, w_i, h_i), als auch in 3D-Fahrzeugkoordinaten (x_v, y_v, z_v) an, von denen jedoch nur die x-y-Position innerhalb der Grundebene betrachtet wird. Um die für diese Arbeit relevante Körper- und Kopfpose des Fußgängers in den Videodaten

Tabelle 6.1: Genauigkeit der Fußgängerposition in der x-y-Grundebene.

	MW	Min	Max
σ_{x_v}	0,50 m	0,06 m	6,13 m
σ_{y_v}	0,14 m	0,00 m	2,56 m

erkennen zu können, werden nur Fußgängerdetektionen verwendet, deren Bounding Box eine Fläche von 3.000 Pixeln initial überschreitet und nach der Erstdetektion eine Grenze von 1.300 Pixel nicht unterschreitet. Wie die Trajektorien der Fußgänger in Abbildung 6.2 zeigen, wird damit ein Bereich von bis zu 35 m vor dem Fahrzeug betrachtet.

Insgesamt beinhalten der erstelle Datensatz 639 Fußgänger, die von dem integrierten Detektionssystem erkannt werden und die beschriebenen Filterbedingungen erfüllen. Da jeder Fußgänger nach der initialen Detektion über mehrere Frames getrackt wird ($\mu = 42$, $\sigma = 37$, $min = 1$, $max = 257$), ergeben sich insgesamt $N = 27.062$ Samples, die zum Training der SVR und zur Evaluation zur Verfügung stehen. Das entspricht theoretisch einer Beobachtungszeit von 25,1 Minuten. Da jedoch häufig mehrere Fußgänger gleichzeitig auftreten, beinhaltet das 40,1 Minuten lange Videomaterial insgesamt 15,2 Minuten, in denen mindestens ein Fußgänger detektiert wird.

Einige der in Abbildung 6.2 gezeigten Trajektorien weisen für Fußgänger untypische Sprünge auf. Dies zeigt, dass vor allem die vom Detektionssystem bestimmte 3D-Position des Fußgängers fehlerbehaftet ist. Messprinzip-bedingt wird hierbei die Position des Fußgängers in lateraler Richtung deutlich ungenauer erkannt, als in longitudinaler Richtung (s. Tab. 6.1). Die durchschnittliche, minimale und maximale Größe der Bounding Boxes der detektierten Fußgänger kann schließlich Tabelle 6.2 entnommen werden. Mit einem durchschnittlichen Seitenverhältnis von 1 : 2,7 sind die vom Detektionssystem ausgegebenen Bounding Boxes im Vergleich zu den aus der Literatur bekannten Größen (Dollár et al., 2012) verhältnismäßig schmal.

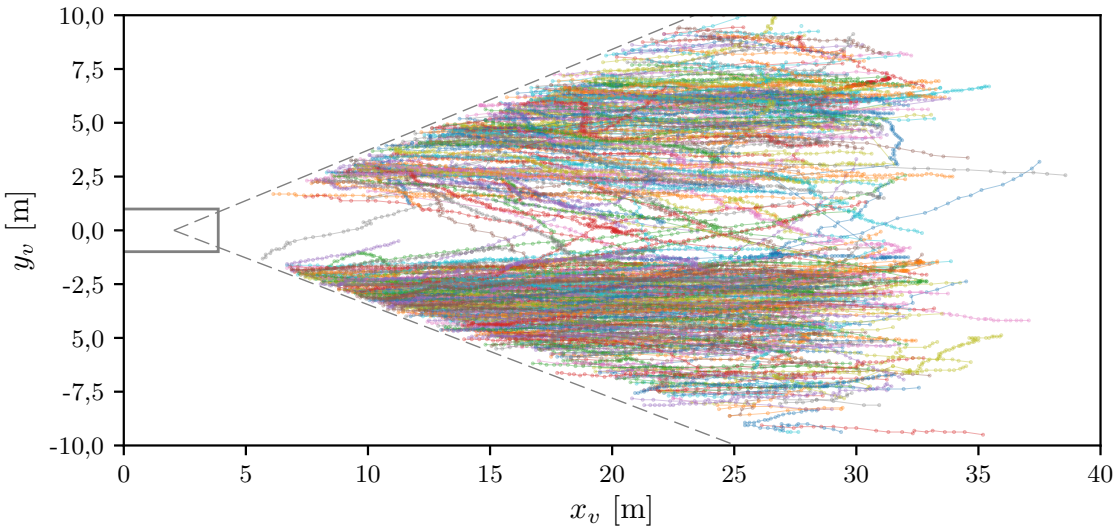


Abbildung 6.2: Trajektorien aller im Datensatz vorhandenen Fußgänger (ohne Kompensation der Bewegung des Ego-Fahrzeugs).

Tabelle 6.2: Bounding Box Größe der detektierten Fußgänger.

	MW	Min	Max
w_i	52 px	22 px	184 px
h_i	139 px	60 px	491 px

Tabelle 6.3: Bounding Box Größe der gelabelten Fußgängerköpfe.

	MW	Min	Max
w_h	25 px	7 px	94 px
h_h	25 px	9 px	96 px

Körper- und Kopforientierung sowie Kopfposition

Neben der Position des Fußgängers wird für das in dieser Arbeit entwickelte kontextbasierte Verfahren die Körperorientierung des Fußgängers relativ zum Fahrzeugkoordinatensystem vorausgesetzt. Gleiches gilt für die Kopforientierung sowie für die Kopfposition in Bildkoordinaten, die von einigen der betrachten posenbasierten Verfahren verwendet werden. Da das in dieser Arbeit zur Verfügung stehende Detektionssystem diese Informationen nicht bereitstellt, werden diese unter Verwendung teilweise selbst entwickelter Labeltools für jeden Fußgänger manuell bestimmt (s. Anhang C.2). Mögliche Umsetzungsformen für automatische Erkennungssysteme sind in (Rosbach, 2016; Flohr et al., 2014; Rehder et al., 2014; Schulz et al., 2011) zu finden.

Abbildung 6.3 zeigt die Verteilung der Fußgänger nach ihrer in acht Orientierungsklassen unterteilten Körper- und Kopforientierungen. Mit insgesamt 58 % bewegen sich die meisten Fußgänger parallel zur Fahrbahn; 35 % befinden sich dabei mit dem Rücken zum Fahrzeug ($\theta_B = 180°$) und 23 % frontal zum Fahrzeug ($\theta_B = 0°$). Nur 14 % aller Fußgänger ändern ihre Körperorientierung während des Beobachtungszeitraums über die Grenze einer Orientierungsklasse hinaus. Ähnliches gilt auch für die Verteilung der Kopforientierung. Im Unterschied zur Körperorientierung ändern die Fußgänger ihre Kopforientierung mit 27 % jedoch deutlich häufiger über die Grenzen der Orientierungsklassen hinaus. Der Datensatz weist bezüglich der Körper- und Kopforientierung der Fußgänger somit eine marginale bis mäßige *within-class-imbalance* auf (s. Abschn. 3.2.3).

Die durchschnittliche, minimale und maximale Größe der Kopf Bounding Box (x_h, y_h, w_h, h_h) kann schließlich Tabelle 6.3 entnommen werden.

Querungsintention

Wie in Abschnitt 4.3 diskutiert, wird in dieser Arbeit als Ground Truth für die Querungsintention eines Fußgängers der Mittelwert der Urteile von sieben unabhängigen Beobachtern verwendet. Abbildung 6.5 zeigt die samplebasierte Verteilung der Labels bei verschiedenen Quantisierungsstufen. Die hierzu verwendeten Intervallgrenzen werden in Abbildung 6.4 abgebildet. Dabei entspricht die fünfstufige Einteilung den

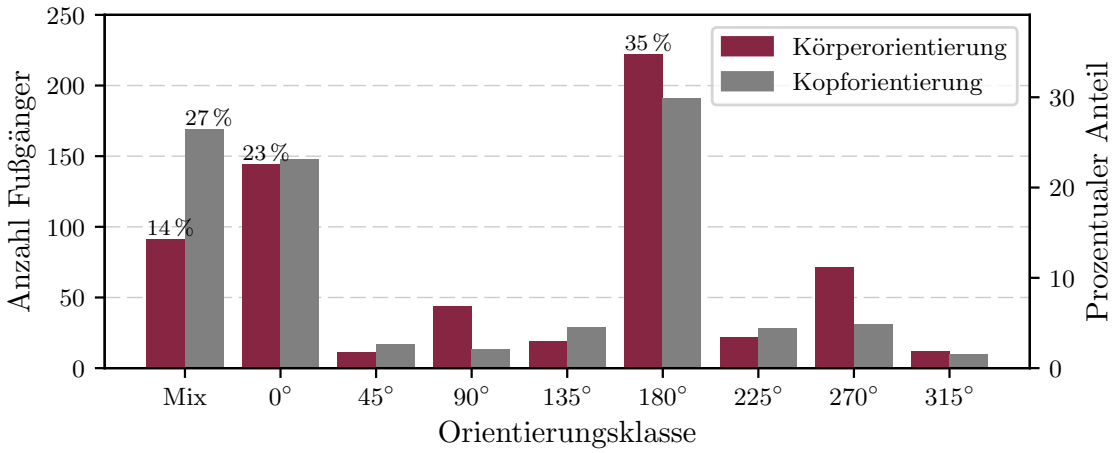


Abbildung 6.3: Verteilung der Fußgänger nach ihrer in acht Orientierungsklassen unterteilten Körper- und Kopforientierungen. *Mix* bezeichnet Fußgänger, deren Orientierung sich während des Beobachtungszeitraums über die Grenze einer Orientierungsklasse hinaus verändert.

Stufen der im Rahmen der Referenzbildung eingesetzten Ratingskala (s. Abschn. 4.1). Abbildung 6.6 zeigt Beispiele der verschiedenen Labelklassen unter Verwendung der fünfstufigen Quantisierung.

Es zeigt sich, dass der Datensatz eine marginale bis mäßige *between-class-imbalance* aufweist. Samples ohne Querungsintention (C_{Neg}) sind in dem Datensatz etwa 4,5 mal so oft vorhanden, wie Samples mit einer Querungsintention (C_{Pos}) (s. Abb. 6.5a). Auch unter Betrachtung der, bei der Referenzbildung zur Beurteilung der Querungsintention verwendeten, fünfstufigen Unsicherheitslevel (s. Abschn. 4.1) überwiegen die Samples mit eindeutig keiner Querungsintention (C_{--}) mit 77 % (s. Abb. 6.5b). Insgesamt besteht nur bei 10 % der Samples eine Unsicherheit über die Ausprägung einer Querungsintention (C_{-}, $C_{?}$ und C_{+}).

Durch die Bildung der Mittelwerte der sieben Beobachter kann die als kontinuierlich zu betrachtende Unsicherheit über die Ausprägung einer Querungsintention theoretisch $7^5 = 16.807$ Werte annehmen. Auf Grund der guten Übereinstimmung der Beobachterurteile beinhaltet der vorliegende Datensatz jedoch nur 25 verschiedene Ausprägungsstufen. Abbildung 6.5c zeigt die Verteilung der Samples über diese Stu-

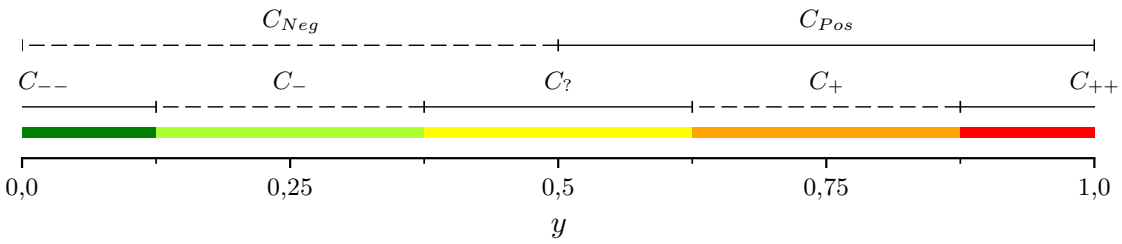


Abbildung 6.4: Zur Quantisierung der Labels verwendete Intervallgrenzen.

Tabelle 6.4: Zuordnung der Labelklassen zu den Objektklassen.

	P_{Neg}	P_{Neg_Unc}	P_{Unc}	P_{Pos_Unc}	P_{Pos}	P_{Mix}
$y_i \in \{\cdots\}$	C_{--}	$C_{--}, C_{-}, C_{?}$	$C_{?}$	$C_{++}, C_{+}, C_{?}$	C_{++}	$C_{--}, C_{-}, C_{?}, C_{+}, C_{++}$

fen. Auch hier zeigt sich eine deutliche Verschiebung der im Datensatz vorhandenen Ausprägungsstufen in Richtung Fußgänger ohne Querungsintention.

Abbildung 6.7 zeigt schließlich die objektbasierte Verteilung der Labels. Diese basiert auf der oben beschriebenen, fünfstufigen Quantisierung und der in Tabelle 6.4 angegebenen Zuordnung der Labelklassen zu den sechs Objektklassen. P_{Pos} und P_{Neg} beschreiben somit Fußgänger, deren Samples alle eindeutig positiv oder eindeutig negativ sind. Die 527 Fußgänger mit solch einer eindeutigen Ausprägung der Querungsintention bilden mit 82 % die überwiegend auftretende Objektklasse ab. Fußgänger, deren Samples eine einseitige Unsicherheit aufweisen, werden mit P_{Neg_Unc} und P_{Pos_Unc} bezeichnet. Entsprechend beschreibt P_{Unc} Fußgänger, die ausschließlich unsichere Labels, ohne Tendenz zu einer positiven oder negativen Ausprägung einer Querungsintention haben. Insgesamt weisen 56 Fußgänger (9 %) solch einseitige Unsicherheiten auf und nur bei drei Fußgängern liegt überhaupt keine Tendenz bezüglich der Querungsintention vor. P_{Mix} umfasst schließlich alle Fußgänger, deren Querungsintention sich über den Beobachtungszeitraum ändert und deren Samples somit sowohl eine negative als auch eine positive Ausprägung haben.

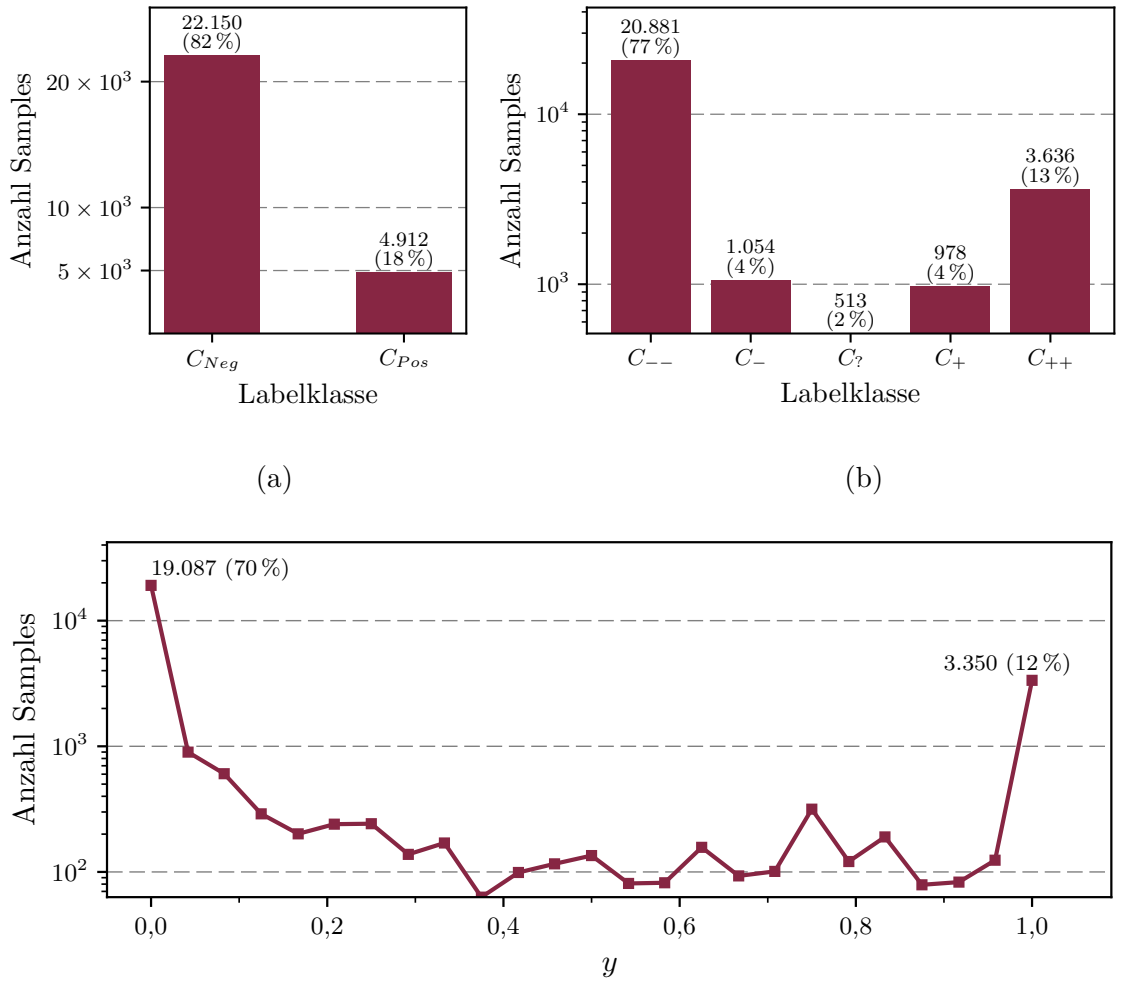


Abbildung 6.5: Samplebasierte Verteilung der Labels mit unterschiedlichen Quantisierungsstufen: (a) Betrachtung als binäres Klassifikationsproblem. (b) Fünfstufige Quantisierung. Entspricht der zur Referenzbildung eingesetzten Ratingskala. (c) Verteilung über den gesamten Wertebereich.

Abbildung 6.6: Beispiele der verschiedenen Labelklassen unter Verwendung der fünfstufigen Quantisierung.

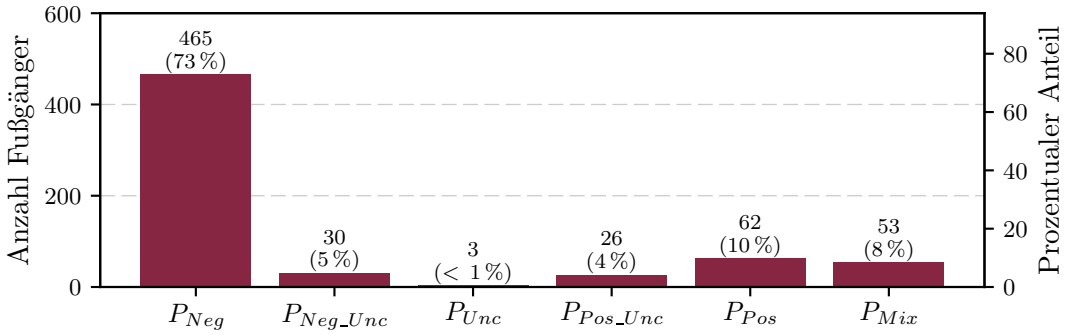
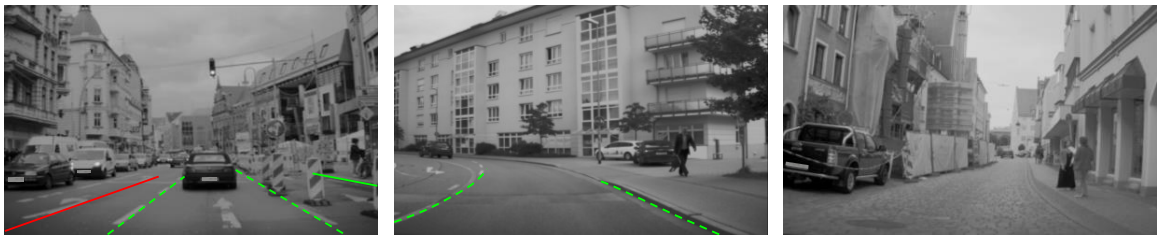


Abbildung 6.7: Objektbasierte Verteilung der Labels.



(a) (b) (c)

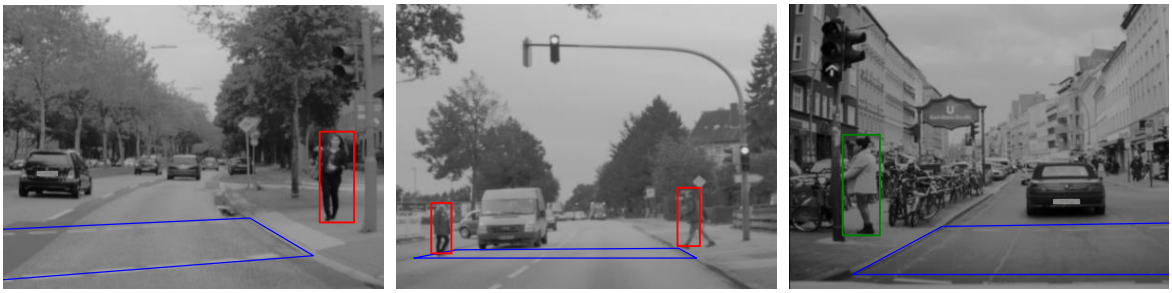
Abbildung 6.8: Situationen, in denen das verwendete Fahrstreifenerkennungssystem die Fahrbahnbegrenzung nicht korrekt erkennt: (a) Fehlerhafte Erkennung des Nebenfahrstreifens (rot). (b) Keine Erkennung des abgesenkten Bordsteins. (c) Keine Erkennung der unterschiedlichen Pflasterung und der Bauzäune.

6.1.2 Fahrstreifenerkennung

Neben der Detektion von Fußgängern ist das verwendete Kamerasystem in der Lage, Fahrstreifenmarkierungen und Fahrbahnbegrenzungen zu erkennen. Diese liegen in Form von Klothoiden (Fatemi et al., 2014) vor und besitzen Attribute, die sowohl die Position (z.B. *EgoRechts*, *NebenLinks*), als auch den Begrenzungstyp (z.B. *Linie gestrichelt*, *Fahrbahnrand*, *Bordstein*) beschreiben. Da das Erkennungssystem noch nicht auf die Anwendung in innerstädtischen Bereichen ausgelegt ist, werden Fahrbahnbegrenzungen, die über sehr niedrige Bordsteine oder durch eine unterschiedliche Pflasterung markiert sind, nicht erkannt (s. Abb. 6.8). In diesen Fällen werden die benötigten Fahrstreifenbegrenzungen unter Verwendung eines selbst entwickelten Labeltools, welches auf Basis der Videodaten arbeitet, manuell nachgetragen (s. Anhang C.2)

6.1.3 Szenenelemente

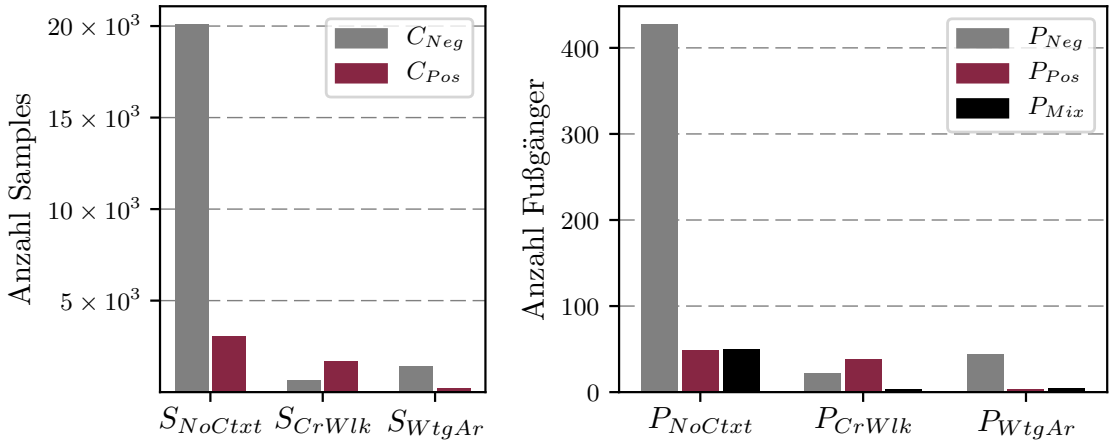
Wie in den Abschnitten 5.2.3 f. beschrieben, sind zur Erstellung des kontextbasierten Merkmalsvektors $\boldsymbol{x}_{Ctxt}$ die Positionen von Szenenelementen wie Fußgängerüberwege und Wartebereiche in Fahrzeugkoordinaten notwendig. Solche Szenenelemente werden von dem verwendeten Detektionssystem jedoch nicht erkannt. Zudem besteht keine Möglichkeit, die benötigten Positionen anhand digitaler Karten (z.B. (HERE, 2017)) zu bestimmen, da in einem Teil des Datensatzes die globale Position des Fahrzeugs nicht



(a)

(b)

Abbildung 6.9: Situationen mit nachgetragenen Szenenelementen: (a) Fußgängerüberwege. (b) Bushaltestellen.



(a) (b)

Abbildung 6.10: Verteilung der Labels abhängig von der Präsenz der Szenenelemente: (a) Samplebasiert. (b) Objektbasiert.

bekannt ist. Daher werden die in dem Datensatz vorhandenen Fußgängerüberwege und Wartebereiche, analog zu den fehlenden Fahrbahnbegrenzungen, manuell nachgetragen (s. Anhang C.2). Abbildung 6.9 zeigt beispielhaft sechs Situationen mit entsprechend markierten Szenenelementen.

Für eine Anwendung des Systems im Realfahrzeug ist neben der Verwendung digitaler Karten auch eine Bestimmung der Szenenelemente über bildbasierte Detektionsverfahren, wie in (Haselhoff und Kummert, 2010; Fritsch et al., 2014) beschrieben, möglich.

Abbildung 6.10 zeigt die sample- sowie die objektbasierte Verteilung der Labels in Abhängigkeit von der Präsenz der betrachteten Szenenelemente. Der Index $CrWlk$ beschreibt dabei Fußgänger, die sich im Bereich eines Fußgängerüberwegs aufhalten. Der Index $WtgAr$ umfasst alle Fußgänger, die sich im Einflussbereich einer Bushaltestelle befinden, und der Index $NoCtxt$ bezieht sich schließlich auf alle Fußgänger in Situationen ohne zusätzlicher Szenenelemente. Aus der Abbildung geht hervor, dass auch hier eine *within-class-imbalance* besteht, da mit 20.078 Samples ($\hat{=}$ 74 %) deutlich mehr Situationen betrachtet werden, die keine Präsenz zusätzlicher Szenenelemente aufweisen. Zudem weisen Situationen mit Szenenelementpräsenz eine *between-class-imbalance* auf, denn Fußgänger, die sich im Bereich von Fußgängerüberwegen aufhalten, haben deutlich häufiger eine Querungsintention (1.654 Samples $\hat{=}$ 72 %) als Fußgänger, die an Bushaltestellen und anderen Wartebereichen stehen. Diese weisen fast ausschließlich keine Querungsintention auf (1.431 Samples $\hat{=}$ 88 %).

6.1.4 Situationskennung

Der Datensatz beinhaltet 211 der 1.920 theoretisch über die Situationskennung differenzierbaren Situationen (vgl. Tabelle 5.2 in Abschnitt 5.4.1). Abbildung 6.11 zeigt die Anzahl der Samples der zehn am häufigsten vertretenen Situationen. Mit 8.135 Samples ist die am häufigsten auftretende Situation die $S_{Id} = 00000$. Somit repräsentieren 30 % aller Samples Fußgänger, die sich mit dem Rücken zum Fahrzeug, auf einem Gehweg ohne zusätzlicher Szenenelemente bewegen und keine Querungsintention haben. Fußgänger mit einer Querungsintention sind hingegen am häufigsten frontal zur Straßenkante ausgerichtet und stehen oder bewegen sich innerhalb der Gehweg-Zone im

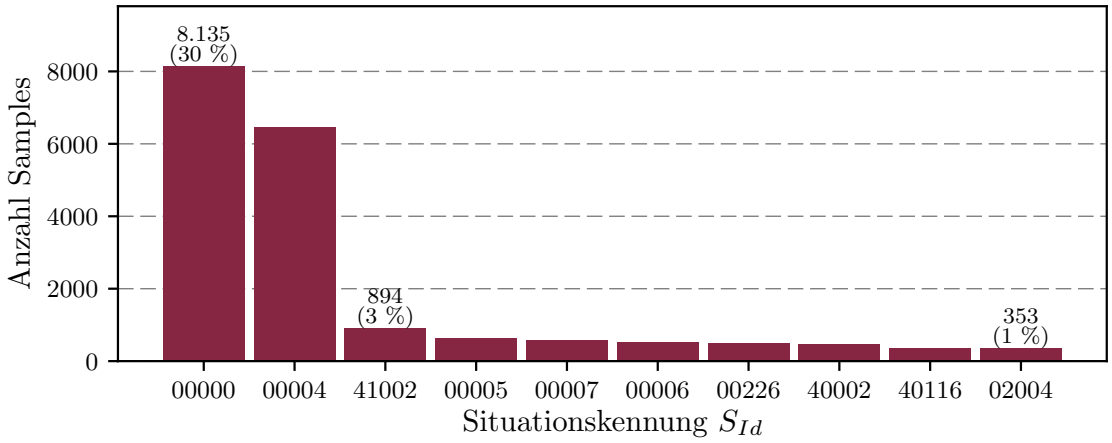


Abbildung 6.11: Anzahl der Samples der zehn häufigsten Situationen.

Umfeld eines Fußgängerüberwegs ($S_{Id} = 41002$). Die wenigsten Samples weisen Situationen auf, bei denen der Fußgänger die Zone zwischen zwei Zeitschritten gewechselt hat. 42 dieser Situationen sind durch jeweils nur ein Sample im Datensatz vertreten.

6.2 Evaluationsmethodik

Im Rahmen der Evaluation werden zunächst die Parameter des in Abschnitt 5.2 vorgestellten, kontextbasierten Ansatzes zur Erkennung der Querungsintention von Fußgängern optimiert und die mit dem neuen Ansatz erreichten Ergebnisse diskutiert. Anschließend wird die Kombination des optimierten, kontextbasierten Merkmalsvektors mit den in Abschnitt 5.3 vorgestellten, posenbasierten Merkmalen evaluiert. Das Ziel ist hierbei ein Merkmal zu finden, das die Körper- und Kopfpose hinreichend beschreibt und so den kontextbasierten Ansatz optimal ergänzt. Für alle durchgeführten Evaluationen wird die im Folgenden beschriebene Methodik angewendet.

6.2.1 Kreuzvalidierung

Um sicherzustellen, dass bei der trainierten SVR keine Überanpassung an die Trainingsdaten besteht, erfolgt die Evaluation über eine k-fold Kreuzvalidierung mit $k = 10$. Wie in Abschnitt 3.5 diskutiert, fordert die bestehende zeitliche und situative Korrelation zwischen den einzelnen Samples sowie die bestehende Unausgewogenheit der verwendeten Daten hierbei eine Kombination aus der Stratified- und der Group k-fold-Methode. Nur so kann sichergestellt werden, dass auch unter Berücksichtigung der bestehenden Abhängigkeiten, jeder Teildatensatz eine repräsentative Teilmenge der gesamten Beispieldaten darstellt.

Zur Umsetzung der beschriebenen Anforderungen wurde der in Abbildung 6.12 gezeigte algorithmische Ablauf entwickelt. Zur Berücksichtigung der bestehenden zeitlichen und situativen Korrelationen zwischen den Samples werden hier zunächst alle Samples einer Videosequenz zu einer Gruppe zusammengefasst. Anschließend werden die Gruppen iterativ einem der k Teildatensätze zugeordnet. Dadurch wird sichergestellt, dass sich bei der Evaluation keine Samples desselben Fußgängers oder derselben Situation sowohl im Test- als auch im Trainingsdatensatz befinden.

Auf Grund der bestehenden *within-* und *in-between-class-imbalances* sollten bei der Zuordnung zudem die in den Daten abgebildeten Situationen möglichst gleichmäßig über alle Teildatensätze verteilt werden. Hierzu wird auf die in Abschnitt 5.4.1 beschriebene Situationskennung S_{Id} zurückgegriffen. Aus dieser wird für jedes Sample ein

Eingaben

1 $S_{Id} = [s_{Id_1}, \ldots, s_{Id_N}]$ // *Situationskennung jedes Datensamples*
2 $V_{Id} = [v_{Id_1}, \ldots, v_{Id_N}]$ // *Videosequenzkennung jedes Datensamples*
3 k // *Anzahl der Teildatensätze*

Algorithmus

Schritt 1: **// Bestimmung Gruppengewicht**

4 G_{Id} := unique(V_{Id}) // *Bestimme alle eindeutigen Gruppenkennungen*
5 W_g := zeros(1, N_g) // *Initialisiere Gruppengewichte*
6 for g := 1 to length(G_{Id}) do // *Für jede Gruppe*
7 $S_{Id,g}$:= S_{Id}[V_{Id} == G_{Id}[g]] // *Extrahiere Situationskennungen der Gruppe*
8 for i := 1 to length($S_{Id,g}$) do // *Für jedes Gruppensample*
9 w_s := N/sum(S_{Id} == S_{Id}[i]) // *Bestimme Samplegewicht*
10 W_g[g] += w_s // *Addiere Samplegewicht zu Gruppengewicht*
11 sort(G_{Id} by W_g in descending order) // *Sortiere nach absteigendem Gewicht*

Schritt 2: **// Zuordnung der Gruppen zu den k folds**

12 F := zeros(1, k) // *Initialisiere Rückgabe*
13 C_s := hist(S_{Id}) // *Bestimme existierende Anzahl an Samples pro Situation*
14 H := zeros(k, length(C_s) // *Initialisiere Verteilung der zugeordneter Samples*
15 for g := 1 to length(G_{Id}) do // *Für jede Gruppe*
16 $H_T = H$ // *Kopie der Sampleverteilung für virtuelle Zuordnung*
17 E := zeros(1, k) // *Initialisiere Fehlerarray*
18 for l := 1 to k do // *Für jeden fold*
19 $S_{Id,g}$:= S_{Id}[V_{Id} == G_{Id}[g]] // *Extrahiere Situationskennungen der Gruppe*
20 H_T[l,:] += hist($S_{Id,g}$) // *Ordne alle Gruppensamples virtuell zu fold l*
21 for j := 1 to length(C_s) do // *Für jede Situation*
22 w_e := 1/C_s[i] // *Gewichte Fehler antiproportional zur Situationshäufigkeit*
23 E[l] := $w_e \cdot$ abs(H_T[l,j]$-C_s$[j]/k) // *Bestimme gewichtete Fehlerdifferenz*
24 k_{idx} := index(min(E)) // *Bestimme fold mit kleinstem Fehler*
25 H[k_{idx},:] += hist($S_{Id,g}$) // *Ordne alle Gruppensamples dem besten fold zu*
26 F[V_{Id} == G_{Id}[g]] := k_{idx} // *Speicher Zuordnung*

Rückgabe

27 F // *Dictionary, mit der Zuordnung jedes Samples zu einem der k folds*

Abbildung 6.12: Algorithmischer Ablauf der entwickelten Kreuzvalidierung.

Gewichtungsfaktor w_s abgeleitet, der antiproportional zur Auftrittswahrscheinlichkeit der Situation ist. Für jede Gruppe wird anschließend, über die Aufsummierung der einzelnen Samplegewichte, eine Gruppengewichtung w_g bestimmt. Die Gruppengewichtung dient schließlich der Bestimmung der Reihenfolge bei der iterativen Zuweisung der Gruppen zu einem der k Teildatensätze. Durch eine Sortierung der Gruppen nach absteigendem Gruppengewicht wird sichergestellt, dass zunächst die Gruppen mit vielen oder mit seltenen Samples gleichmäßig auf die Teildatensätze verteilt werden und erst am Ende einzelne, häufig auftretende Samples zum Auffüllen der Teildatensätze verwendet werden.

Die angestrebte Gleichverteilung der Situationen über die k Teildatensätze wird schließlich über ein zweistufiges Verfahren realisiert, welches auf der angestrebten Anzahl an Samples pro Situation und Teildatensatz basiert mit $N_{S_{Id}} = hist(S_{Id})/k$. Hierbei werden die Samples der aktuellen Gruppe zunächst virtuell jedem der k Teildatensätze zugewiesen, um anschließend diejenige Zuordnung zu bestimmen, die den besten Beitrag zu der Gleichverteilung der Situationen über alle Teildatensätze leistet. Das dazu verwendete Fehlermaß leitet sich aus der Differenz zwischen der Sampleanzahl pro Situation und Teildatensatz zur Zielanzahl ab. Die Differenzen werden schließlich mit einem Gewichtungsfaktor w_e, der antiproportional zur Zielanzahl ist, multipliziert und zu dem Fehlermaß e_k aufsummiert. Die Gewichtung verhindert dabei, dass ausschließlich Situationen, die in dem Datensatz häufig vorkommen, einen Einfluss auf die Zuordnung haben, wodurch nur diese über die Teildatensätze gleichverteilt wären.

Im zweiten Schritt wird schließlich der k-te Teildatensatz mit dem geringsten Fehler e_k bestimmt, um alle Samples der aktuellen Gruppe diesem Teildatensatz zuzuordnen. Weisen mehrere Teildatensätze den gleichen Fehlerwert auf, wird einer der Kandidaten zufällig ausgewählt.

Die beschriebene Zuordnung von Samples zu Teildatensätzen wird nur einmalig vorgenommen und in Form einer Lookup-Tabelle gespeichert. So ist gewährleistet, dass alle evaluierten Merkmalsvektor-SVR-Kombinationen auf denselben Samplekombinationen trainiert und getestet werden und die Ergebnisse somit vergleichbar sind. Zur besseren Übersicht skizziert Abbildung 6.13 schließlich die Struktur des zur Evaluation eingesetzten Datensatzes.

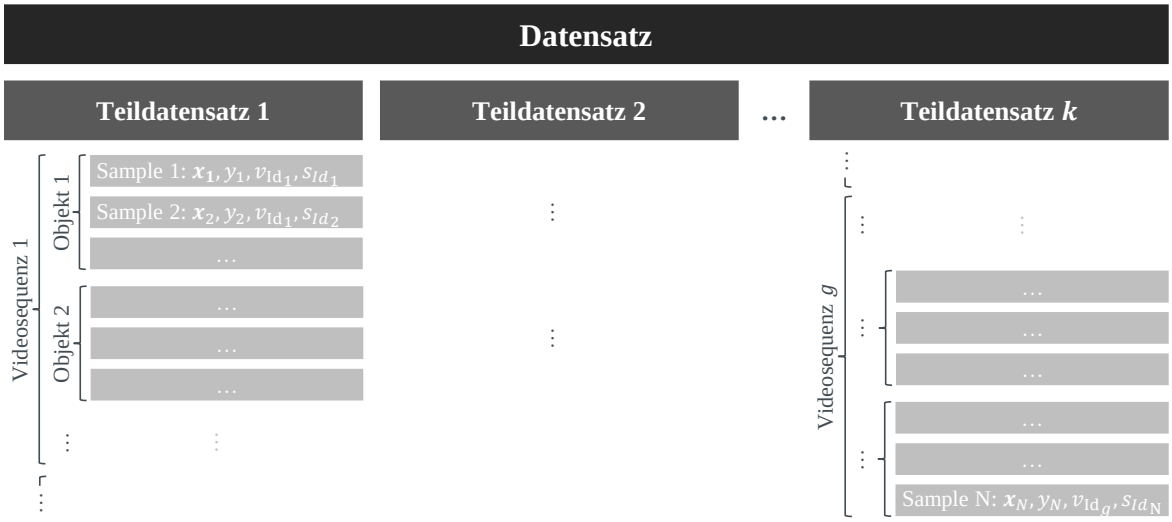


Abbildung 6.13: Struktur des zur Evaluation verwendeten Datensatzes.

6.2.2 Samplebasierte Evaluation

Um die optimale Parametrisierung der zu evaluierenden Merkmalsvektoren in Kombination mit der ebenfalls über Parameter zu optimierenden SVR zu bestimmen, wird zunächst eine samplebasierte Evaluation durchgeführt. Bei dieser wird die grundsätzliche Leistung der betrachteten Merkmalsvektor-SVR-Kombinationen ausgewertet, indem die Güte der Prädiktion objektunabhängig, d.h. unabhängig von der Zuordnung der Samples zu einem Fußgänger, bestimmt wird. Gemäß der Diskussion in Abschnitt 3.5 wird als zu optimierendes Gütemaß der gewichtete Root Mean Squared Error $RMSE_w$ verwendet. Um einen vollständigen Eindruck der Leistung des Regressionsmodells zu bekommen, werden an relevanten Stellen zudem weitere der in Abschnitt 3.3.3 vorgestellten Beurteilungsmetriken der Regression angegeben und diskutiert. Zur Abbildung des Einflusses der verschiedenen Parameterausprägungen auf die Klassifikationsgüte des Regressionsmodells werden bei der Parameteroptimierung zusätzlich die TPR, TNR und das HM betrachtet. Der hierbei angewendete Schwellwert zur Binarisierung des Regressionsproblems wird auf 0,5 gesetzt.

Um die Ergebnisse, die unter Betrachtung eines binären Klassifikationsproblems erreicht werden, eindeutig von denen unterscheiden zu können, bei denen die bei der Querungsintention bestehende Unsicherheit berücksichtigt wird, wird im Folgenden

für das Erstgenannte der Begriff „Klassifikation“ und für das Letztgenannte der Begriff „Prädiktion“ verwendet.

Evaluierte Parameter: Merkmalsvektoren

Allein der kontextbasierte Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ hat zehn Parameter, die die Eigenschaften des Merkmals verändern und damit einen Einfluss auf die Güte des Gesamtsystems haben. Um den Aufwand der Parameterevaluation in einem vertretbaren Rahmen zu halten und den Einfluss der einzelnen Parameter auf das Regressionsergebnis untersuchen zu können, wird die Parameterevaluation iterativ durchgeführt. Hierbei wird von dem als Experte fungierenden Autor dieser Arbeit ein initialer Wert für jeden Parameter festgesetzt, um anschließend sequentiell jeweils nur einen Parameter oder eine zusammengehörige Parametergruppe zu variieren. Der Wert oder die Werte, die zu dem kleinsten $RMSE_w$ führen, werden als optimal angenommen und für alle folgenden Evaluationsschritte festgesetzt. Die Reihenfolge der für $\boldsymbol{x}_{Ctxt}$ evaluierten Parameter sowie der Initial- als auch der als optimal angenommene Endwert sind Tabelle 6.5 in Abschnitt 6.4.1 zu entnehmen.

Bei den im Rahmen der posenbasierten Erweiterung evaluierten Merkmalen handelt es sich fast ausschließlich um Deskriptoren, die bereits aus dem Stand der Technik bekannt sind. Daher wird hier keine ausführliche Parameterevaluation durchgeführt, sondern die in Abschnitt 5.3.2 beschriebene, aus dem Stand der Technik bekannte und an die vorliegende Datenbasis angepasste Parametrisierung verwendet. Eine Ausnahme bildet die zur Beschreibung der Kopfbewegung betrachtete Historie. Bei dieser werden verschiedene, aus dem Stand der Technik sinnvoll erscheinenden Parametrisierungen evaluiert. Auf Details wird im Rahmen der Vorstellung der erreichten Ergebnisse in Abschnitt 6.5.2 eingegangen.

Evaluierte Parameter: SVR

Die in dieser Arbeit verwendete SVR mit RBF-Kernel ist über den Gewichtungsfaktor C und den Skalierungsfaktor γ parametrisierbar (s. Abschn. 3.2.1, S. 62 ff.). Auf Grund der starken Interaktion zwischen C und γ wird die beste Parameter-

kombination über ein Grid Search-Verfahren mit logarithmisch steigenden Werten durchgeführt. Hierbei wird mit dem in der Literatur oft zu findenden Wertebereichen von $C = [1, 1e1, 1e2, 1e3]$ und $\gamma = [1e{-}3, 1e{-}4]$ begonnen (Pedregosa et al., 2011), um diesen sukzessiv in Richtung des minimalen $RMSE_w$ zu erweitern, bis ein Minimum gefunden wurde.

Da eine Änderung der Parameter des Merkmalsvektors zu einer veränderten Datenstruktur führt und für die neue Datenstruktur unterschiedliche C- und γ-Werte der SVR optimal sein können, wird das Grid Search-Verfahren für jeden evaluierten Merkmalsvektor durchgeführt. Ausgangspunkt der evaluierten SVR-Parameter ist dabei jeweils das Parameter-Paar, das in dem vorangegangenen Evaluierungsschritt über den $RMSE_w$ Wert als optimal bewertet wurde.

6.2.3 Objektbasierte Evaluation

Zur Ermittlung der Gesamtleistung der betrachteten Ansätze und zur weiteren Analyse der Situationen, die gut bzw. nicht so gut gehandhabt werden können, wird eine objektbasierte Evaluation, unter der Berücksichtigung der Zuordnung der Samples zu einem Fußgänger, durchgeführt. Basis der objektbasierten Evaluation ist die auf Seite 127 in Tabelle 6.4 dargestellte Zuordnung der fünfstufigen Labelklassen zu den sechs Objektklassen. Die objektbasierte Evaluation erfolgt nur für die bereits optimierten Merkmalsvektor-SVR-Kombinationen.

6.3 Ergebnisse: Kreuzvalidierung

6.3.1 Samplebasierte Ergebnisse

Die in Abschnitt 6.2.1 vorgestellte Methodik zur Kreuzvalidierung zielt auf eine möglichst gleichmäßige Verteilung der Samples und Situationen über die zur Validierung verwendeten zehn Teildatensätze. Abbildung 6.14 enthält die erreichte Anzahl der Samples pro Teildatensatz. Mit einem angestrebten Mittelwert von $\mu = \frac{N}{k} = 2.706$ weist die Verteilung eine Standardabweichung von $\sigma = 249$ auf.

Abbildung 6.15 zeigt weiter die mit der neuen Methode erreichte Verteilung für die fünf am häufigsten auftretenden Situationen. Mit 16.677 Samples bilden die fünf gezeigten Situationen 62 % des gesamten Datensatzes ab. Aus der Abbildung geht hervor, dass vier der fünf Situationen über die zehn Teildatensätze annähernd gleichverteilt sind. Der Teildatensatz 1 beinhaltet mit 799 Samples beispielsweise die wenigsten Samples der Situation $S_{Id} = 00000$. Dem Teildatensatz 4 sind mit 844 Samples die meisten Samples dieser Situation zugewiesen. Beide Werte sind im Rahmen der angestrebten 10 % der Gesamtanzahl an Samples pro Situation. Die Situation $S_{Id} = 41002$ weist hingegen einzelne Ausreißer auf. So beinhaltet Teildatensatz 3 mit 18 Samples nur 2 % aller Samples dieser Situation, Teildatensatz 6 mit 255 Samples hingegen 29 % der Samples. Diese ungleichmäßigere Verteilung ist auf die Gruppenzugehörigkeit der

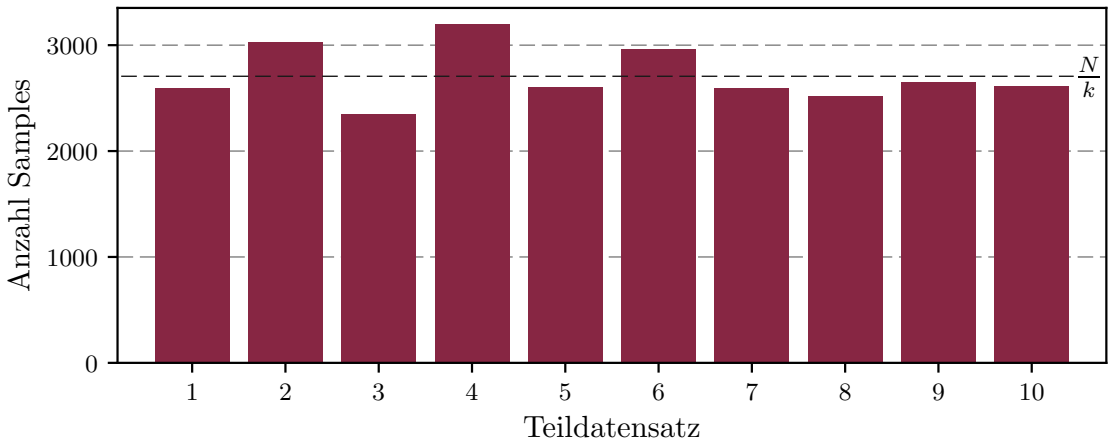


Abbildung 6.14: Verteilung der Samples über die zehn Teildatensätze.

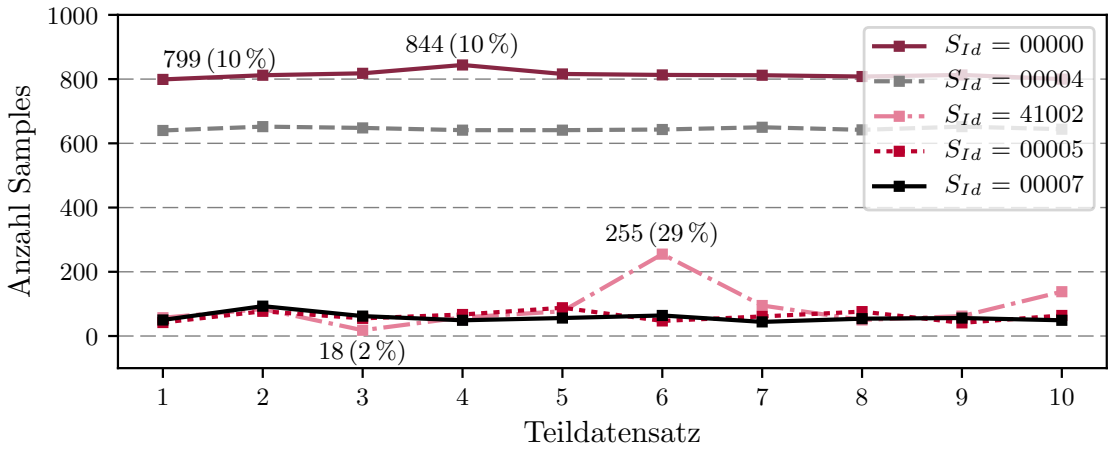


Abbildung 6.15: Verteilung der fünf am häufigsten auftretenden Situationen über die zehn Teildatensätze.

Samples innerhalb dieser Situation zurückzuführen. So sind die 894 Samples mit der Situationskennung $S_{Id} = 41002$ insgesamt in nur 17 verschiedene Gruppen gruppiert. Die größte dieser Gruppen umfasst dabei 255 Samples, die somit ganzheitlich demselben Teildatensatz zugewiesen werden müssen.

Eine solche Gruppenabhängigkeit verhindert vor allem bei seltenen Situationen eine gleichmäßige Verteilung der Samples über die zehn Teildatensätze. Da seltene Situationen häufig nur in einer einzelnen Videosequenz auftreten, gehören ihre Samples in der Regel zu einer Gruppe. Dadurch können 112 von insgesamt 211 Situationen nicht über mehr als einen Teildatensatz verteilt werden, wodurch das trainierte System diese Situationen bewerten muss, ohne Beispiele dieser Situation im Trainingsprozess jemals gesehen zu haben. Mit 1.119 Samples bilden diese Situationen jedoch nur 4 % des gesamten Datensatzes ab.

6.3.2 Diskussion und Bewertung

Ziel des in Abschnitt 6.2.1 vorgestellten Verfahrens zur Kreuzvalidierung ist die Erzeugung einer gleichmäßigen Verteilung der Samples und Situationen über die zehn Teildatensätze. Die im vorangegangenen Abschnitt vorgestellten Ergebnisse zeigen, dass

das Ziel mit dem zur Erstellung der Teildatensätze entwickelten Verfahren erreicht wurde. Mit einer relativen Standardabweichung von unter 1 % kann die Anzahl der Samples als gleichmäßig verteilt betrachtet werden. Ebenso sind die häufig auftretenden Situationen ausgewogen über die zehn Teildatensätze verteilt. Ausreißer entstehen hier nur bei selten auftretenden Situationen, bei denen auf Grund der im Datensatz bestehenden Abhängigkeiten keine weitere Aufteilung der Gruppen möglich ist. Ohne die bestehenden Gruppenabhängigkeiten zu verletzen, lässt sich eine noch gleichmäßigere Verteilung der Situationen somit nicht erreichen. Abhilfe könnte hier eine systematische Erweiterung des Datensatzes schaffen, bei der vornehmlich Videosequenzen der aktuell selten auftretenden Situationen ergänzt werden; was jedoch mit einen großen Aufzeichnungs- und Labelaufwand verbunden ist.

Zudem arbeitet das zur Kreuzvalidierung entwickelte Verfahren sehr generisch. Daher ist es nicht nur auf den in dieser Arbeit betrachteten Anwendungsfall beschränkt. Es kann vielmehr bei allen Problemstellungen eingesetzt werden, bei denen unausgewogene Daten sowie eine Korrelationen zwischen einzelnen Samples vorliegen und somit eine Kombination aus der Stratified- und der Group k-fold Kreuzvalidierung gefordert ist. Einzig die Bestimmung der Samplegewichte ist abhängig vom jeweiligen Anwendungsfall und muss entsprechend angepasst werden.

6.4 Ergebnisse: Kontextbasierter Ansatz

6.4.1 Samplebasierte Ergebnisse

Im Folgenden werden zunächst die Ergebnisse der Parameterevaluation vorgestellt, um anschließend im Detail auf die samplebasierten Ergebnisse des besten Parametersatzes einzugehen. In dem Rahmen wird auch der Einfluss weiterer Faktoren des Gesamtsystems, wie die in Abschnitt 5.4 vorgestellte situationsspezifische Gewichtung der Samples während des Trainingsprozesses sowie die Genauigkeit der erkannten Körperorientierung des Fußgängers untersucht.

Parameterevaluation

Die Parameter des kontextbasierten Merkmalsvektors $\boldsymbol{x}_{Ctxt}$ werden über die in Abschnitt 6.2.2 beschriebene Methodik sowie der in Tabelle 6.5 angegebenen Reihenfolge und Initialparametrisierung evaluiert.

Abbildung 6.16 verdeutlicht zunächst den Einfluss der SVR-Parameter C und γ auf die Güte des Gesamtsystems anhand des initialen Parametersatzes.

Laut oberer linker Grafik erreicht eine SVR mit $C = 1\text{e}6$ und $\gamma = 1\text{e}{-4}$ mit einem $RMSE_w$ von 0,3511 für den initiale Parametersatz das beste Ergebnis. Entsprechend dem $RMSE$-Wert führt diese Parametrisierung zu einem Modell, dessen Prädiktionen im Durchschnitt um 0,27 vom tatsächlichen Wert abweichen. Da die TPR mit 64 % deutlich unter der TNR von 95 % liegt ($HM = 0{,}764$), scheinen diese Abweichungen vornehmlich in einer Unterschätzung der Ausprägung einer Querungsintention zu beruhen.

Die Grafiken unterstreichen zudem die Bedeutung des zur Bestimmung der SVR-Parameter C und γ durchgeführten Grid-Search Verfahrens. So führen steigende γ Werte beispielsweise sehr schnell zu einer deutlich schlechteren Trennbarkeit der Daten; eine Änderung zu $\gamma = 1\text{e}{-2}$ hat beispielsweise eine Reduktion der TPR auf 15 % zur Folge. Ebenso führt ein niedrigerer C-Wert zu einem deutlichen Overfitting auf die negativen Daten, was sich ab einem Wert von $C = 1e2$ in einer TPR von 13 % bis 0 %

Tabelle 6.5: Evaluierte Parameter für $\boldsymbol{x}_{Ctxt}$.

	Parameter		Einheit	Initialwert	Endwert
CMHI: Normalisierter Szenenausschnitt					
1.	Auflösung	r_s	px/m	10	5
2.	Größe	$M_p \times N_p$	m	10×10	8×8
CMHI: Movement History Image					
3.	Zeitfaktor	δ	-	1	1
3.	Zerfallsvariable	τ	-	200	50
CHOG					
4.	Zellengröße	$M_c \times N_c$	px	10×10	20×20
4.	Anzahl Bins	b	-	9	9
4.	Wertebereich		Grad	$[0° - 180°)$	$[0° - 180°)$
5.	Normalisierungsmethode		-	L2-Norm	L2-Hys
CO					
6.	Auflösung	r_o	Zellen/m	1	1
WAO					
7.	Auflösung	r_w	Zellen/m	1	0,5
Dimension d des Merkmalsvektors:				1.104	120

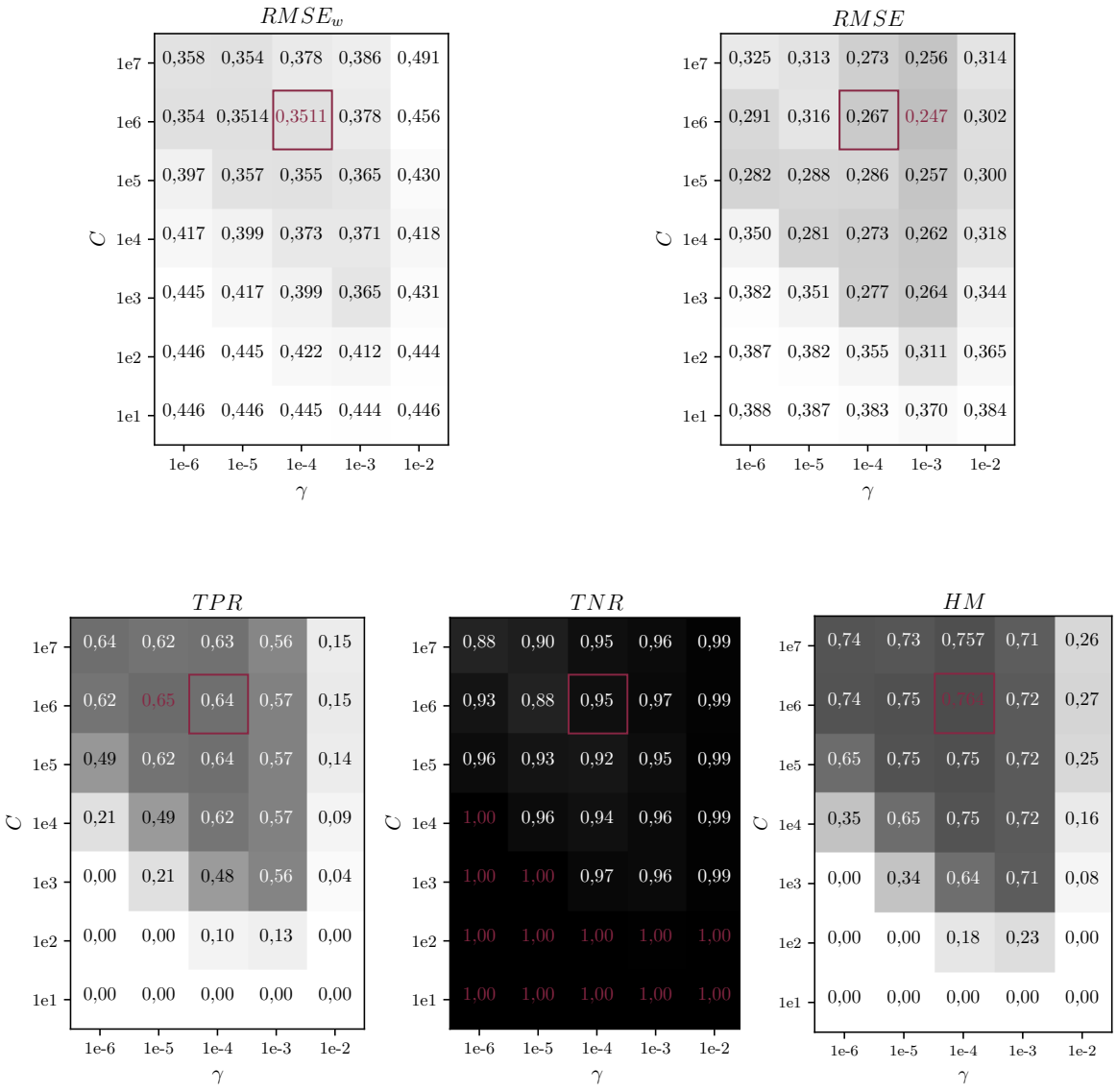


Abbildung 6.16: Einfluss der SVR-Parameter C und γ unter Verwendung des initialen Parametersatzes für den Merkmalsvektors $\boldsymbol{x}_{Ctxt}$.

Tabelle 6.6: Einfluss der Auflösung r_s des CMHI in px/m.

r_s	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
12	1.500	0,360	0,298	0,62	0,93	0,74	1e0	1e−5
10	1.104	0,351	0,267	0,64	0,95	0,76	1e6	1e−4
8	780	0,347	0,274	0,62	0,94	0,75	1e0	1e−6
6	528	0,336	0,258	0,67	0,94	0,78	1e2	1e−5
5	429	0,319	0,246	0,66	0,95	0,78	1e7	1e−7
4	348	0,324	0,253	0,64	0,94	0,76	1e−1	1e−5
3	285	0,362	0,251	0,53	0,97	0,69	1e6	1e−2
2	240	0,356	0,283	0,51	0,92	0,66	1e6	1e−2

zeigt, bei einer TNR von 100 %. Die RMSE-Werte nehmen in diesen Fällen mit $RMSE_w = 0{,}45$ und $RMSE = 0{,}39$ ihre Maximalwerte an.

Wie in Tabelle 6.5 angegeben, wird im Rahmen der Parameterevaluation des kontextbasierten Merkmalsvektors zunächst die Auflösung des normalisierten Szenenausschnitts des CMHI variiert. Tabelle 6.6 enthält die Ergebnisse jeweils für die SVR-Parametrisierung, die zu einem minimalen $RMSE_w$-Wert führt. In dieser und den nächsten Ergebnistabellen kennzeichnet die grau gestrichelt umrandete Zeile die jeweilige Initialparametrisierung und die schwarz umrandete Zeile die als am besten bewertete Parametrisierung, welche wiederum als Initialparameterisierung für den nächsten Evaluierungsschritt dient.

Entsprechend den angegebenen $RMSE_w$-Werten führt eine Reduktion der Auflösung auf $r_s = 5$ px/m mit einem $RMSE_w$ von 0,319 zu dem besten Ergebnis. Durch die Reduzierung der Auflösung und der daraus resultierenden Reduktion der Größe des Merkmalsvektors auf $d = 429$ Elemente, sinkt der durchschnittliche Vorhersagefehler auf $RMSE = 0{,}246$. Bezüglich der binären Klassifikation steigt einzig die TPR um 2 % auf $TPR = 0{,}66$. Die Dimensionsreduktion des Merkmalsvektors führt somit zu einer verbesserten Genauigkeit der Vorhersage bezüglich der Unsicherheit über die Ausprägung einer Querungsintention sowie zu einer leichten Verbesserung der binären Entscheidung. Ab $r_s = 4$ px/m übersteigt der durch die geringere Auflösung hervor-

Tabelle 6.7: Einfluss der Größe $M_p \times N_p$ des CMHI in m.

$M_p \times N_p$	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
18×18	1.381	0,341	0,268	0,64	0,94	0,76	1e9	1e−8
16×16	1.092	0,346	0,254	0,61	0,95	0,75	1e6	1e−6
14×14	837	0,326	0,269	0,69	0,92	0,79	1e9	1e−8
12×12	616	0,329	0,255	0,65	0,94	0,77	1e6	1e−6
10×10	429	0,319	0,246	0,66	0,95	0,78	1e7	1e−7
8×8	276	0,315	0,247	0,63	0,94	0,76	1e4	1e−4
6×6	157	0,322	0,271	0,63	0,92	0,75	1e9	1e−5
4×4	72	0,319	0,288	0,63	0,92	0,75	1e6	1e−4
2×2	45	0,339	0,290	0,52	0,91	0,66	1e9	1e−6

gerufene Informationsverlust jedoch die vorteilhafte Dimensionsreduktion, was sich in einer Verschlechterung der Ergebnisse zeigt.

Als zweiter Parameter wird der Einfluss der Größe $M_p \times N_p$ des CMHI evaluiert. Die Ergebnisse in Tabelle 6.7 zeigen, dass eine Reduktion der betrachteten Szene auf einen Bereich von 8×8 m zu einem verbesserten $RMSE_w$ führt. Dies hat jedoch zur Folge, dass der Anteil der Fußgänger, die richtig als querungswillig klassifiziert werden, um 3 % sinkt ($TPR = 0{,}63$). Bemerkenswert ist, dass selbst ein deutlich reduzierter Szenenbereich von 4×4 m zu einer ähnlich guten Klassifikationsleistung führt. Erst bei einer Einschränkung des kontextuellen Einflussbereichs auf 2×2 m fällt diese deutlich ab.

Die dritte Größe, die Einfluss auf die Gestaltung des CMHI hat, ist die über den Zeitfaktor δ und die Zerfallsvariable τ parametrierbare Historie der Fußgängerbewegung. Tabelle 6.8 zeigt die Ergebnisse der durchgeführten Untersuchung. Eine Historie von $t = 50$ Samples ($\hat{=}$ 2,8 s) führt mit einem $RMSE_w$ von 0,309 zu den besten Ergebnissen. Mit $\mu = 42$ entspricht dieser Wert in etwa der durchschnittlichen Zeit, die ein Fußgänger in dem verwendeten Datensatz präsent ist (s. Abschn. 6.1.1). Die Abbildung längerer Bewegungshistorien scheint keinen Mehrwert für die Erkennung der Querungsintention eines Fußgängers zu bieten. Die Ergebnisse werden sogar leicht

Tabelle 6.8: Einfluss des Zeitfaktors δ und der Zerfallsvariable τ des CMHI.

δ	τ	t	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
200	1	200	0,315	0,247	0,63	0,94	0,76	1e4	1e−4
100	1	100	0,317	0,250	0,63	0,95	0,76	1e5	1e−5
60	1	60	0,320	0,254	0,65	0,95	0,77	1e5	1e−4
50	1	50	0,309	0,237	0,67	0,93	0,78	1e5	1e−4
40	1	40	0,319	0,252	0,65	0,95	0,77	1e5	1e−4
10	1	10	0,327	0,242	0,67	0,95	0,79	1e5	1e−4
5	1	5	0,331	0,245	0,66	0,95	0,78	1e5	1e−4
2	1	2	0,344	0,243	0,58	0,93	0,71	1e6	1e−4
1	1	1	0,344	0,243	0,58	0,93	0,71	1e6	1e−4
250	5	50	0,319	0,251	0,64	0,94	0,76	1e5	1e−5
250	25	10	0,333	0,242	0,67	0,95	0,79	1e5	1e−4
250	250	1	0,344	0,243	0,58	0,93	0,71	1e6	1e−4

schlechter. Dies kann damit begründet werden, dass sich bei einer längeren Bewegungshistorie neue in das CMHI eingezeichnete Bewegungsmuster nicht so deutlich von der bereits eingezeichneten Historie abzeichnen, wie bei der Abbildung kürzerer Bewegungshistorien, und die neuen Informationen somit verschleiert werden.

Zudem geht aus den Ergebnissen hervor, dass selbst mit einem kurzen Zeitfenster von fünf Samples ähnlich gute Ergebnisse erreicht werden können, wie mit der als optimal angenommenen Historie von 50 Samples. Dies gilt vor allem in Bezug auf die Klassifikation. Eine reine Darstellung der aktuellen kontextuellen Situation ohne Bewegungshistorie ($t = 1$) führt hingegen zu deutlich schlechteren Ergebnissen. Mit einem $RMSE$ von 0,243 und einem HM von 0,71 kann aber selbst damit ein überzufälliges Ergebnis erreicht werden. Entsprechend den unteren drei Zeilen der Tabelle 6.8, führt eine Erhöhung des Kontrasts zwischen den einzelnen Zeitschritten durch eine erhöhte Zerfallsvariable τ schließlich zu keiner Ergebnisverbesserung.

Für das auf Basis des CMHI erstellte CHOG wird zunächst der Einfluss der Zellengröße $M_c \times N_c$, der Anzahl der Bins b sowie der Einfluss des Wertebereichs der

Tabelle 6.9: Einfluss der Zellengröße $M_c \times N_c$ in px und der Anzahl der Bins b des CHOG, bei einem Wertebereich von $[0°, 180°)$.

$M_c \times N_c$	b	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
40×40	9	141	0,361	0,291	0,60	0,90	0,72	1e7	1e−5
20×20	9	168	0,308	0,250	0,72	0,93	0,81	1e5	1e−4
10×10	9	276	0,309	0,237	0,67	0,93	0,78	1e5	1e−4
8×8	9	357	0,336	0,259	0,66	0,94	0,78	1e5	1e−5
5×5	9	708	0,321	0,256	0,70	0,94	0,80	1e5	1e−4
4×4	9	1.032	0,316	0,272	0,67	0,94	0,78	1e5	1e−4
40×40	6	138	0,351	0,274	0,62	0,91	0,74	1e7	1e−5
20×20	6	156	0,316	0,262	0,74	0,92	0,82	1e7	1e−5
10×10	6	228	0,326	0,251	0,66	0,93	0,77	1e5	1e−4
40×40	4	136	0,369	0,284	0,62	0,91	0,74	1e7	1e−6
20×20	4	148	0,317	0,267	0,72	0,93	0,81	1e6	1e−4
10×10	4	196	0,335	0,252	0,60	0,95	0,74	1e6	1e−6

Tabelle 6.10: Einfluss des Wertebereichs und der Anzahl der Bins b des CHOG, bei einer Zellengröße von 20×20 px.

Wertebereich	b	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
$[0°, 180°)$	9	168	0,308	0,250	0,72	0,93	0,81	1e5	1e−4
$[0°, 180°)$	6	156	0,316	0,262	0,74	0,92	0,82	1e7	1e−5
$[0°, 180°)$	4	148	0,317	0,267	0,72	0,93	0,80	1e6	1e−4
$[0°, 360°)$	18	204	0,313	0,252	0,73	0,93	0,82	1e5	1e−5
$[0°, 360°)$	12	180	0,332	0,262	0,71	0,93	0,80	1e6	1e−5
$[0°, 360°)$	8	164	0,323	0,262	0,71	0,93	0,80	1e6	1e−5

Tabelle 6.11: Einfluss der Normalisierungsmethode des CHOG.

Methode	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
L2-Norm	0,308	0,250	0,72	0,93	0,81	1e5	1e−5
L2-Hys	0,294	0,237	0,75	0,94	0,83	1e5	1e−4
L1-Norm	0,307	0,251	0,75	0,93	0,83	1e5	1e−4
L1-Sqrt	0,298	0,247	0,76	0,93	0,84	1e5	1e−4
Keine	0,310	0,248	0,72	0,94	0,82	1e5	1e−4

Gradienten betrachtet. Gemäß Tabelle 6.9 führt eine Aufteilung des 40×40 px großen CMHI in vier 20×20 px große Zellen zu den besten Ergebnissen. Mit allen drei getesteten Bin-Variationen wird erstmalig eine TPR von über 70 % und ein HM von über 80 % erreicht. Der insgesamt niedrigste $RMSE_w$-Wert wird schließlich mit einer Einteilung der Gradienten in ein Histogramm mit 9 Bins erzielt. Tabelle 6.10 zeigt zudem, dass die Betrachtung der Gradientenvorzeichen (Wertebereich $[0°, 360°)$) zu keiner Verbesserung der Ergebnisse führt.

Die zur Normalisierung der Zellenhistogramme eingesetzte Methode hat einen weiteren Einfluss auf das $CHOG$. Tabelle 6.11 enthält die Ergebnisse für die in der Literatur üblicherweise zu findenden Methoden. Mit einem $RMSE_w$ von 0,294 führt die als L2-Hys bezeichnete Methode zu den besten Ergebnissen. Bei dieser werden ie L2-normalisierten Zellenhistogramme zunächst auf einen Maximalwert von 0,2 begrenzt, um anschließend eine erneute L2-Normalisierung durchzuführen. Der schlechteste $RMSE_w$ wird erreicht, wenn die Zellenhistogramme nicht normalisiert werden.

Im letzten Evaluierungsschritt wird, über die Variation der Auflösung r_o bzw. r_w und der daraus resultierenden Anzahl der Zellen der CO und der WAO, der Einfluss der Positionsgenauigkeit gegebenenfalls präsenter Fußgängerüberwege sowie Wartebereiche untersucht. Gemäß Tabelle 6.12 führt die initial angenommene Auflösung von $r_o = 1$ Zelle/m bei der Beschreibung der Position von Fußgängerüberwegen zu den besten Ergebnissen. Bei der Beschreibung von Wartebereichen führt hingegen eine gröbere Auflösung von $r_w = 0{,}5$ Zellen/m zu einer weiteren Verbesserung der Er-

gebnisse (s. Tab. 6.13). Bemerkenswert ist, dass sich die Veränderung der Ergebnisse sowohl bei der CO als auch bei der WAO fast ausschließlich in dem situationsspezifisch gewichteten $RMSE_w$ zeigt.

Tabelle 6.14 fasst schließlich die mit dem optimalen Parametersatz erreichten Ergebnisse zusammen und stellt diese den Ergebnissen der Initialparametrisierung gegenüber.

Tabelle 6.12: Einfluss der Auflösung r_o in Zellen/m und der daraus resultierenden Anzahl an Zellen $M_o \times N_o$ der CO.

r_o	$M_o \times N_o$	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
2	16×16	360	0,301	0,238	0,74	0,94	0,83	1e6	1e−5
1	8×8	168	0,294	0,237	0,75	0,94	0,83	1e5	1e−4
0,5	4×4	120	0,302	0,242	0,74	0,94	0,83	1e6	1e−4
0,25	2×2	108	0,313	0,247	0,75	0,94	0,83	1e5	1e−4
0,125	1×1	105	0,312	0,248	0,74	0,94	0,83	1e6	1e−5

Tabelle 6.13: Einfluss der Auflösung r_w in Zellen/m und der daraus resultierenden Anzahl an Zellen $M_w \times N_w$ der WAO.

r_w	$M_w \times N_w$	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
2	16×16	360	0,353	0,240	0,73	0,94	0,82	1e4	1e−3
1	8×8	168	0,294	0,237	0,75	0,94	0,83	1e5	1e−4
0,5	4×4	120	0,290	0,237	0,75	0,94	0,83	1e5	1e−4
0,25	2×2	108	0,305	0,245	0,72	0,94	0,82	1e5	1e−4
0,125	1×1	105	0,328	0,246	0,72	0,93	0,81	1e5	1e−4

Tabelle 6.14: Ergebnisse des optimalen Parametersatzes im Vergleich zur Initialparametrisierung.

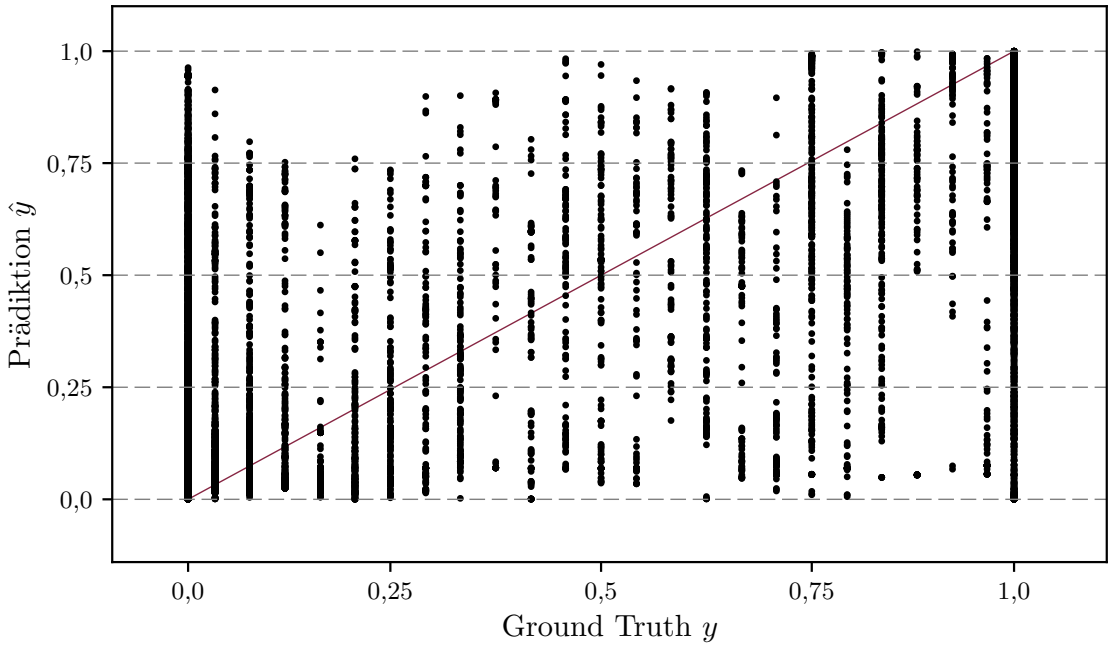
Parametrisierung	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
Optimal	120	0,290	0,237	0,75	0,94	0,83	1e5	1e−4
Initial	1.104	0,351	0,267	0,64	0,95	0,76	1e6	1e−4

Optimaler Parametersatz

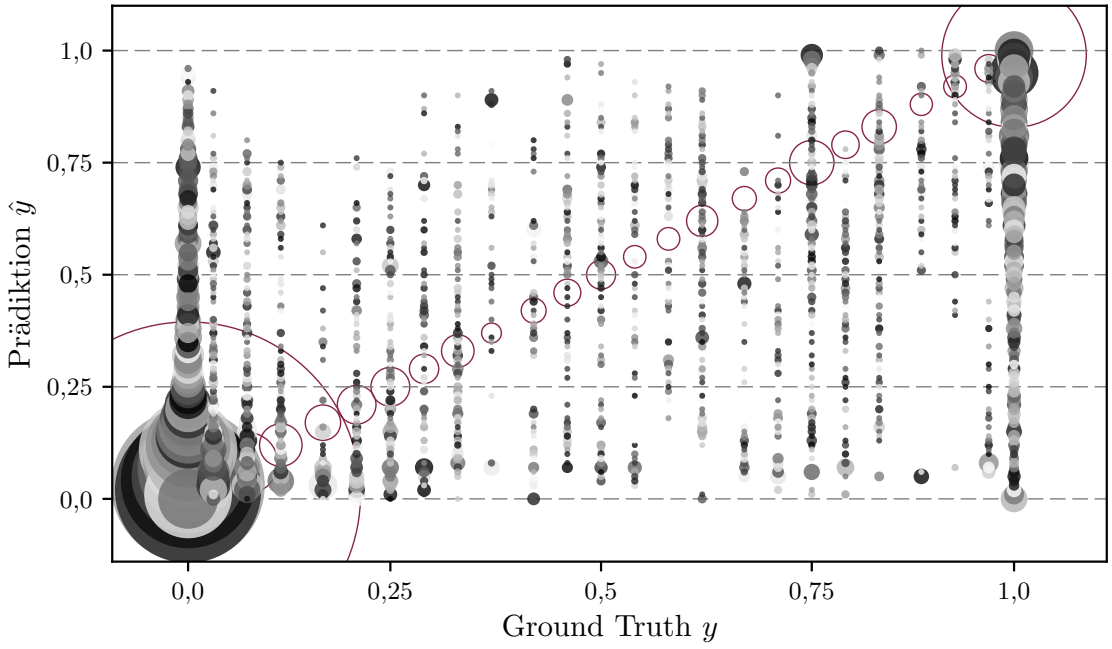
Wie im obigen Abschnitt beschrieben, kann die Querungsintention eines Fußgängers über den optimal parametrisierten, kontextbasierten Merkmalsvektor $\boldsymbol{x}_{ctxt}$ und einer SVR mit einem RBF-Kernel mit $C = 1e5$ und $\gamma = 1e-4$, mit einer durchschnittlichen Abweichung von $RMSE = 0{,}237$ vorhergesagt werden.

Abbildung 6.17 vermittelt einen qualitativen Eindruck der Verteilung des Prädiktionsfehlers entlang des gesamten Wertebereichs der Ground Truth. Während das Streudiagramm in Abbildung 6.17a dem aus der Literatur bekannten, klassischen Streudiagramm entspricht (s. Abschn. 3.3.3), zeigt Abbildung 6.17b eine an die große Menge an Datenpunkten angepasste Darstellung, bei der die Größe eines jeden Kreises die Anzahl der Samples mit demselben $(y, \hat{y})$-Wertepaares repräsentiert, gerundet auf zwei Dezimalstellen. Die rote Linie in Abbildung 6.17a bzw. die roten Kreise in Abbildung 6.17b zeigen jeweils die angestrebte Darstellung einer perfekten Prädiktion ohne Fehler ($\hat{y} = y$).

Es zeigt sich, dass die abgebildeten Wertepaare prinzipiell den gesamten Wertebereich abdecken und somit kein systematischer Fehler, wie beispielsweise ein Overfitting auf den Mittelwert, vorliegt. Weiter geht aus den Abbildungen hervor, dass Beispiele, bei denen der Fußgänger keine Querungsintention hat, mit einem geringeren Fehler prädiziert werden, als solche, mit Querungsintention. So weisen Beispiele mit einer Ground Truth von $y \leq 0{,}25$ eine deutliche Häufung im Bereich $\hat{y} \leq 0{,}25$ auf und werden selten bis nie auf Werte über 0,75 prädiziert. Dies gilt vor allem für Samples mit eindeutig keiner Querungsintention ($y = 0{,}0$). Diese werden vornehmlich auch als solche prädiziert. Zudem sinkt die Anzahl der Falschprädiktionen mit steigendem Fehlerwert kontinuierlich.



(a)



(b)

Abbildung 6.17: Grafische Darstellung der Regressionsergebnisse für $\boldsymbol{x}_{Ctxt}$.
(a) Klassisches Streudiagramm. (b) An die Datenmenge angepasste Darstellung.

Bei Beispielen von Fußgängern mit Querungsintention kann dieses in ähnlicher Form bei Ground Truth Werten bis zu $y \geq 0{,}88$ beobachtet werden. Insbesondere die Prädiktionen von Samples mit einer eindeutig ausgeprägten Querungsintention ($y = 1{,}0$) häufen sich auch hier in Bereichen mit niedrigen Fehlerwerten. Mit steigendem Fehlerwert sinkt die Anzahl der Falschprädiktionen jedoch nicht so stark ab, wie bei den Samples mit eindeutig keiner Querungsintention.

Beispiele, bei denen eine Unsicherheit über die Ausprägung einer Querungsintention herrscht, zeigen hingegen keine Häufung der Prädiktionen $\hat{y}$ im Bereich des Ground Truth Werts y. Die Prädiktionen sind vielmehr gleichmäßig über den gesamten Unsicherheitsbereich verteilt, nehmen dabei jedoch selten Werte nah am Mini- oder Maximum ein. Somit wird bei den Samples mit Unsicherheit nur selten eine eindeutige Aussage über die Ausprägung einer Querungsintention suggeriert, der genaue Unsicherheitswert wird jedoch auch nur selten korrekt vorhergesagt.

Aus den in Abbildung 6.18 dargestellten Labelklassen-spezifischen RMSE-Werte geht hervor, dass die Samples der unsicheren Klassen ($C_-, C_?, C_+$) dennoch mit einem durchschnittlich geringeren Fehler prädiziert, als Samples mit einer eindeutig positiven Querungsintention (Klasse C_{++}). Im Detail steigt die durchschnittliche Abweichung der Prädiktionen von der Ground Truth in Richtung der Klassen, die eine positiv ausgeprägte Querungsintention repräsentieren, kontinuierlich an. Dabei werden Beispiele der Klasse C_{--} mit einem RMSE von 0,190 mit dem durchschnittlich geringsten Fehler prädiziert; Prädiktionen der Klasse C_{++} weisen mit einem RMSE von 0,386 den höchsten Fehler auf.

Ein detaillierter Blick auf die Konfusionsmatrix in Abbildung 6.19 zeigt dabei, dass für alle fünf Klassen, die meisten Beispiele in die richtige Klasse oder in eine direkte Nachbarklasse prädiziert werden. Mit 72 % wird die Klasse C_{--} am häufigsten richtig prädiziert. Bei der Klasse C_- wird die Unsicherheit in der Beobachtung hingegen nur zu 27 % richtig vorhergesagt, während 51 % der Samples in Richtung der Klasse C_{--} überschätzt werden. Bei den Samples mit Querungsintention werden die eindeutigen Beispiele (Klasse C_{++}) mit 33 % zu 39 % hingegen häufiger als unsicherheitsbehaftet prädiziert. Samples der Klasse $C_?$, bei denen keine Aussage über die Präsenz einer Querungsintention gemacht werden kann, werden mit 30 % ebenfalls am häufigsten

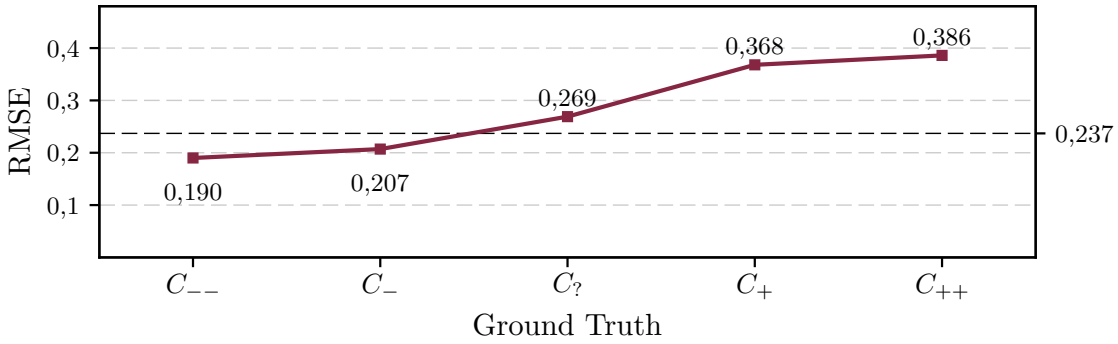


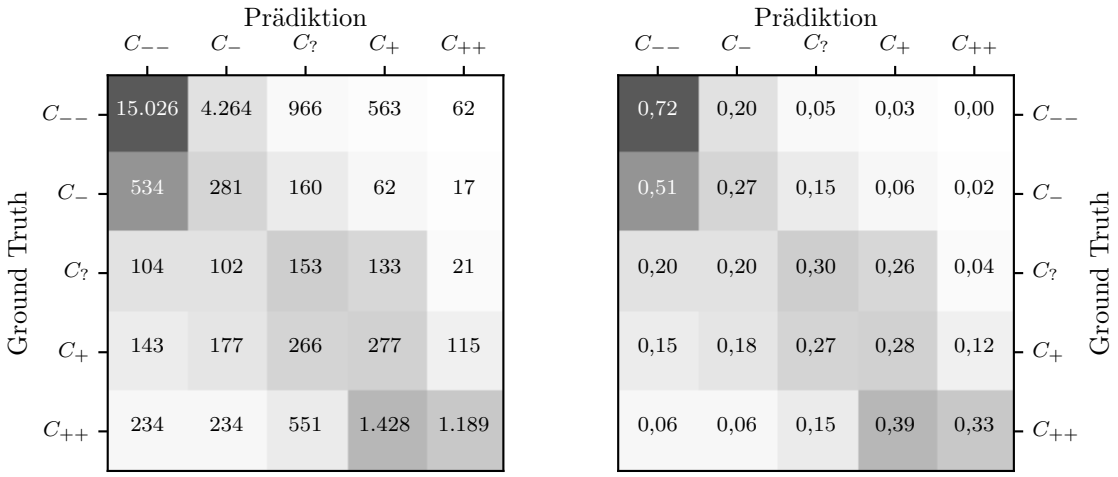
Abbildung 6.18: Labelklassen-spezifische RMSE-Werte für $\boldsymbol{x}_{Ctxt}$.

in diese Klasse prädiziert. Bis auf die Klasse C_{++} sind die anderen vier Klassen mit 20–26 % aber ähnlich wahrscheinlich, was die obigen Beobachtungen bezüglich der Vorhersage des genauen Unsicherheitswerts bestätigt.

Wie im Rahmen der Parameterevaluation oben gezeigt, wird unter der Betrachtung eines binären Klassifikationsproblems und einem Schwellwert von 0,5 eine TPR von 75 % und eine TNR von 94 % erreicht. Das daraus resultierende harmonischen Mittel liegt bei $HM = 0{,}83$. Entsprechend der in Abbildung 6.20 gezeigten ROC-Kurve ist es möglich, das harmonische Mittel, über eine Verschiebung des Schwellwerts zu 0,3, auf $HM = 0{,}87$ zu erhöhen. Mit einer TPR von 83 % und einer TNR von 90 % wirkt sich diese Verschiebung zu Gunsten der Beispiele mit Querungsintention aus. Die, über die AUROC angegebene Wahrscheinlichkeit, dass ein zufälliges positives Testbeispiel höher eingestuft wird als ein zufälliges negatives Testbeispiel, liegt bei 91 %.

Situationsspezifische Ergebnisse

Abbildung 6.21a erlaubt eine situationsspezifische Betrachtung der Ergebnisse an Hand einer Kategorisierung der Samples in Abhängigkeit der Präsenz zusätzlicher Szenenelemente (s. Abschn. 6.1.3). Es zeigt sich, dass sowohl Beispiele von Fußgängern, die sich im Einflussbereich eines Wartebereichs befinden (S_{WtgAr}), als auch Beispiele von Fußgängern bei denen keine zusätzlichen Szenenelemente im betrachteten Einflussbereich vorhanden sind, einen niedrigeren Fehlerwert aufweisen, als der gesamte Datensatz



(a) (b)

Abbildung 6.19: Fünfklassige Konfusionsmatrix für $\boldsymbol{x}_{Ctxt}$. (a) Absolute Anzahl an Samples. (b) Prozentualer Anteil an Samples in Bezug auf die Ground Truth.

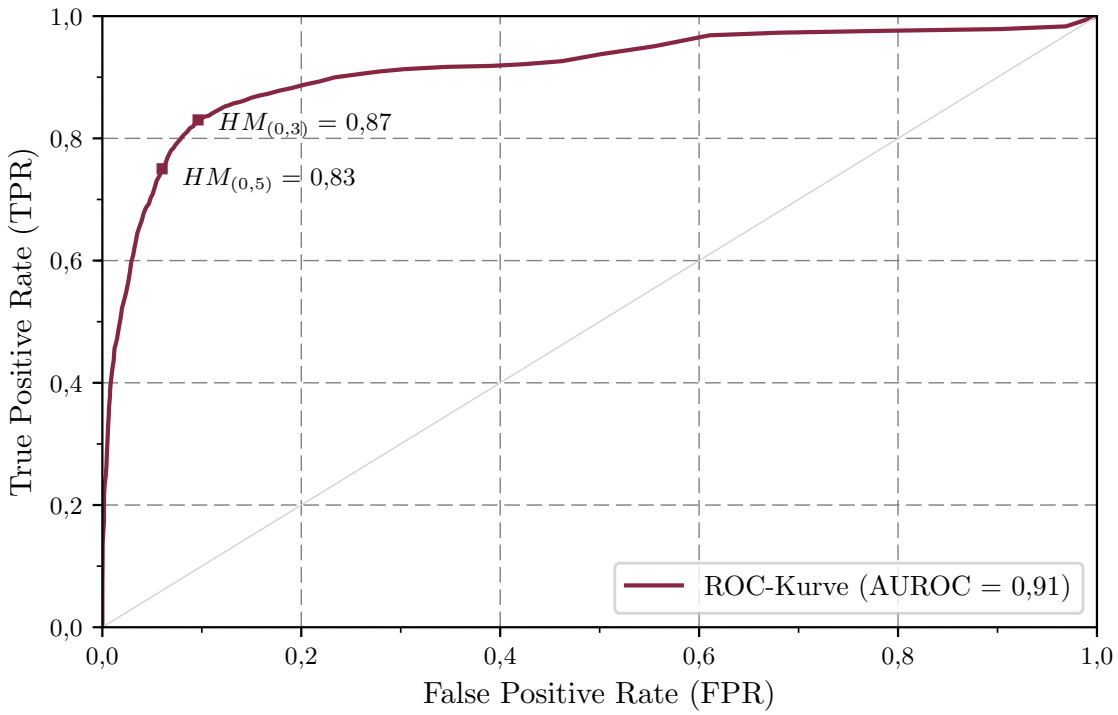


Abbildung 6.20: ROC-Kurve für $\boldsymbol{x}_{Ctxt}$.

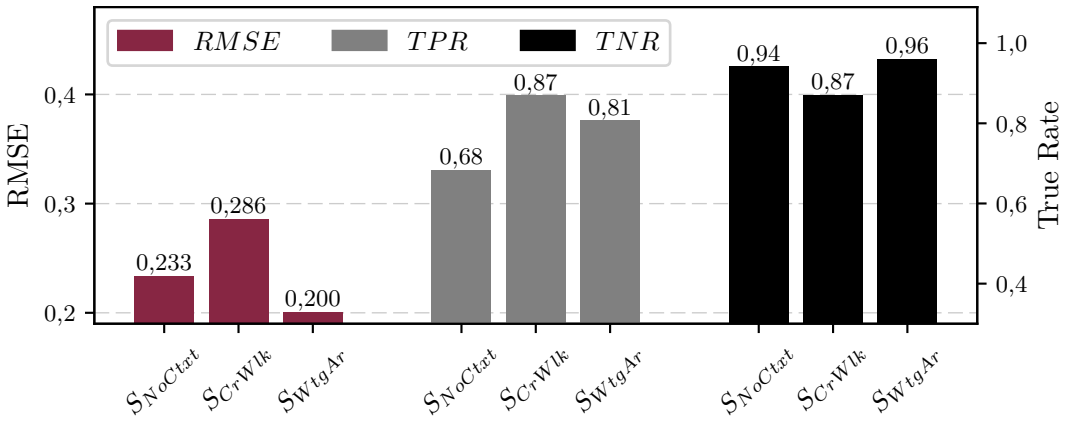
($RMSE = 0{,}237$). Einzig die Samples von Fußgängern, die im Einflussbereich eines Fußgängerüberwegs positioniert sind, werden mit einem $RMSE$ von 0,286 weniger genau prädiziert.

Der größere Prädiktionsfehler scheint vor allem auf die Fußgänger zurückzuführen zu sein, die in der Nähe eines Fußgängerüberwegs keine Absicht haben, die Straße zu queren. Denn während die positiven Samples der Kategorie S_{CrWlk} mit einer TPR von 87 %, im Vergleich zu den anderen zwei Kategorien, am häufigsten korrekt positiv klassifiziert werden, erreichen die negativen Samples mit einer TNR von ebenfalls 87 % nur den niedrigsten Wert.

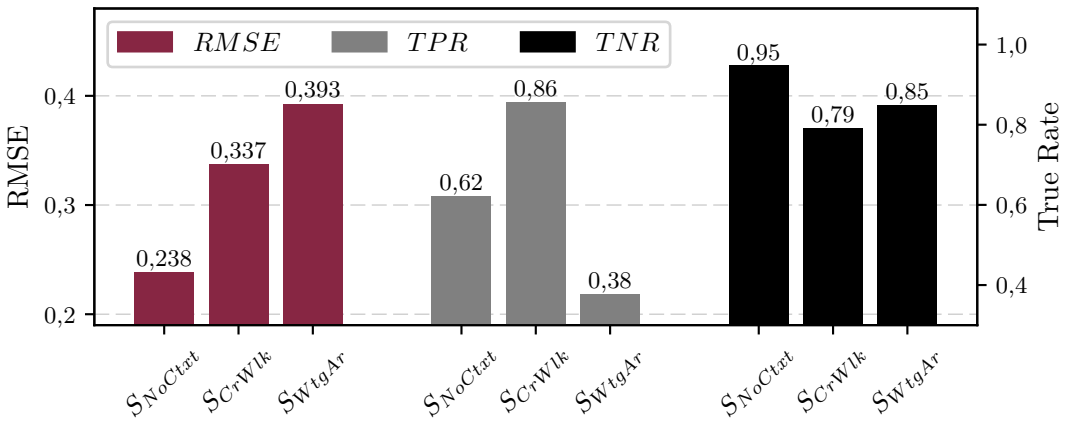
Dass trotz der ungleichverteilten Datenmengen auch die Situationen mit wenig Trainingsbeispielen gut vorhergesagt werden können, ist vor allem auf die in Abschnitt 5.4 vorgestellte, situationsspezifische Gewichtung der Samples während des Trainingsprozesses zurückzuführen. Gemäß Abbildung 6.21b führt ein Training der SVR ohne der zusätzlichen Gewichtung bei selten auftretenden Samples zu einer deutlichen Verschlechterung der Prädiktionsgüte, wie beispielsweise bei Fußgängern, die in einem Wartebereich eine positiv ausgeprägte Querungsintention haben. Ebenfalls werden Samples von Fußgängern, die im Einflussbereich eines Fußgängerüberwegs keine Querungsintention aufweisen, ohne Gewichtung der Trainingsbeispiele deutlich schlechter vorhergesagt. Dies spiegelt die auf Seite 131 in Abbildung 6.10 gezeigte, Szenenelementabhängige Verteilung der Labels wider. Der RMSE-Wert des gesamten Datensatzes beträgt unter Verwendung des ungewichteten Trainingsprozesses 0,260.

Genauigkeit der Körperorientierung

Wie in Abschnitt 6.1.1 erläutert, wird in dieser Arbeit die, für den kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ vorausgesetzte, Körperorientierung des Fußgängers θ_B über ein manuelles Labeling bestimmt. Um dennoch die Leistung des kontextbasierten Verfahrens unter Verwendung eines realistischen Erkennungssystems abschätzen zu können, wird im Folgenden den Körperorientierungslabels mittelwertfreies, normalverteiltes Sensorrauschen mit einer variierenden Standardabweichung hinzugefügt. Gemäß Abbildung 6.22a führt das zusätzliche Sensorrauschen zu einer Verschlechterung der



(a)



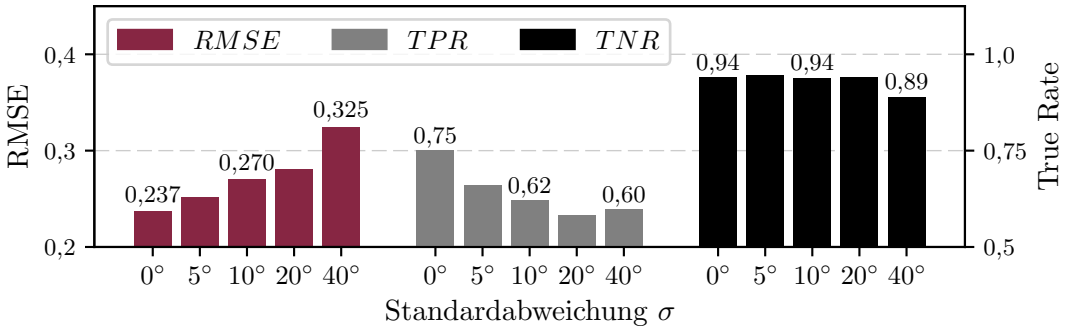
(b)

Abbildung 6.21: Situationsspezifische Ergebnisse und Einfluss der Samplegewichtung während des SVR-Trainingsprozesses. (a) Situationsspezifische Gewichtung. (b) Keine Gewichtung.

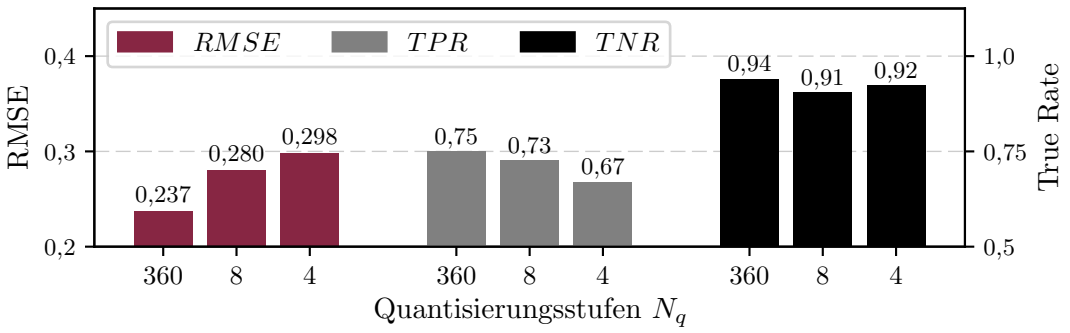
Ergebnisse. Neben einem steigendem RMSE verschlechtert sich vor allem die Klassifikationsleistung bei Beispielen mit einer positiv ausgeprägten Querungsintention. So führt ein zusätzliches Rauschen mit einer Standardabweichung von $\sigma = 10°$ beispielsweise zu einer Verschlechterung der TPR auf 62 %, bei einer konstanten TNR von 94 %.

Die in dieser Arbeit verwendeten Labels können, bei einer theoretischen Genauigkeit von 1°, maximal 360 Werte annehmen. Die meisten aus dem Stand der Technik bekannten, bildbasierten Ansätze zur Erkennung der Körperorientierung eines Fußgängers begrenzen die Erkennungsleistung jedoch auf $N_q = 8$ oder $N_q = 4$ Quantisierungsstufen.

Abbildung 6.22b zeigt den Einfluss einer solchen Quantisierung auf die Leistung des kontextbasierten Verfahrens. Wie die steigenden $RMSE$-Werte bei nur leicht niedrigeren TPR- und TNR-Werten zeigen, führt eine Reduzierung der Quantisierungsstufen auf $N_q = 8$ vornehmlich zu einer verschlechterten Abbildung des genauen Unsicherheitslevels, bei einer nur leichten Verschlechterung der Klassifikationsleistung. Eine weitere Reduzierung der Quantisierungsstufen auf $N_q = 4$ führt mit einem RMSE von 0,298 und einer TPR von 67 % schließlich zu einer deutlicheren Verschlechterung der Vorhersagequalität, vor allem in Bezug auf die positiven Beispiele. Im Vergleich zum additiven Sensorrauschen führt eine Einschränkung der kontinuierlichen Körperorientierung auf $N_q = 8$ Klassen bezüglich der Vorhersage des Unsicherheitslevels somit zu einer ähnlichen Verschlechterung der Ergebnisse, wie ein Sensor mit einem Sensorrauschen mit $\sigma = 20°$.



(a)



(b)

Abbildung 6.22: Einfluss der Genauigkeit der erkannten Körperorientierung eines Fußgängers. (a) Additives Rauschen mit Variation der Standardabweichung σ. (b) Variation der Quantisierungsstufen N_q.

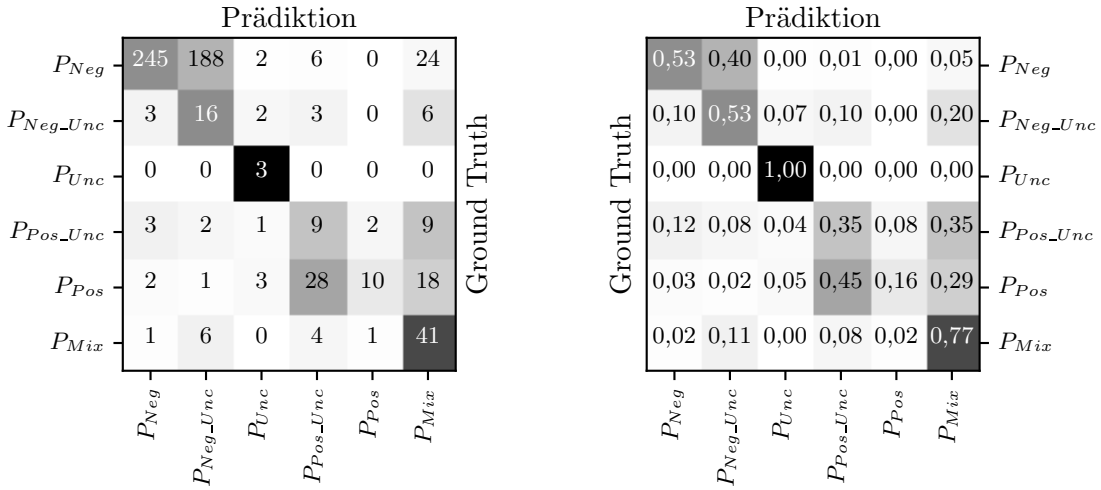


Abbildung 6.23: Objektbasierte Konfusionsmatrix für $\boldsymbol{x}_{Ctxt}$. (a) Absolute Anzahl an Objekten. (b) Relativer Anteil an Objekten in Bezug auf die Ground Truth.

6.4.2 Objektbasierte Ergebnisse

Die Konfusionsmatrizen in Abbildung 6.23 enthalten die Ergebnisse der objektbasierten Evaluation. Analog zu den samplebasierten Ergebnissen zeigt sich, dass für alle Objektkategorien, außer der Kategorie P_{Pos}, alle Samples eines Fußgängers überwiegend in die richtige Klasse prädiziert werden, und Fehlprädiktionen meistens in die Grenzen einer der Nachbarklassen fallen.

Objektkategorie P_{Neg}

Die besten Ergebnisse werden bei Fußgängern ohne Querungsintention erreicht: Bei 53 % der Fußgänger der Objektkategorie P_{Neg} werden alle Samples korrekt prädiziert und zusammen mit der Objektkategorie P_{Neg_Unc} ist bei 93 % der Fußgänger ohne Querungsintention zumindest die Tendenz der Vorhersage korrekt. Abbildung 6.24 zeigt beispielhaft neun dieser Fußgänger einschließlich der aus der Situation resultierenden CMHI, CO und WAO. Die Farbcodierung in dieser sowie den nachfolgenden

Abbildung 6.24: Beispiele der als P_{Neg} oder P_{Neg_Unc} prädizierten Fußgänger der Objektkategorie P_{Neg}, mit den aus der Situation resultierenden CMHI, CO und WAO (von oben nach unten).

Abbildungen entspricht der in Abbildung 6.4 gezeigten, fünfstufigen Quantisierung der Samples.

Es zeigt sich, dass das System in der Lage ist, in unterschiedlichen Situationen Fußgänger der Objektkategorie P_{Neg} richtig oder mindestens tendenziell richtig zu prädizieren. So beinhalten die als P_{Neg} oder P_{Neg_Unc} prädizierte Objekte Fußgänger mit verschiedenen Körperorientierungen und Bewegungsmustern, sowie verschiedenen Zonenzugehörigkeiten und kontextuellen Elementen.

Bemerkenswert ist, dass keiner der Fußgänger der Kategorie P_{Neg} durchgehend als Fußgänger mit einer positiven Querungsintention (P_{Pos}) prädiziert wird. 5 % der P_{Neg} Fußgänger weisen jedoch zumindest teilweise falsch positiv prädizierte Samples auf und werden somit als P_{Mix} kategorisiert.

Abbildung 6.25 enthält für jeden dieser P_{Mix} Fußgänger den Verlauf von Ground Truth (obere Hälfte jeder Zeile) und Prädiktion (untere Hälfte jeder Zeile) über den gesamte Beobachtungszeitraum. Die dünnen, weißen Linien markieren jeweils den bei der binären Klassifikation verwendeten Grenzwert von 0,5; die unterliegende Farbcodierung entspricht auch hier der fünfstufigen Quantisierung der Samples.

Aus der Abbildung geht hervor, dass bei den meisten P_{Mix} Objekte die in den Wertebereich der Klasse C_{-} prädizierten Samples überwiegen. Die Tendenz der Prädiktion ist somit in der Regel korrekt, nicht aber der prädizierte Ausprägungsgrad der Querungsintention. Zudem zeigt sich, dass Falschprädiktionen fast ausschließlich über mehrere Samples hinweg auftreten und es sich nur in seltenen Fällen um einzelne Ausreißer handelt.

Eine genaue Betrachtung der teilweise falsch prädizierten Fußgänger zeigt weiter, dass es sich bei elf der 24 als P_{Mix} prädizierten P_{Neg} Objekte (Objekt Nr. 442–452) um Fußgänger handelt, die mit anderen Personen oder Objekten interagieren und dabei in Richtung der Fahrbahnkante ausgerichtet sind. Wie in Abbildung 6.26 gezeigt, unterhält sich der Fußgänger mit der Objekt Nr. 448 beispielsweise mit einer Gruppe anderer Fußgänger: Obwohl sein Körper in Richtung Straße ausgerichtet ist, lässt sich über die Interaktion mit den anderen Fußgängern darauf schließen, dass er keine Querungsintention hat. Ähnliches gilt für die ebenfalls in Abbildung 6.26 enthaltenen Fußgänger mit den Objekt Nr. 443, 449, 451 und 452: Obwohl sich alle vier Fußgänger innerhalb des Ego-Fahrstreifens befinden, lässt bei den ersten drei Fußgängern die Interaktion mit dem LKW bzw. PKW und beim vierten Fußgänger die Interaktionen mit dem, auf dem Gehweg laufenden zweiten Fußgänger, einen Rückschluss auf die nicht vorhandene Querungsintention zu. Die Sprünge der Prädiktionen während des Beobachtungszeitraums lassen sich jeweils über Änderungen der Körperorientierungen des Fußgängers sowie über Sprünge in der, vom verwendeten Fußgängerdetektionssystem ausgegebene x-y-Position des Fußgängers erklären: Beides führt zu Änderungen im CMHI.

Weitere Situationen, bei denen als P_{Mix} prädizierte P_{Neg} Objekte auftreten, finden sich bei Fußgängern, die an Bushaltestellen warten (Objekt Nr. 453 und 454). Während bei Objekt Nr. 453 die Bushaltestelle in den ersten falsch prädizierten Samples

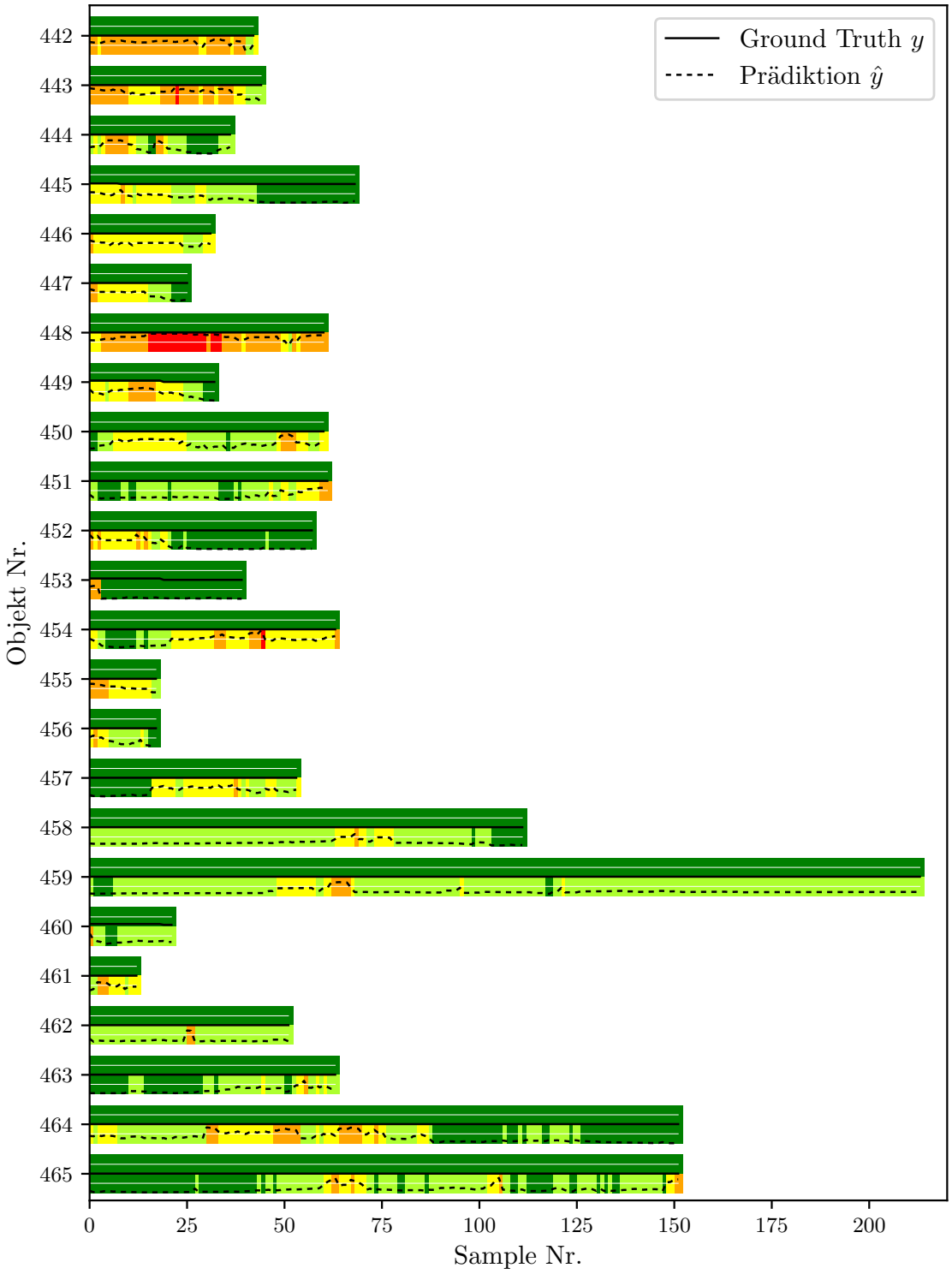


Abbildung 6.25: Verlauf von Ground Truth (obere Hälfte jeder Zeile) und Prädiktion (untere Hälfte jeder Zeile) aller als P_{Mix} prädizierten Fußgänger der Objektkategorie P_{Neg}.

Abbildung 6.26: Beispiele der als P_{Mix} prädizierten Fußgänger der Objektkategorie P_{Neg}, mit den aus der Situation resultierenden CMHI, CO und WAO (von oben nach unten).

nicht im Einzugsbereich der WAO ist, lässt sich bei Objekt Nr. 454 die sprunghafte Fehlprädiktion nicht erklären, denn wie in Abbildung 6.26 zu sehen ist, bildet die WAO die Präsenz der Bushaltestelle hier eindeutig ab, und auch die anderen zwei Fußgänger, die an derselben Bushaltestelle stehen, werden korrekt als P_{Neg} prädiziert.

Teilweise falsch prädizierte Fußgänger finden sich zudem in räumlicher Nähe zu einer Ampelanlage (Objekt Nr. 455–457). Während Objekt Nr. 448 und 449 mit dem Rücken zur Ampel stehen und das System hier erst mit steigendem Beobachtungszeitraum zu der Entscheidung tendiert, dass diese Fußgänger keine Querungsintention haben, läuft Objekt Nr. 457 parallel zur Fahrbahnkante auf eine Fußgängerampel zu (s. Abb. 6.26, Zeile 4). Der zur Ampel gehörende Fußgängerüberweg füllt ab dem Zeitpunkt der beginnenden Fehlprädiktion (Sample Nr. 25) 12,5 % des Einzugsbereichs der WAO aus. Im Gegensatz zum Prädiktionssystem haben die bei der Referenzbildung eingesetzten menschlichen Beobachter jedoch die Möglichkeit, dass der Fußgänger die Ampel zur Querung verwenden möchte, nicht als wahrscheinlich erachtet.

Bei den restlichen acht als P_{Mix} prädizierten P_{Neg} Objekten (Objekt Nr. 458–465) handelt es sich schließlich um Fußgänger, die auf einem Fußgängerweg parallel zur Straße laufen und bei denen leichte Sprünge bei der Fahrstreifen- oder der Fußgängerpositionserkennung zu beobachten sind. Diese werden im CMHI wie schnelle Bewegungen des Fußgängers in Richtung Fahrbahnkante abgebildet, was für das System wiederum ein Indikator für eine Querungsintention ist (s. Abb. 6.26, Zeile 5).

Objektkategorie P_{Pos}

Im Gegensatz zu der Kategorie P_{Neg} werden Fußgänger der Kategorie P_{Pos} nur zu 16 % auch als solche prädiziert. Bei diesen zehn durchgehend korrekt prädizierten Objekten handelt es sich um Fußgänger, die die Straße an einem, im Einzugsbereich der WAO liegenden, Fußgängerüberweg queren wollen (s. Abb. 6.27, Zeile 1).

Mit 45 % wird der größte Teil der P_{Pos} Fußgänger stattdessen als P_{Pos_Unc} prädiziert. Das heißt, statt einer durchgehend eindeutig ausgeprägten Querungsintention wird ein Teil der Samples als tendenziell positiv, aber mit Unsicherheiten behaftet, prädiziert.

Bei knapp einem Drittel dieser als P_{Pos_Unc} prädizierten Objekte handelt es sich um Fußgänger, die sich entweder einem Fußgängerüberweg nähern oder bereits an einem solchen stehen und auf eine sichere Möglichkeit zur Querung der Straße warten (s. Abb. 6.27, Zeile 2). In der erstgenannten Situation steigt die vom System prädiziert Wahrscheinlichkeit einer Querungsintention mit der Annäherung an den Überweg an. Bei der zweitgenannten Situation handelt es sich bei den falsch prädizierten Samples stets um einzelne Ausreißer.

Die restlichen 20 der als P_{Pos_Unc} prädizierten Fußgänger haben alle die Absicht, die Straße ohne Querungshilfe zu überqueren und stehen dazu am Fahrbahnrand, mit einer zur Fahrbahn hin ausgerichteten Körperorientierung (s. Abb. 6.27, Zeile 3). Es zeigt sich, dass die Samples dieser Fußgänger mit durchschnittlich 78 % überwiegend als C_+ prädiziert werden, was darauf hindeutet, dass das System die Querungsintention von Fußgängern, die ohne Querungshilfe die Straße queren wollen, in der Regel um eine Ausprägungsstufe unterschätzt.

Mit jeweils drei Objekten gibt das System bei insgesamt 10 % der P_{Pos} Fußgänger keine oder eine tendenziell falsche Prognose über die Querungsintention der Fußgänger ab. Gemäß Abbildung 6.28 handelt es sich bei der Hälfte dieser Fälle um Fußgänger, die so weit von der Fahrbahnbegrenzung entfernt sind, dass diese noch nicht im CMHI erfasst wird. Bei den anderen drei Fußgängern ist die Fahrbahnbegrenzung deutlich im CMHI abgebildet, wodurch es keine eindeutige Erklärung für die Unterschätzung der ausgeprägten Querungsintention durch das Prädiktionssystem gibt.

29 % der Fußgänger mit Querungsintention weisen neben richtig prädizierten Samples schließlich auch negative Samples auf und werden somit als P_{Mix} kategorisiert. Nach Abbildung 6.29 überwiegt hier der Anteil der tendenziell richtig positiv prädizierten Samples bei elf der 18 Fußgänger. Im Durchschnitt werden pro Objekt 56 % der Samples mit einer richtigen Tendenz prädiziert.

Bei sieben dieser als P_{Mix} prädizierten Objekte (Objekt Nr. 569–575) handelt es sich um Fußgänger, die an einer Fußgängerampel stehen und bei denen Fehler oder Sprünge in der Position des Fußgängers oder des Fußgängerüberwegs dazu führen, dass die im CO abgebildete Fläche des Fußgängerüberwegs für das System nicht ausreicht, um auf eine positiv ausgeprägte Querungsintention zu schließen. Dies gilt insbesondere für den

Abbildung 6.27: Beispiele der als P_{Pos} oder P_{Pos_Unc} prädizierten Fußgänger der Objektkategorie P_{Pos}, mit den aus der Situation resultierenden CMHI, CO und WAO (von oben nach unten).

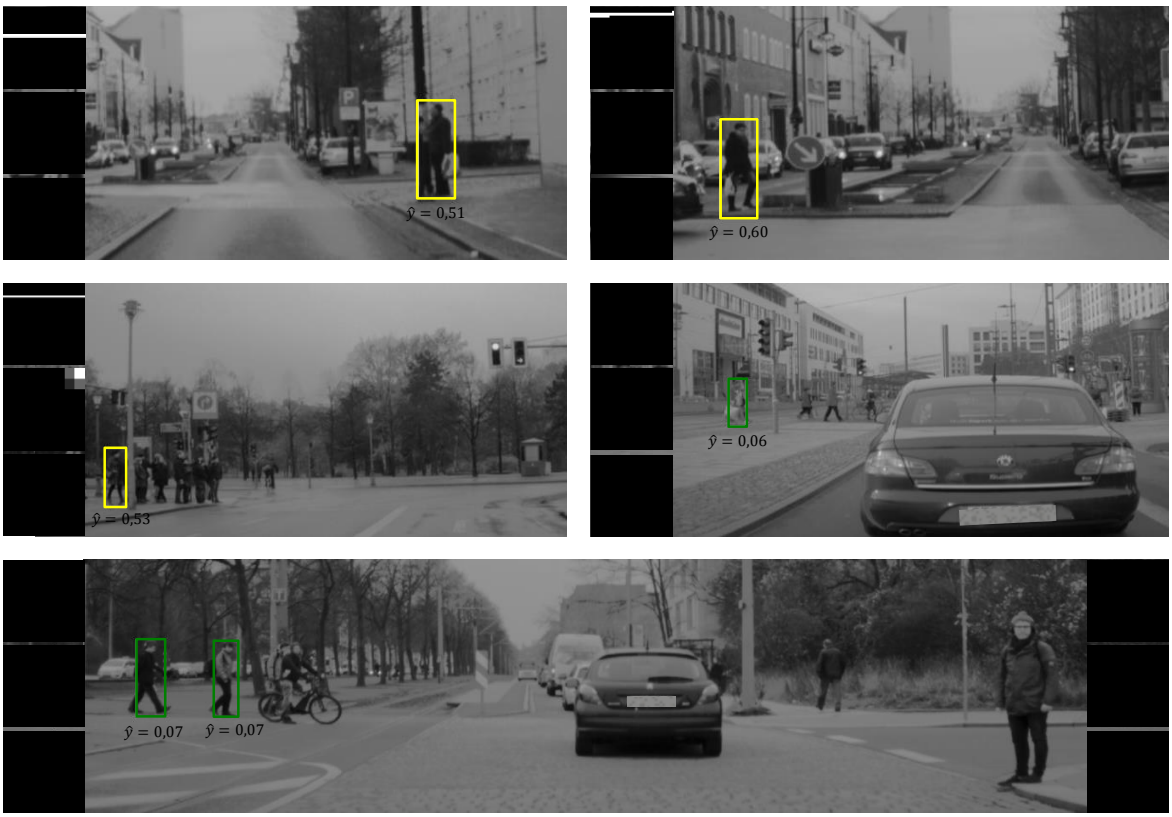


Abbildung 6.28: Die als mindestens P_{Unc} falsch prädizierten Fußgänger der Objektkategorie P_{Pos}, mit den aus der Situation resultierenden CMHI, CO und WAO (von oben nach unten).

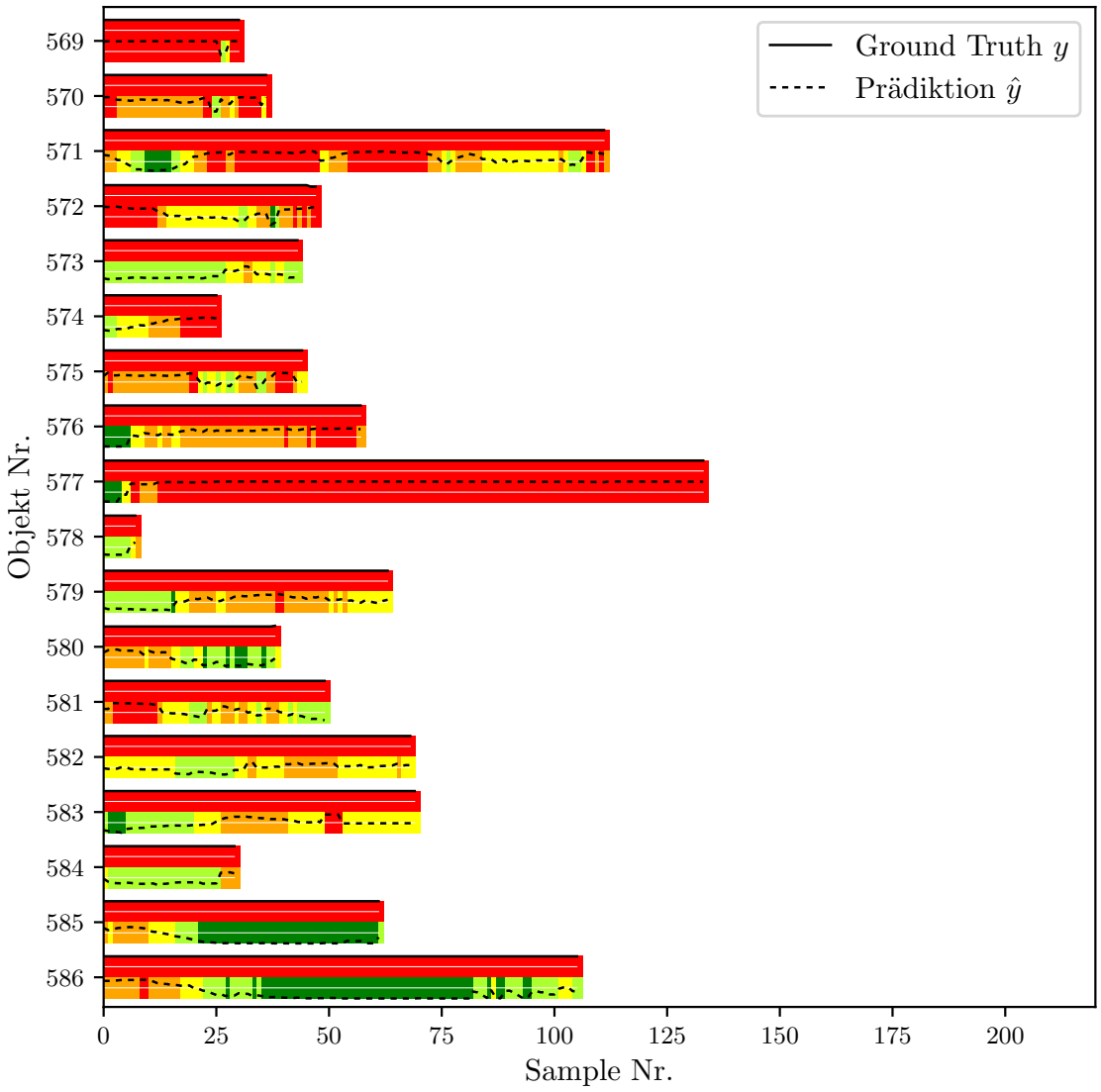


Abbildung 6.29: Verlauf von Ground Truth (obere Hälfte jeder Zeile) und Prädiktion (untere Hälfte jeder Zeile) aller als P_{Mix} prädizierten Fußgänger der Objektkategorie P_{Pos}.

Abbildung 6.30: Beispiele der als P_{Mix} prädizierten Fußgänger der Objektkategorie P_{Pos}, mit den aus der Situation resultierenden CMHI, CO und WAO (von oben nach unten).

Fußgänger mit der Objekt Nr. 573. Wie Abbildung 6.30 zeigt, wartet dieser Fußgänger an einer Fußgängerampel auf grün und ist dabei parallel zur Fahrbahn ausgerichtet, was ohne die Zusatzinformation der Fußgängerampel keine typische Ausrichtung eines Fußgängers mit Querungsintention ist.

Weitere drei Objekte (Objekt Nr. 576–578 in Abb. 6.29) sind am Anfang der Beobachtung noch so weit von der Fahrbahnbegrenzung entfernt, dass diese nicht im CMHI abgebildet wird. Abbildung 6.30 zeigt für den Fußgänger mit der Objekt Nr. 577 jeweils die Samples, an denen die Prädiktion von C_{--} auf $C_?$ und anschließend auf C_{++} wechselt.

Bei den restlichen acht der als P_{Mix} prädizierten P_{Pos} Objekte (Objekt Nr. 579–586 in Abb. 6.29) handelt es sich schließlich um Fußgänger, die schräg auf die Fahrbahn zulaufen, um diese teilweise auch schräg zu queren. Trotz einer genauen Betrachtung der einzelnen Samples dieser acht Fußgänger ist hier kein systematischer Fehler erkennbar: Bei einem Teil der Fußgänger treten die Fehlprädiktionen bereits vor dem Betreten der Straße auf (z.B. Objekt Nr. 582 und Nr. 583) und bei einem anderen Teil erst während der Querung (z.B. Objekt Nr. 586). Abbildung 6.30 enthält in Zeile 3 beispielhafte Samples dieser drei Fußgänger.

Objektkategorie P_{Mix}

Fußgänger mit einer wechselnden Querungsintention werden zu 77 % auch als solche prädiziert. Dies sagt jedoch noch nicht viel über die Qualität der Prädiktion aus. Daher enthält Abbildung 6.31 analog zu den vorherigen Abbildungen den Verlauf von Ground Truth und Prädiktion, zunächst für jeden Fußgänger, dessen Querungsintentionswechsel mit einem Verlassen des Ego-Fahrstreifens verknüpft ist ($Z_{ped,t_{-1}} = Z_{ego} \wedge Z_{ped,t_0} \neq Z_{ego}$). Somit wird hier zunächst der Intentionswechsel von einer positiv ausgeprägten Querungsintention (C_{++}) bis hin zu keiner Querungsintention (C_{--}) betrachtet.

Die Fußgänger mit den Objekt Nr. 599–608 queren den Ego-Fahrstreifen dabei jeweils von links nach rechts, wobei die Objekte mit den Nr. 599–606 einen Zonenwechsel von Z_{ego} zu Z_{swk} vollziehen und die Objekte Nr. 607 und 608 von Z_{ego} zu Z_{mix}. Die

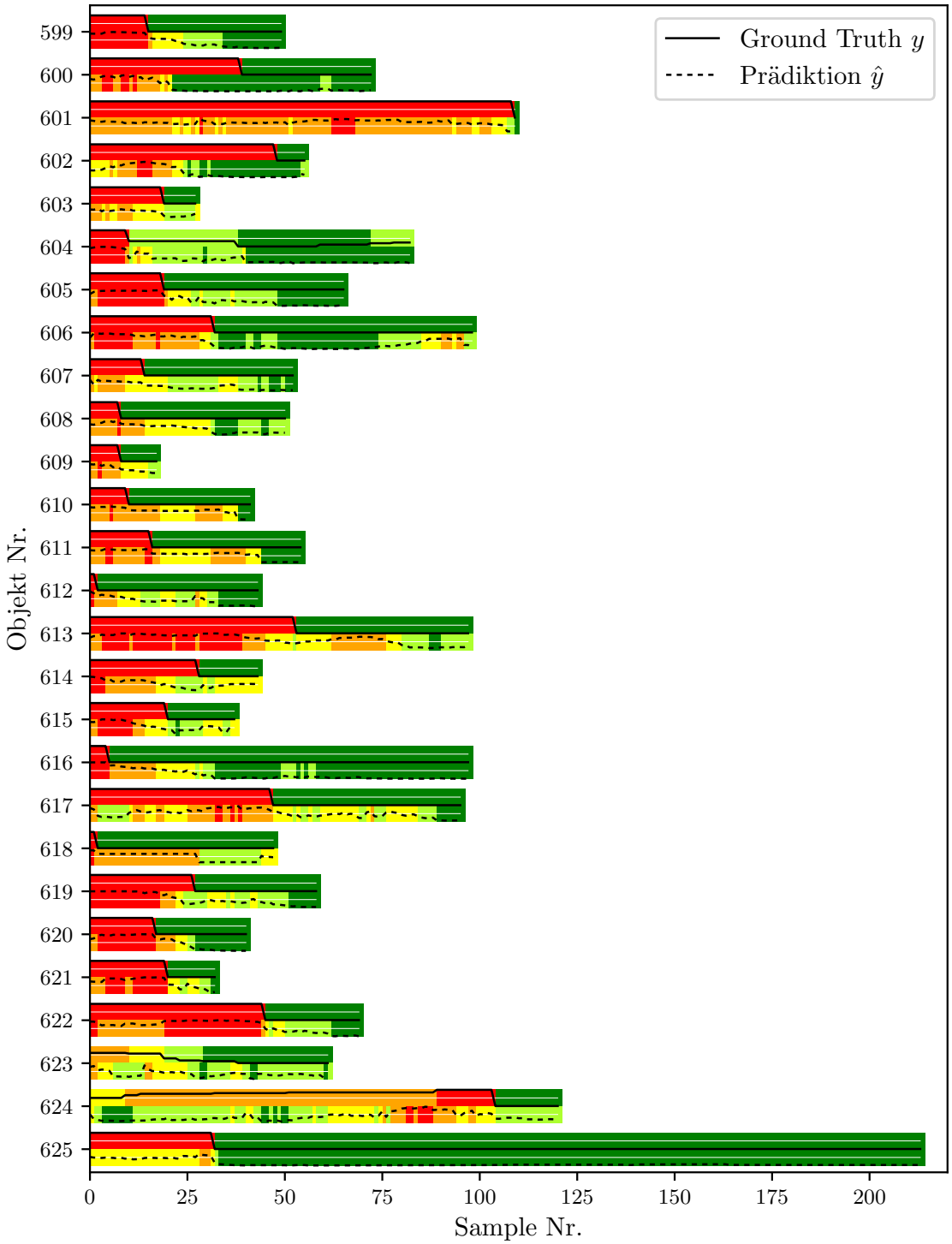


Abbildung 6.31: Verlauf von Ground Truth (obere Hälfte jeder Zeile) und Prädiktion (untere Hälfte jeder Zeile) der richtig als P_{Mix} prädizierten Fußgänger, bei denen der Wechsel der Querungsintention mit dem Verlassen des Ego-Fahrstreifens verknüpft ist.

restlichen 17 Fußgänger queren den Ego-Fahrstreifen entsprechend von rechts nach links und während Objekte Nr. 609–621 jeweils eine mehrspurige Straße queren und somit vom Ego-Fahrstreifen Z_{ego} zu Z_{str} wechseln, queren die Objekte mit den Nr. 622–625 eine Einbahnstraße und wechseln somit von Z_{ego} zu Z_{swk}.

Es zeigt sich, dass das System prinzipiell in der Lage ist, einen auf das Verlassen des Ego-Fahrstreifens zurückzuführenden Wechsel der Querungsintention zu erkennen. Bei allen Fußgänger sinkt der prädizierte Wert nach dem Wechsel tendenziell und bei 63 % reduziert sich zum Zeitpunkt des Intentionswechsels (von C_{++} auf C_{--}) die Prädiktion um mindestens eine Quantisierungsstufe. Hierbei ist jedoch zu beobachten, dass das System die Querungsintention genau nach dem Wechsel oft noch überschätzt und zunächst Werte im Bereich $C_?$ oder C_- prädiziert. Zudem ist zu beobachten, dass das System während der Querung des Ego-Fahrstreifens bei 55 % der Objekte die ausgeprägte Querungsintention unterschätzt und überwiegend Werte im Bereich der Kategorie C_+, oder in einzelnen Fällen, der Kategorie $C_?$ prädiziert. Das deckt sich mit den Beobachtungen bei der samplebasierten Evaluation und den Fußgängern der Objektkategorie P_{Pos}. Eine Abhängigkeit der Prädiktionsqualität von der Querungsrichtung oder dem Zonenwechsel kann nicht beobachtet werden.

Abbildung 6.32 zeigt beispielhaft drei Fußgänger, die den Ego-Fahrstreifen aus unterschiedlichen Richtungen oder in unterschiedliche Zonen verlassen. Die jeweils drei ausgewählten Samples entsprechen den Zeitpunkten, an denen sich die Ground Truth y durch das Verlassen der Zone Z_{ego} zu C_{--} oder die Prädiktion $\hat{y}$ entsprechend ändert. Dabei repräsentiert die Farbe der oberen Hälfte der Bounding Box die Ground Truth und die Farbe der unteren Hälfte die Prädiktion.

Auch bei den fünf als P_{Pos} oder P_{Pos_Unc} prädizierten Objekte der Kategorie P_{Mix} handelt es sich um Fußgänger, deren Querungsintentionswechsel auf das Verlassen des Ego-Fahrstreifen zurückzuführen ist und die kurz nach dem Zonenwechsel den Sichtbereich des Kamerasystems verlassen. Die nach dem Zonenwechsel zur Verfügung stehende Beobachtungszeit reicht dem System hier somit nicht aus, um den Wechsel der Querungsintention zu erkennen.

Bei den weiteren 14 Fußgängern der richtig als Kategorie P_{Mix} prädizierten Objekte liegt hingegen kein Wechsel der Zone Z_{ego} zu einer der anderen Zonen vor. Bei diesen

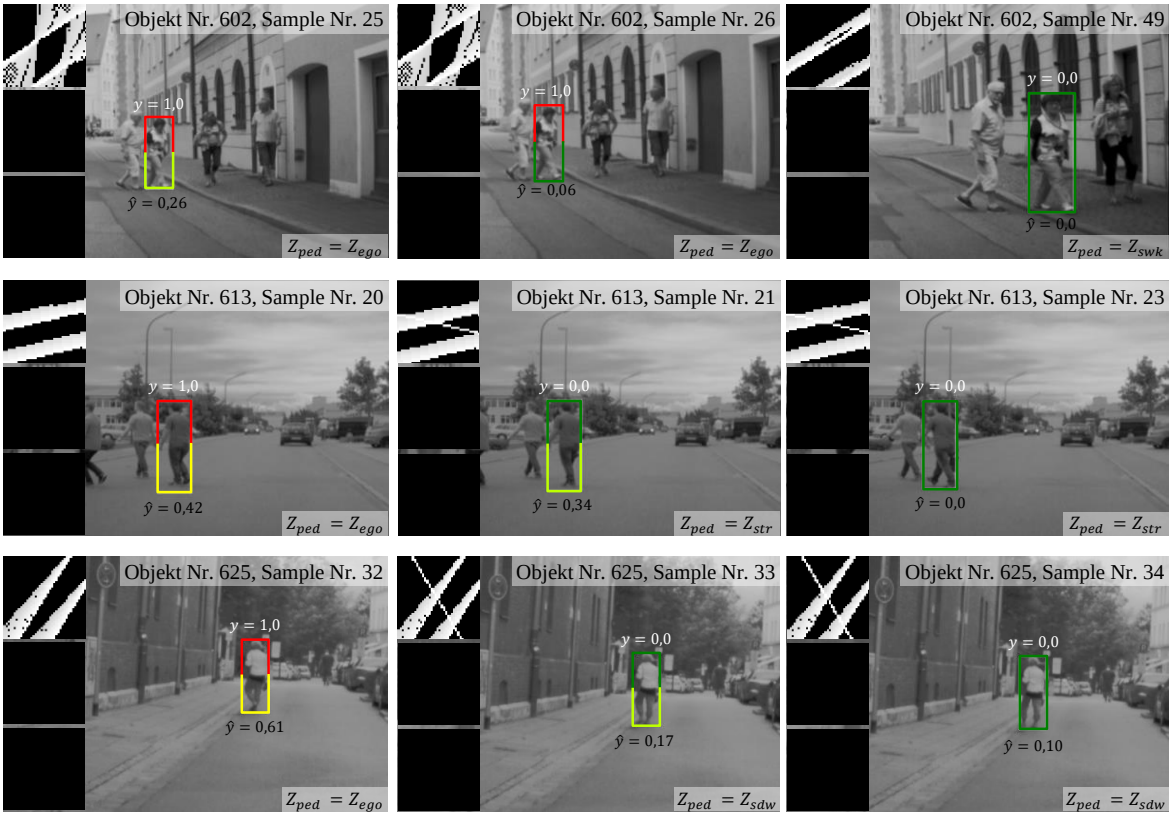


Abbildung 6.32: Beispiele der richtig als P_{Mix} prädizierten Fußgänger, bei denen der Querungsintentionswechsel mit einem Verlassen des Ego-Fahrstreifens verknüpft ist.

Fußgängern resultiert der beobachtete Intentionswechsel vornehmlich aus dem Anzeigen eines Absicherungsverhalten, über das auf eine Querungsabsicht geschlossen wird. Wie Abbildung 6.33 zeigt, wird hier somit hauptsächlich der Intentionswechsel von keiner Querungsintention (C_{--}) zu einer Querungsintention (C_{++}) betrachtet. Ausnahmen bilden die Objekte Nr. 637–639. Bei diesen führt ein Abbrechen des Absicherungsverhaltens dazu, dass die Beobachter im Laufe des Videos ihre Einschätzung ändern und schließlich nicht mehr davon ausgehen, dass der Fußgänger eine Querungsintention hat. In allen Fällen ist der Intentionswechsel mit einer Phase der Unsicherheit verknüpft, bei der die Querungsintention im Wertebereich der Klassen C_{-}, $C_{?}$ oder C_{+} liegt.

Der in Abbildung 6.33 dargestellte Vergleich der Prädiktionswerte mit der Ground Truth zeigt, dass bei allen Objekten dieser Kategorie der auf ein Absicherungsverhalten zurückzuführende Wechsel der Querungsintention prinzipiell erkannt wird. So ist bei allen 14 Objekten zeitnah zum Querungsintentionswechsel auch eine Änderung der Prädiktionswerte um mindestens eine Quantisierungsstufe zu beobachten. Ebenso wird der Verlauf der Unsicherheiten bei den meisten Objekten, analog zur Ground Truth, über die Stufen C_{-}, $C_{?}$ und C_{+} prädiziert. Der Zeitpunkt des prädizierten Intentionswechsels stimmt jedoch bei keinem der 14 Objekte genau mit der Ground Truth überein. Während das System den Beginn einer Unsicherheitsphase, in der Regel eingeleitet durch einen Wechsel von C_{--} zu C_{-}, bei den meistens Objekten (Objekt Nr. 627–628, 633–635, 637–639) früher prädiziert, als ein menschlicher Beobachter, erfolgt die Vorhersage des eigentlichen Intentionswechsels zu C_{+} oder C_{++} in fast allen Fällen verzögert. Zudem finden sich, analog zu den bisherigen Ergebnissen, auch hier nur selten Prädiktionen im Wertebereich der Kategorie C_{++}.

Abbildung 6.34 verdeutlicht am Beispiel des Objekts Nr. 628 den oben beschriebenen, typischen Verlauf von Prädiktion und Ground Truth während des beobachtbaren Absicherungsverhaltens. Abbildung 6.35 zeigt zudem für drei weitere Fußgänger jeweils drei Samples von Zeitpunkten, an denen sich die Ground Truth oder die Prädiktion während des Absicherungsprozesses um eine Kategorie ändert.

Wie die Samples des Objekts Nr. 628 in Abbildung 6.34 und des Objekts Nr. 635 in Abbildung 6.35 zeigen, reagiert das System bereits auf leichte Änderungen der Körper-

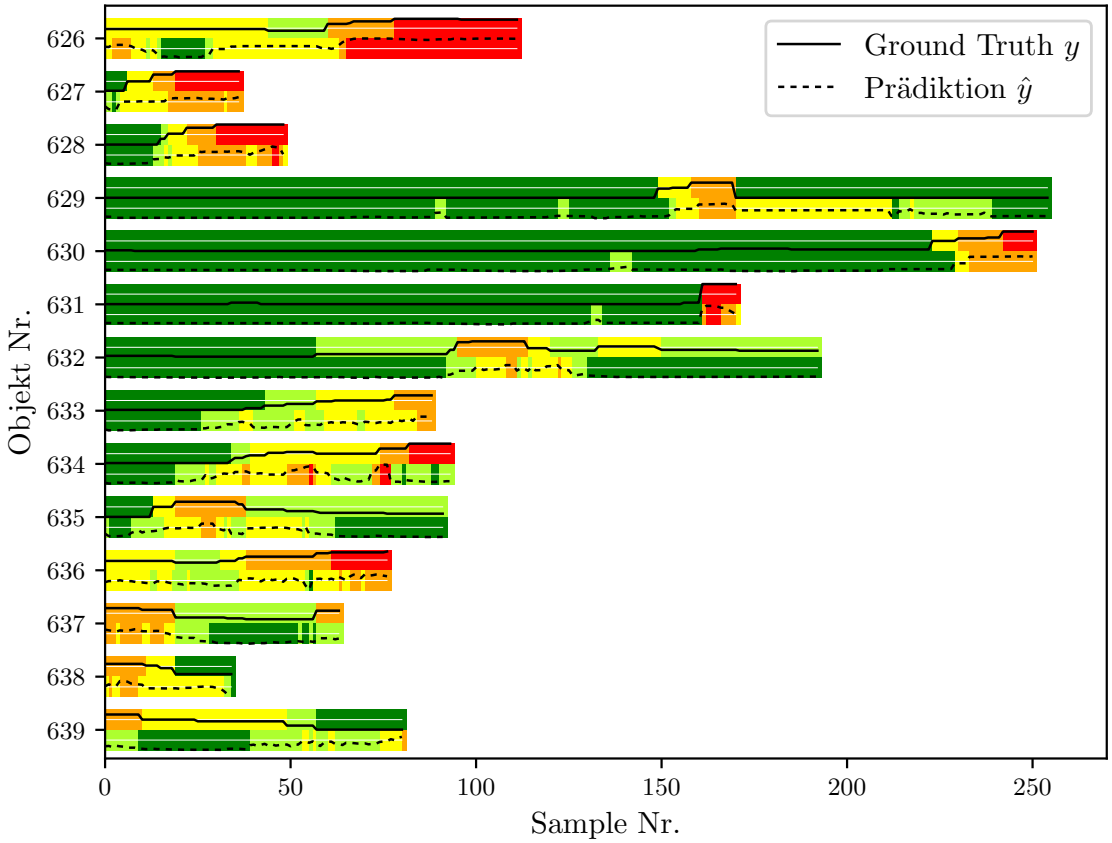


Abbildung 6.33: Verlauf von Ground Truth (obere Hälfte jeder Zeile) und Prädiktion (untere Hälfte jeder Zeile) der richtig als P_{Mix} prädizierten Objekte, bei denen der Querungsintentionswechsel aus dem Anzeigen eines Absicherungsverhaltens resultiert.

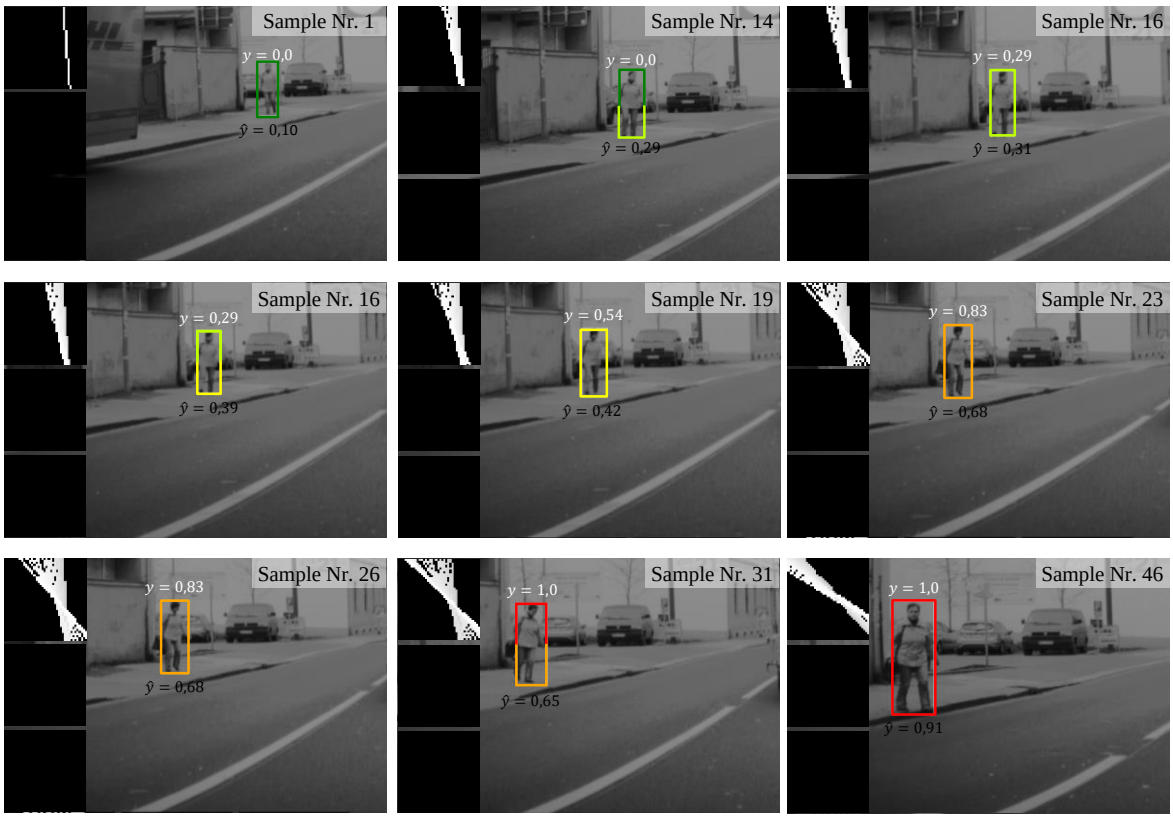


Abbildung 6.34: Samples des Objekts Nr. 628, zur Illustration des typischen Verlaufs der Prädiktion bei Fußgängern der Kategorie P_{Mix}, deren Querungsintentionswechsel auf ein Absicherungsverhalten zurückzuführen ist.

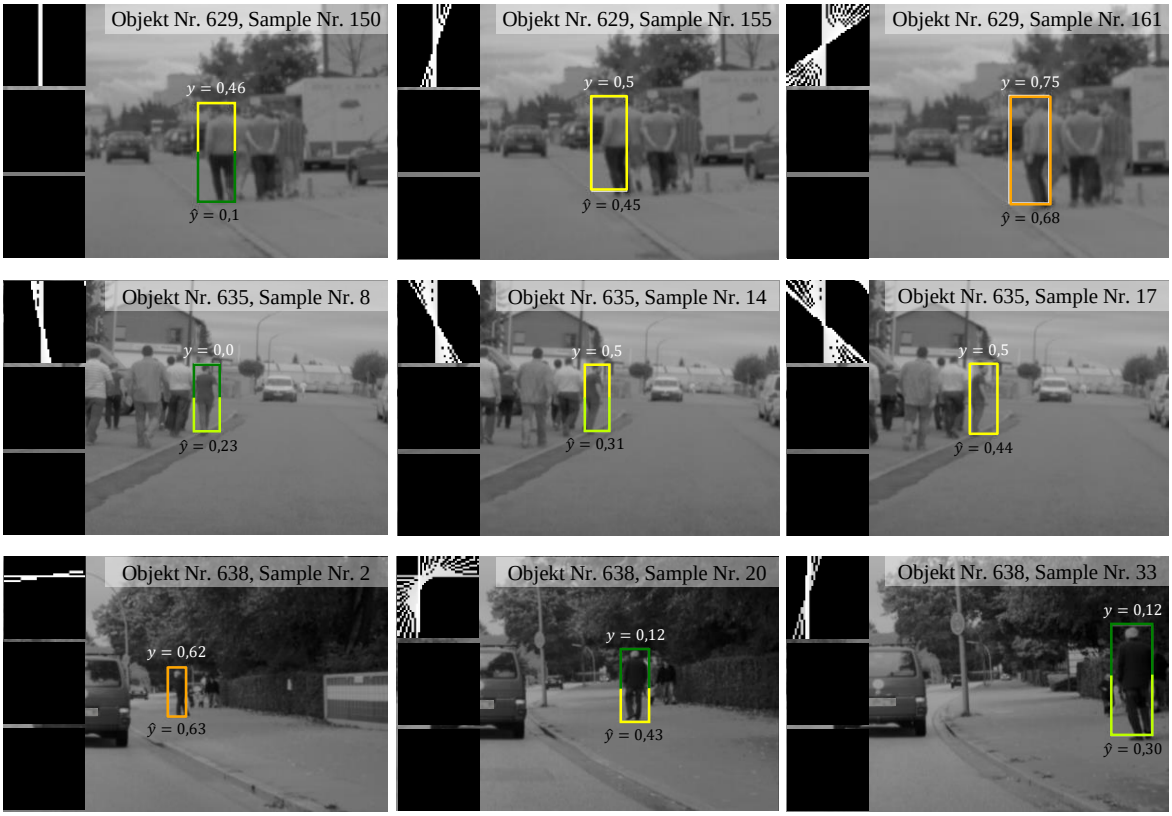


Abbildung 6.35: Beispiele der richtig als P_{Mix} prädizierten Fußgänger, bei denen der Querungsintentionswechsel aus dem Anzeigen eines Absicherungsverhaltens resultiert.

orientierung mit Prädiktionen im Wertebereich der unsicherheitsbehafteten Kategorie C_{-}, was zu dem oben beschriebenen, verfrüht prädizierten Start der Unsicherheitsphase führt.

Bei Fußgängern, die ihr Absicherungsverhalten mit einer Kopfdrehung initiieren, ohne dabei ihre Körperorientierung zu ändern (s. Objekt Nr. 629 in Abbildung 6.35), reagiert das System hingegen langsamer als der menschliche Beobachter, da der kontextbasierte Merkmalsvektor keinerlei Informationen über die Kopforientierung oder -bewegung beinhaltet.

Die oben beschriebenen leichten Änderungen in der Körperorientierung sind nur mit einem sehr genau arbeitenden Erkennungssystem wahrzunehmen. Wie bereits in Abschnitt 6.4.1 erwähnt, haben die in dieser Arbeit verwendeten Labels eine theoretische Genauigkeit von 1°, während die meisten aus dem Stand der Technik bekannten Ansätze die Erkennung der Körperorientierung auf $N_q = 8$ Quantisierungsstufen begrenzen. Abbildung 6.36 zeigt den Einfluss einer solchen Quantisierung auf die Erkennung der Querungsintentionswechsel, die aus dem Anzeigen eines Absicherungsverhalten resultieren. Gemäß der Abbildung führt eine Quantisierung der Körperorientierung auf $N_q = 8$ Quantisierungsstufen dazu, dass das System den Querungsintentionswechsel in fast allen Fällen deutlich später oder gar nicht prädiziert. Dies ist darauf zurückzuführen, dass das auf der Körperorientierung basierende CMHI das Absicherungsverhalten des Fußgängers hier erst ab einer Änderung der Körperorientierung von bis zu 45° abbildet.

Bei den sieben als P_{Neg} oder P_{Neg_Unc} prädizierten Objekte der Kategorie P_{Mix} handelt es sich schließlich um Fußgänger, die ein beobachtbares Absicherungsverhalten kurz vor dem Verlassen des Sichtfelds der Kamera beginnen. Das System prädiziert bei diesen Fußgängern somit maximal den Beginn der Unsicherheitsphase, nicht aber den tatsächlichen Querungsintentionswechsel.

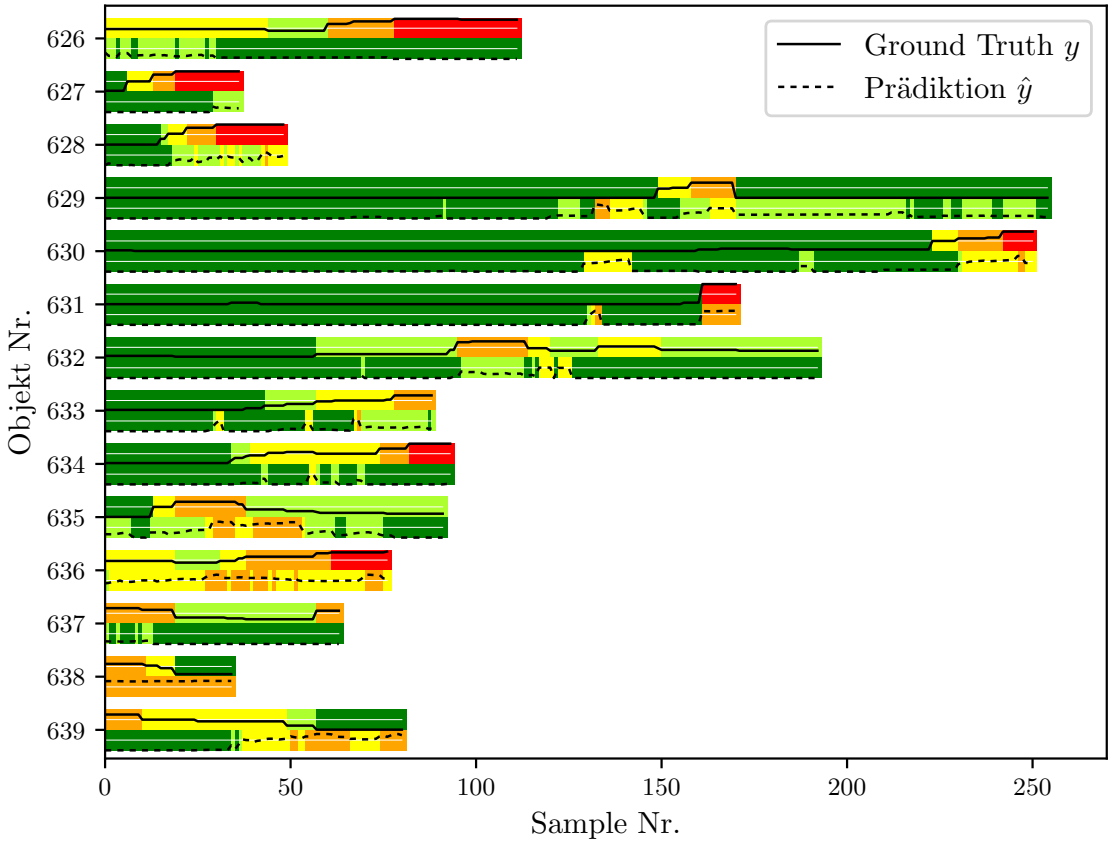


Abbildung 6.36: Verlauf von Ground Truth (obere Hälfte jeder Zeile) und Prädiktion (untere Hälfte jeder Zeile) der in Abbildung 6.33 gezeigten Fußgänger der Objektkategorie P_{Mix} unter Verwendung einer in $N_q = 8$ Stufen quantisierten Körperorientierung.

6.4.3 Diskussion und Bewertung

Die in den Abschnitten 6.4.1 und 6.4.2 vorgestellten Ergebnisse zeigen, dass die Erkennung der Querungsintention von Fußgängern mit dem im Rahmen dieser Arbeit entwickelten, kontextbasierten Ansatz prinzipiell möglich ist. Die in Abschnitt 2.5 aufgestellte Hypothese, dass das kontextuelle Bewegungsverhalten eines Fußgängers ein Indikator für seine Querungsintention ist, wird somit verifiziert. Zudem bestätigen die Ergebnisse, dass der entwickelte Merkmalsvektor in der Lage ist, dieses Bewegungsverhalten abzubilden.

Im Rahmen der samplebasierten Evaluation zeigt sich zunächst, dass die bei der **Parameterevaluation** durchgeführte, sorgfältige Auswahl des optimalen Parametersatzes zu einer deutlichen Leistungssteigerung führt, bei einer gleichzeitig starken Reduktion der Merkmalsvektordimension (s. Tab. 6.14). Das belegt, dass die aufwendige Parameterevaluation als relevanter Teil der Entwicklung des kontextbasierten Merkmalsvektors $\boldsymbol{x}_{Ctxt}$ zu bewerten ist.

Insgesamt hat die Parameterevaluation vor allem zu einer Leistungssteigerung bei der im Datensatz unterrepräsentierten Klasse der Fußgänger mit positiv ausgeprägter Querungsintention geführt (TPR +11 %), während die Klassifikationsleistung bei den Fußgängern ohne Querungsintention annähernd gleich geblieben ist (TNR –1 %). Das bedeutet, dass durch die Verwendung des **gewichteten RMSE** als zu optimierendes Gütemaß, ein Overfitting auf die häufig vorkommenden Samples vermieden wird. Somit ist der $RMSE_w$, in Kombination mit der entwickelten Gewichtungsfunktion, als ein für die Parameterevaluation geeignetes Gütemaß zu bewerten. Dies wird von den Ergebnissen bei der Evaluierung der Anzahl der Zellen der CO und der WAO bestätigt, da sich die Veränderung der Prädiktionsleistung hier fast ausschließlich in den $RMSE_w$-Werten zeigen.

Die sich im Rahmen der Parameterevaluation stetig verändernden optimalen SVR-Parameter C und γ bestätigen schließlich die in Abschnitt 6.2.2 aufgestellte Vermutung, dass eine Änderung der Parameter des Merkmalsvektors zu einer veränderten Datenstruktur führt und somit unterschiedliche C- und γ-Werte optimal sein können. Daher ist die gewählte Methode, für jeden evaluierten Merkmalsvektor ein erneutes

Grid Search-Verfahren durchzuführen, trotz des hohen Aufwands als notwendig zu bewerten.

Wie oben angesprochen, resultiert die durchgeführte Parameterevaluation in einer Reduktion der **Merkmalsvektordimension** auf etwa 10 % der initialen Größe. Dadurch stellt der entwickelte kontextbasierte Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ mit $d = 120$ eine sehr kompakte, speichereffiziente Form zur Beschreibung der Bewegungshistorie des Fußgängers in Bezug zu den relevanten Szenenelementen dar.

Den größten Einfluss auf die Merkmalsdimension hat die Auflösung des normalisierten Szenenausschnitts bei der Erstellung des CMHI. Im Rahmen der Parameterevaluation zeigt sich, dass hier eine Auflösung von 5 px/m ausreicht, um einen optimalen Kompromiss zwischen Informationsmenge und Merkmalsdimension zu erreichen. Hieraus lässt sich ableiten, dass zur Erkennung der Querungsintention von Fußgängern, die absolute **Fußgängerposition** nicht zentimetergenau bekannt sein muss, da vor allem die tendenzielle Bewegung des Fußgängers in Richtung der Fahrbahnkante von Bedeutung ist. Dies ist eine wesentliche Erkenntnis für die Genauigkeitsanforderung an zukünftige Fußgängerdetektionssysteme.

Die durchgeführte Parameterevaluation zeigt weiter, dass bei der Erstellung des CMHI ein **Szenenausschnitt** mit einer **Größe** von 8×8 m zu den besten Ergebnissen führt. Wie die sinkende TPR bei der samplebasierten Evaluation (s. Tab. 6.7) sowie die Analyse der falsch prädizierten P_{Pos} Objekte bei der objektbasierten Evaluation (s. Abschn. 6.4.2, S. 166 ff.) zeigt, werden einige Fußgänger mit Querungsintention hierdurch jedoch falsch prädiziert, da sie weiter als 4 m von der Fahrbahn entfernt sind und die Fahrbahnkante somit nicht im CMHI erfasst wird.

Eine Möglichkeit, einen größeren Bereich zu betrachten, ohne den Merkmalsvektor übermäßig zu vergrößern, ist die Verwendung einer abstandsabhängigen Auflösung. Bei dieser kann der Fernbereich des Fußgängers mit einer niedrigeren Auflösung erfasst werden als der Nahbereich. Da zu vermuten ist, dass die Anforderungen an die Positionsgenauigkeit im Nahbereich höher sind als im Fernbereich, hält sich der dadurch generierte Informationsverlust wahrscheinlich in Grenzen. Die Verwendung einer solchen nicht konstanten Auflösung bei der Erstellung des normalisierten Szenenausschnitts hat jedoch den Nachteil, dass der Abstand der Linien im CMHI nicht mehr di-

rekt die Bewegungsgeschwindigkeit des Fußgängers beschreibt, sondern auch durch die variierende Auflösung beeinflusst wird. Ob die Vorteile einer solchen abstandsabhängigen Auflösung die Nachteile überwiegen, muss daher in einer weiteren umfassenden Evaluation bewertet werden.

Im Rahmen der Parameterevaluation wurde weiter festgestellt, dass die **Bewegungshistorie** des Fußgängers ein relevanter Faktor bei der Bestimmung der Querungsintention ist (s. Tab. 6.8). Selbst kurze Zeitfenster von nur $t = 5$ Samples ($\hat{=}$ 0,28 s) bilden einen deutlichen Mehrwert gegenüber der reinen Darstellung der aktuellen Situation. Dieses Ergebnis steht im Einklang mit den Ergebnissen von Voelz et al. (2015) (s. Abschn. 2.3.3), bei denen mit kontextuellen Merkmalen des aktuellen Zeitschritts in Kombination mit den Merkmalen des Zeitschritts $t - 4$ die besten Ergebnisse erreicht werden. Längere Historien wurden von Voelz et al. jedoch nicht evaluiert. Dabei zeigen die Ergebnisse dieser Arbeit, dass ein kurzes Zeitfenster zwar in Bezug auf die Klassifikation bereits ähnlich gute Ergebnisse wie die als optimal bewertete Historie von $t = 50$ Samples ($\hat{=}$ 2,8 s) erreicht und mit dem entwickelten Ansatz somit bereits nach kurzen Beobachtungszeiten eine erste Aussage über die Querungsintention eines Fußgängers getroffen werden kann. Wie die verringerten $RMSE$-Werte zeigen, wird das genaue Ausprägungsmaß mit einem steigenden Beobachtungszeitraum jedoch exakter prädiziert.

Die schließlich in den Tabellen 6.9 und 6.10 gezeigten Ergebnisse bezüglich der Parametrisierung der **Zellenhistogramme** des CHOG entsprechen dem Erwarteten. Die besten Ergebnisse werden mit einer groben Einteilung des CMHI in 4×4 Zellen erreicht. Dieses unterstreicht, dass das CMHI – anders als bei den üblichen Anwendungsbereichen des HOG, wie der Fußgängerdetektion – wenig feine und lokale Muster aufweist und daher eine Einteilung in wenige Zellenhistogramme ausreicht, um die charakteristischen Strukturen abzubilden. Ebenso bringt die Betrachtung der Vorzeichen der Gradienten durch die Erweiterung des Wertebereichs auf $[0°, 360°)$ keinen Mehrwert.

Mit dem **optimalen Parametersatz** wird schließlich eine durchschnittliche Prädiktionsgenauigkeit von $RMSE = 0{,}237$ erreicht (s. Abschn. 6.4.1, S. 152). Das entspricht annähernd der Genauigkeit, die ein einzelner Beobachter mit der zur Referenz-

bildung eingesetzten fünfstufigen Ratingskala erzielt (s. Abschn. 4.1). Die erreichte durchschnittliche Prädiktionsgenauigkeit ist somit als gut zu bewerten.

Aus der detaillierten Analyse der samplebasierten Ergebnisse geht hervor, dass der durchschnittliche Fehler jedoch nicht homogen über den gesamten Wertebereich verteilt ist. Die positive Ausprägung einer Querungsintention wird dabei tendenziell unterschätzt. So liegt der Labelklassen-spezifische $RMSE$-Wert der Klasse C_{++} mit $RMSE = 0{,}386$ beispielsweise über einem als gut zu bewertenden Ergebnis. Mit einem Verhältnis von $TPR = 0{,}75$ zu $TNR = 0{,}94$ spiegelt sich diese ungleiche Prädiktionsleistung auch in den erreichten Klassifikationsergebnissen wider. Mit einer Genauigkeit von AC = 0,91 ist die Klassifikationsleistung jedoch mit der von Voelz et al. (2015) vergleichbar[1]. Wobei Voelz et al. nur die Querungsintention von Fußgängern an Fußgängerüberwegen betrachten und der in dieser Arbeit entwickelte Ansatz somit mit einer deutlich größeren Situationsvielfalt umzugehen hat.

Dass der entwickelte kontextbasierte Ansatz dieser Herausforderung prinzipiell gewachsen ist, zeigen die in Abschnitt 6.4.1 ab Seite 155 beschriebenen **situationsspezifischen Ergebnisse**. Das System ist intelligent genug, um nicht ausschließlich von der Präsenz eines Szenenelements auf die Querungsintention zu schließen. So werden beispielsweise auch querungswillige Fußgänger, die im Bereich einer Bushaltestelle sind, zu 81 % korrekt detektiert. Ebenso werden Fußgänger, die sich im Umfeld eines Fußgängerüberwegs bewegen, diesen aber nicht nutzen wollen, zu 87 % korrekt klassifiziert. Der bei der situationsspezifischen Evaluierung gezeigte Vergleich der Ergebnisse, die mit bzw. ohne einer Gewichtung der Samples während des SVR-Trainingsprozesses erreicht wurden, betont schließlich die Notwendigkeit des über die Situationskennung realisierten Verfahrens zur Gewichtung der selten auftretenden Situationen.

Neben der oben diskutierten Unterschätzung von Fußgängern mit Querungsintention weist auch die Vorhersage des genauen **Unsicherheitslevels** Verbesserungspotential auf. Nach der in Abschnitt 6.4.1 ab Seite 154 diskutierten und in Abbildung 6.19 dargestellten Konfusionsmatrix wird die Mehrheit der Beispiele zwar tendenziell richtig vorhergesagt, mit Ausnahme der Klasse C_{--} werden jedoch bei keiner der anderen

[1] Klassenspezifische TPR- oder TNR-Werte werden in (Voelz et al., 2015) nicht angegeben.

vier Klassen mehr als 33 % der Samples genau in die richtige Klasse prädiziert. Das lässt vermuten, dass die in dem kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ erfasste Bewegungsinformation des Fußgängers nicht ausreicht, um die bei menschlichen Beobachtern bestehenden Unsicherheiten bezüglich der Querungsintention von Fußgängern abzubilden.

Zudem unterstreichen die Labelklassen-spezifischen Ergebnisse, dass auf Grund der ungleichverteilten Daten und der überdurchschnittlich guten Prädiktion der häufig vertretenen Klasse, eine reine Angabe der durchschnittlichen Prädiktionsgenauigkeit die wahre Leistung des Ansatzes nicht widerspiegelt. Die in dieser Arbeit durchgeführte ausführliche Evaluation ist daher notwendig, um die reale Leistungsfähigkeit des entwickelten Ansatzes bewerten zu können.

Die im Rahmen der **objektbasierten Evaluation** (s. Abschn. 6.4.2) durchgeführte Analyse der mit dem kontextbasierten Ansatz nicht handhabbaren Situationen erlaubt schließlich die folgende Diskussion möglicher Verbesserungsmaßnahmen.

Zunächst ist festzustellen, dass es sich bei den Fehlprädiktionen selten um einzelne Ausreißer handelt. Zwar werden auch nur in einzelnen Fällen alle Samples eines Fußgängers falsch prädiziert, aber auch bei den als P_{Mix} vorhergesagten Fußgängern bilden die falsch prädizierten Samples in der Regel größere zusammenhängende Zeitbereiche (s. Abb. 6.25, Abb. 6.29, Abb. 6.31 und Abb. 6.33). Daher verspricht eine, bei Zeitreihen oft angewendete, Berücksichtigung der Historie zur Glättung der aktuellen Prädiktion, zum Beispiel über einen gleitenden Mittelwert (Smith, 1997), keine verbesserten Ergebnisse.

Vielversprechend ist jedoch die Beobachtung, dass es sich bei dem Großteil der falsch prädizierten P_{Neg} Objekte um Fußgänger handelt, die mit anderen Personen oder Objekten interagieren und dazu in Richtung der Fahrbahnkante ausgerichtet sind (s. Abschn. 6.4.2, S. 163). Durch die Interaktion ist einem menschlichen Beobachter klar, dass die Fußgänger keine Querungsintention haben. Eine Erweiterung des kontextbasierten Merkmalsvektors $\boldsymbol{x}_{Ctxt}$ um die Information, dass im näheren Umfeld des Fußgängers Personen oder Objekte sind, mit denen der Fußgänger potenziell interagiert, kann somit zu einer Verbesserung der Prädiktionsleistung, insbesondere in Bezug auf die FPR, führen. Eine Umsetzung ist in ähnlicher Form wie bei der CO und WAO denkbar.

Eine weitere Verbesserung ist durch den Einsatz einer weniger sprunghaften Detektion der Fahrstreifen- und Fußgängerposition zu erwarten. Da jeder Positionssprung im CMHI wie eine schnelle Bewegung des Fußgängers repräsentiert wird, ist eine möglichst stabile Position des Fußgängers relativ zur Fahrbahnkante für die Erstellung des kontextbasierten Merkmalsvektors von großer Bedeutung. Neben der Verwendung von Kartendaten zur Verbesserung der Fahrstreifenposition (HERE, 2017) ist auch der Einsatz eines bildgestützten Systems denkbar, welches die bei aktuellen Systemen einzeln bestimmten Positionen von Fahrstreifen und Fußgänger, bildbasiert zueinander verifiziert und so ungerechtfertigte Positionssprünge erkennt.

Die Analyse der P_{Pos} Fußgänger zeigt, dass vor allem bei Fußgängern, die ohne Querungshilfe die Straße queren wollen, die Notwendigkeit besteht, die Prädiktionsergebnisse zu verbessern. Während der entwickelte kontextbasierte Ansatz Fußgänger, die im Erfassungsbereich der CO einen Fußgängerüberweg haben, in der Regel als sicher querungswillig prädiziert, werden Fußgänger, die keine Querungshilfe im näheren Umfeld haben, stets mit einem Restzweifel in die Kategorie P_{Pos_Unc} prädiziert. Abhilfe könnte hier die Ergänzung von, in Abschnitt 2.5 bereits diskutierten, posenbasierten Informationen schaffen. Insbesondere die Kopfbewegung des Fußgängers ist hier vielversprechend, da Fußgänger vor allem freie Querungen visuell absichern und sich in dieser Situation die Kopfbewegung eines Fußgängers mit Querungsintention deutlich von denen ohne Querungsintention unterscheidet (s. Abschn. 2.2.2).

Bei Fußgängern mit einer wechselnden Querungsintention (P_{Mix}) sollte der Fokus der Verbesserungsmaßnahmen vor allem auf einer verbesserten Vorhersage des genauen Intentionswechsels liegen, da hier überwiegend temporale und weniger systematische Systemfehler zu beobachten sind (s. Abschn. 6.4.2, S. 172 ff.). Intentionswechsel, die mit einem Verlassen des Ego-Fahrstreifens verknüpft sind, werden, obwohl der kontextbasierte Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ die aktuelle sowie die vorangegangene Zone des Fußgängers Z_{ped} beinhaltet, in der Regel nur verspätet detektiert. Abhilfe könnte hier die oben bereits vorgeschlagene, abstandsabhängige Auflösung des CMHI schaffen. Denn durch eine höhere Auflösung im Nahbereich ist das Überschreiten der Fahrstreifenbegrenzung durch den Fußgänger im CMHI deutlich klarer erkennbar, da der abgebildete Fahrstreifen schneller einen größeren Abstand zur Mitte aufweist.

Auch bei Intentionswechseln, die vornehmlich aus dem Anzeigen eines Absicherungsverhaltens resultieren, besteht eine zeitliche Diskrepanz zwischen der Vorhersage des Systems und den Angaben der menschlichen Beobachter. Auch hier ist die Ergänzung posenbasierter Informationen vielversprechend. So verspricht eine Erweiterung um Informationen über die Kopfbewegung des Fußgängers eine frühere Erkennung des Intentionswechsels. Dies gilt vor allem bei Fußgängern, die ihr Absicherungsverhalten mit einer Kopfdrehung initiieren, ohne dabei ihre Körperorientierung zu ändern.

Zudem liegt die in dieser Arbeit angenommene Genauigkeit der Körperorientierung deutlich über der aktuell im Stand der Technik erreichten Genauigkeit. Auch hier könnten posenbasierte Merkmale, die die Körperbewegung des Fußgängers beschreiben, die mit fehlerbehafteten oder ungenauen Körperorientierungen erreichbaren Ergebnisse noch deutlich verbessern. Nach den objektbasierten Ergebnissen der Kategorie P_{Mix} in Abbildung 6.36 sowie den samplebasierten Ergebnissen in Abbildung 6.22 ist dies für eine praktische Anwendung des entwickelten Ansatzes von großer Bedeutung. So führt die Verwendung eines Systems, welches die Körperorientierung von Fußgängern mit einer realistischen Standardabweichung von $\sigma = 10°$ angibt, beispielsweise zu einer Reduzierung der TPR auf 62 % (-13 %). Die durch eine Quantisierung der Körperorientierung auftretende deutliche Verschlechterung des $RMSE$-Werts, bei nur leicht schlechteren TPR- und TNR-Werten, betont schließlich, dass die Körperorientierung und -bewegung eines Fußgängers ein relevanter Faktor bei der Abbildung des Unsicherheitslevels ist.

6.5 Ergebnisse: Posenbasierte Erweiterung

6.5.1 Samplebasierte Ergebnisse: MCHOG, PAF, HOG

Im Folgenden werden zunächst die Ergebnisse vorgestellt, die durch eine Erweiterung des kontextbasierten Merkmalsvektor $\boldsymbol{x}_{Ctxt}$ über eines der drei, die Körperhaltung und/oder die Körperbewegung beschreibenden Merkmale erreicht werden. Hierbei handelt es sich um die Merkmale:

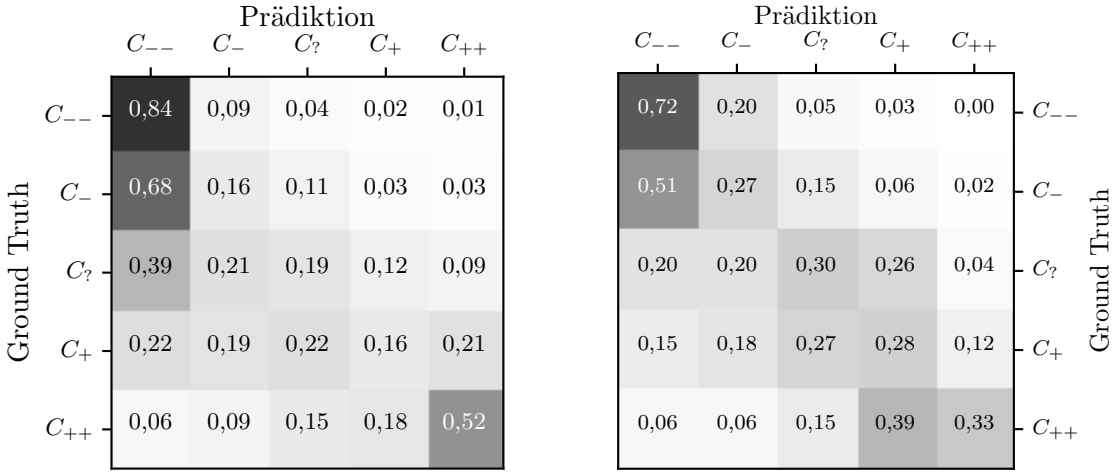
- Motion Contour HOG $\boldsymbol{x}_{Pose,MCHOG}$
- Posture Appearance Feature $\boldsymbol{x}_{Pose,PAF}$
- Histograms of Oriented Gradients $\boldsymbol{x}_{Pose,HOG}$.

Tabelle 6.15 gibt einen Überblick über die samplebasierten Ergebnisse. Es zeigt sich, dass alle drei der hier betrachteten posenbasierten Erweiterungen nur einen geringen Einfluss auf die erzielten Ergebnisse haben, und nur mit einer der drei Kombinationen eine Verbesserung einzelner Teilaspekte erzielt wird.

Mit einem $RMSE$ von 0,231 führt die Ergänzung der MCHOG zu einer leicht besseren Vorhersage des genauen Zielwerts als der rein kontextbasierte Ansatz. Diese Verbesserung geht jedoch auf Kosten der Klassifikationsleistung. Mit einer TPR von 69 % liegt der Anteil der richtig positiv bewerteten Samples 6 % unter dem vom kontextbasierten Ansatz erreichten Wert. Gemäß Abbildung 6.37 ist dies vor allem auf die Prädiktionsleistung bei unsicherheitsbehafteten Samples zurückzuführen. Während mit der

Tabelle 6.15: Samplebasierte Ergebnisse einer posenbasierten Erweiterungen durch die MCHOG, das PAF und die HOG.

Merkmalsvektor	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
$\boldsymbol{x}_{Ctxt}$	120	0,297	0,237	0,75	0,94	0,83	1e5	1e−4
$\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,MCHOG}$	1.344	0,292	0,231	0,69	0,95	0,80	1e7	1e−4
$\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,PAF}$	132	0,301	0,240	0,74	0,94	0,83	1e6	1e−4
$\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,HOG}$	3.036	0,327	0,253	0,68	0,94	0,79	1e7	1e−5



(a) (b)

Abbildung 6.37: Fünfklassige Konfusionsmatrix für (a) $\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,MCHOG}$. (b) $\boldsymbol{x}_{Ctxt}$. Jeweils gezeigt ist der prozentuale Anteil an Samples in Bezug auf die Ground Truth.

Merkmalsvektorkombination $\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,MCHOG}$ die Samples, bei denen keine Unsicherheit über die Ausprägung einer Querungsintention besteht (Labelklasse C_{--} und C_{++}), deutlich häufiger in die korrekte Klasse prädiziert werden, als mit dem einzelnen Merkmalsvektor $\boldsymbol{x}_{Ctxt}$, werden unsicherheitsbehaftete Samples der Klassen C_{-}, $C_{?}$ und C_{+} deutlich ungenauer prädiziert, mit einer Tendenz in Richtung der Klasse C_{--}. Dieses führt, trotz der verbesserten Prädiktion der C_{++} Klasse, zu der reduzierten TPR.

Im Gegensatz zum MCHOG führt eine Erweiterung des kontextbasierten Ansatzes über das PAF oder die HOG zu keinen verbesserten Ergebnissen. Während das niedrigdimensionale PAF nur einen geringen Einfluss auf die Vorhersageleistung hat, führen die hochdimensionalen HOG zu einer deutlichen Verschlechterung der Ergebnisse.

Um zu evaluieren, ob die drei betrachteten posenbasierten Merkmale überhaupt einen Beitrag zur Erkennung der Querungsintention von Fußgängern leisten können, zeigt Tabelle 6.16 die samplebasierten Ergebnisse einer rein posenbasierten Erkennung, ohne Verwendung des kontextbasierten Merkmalsvektors. Es zeigt sich, dass keines der drei posenbasierten Merkmale eine mit dem kontextbasierten Ansatz vergleichbare Leistung

Tabelle 6.16: Samplebasierte Ergebnisse unter Verwendung der rein posenbasierten Merkmale MCHOG, PAF und HOG (ohne den Merkmalsvektor $\boldsymbol{x}_{Ctxt}$).

Merkmalsvektor	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
$\boldsymbol{x}_{Ctxt}$	120	0,290	0,237	0,75	0,94	0,83	1e5	1e−4
$\boldsymbol{x}_{Pose,MCHOG}$	1.224	0,472	0,391	0,39	0,77	0,52	1e4	1e−3
$\boldsymbol{x}_{Pose,PAF}$	12	0,519	0,389	0,38	0,68	0,49	1e5	1e−6
$\boldsymbol{x}_{Pose,HOG}$	2.916	0,436	0,369	0,34	0,85	0,48	1e7	1e−5

erreicht. Der im Vergleich zum $RMSE$-Wert deutlich stärker gestiegene $RMSE_w$-Wert sowie eine TPR von unter 40 % zeigen, dass vor allem die selten auftretenden positiven Beispiele überwiegend falsch prädiziert werden. Aber auch die TNR liegt bei allen drei betrachteten Merkmalen deutlich unter der des kontextbasierten Ansatzes. Nach den ROC-Kurven in Abbildung 6.38 erreichen alle drei posenbasierten Merkmalsvektoren dennoch eine überzufällige Klassifikationsleistung, wobei die HOG den MCHOG und dem PAF deutlich überlegen sind.

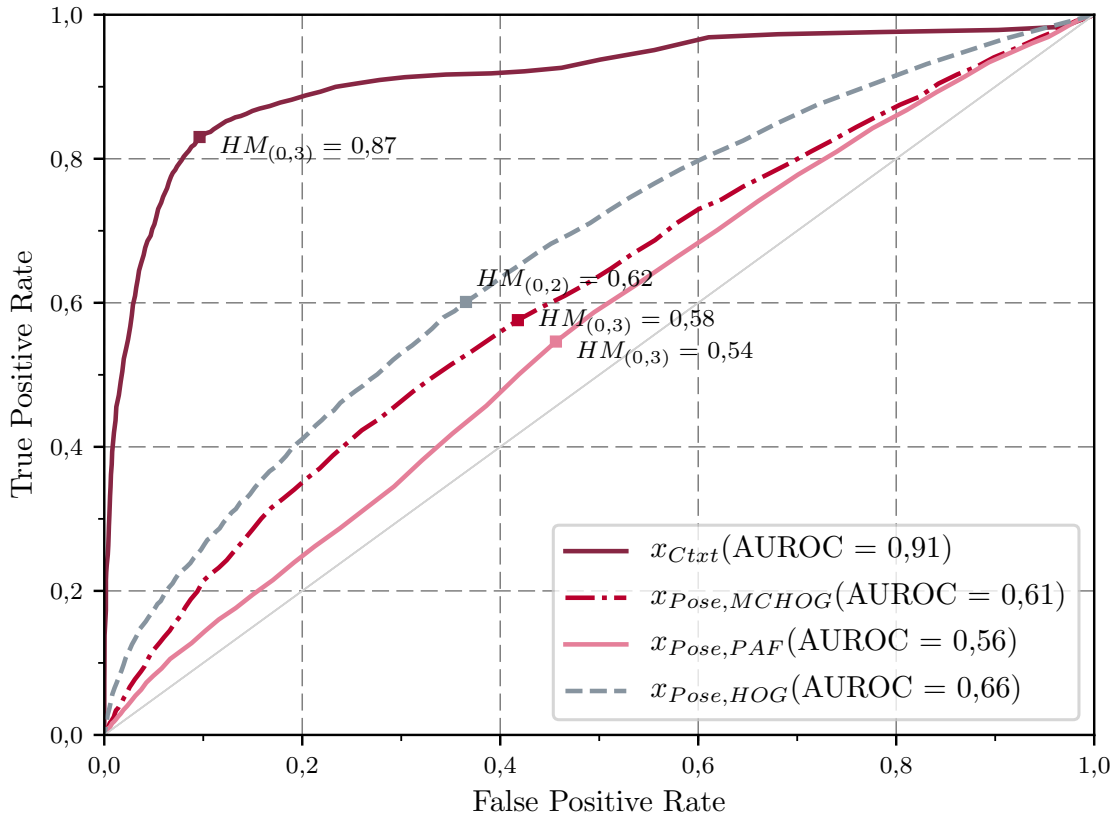


Abbildung 6.38: ROC-Kurven der posenbasierten Merkmale MCHOG, PAF und HOG ohne den kontextbasierten Merkmalsvektor.

6.5.2 Samplebasierte Ergebnisse: OF, LBP

Die im Folgenden vorgestellten Ergebnisse des

- Orientation Features $\boldsymbol{x}_{Pose,OF_{(n,l)}}$

und der

- Local Binary Patterns $\boldsymbol{x}_{Pose,LBP}$

geben Aufschluss über den Beitrag, den die Kopfpose bei der Erkennung der Querungsintention eines Fußgängers leisten kann.

Tabelle 6.17 zeigt zunächst den über das OF evaluierten Mehrwert, den eine theoretisch perfekte Erkennung der Kopforientierung und eine darüber beschriebene Kopfbewegung bieten kann, unter Variation der berücksichtigten Historie.

Die Ergebnisse zeigen, dass ergänzende Informationen über die Kopfbewegung des Fußgängers zu einer Systemverbesserung führen. Mit einem $RMSE_w$ von 0,279 führt dabei eine mit $l = 2$ fein gesampelte Historie von insgesamt $t = 5$ Frames ($\hat{=}$ 0,28 s) zu den besten Ergebnissen. Aus dem nur leicht verbesserten $RMSE$-Wert und der Steigerung der TPR um 4 % lässt sich ableiten, dass die Ergänzung der Kopfbewegung vor allem bei den selten auftretenden, positiven Samples einen Mehrwert bietet. Die Betrachtung längerer Historien führt dabei zu keiner weiteren Verbesserung der Vorhersagequalität.

Die Verwendung von LBPs zur Erfassung der Kopfpose scheint hier jedoch nicht zielführend zu sein. Entsprechend den Ergebnissen in Tabelle 6.18 führt die Erweiterung des kontextbasierten Ansatzes mit LBPs, die dieselbe Historie abbilden, wie das am besten performende OF, zu schlechteren Ergebnissen.

6.5.3 Diskussion und Bewertung

Ziel der Evaluation der posenbasierten Ansätze ist das Finden eines Merkmals, das die Körper- und Kopfpose hinreichend beschreibt, um einen Großteil der in Abschnitt 6.4.3 diskutierten Schwächen des kontextbasierten Ansatzes zu überwinden. Wie die vorgestellten Ergebnisse zeigen, wird dieses Ziel jedoch mit keiner der evaluierten Merkmalskombinationen zufriedenstellend erreicht.

Tabelle 6.17: Samplebasierte Ergebnisse der Merkmalsvektorkombination $\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,OF_{(n,l)}}$ unter Variation der Historienparameter n und l.

n	l	t	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
1	1	1	122	0,292	0,236	0,74	0,94	0,83	1e5	1e−4
2	4	5	124	0,281	0,234	0,77	0,94	0,85	1e5	1e−4
3	2	5	126	0,279	0,233	0,79	0,94	0,86	1e5	1e−4
5	2	9	132	0,288	0,236	0,76	0,94	0,84	1e5	1e−4
13	4	49	146	0,315	0,248	0,72	0,94	0,82	1e6	1e−5
25	2	49	170	0,317	0,250	0,72	0,93	0,82	1e7	1e−6
$\boldsymbol{x}_{Ctxt}$		50	120	0,290	0,237	0,75	0,94	0,83	1e5	1e−4

Tabelle 6.18: Samplebasierte Ergebnisse unter Verwendung von LBPs zur Beschreibung der Kopfpose.

Merkmalsvektor	d	$RMSE_w$	$RMSE$	TPR	TNR	HM	C	γ
$\boldsymbol{x}_{Ctxt}$	120	0,290	0,237	0,75	0,94	0,83	1e5	1e−4
$\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,OF_{(3,2)}}$	126	0,279	0,233	0,79	0,94	0,86	1e5	1e−4
$\boldsymbol{x}_{Ctxt} + \boldsymbol{x}_{Pose,LBP_{(3,2)}}$	1.713	0,307	0,253	0,70	0,96	0,81	1e7	1e−4

Keines der drei die Körperpose beschreibenden Merkmale führt zu einer deutlichen Verbesserung der Prädiktionsergebnisse; eine Erweiterung des kontextbasierten Merkmalsvektors durch das PAF oder die HOG führt, vermutlich auf Grund der Dimensionserhöhung, sogar zu schlechteren Ergebnissen (s. Abschn. 6.5.1). Aus den von den posenbasierten Merkmalen einzeln erreichten überzufälligen Ergebnisse geht jedoch hervor, dass alle drei Merkmalsvektoren einen gewissen Informationsgehalt aufweisen; dieser liegt allerdings deutlich unter dem des kontextbasierten Ansatzes. Unter der Annahme, dass die in Abschnitt 2.5 aufgestellte und in Abschnitt 6.4.3 diskutierte Hypothese – eine Kombination aus kontext- und posenbasierten Informationen bildet die Basis zur Erkennung der Querungsintention von Fußgängern – gültig ist, zeigen die Ergebnisse, dass eine Verbesserung der posenbasierten Beschreibung notwendig ist.

Im Vergleich zu den aus dem Stand der Technik bekannten Originalimplementierungen werden die bildbasierten Merkmale in dieser Arbeit aus deutlich niedriger aufgelösten Daten, mit einer geringeren Bildwiederholrate und ohne die Möglichkeit, den Hintergrund über ein statisches Hintergrundbild oder über Tiefendaten zu filtern, extrahiert. Somit ist eine Erhöhung der Bildqualität vielversprechend, um den Informationsgehalt der posenbasierten Merkmale zu steigern. Genaue Anforderungen, beispielsweise an die Auflösung des Kamerasystems, können ohne weitere, abstandsabhängige Evaluationen nicht abgeleitet werden. Zudem sind die vom verwendeten Detektionssystem bereitgestellten Fußgängerpositionen über die manuell markierten Kopfpositionen zu stabilisieren, um überhaupt in der Lage zu sein, die Körperbewegung des Fußgängers abbilden zu können (s. Abschn. 5.3.2). Dies unterstreicht die Bedeutung eines stabilen Trackings der erkannten Fußgänger für zukünftige Fußgängerdetektionssysteme.

Doch die posenbasierten Merkmale weisen bereits jetzt einen gewissen Informationsgehalt auf. Demnach ist das trainierte System nicht in der Lage, den Mehrwert, den die posenbasierten Informationen in einzelnen Situationen bieten, zu erlernen. Das einfache Zusammenfügen der unterschiedlichen Informationsquellen zu einem Merkmalsvektor scheint hier nicht genügend zielführend zu sein, um das komplexe Zusammenspiel aus Kontext und Pose abzubilden.

Eine Möglichkeit zur Verbesserung ist die Reduktion der vom System zu erfassenden Komplexität. Dies kann, ähnlich wie bei Bonnin et al. (2014a), durch ein System erreicht werden, das nicht generisch für alle möglichen Situationen trainiert wird, sondern aus mehreren Situationsexperten besteht. Hierzu ist hierarchisch zunächst die aktuelle Situation zu erkennen, um anschließend das in der jeweiligen Situation typischerweise gezeigte Verhalten querungswilliger Fußgänger spezifisch zu erlernen. Zur Klassifikation der aktuellen Situation kann beispielsweise der in dieser Arbeit entwickelte kontextbasierte Merkmalsvektor herangezogen werden, um anschließend posenbasierte Prädiktoren anzuwenden, die auf die situationsspezifische Körpersprache des Fußgängers trainiert sind.

Eine weitere Alternativ bieten Ansätze, die im Vergleich zu der in dieser Arbeit eingesetzten SVR noch besser in der Lage sind, sehr komplexe Zusammenhänge zu erlernen. Hier sind vor allem die in Abschnitt 2.5 bereits diskutierten Methoden aus dem Be-

reich des Deep Learning vielversprechend. Die in dem Abschnitt genannten Nachteile der Methoden, wie die zum Training benötigte Datenmenge sowie die Herausforderung, den tiefen neuronalen Netzen Informationen über den Kontext bereitzustellen und trotzdem eine Aussage über die Intention eines einzelnen Fußgängers zu erhalten, bleiben jedoch weiter bestehen.

Die über das OF erreichten und in Abschnitt 6.5.2 vorgestellten Ergebnissen verifizieren schließlich die auf Basis der Literaturrecherche zu beobachtbarem Fußgängerverhalten aufgestellte (s. Abschn. 2.5) und über eigene Beobachtungen verstärkte (s. Abschn. 6.4.3) Hypothese, dass die Kopfbewegung eines Fußgängers ein wichtiger Indikator für seine Querungsintention ist. Die durchgeführte Evaluation zeigt weiter, dass bereits die Betrachtung einer kurzen Bewegungshistorie von 0,28 s zu einer verbesserten Intentionserkennung führt, wobei hier kurze Abtastzeitpunkte zur Beschreibung der Bewegung von großer Bedeutung sind. Das betont wiederum die oben bereits angesprochene Relevanz einer hohen Abtastrate für zukünftige Detektionssysteme.

Die Beschreibung der Kopfpose über LBPs ist, analog zu den bildbasierten Merkmalen zur Beschreibung der Körperpose, jedoch nicht zielführend. Auch hier ist zu vermuten, dass eine Verbesserung der Bildqualität, vor allem der Auflösung, einen Mehrwert bietet.

Kapitel 7

Schlussfolgerung und Ausblick

Mit der Entwicklung eines Systems zur Erkennung der Querungsintention von Fußgängern soll die vorliegende Arbeit einen Beitrag zur Übertragung des menschlichen Situationsbewusstseins auf ein technisches System leisten.

Zur Erreichung dieses Ziels wird zunächst eine ausführliche Auseinandersetzung mit dem Begriff „Intention“ geführt und bestehende handlungstheoretische Modelle zur Intention werden analysiert (s. Abschn. 2.1). Die Ergebnisse bestätigen das Verständnis des Begriffs „Intention“ als einen kognitiven, nicht direkt beobachtbaren Prozess, der Ursache menschlichen Handelns ist. Aus theoretischer Sicht ist somit zu schlussfolgern, dass eine Intentionserkennung zu einer verbesserten Vorhersage der zukünftigen Situation führt. Diese Erkenntnis vermag zu weiterer umfänglicher Entwicklungsarbeit im Bereich Intentionserkennung zu motivieren.

Zudem wird in dieser Arbeit, auf Basis der theoretischen Auseinandersetzung, erstmalig der Begriff „Fußgängerintention“ eindeutig definiert als die *prinzipielle Absicht eines Fußgängers eine Handlung auszuführen.* Im Hinblick auf das automatisierte Fahren lässt sich dieser Begriff noch weiter auf die „Querungsintention“ eines Fußgängers als seine *prinzipielle Absicht, den Fahrstreifen zu queren* einengen. Dadurch lassen sich jetzt die im bisherigen Stand der Technik oft fälschlicherweise als Intentionserkennung bezeichneten Verfahren zur Aktionserkennung oder Trajektorienprädiktion eindeutig abgrenzen. Diese Arbeit leistet somit einen maßgeblichen Beitrag zum theoretischen Diskurs im Bereich der operationalen Verhaltenserkennung. Die hierauf aufbauende

Vorstellung des Stands der Technik zur Erkennung und Vorhersage von Fußgängerverhalten (s. Abschn. 2.3) ist, durch die eindeutige Differenzierung der Begrifflichkeiten und die dadurch erreichte Strukturierung der bekannten Ansätze, zudem die erste dieser Art.

Mit dem Jordan-Modell von Diederichs (2017) wird ein aus dem Stand der Technik bekanntes, beobachtungsbasiertes Modell der Intention vorgestellt, das bisher nur auf Basis von Fahrsimulatorstudien für die Beschreibung von Fahrerintentionen validiert ist. Auf Basis der in Abschnitt 2.5 analysierten Erkenntnisse über das beobachtbare Querungsverhalten von Fußgängern wird in dieser Arbeit gezeigt, dass das Jordan-Modell in einer angepassten Form auch auf die Querungsintention von Fußgängern angewendet werden kann. Damit leistet die vorliegende Arbeit einen weitergehenden Beitrag zur systematischen Beschreibung intendierter Verhaltensweisen. Zudem zeigt die zur Modellvalidierung verwendete Methode eine Möglichkeit auf, die Anwendbarkeit des Jordan-Modells auch für andere Verkehrsteilnehmer zu überprüfen, wie z.B. Fahrradfahrer.

Aus der Analyse des beobachtbaren Querungsverhaltens von Fußgängern geht weiter hervor, dass sich die Querungsintention von Fußgängern in kontext- sowie posenbasierten Verhaltensweisen zeigt. Daraus ist zu folgern, dass diese beiden Informationsquellen auch die konzeptionelle Basis zukünftiger technischer Systeme zur Erkennung der Querungsintention von Fußgängern bilden sollten.

Da die nicht direkte Beobachtbarkeit der Intention, entsprechend der Darstellung in Abschnitt 2.4, neue Ansätze zur Referenzbestimmung fordert, wird in der vorliegenden Arbeit erstmalig das Urteil menschlicher Beobachter als Referenz für die Beurteilung der Fußgängerintention verwendet. In Kombination mit der zur Referenzbildung entwickelten und als reliabel bewerteten, beobachterbasierten Videoannotationsmethode (s. Kapitel 4) stellt dieses eine deutliche Erweiterung des bisherigen Stands der Technik dar.

Zudem wird mit der Analyse der Beobachterurteile bewiesen, dass bei der Erkennung der Querungsintention von Fußgängern Unsicherheiten in der Beobachtung bestehen. Dies stellt eine entscheidende Erkenntnis für die Entwicklung und Verwendung intentionserkennender Systeme dar.

Das schließlich zur Erkennung der Querungsintention entwickelte System ist der erste Ansatz, der die Querungsintention eines Fußgängers situationsunabhängig erkennt und die Intention als unsicherheitsbehafteten Wert vorhersagt. Zur Umsetzung werden merkmalsbasierte Methoden des maschinellen Lernens unter Verwendung der Support Vector Regression eingesetzt. Es wird gezeigt, dass die Erkennung der Querungsintention mit dem in dieser Arbeit entwickelten Ansatz zur Abbildung kontextbasierter Informationen prinzipiell möglich ist. Der entwickelte kontextbasierte Merkmalsvektor stellt damit eine erneute Erweiterung des bisherigen Stands der Technik dar. Zudem bestätigen die Ergebnisse, dass das kontextuelle Bewegungsverhalten von Fußgängern als Indikator für die Querungsintention geeignet ist.

Die mit der optimierten Parametrisierung erreichte Genauigkeit des Systems entspricht annähernd der Genauigkeit, die ein einzelner Beobachter mit der zur Referenzbildung eingesetzten fünfstufigen Ratingkala erzielt. Auf Grund des über den Wertebereich ungleich verteilten, durchschnittlichen Fehlers sind die Ergebnisse jedoch vor allem in Bezug auf die Erkennung einer positiv ausgeprägten Querungsintention sowie bezüglich der Vorhersage des genauen Unsicherheitswerts nicht zufriedenstellend. Der mit dem kontextbasierten Ansatz jetzt erreichte Stand der Technik benötigt daher noch weitere Entwicklungsarbeit. Die ausführliche objektbasierte Evaluation ermöglicht hierzu eine genaue Identifikation der Grenzen rein kontextbasierter Ansätze und leistet somit einen Beitrag zur Feststellung, welche Informationskategorien zur Erkennung der Querungsintention relevant sind.

Eine dieser Kategorien bilden posenbasierte Informationen, die konkret die Körperhaltung und -bewegung sowie die Kopforientierung und -bewegung beschreiben. Die Arbeit belegt, dass eine Erweiterung kontextbasierter Daten mit Information über die Kopfbewegung zu einer besseren Intentionserkennung führt. Mit der Ergänzung posenbasierter Informationen, die über bildbasierte, aus dem Stand der Technik bekannte Merkmale beschrieben werden, konnte jedoch keine relevante Verbesserung der Ergebnisse erreicht werden.

Aus dieser Feststellung lässt sich weiter ableiten, dass neben dem Bedarf an einer gesteigerten Bildqualität, die vom System zu erfassende Komplexität für den gewählten Ansatz zu hoch ist. Als weiteres Ergebnis der vorliegenden Arbeit ergibt sich für zu-

künftige Forschungs- und Entwicklungsprojekte daher die Empfehlung, die Querungsintention von Fußgängern nicht in allen möglichen Situationen mit einem einzigen, generischen Prädiktor zu bestimmen. Ein Lösungsansatz ist die Verwendung mehrerer, für einzelne Situationen spezifisch trainierter, Situationsexperten, die in Kombination mit einer Situationsklassifikation ein Gesamtsystem zur generischen Erkennung der Querungsintention von Fußgängern bilden. So ist beispielsweise ein Prädiktor denkbar, der auf das Verhalten querungswilliger Fußgänger an Zebrastreifen spezialisiert ist und einer, der als Experte für Fußgänger an Bushaltestellen fungiert.

Nach der ausführlichen Analyse des verwendeten Datensatzes (s. Abschn. 6.1) besteht bei im realen Straßenverkehr natürlich aufgenommenen Daten sowohl eine *between-class* als auch eine *within-class imbalance*. Die Entwicklung einzelner Situationsexperten ist auch hier vorteilhaft, da zumindest die *within-class imbalance* beim Training und der Evaluation der einzelnen Prädiktoren nicht zum Tragen kommt.

Die in dieser Arbeit bestehende Unausgewogenheit der Daten sowie die bei der Verwendung von Videodaten stets bestehende Korrelation der einzelnen Samples fordern ein neues Verfahren zur Kreuzvalidierung. Die hierzu erfolgreich entwickelte Kombination der Stratified- und Group-k-fold-Methode (s. Abschn. 6.2.1) kann generisch für alle Problemstellungen, bei denen eine entsprechende Datenbasis vorliegt, eingesetzt werden und stellt somit eine Erweiterung des Stands der Technik dar.

Zusammenfassend weist die vorliegende Arbeit nach, dass die Erkennung der Querungsintention von Fußgängern prinzipiell möglich ist, und dass eine Intentionserkennung die Vorhersage der zukünftigen Situation in der Theorie verbessert. Im nächsten Schritt müssen diese Erkenntnisse „auf die Straße gebracht“ werden. Für weitere Forschungs- und Entwicklungsprojekte bietet sich, parallel zur Weiterentwicklung der Intentionserkennung, die Entwicklung neuer Prädiktionssysteme an, die das zukünftige Verhalten eines Fußgängers unter Verwendung der erkannten Intention langfristig vorhersagen. Der hierbei bestehende Forschungsbedarf liegt vor allem in der Gestaltung von Systemen, die bei der Prädiktion zusätzlich den Einfluss des Verhaltens des Fahrzeugs auf das Verhalten des Fußgängers berücksichtigen, um so schließlich ein erwartungskonformes und kooperatives System für das automatisierte Fahren im städtischen Umfeld zu erhalten.

Abschließend bleibt zu erwähnen, dass die vorgestellten Ergebnisse zunächst nur für den deutschen Verkehrsraum gelten. Zum einen wurden menschliche Beobachter zur Referenzbildung herangezogen, die langjährige Erfahrungen als aktive Teilnehmer im deutschen Straßenverkehr aufweisen, und zum anderen bilden die verwendeten Videodaten das Verkehrsgeschehen in deutschen Innenstädten ab. Die Übertragbarkeit der Erkenntnisse auf anders beschaffene Verkehrsräume, wie beispielsweise der chinesische Verkehrsraum, ist in interkulturellen Studien zu prüfen.

Literaturverzeichnis

Achtziger, A. und P. M. Gollwitzer (2006). „Motivation und Handeln: Einführung und Überblick“. In: *Motivation und Volition im Handlungsverlauf*. Herausgegeben von J. Heckhausen und H. Heckhausen. 3. Auflage. Heidelberg: Springer. Kapitel 11, Seiten 277–302 (siehe Seite 12).

Ahonen, T., A. Hadid und M. Pietikainen (2006). „Face Description with Local Binary Patterns: Application to Face Recognition“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Band 28, Nr. 12, Seiten 2037–2041 (siehe Seite 80).

Ajzen, I. (1985). „From Intentions to Actions: A Theory of Planned Behavior“. In: *Action Control*. Herausgegeben von J. Kuhl und J. Beckmann. Heidelberg: Springer. Kapitel 2, Seiten 11–39 (siehe Seiten 10, 11).

Akbani, R., S. Kwek und N. Japkowicz (2004). „Applying Support Vector Machines to Imbalanced Datasets“. In: *Proceedings of the 15th European Conference on Machine Learning*. Pisa, Italien, Seiten 39–50 (siehe Seite 67).

Altman, D. G. (1991). *Practical Statistics for Medical Research*. London, UK: Chapman und Hall/CRC (siehe Seite 44).

Anscombe, G. E. M. (1957). *Intention*. New York, USA: Cornell University Press (siehe Seite 10).

Anthony, S. (2016). *The Trollable Self-Driving Car*. In: Slate - Future Tense. URL: `http://www.slate.com/articles/technology/future_tense/2016/03/google_self_driving_cars_lack_a_human_s_intuition_for_what_other_drivers.html` (besucht am 01. 11. 2017) (siehe Seite 1).

Bad architects group (2012). *SHARED-SPACE-KONZEPTE in Österreich, der Schweiz und Deutschland*. Salzburger Institut für Raumordnung & Wohnen (SIR) (siehe Seite 5).

Bandyopadhyay, T., C. Z. Jie, D. Hsu, H. Marcelo, A. Jr, D. Rus und E. Frazzoli (2013). „Intention-Aware Pedestrian Avoidance". In: *Experimental Robotics. Springer Tracts in Advanced Robotics*. Herausgegeben von J. Desai, G. Dudek, O. Khatib und V. Kumar. Band 88. Heildberg: Springer, Seiten 963–977 (siehe Seite 29).

Bar-Shalom, Y., R. Li und T. Kirubarajan (2001). *Estimation with Applications To Tracking and Navigation*. Hoboken, NJ, USA: John Wiley & Sons, Inc. (siehe Seite 25).

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, USA: Springer (siehe Seiten 32, 34, 36).

Bonnin, S., T. H. Weisswange, F. Kummert und J. Schmuedderich (2014a). „General Behavior Prediction by a Combination of Scenario-Specific Models". In: *IEEE Transactions on Intelligent Transportation Systems*, Band 15, Nr. 99, Seiten 1478–1488 (siehe Seite 195).

Bonnin, S., T. H. Weisswange, F. Kummert und J. Schmuedderich (2014b). „Pedestrian Crossing Prediction using Multiple Context-based Models". In: *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*. Qingdao, China, Seiten 378–385 (siehe Seiten 25, 26, 30, 35, 36, 39, 53).

Breniere, Y. und M. Do (1986). „When and how does steady state gait movement induced from upright posture begin?" In: *Journal of Biomechanics*, Band 19, Nr. 12, Seiten 1035–1040 (siehe Seite 17).

Brouwer, N., H. Kloeden, R. H. Rasshofer und C. Stiller (2015). „Intelligenter Fußgängerschutz - Bewertung von Umfeldinformationen für die Fußgängerprädiktion". In: *7. Tagung Fahrerassistenz*. München (siehe Seite 30).

Cambon de Lavalette, B., C. Tijus, S. Poitrenaud, C. Leproux, J. Bergeron und J. P. Thouez (2009). „Pedestrian crossing decision-making: A situational and behavioral approach". In: *Safety Science*, Band 47, Nr. 9, Seiten 1248–1253 (siehe Seite 23).

Chandra, S., R. Rastogi und V. R. Das (2013). „Descriptive and Parametric Analysis of Pedestrian Gap Acceptance in Mixed Traffic Conditions“. In: *KSCE Journal of Civil Engineering*, Band 18, Nr. 1, Seiten 284–293 (siehe Seite 16).

Chen, Z. und N. H. C. Yung (2009). „Improved multi-level pedestrian behavior prediction based on matching with classified motion patterns“. In: *2009 IEEE 12th International Conference on Intelligent Transportation Systems (ITSC)*. St. Louis, MO, USA, Seiten 249–254 (siehe Seite 26).

Chu, X., M. Guttenplan und M. Baltes (2004). „Why People Cross Where They Do: The Role of Street Environment“. In: *Transportation Research Record*, Band 1878, Nr. 1, Seiten 3–10 (siehe Seite 18).

Cortes, C. und V. Vapnik (1995). „Support Vector Networks“. In: *Machine Learning*, Band 20, Nr. 3, Seiten 273–297 (siehe Seite 60).

Dalal, N. und B. Triggs (2005). „Histograms of Oriented Gradients for Human Detection“. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. San Diego, CA, USA, Seiten 886–893 (siehe Seiten 33, 77, 110).

Das, S., C. F. Manski und M. D. Manuszak (2005). „Walk or wait? An Empirical Analysis of street Crossing Decisions“. In: *Journal of Applied Econometrics*, Band 20, Nr. 4, Seiten 529–548 (siehe Seite 22).

Davidson, D. (1963). „Actions, Reasons and Causes“. In: *The Journal of Philosophy*, Band 60, Nr. 23, Seiten 685–700 (siehe Seite 10).

Diederichs, F. (2017). *Entwicklung von verhaltensbasierten Verfahren zur Erkennung von Fahrerintention für die Prädiktion von Fahrmanövern*. Dissertation. Universität Stuttgart. Schriftenreihe zu Arbeitswissenschaft und Technologiemanagement, 36, Fraunhofer Verlag (siehe Seiten 10–15, 39, 47, 49, 198).

Diederichs, F. und G. Pöhler (2014). „Driving Maneuver Prediction Based on Driver Behavior Observation“. In: *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (AHFE)*. Kraków, Polen, Seiten 68–73 (siehe Seiten 44, 96).

Dipietro, C. M. und L. E. King (1970). „Pedestrian Gap-Acceptance“. In: *Highway Research Record*, Nr. 308, Seiten 80–91 (siehe Seite 22).

Dollár, P., C. Wojek, B. Schiele und P. Perona (2012). „Pedestrian Detection: An Evaluation of the State of the Art“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Band 34, Nr. 4, Seiten 743–61 (siehe Seiten 76, 123).

Döring, N. und J. Bortz (2016a). „Datenerhebung“. In: *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Herausgegeben von N. Döring und J. Bortz. Berlin Heidelberg: Springer. Kapitel 10, Seiten 321–577 (siehe Seite 40).

Döring, N. und J. Bortz (2016b). „Operationalisierung“. In: *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Herausgegeben von N. Döring und J. Bortz. Berlin Heidelberg: Springer. Kapitel 8, Seiten 221–289 (siehe Seite 41).

Dougherty, E. R. und R. A. Lotufo (2003). *Hands-On Morphological Image Processing*. Bellingham, WA, USA: Spie Press (siehe Seite 114).

Drucker, H., C. J. C. Burges, L. Kaufman, A. Smola und V. Vapnik (1996). „Support Vector Regression Machines“. In: *Advances in Neural Information Processing Systems (NIPS 1996)*, Band 9, Seiten 155–161 (siehe Seite 65).

Eberhardt, W. und G. Himbert (1977). „Bewegungsgeschwindigkeiten - Versuchsergebnisse nichtmotorisierter Verkehrsteilnehmer“. In: *Der Verkehrsunfall*, Band 15, Nr. 4, Seiten 79–84 (siehe Seite 16).

Egan, C. D., A. Willis, H. Ness und S. Stradling (2008). *Visual gaze behaviour of children and adult pedestrians at signalized and unsignalized road crossings*. Technischer Bericht. Edinburgh: Napier University (siehe Seite 20).

Elektrobit (2017). *EB Assist ADTF*. URL: `https://www.elektrobit.com/products/eb-assist/adtf/` (besucht am 01. 11. 2017) (siehe Seite 101).

Endsley, M. R. (1995). „Toward a Theory of Situation Awareness in Dynamic Systems“. In: *Human Factors*, Band 37, Nr. 1, Seiten 32–64 (siehe Seite 2).

Engel, J. (2010). *Anwendungsorientierte Mathematik: Von Daten zur Funktion*. Berlin: Springer (siehe Seite 74).

Fatemi, M., L. Hammarstrand, L. Svensson und A. F. Garcia-Fernandez (2014). „Road geometry estimation using a precise clothoid road model and observations of moving

vehicles“. In: *2014 17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. Qingdao, China, Seiten 238–244 (siehe Seite 130).

FGSV (2010). *Richtlinien für Lichtsignalanlagen*. Köln: Forschungsgesellschaft für Straßen- und Verkehrswesen e. V. (siehe Seite 17).

FGSV (2015). *Handbuch für die Bemessung von Straßenverkehrsanlagen*. Köln: Forschungsgesellschaft für Straßen- und Verkehrswesen e. V. (siehe Seite 17).

Fishbein, M. und I. Ajzen (1975). *Belief, attitude, intention, and behavior*. Reading, MA, USA: Addison-Wesley, Seite 480 (siehe Seite 10).

Fleiss, J. L., B. Levin und M. C. Paik (1973). *Statistical Methods for Rates and Proportions*. New York, USA: Wiley (siehe Seite 44).

Flohr, F., M. Dumitru-Guzu, J. F. P. Kooij und D. M. Gavrila (2014). „Joint probabilistic pedestrian head and body orientation estimation“. In: *2014 IEEE Intelligent Vehicles Symposium (IV)*. Dearborn, MI, USA, Seiten 617–622 (siehe Seite 125).

Fritsch, J., T. Kuhnl und F. Kummert (2014). „Monocular Road Terrain Detection by Combining Visual and Spatial Information“. In: *IEEE Transactions on Intelligent Transportation Systems*, Band 15, Nr. 4, Seiten 1586–1596 (siehe Seite 132).

Furuhashi, R. und K. Yamada (2011). „Estimation of Street Crossing Intention from a Pedestrian's Posture on a Sidewalk Using Multiple Image Frames“. In: *2011 First Asian Conference on Pattern Recognition (ACPR)*. Beijing, China, Seiten 17–21 (siehe Seiten 34, 53, 110, 116).

Galanis, A. und E. Nikolaos (2012). „Pedestrian Crossing Behaviour in Signalized Crossings in Middle Size Cities in Greece“. In: *17th International Conference on Urban Planning, Regional Development and Information Society (REAL CORP 2012)*. Schwechat, Österreich, Seiten 563–570 (siehe Seiten 16, 23).

Geruschat, D. R., S. E. Hassan und K. A. Turano (2003). „Gaze Behavior while Crossing Complex Intersections“. In: *Optometry and Vision Science*, Band 80, Nr. 7, Seiten 515–528 (siehe Seiten 16, 21, 23).

Goldhammer, M., M. Gerhard, S. Zernetsch, K. Doll und U. Brunsmann (2013). „Early Prediction of a Pedestrian's Trajectory at Intersections“. In: *2013 16th International IEEE Conference on Intelligent Transportation Systems*. Den Haag, Niederlande, Seiten 237–242 (siehe Seite 25).

Goodfellow, I., Y. Bengio und A. Courville (2016). *Deep Learning*. Cambridge, MA, USA: The MIT Press, Seite 800 (siehe Seite 52).

Grayson, G. B. (1975). *Observations of pedestrian behaviour at four sites*. Technischer Bericht. Crowthorne, UK: Transport und Road Research Laboratory, LR 670 (siehe Seite 20).

Guennouni, S., A. Ahaitouf und A. Mansouri (2015). „A Comparative Study of Multiple Object Detection Using Haar-Like Feature Selection and Local Binary Patterns in Several Platforms". In: *Modelling and Simulation in Engineering*, Band 2015, Seiten 1–8 (siehe Seite 80).

Guyon, I., J. Weston, S. Barnhill und V. Vapnik (2002). „Gene Selection for Cancer Classification using Support Vector Machines". In: *Journal of Machine Learning Research*, Band 46, Nr. 1-3, Seiten 389–422 (siehe Seite 37).

Hagen, K., C. Schulze und B. Schlag (2010). „Verkehrssicherheit von schwächeren Verkehrsteilnehmern im Zusammenhang mit dem geringen Geräuschniveau von Fahrzeugen mit alternativen Antrieben". In: *FAT-Schrifenreihe, Forschungsvereinigung Automobiltechnik E.V.* Band 245 (siehe Seiten 17–21, 24).

Hamaoka, H., T. Hagiwara und M. Tada (2013). „A Study on the Behavior of Pedestrians when Confirming Approach of Right/Left-Turning Vehicle while Crossing a Crosswalk". In: *Proceedings of the Eastern Asia Society for Transportation Studies*, Band 9 (siehe Seite 20).

Hamed, M. M. (2001). „Analysis of pedestrians' behavior at pedestrian crossings". In: *Safety Science*, Band 38, Seiten 63–82 (siehe Seite 22).

Hariyono, J. und K.-H. Jo (2015). „Pedestrian Action Recognition using Motion Type Classification". In: *2015 IEEE 2nd International Conference on Cybernetics (CYBCONF)*. Gdynia, Polen, Seiten 129–132 (siehe Seiten 32, 39).

Haselhoff, A. und A. Kummert (2010). „On visual crosswalk detection for driver assistance systems". In: *2010 IEEE Intelligent Vehicle Symposium (IV)*. San Diego, CA, USA, Seiten 883–888 (siehe Seite 132).

Hastie, T., R. Tibshirani und J. Friedman (2009). *The Elements of Statistical Learning*. 2. Auflage. New York, NY, USA: Springer (siehe Seite 61).

Hatfield, J. und S. Murphy (2007). „The effects of mobile phone use on pedestrian crossing behaviour at signalised and unsignalised intersections“. In: *Accident Analysis and Prevention*, Band 39, Nr. 1, Seiten 197–205 (siehe Seite 21).

He, H. und Y. Ma (2013). *Imbalanced Learning: Foundations, Algorithms, and Applications*. Hoboken, NJ, USA: John Wiley & Sons, Inc. (siehe Seiten 66, 67).

Heckhausen, H. und P. M. Gollwitzer (1987). „Thought Contents and Cognitive Functioning in Motivational versus Volitional States of Mind“. In: *Motivation and Emotion*, Band 11, Nr. 2, Seiten 101–120 (siehe Seite 11).

HERE (2017). *Maps for developers*. URL: `https://developer.here.com/` (besucht am 01. 11. 2017) (siehe Seiten 130, 187).

Himanen, V. und R. Kulmala (1988). „An Application of Logit Models in Analysing the Behaviour of Pedestrians and Car Drivers on Pedestrian Crossings“. In: *Accident Analysis and Prevention*, Band 20, Nr. 3, Seiten 187–197 (siehe Seite 23).

Holland, C. und R. Hill (2007). „The effect of age, gender and driver status on pedestrians' intentions to cross the road in risky situations“. In: *Accident Analysis and Prevention*, Band 39, Nr. 2, Seiten 224–237 (siehe Seite 22).

Hoogendoorn, S. P. und P. H. L. Bovy (2004). „Pedestrian route-choice and activity scheduling theory and models“. In: *Transportation Research Part B: Methodological*, Band 38, Seiten 169–190 (siehe Seite 15).

Hyndman, R. J. und G. Athanasopouls (2013). *Forecasting: principles and practice*. Melbourne, Australien: OTexts (siehe Seite 75).

Jain, A., A. Gupta und R. Rastogi (2014). „Pedestrian Crossing Behavior Analysis At Intersections“. In: *International Journal for Traffic and Transport Engineering*, Band 4, Nr. 1, Seiten 103–116 (siehe Seite 23).

James, G., D. Witten, T. Hastie und R. Tibshirani (2013). *An Introduction to Statistical Learning*. New York, USA: Springer (siehe Seiten 60, 63).

Japkowicz (2013). „Assessment Metrics for Imbalanced Learning“. In: *Imbalanced Learning: Foundations, Algorithms, and Applications*. Herausgegeben von H. He und Y. Ma. Hoboken, NJ, USA: John Wiley & Sons, Inc. Kapitel 8, Seiten 187–206 (siehe Seiten 69–73).

Jeni, L. A., J. F. Cohn und F. De La Torre (2013). „Facing Imbalanced Data - Recommendations for the Use of Performance Metrics“. In: *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII)*. Geneva, Schweiz, Seiten 245–251 (siehe Seiten 70–72).

Kadali, B. R. und P. Vedagiri (2013). „Modelling pedestrian road crossing behaviour under mixed traffic condition“. In: *European Transport*, Nr. 55, Seiten 1–17 (siehe Seite 22).

Kalman, R. E. (1960). „A New Approach to Linear Filtering and Prediction Problems“. In: *Journal of Basic Engineering*, Band 82, Nr. 1, Seiten 35–45 (siehe Seite 25).

Karasev, V. und S. Soatto (2016). „Intent-Aware Long-Term Prediction of Pedestrian Motion“. In: *2016 IEEE International Conference on Robotics and Automation (ICRA)*. Stockholm, Schweden (siehe Seite 28).

Keller, C. G. und D. M. Gavrila (2014). „Will the Pedestrian Cross? A Study on Pedestrian Path Prediction“. In: *IEEE Transactions on Intelligent Transportation Systems*, Band 15, Nr. 2, Seiten 494–506 (siehe Seiten 31, 52).

Keller, C. G., C. Hermes und D. M. Gavrila (2011). „Will the pedestrian cross? Probabilistic Path Prediction Based on Learned Motion Features“. In: *DAGM Symposium*. Frankfurt, Seiten 1–10 (siehe Seiten 31, 32, 39–41, 52).

Kitani, K. M., B. D. Ziebart, J. A. Bagnell und M. Hebert (2012). „Activity Forecasting“. In: *Proceedings of European Conference on Computer Vision (ECCV)*. Florenz, Italien, Seiten 201–214 (siehe Seiten 27, 28).

Kloeden, H., N. Brouwer, S. Ries und R. H. Rasshofer (2014). „Potenzial der Kopfposenerkennung zur Absichtsvorhersage von Fußgängern im urbanen Verkehr“. In: *Workshop Fahrerassistenzsysteme*. Walting, Seiten 67–78 (siehe Seiten 4, 20, 25, 30, 35, 50, 116).

Kobiela, F. (2011). *Fahrerintentionserkennung für autonome Notbremssysteme*. Dissertation. Technische Universität Dresden. Verlag für Sozialwissenschaften (siehe Seite 89).

Köhler, S., M. Goldhammer, S. Bauer, K. Doll, U. Brunsmann und K. Dietmayer (2012). „Early Detection of the Pedestrian's Intention to Cross the Street“. In: *2012 15th International IEEE Conference on Intelligent Transportation Systems*

(ITSC). Anchorage, AK, USA, Seiten 1759–1764 (siehe Seiten 32–34, 38, 50, 52, 106, 107, 110, 114).

Köhler, S., B. Schreiner, S. Ronalter, K. Doll, U. Brunsmann und K. Zindler (2013). „Autonomous Evasive Maneuvers Triggered by Infrastructure-Based Detection of Pedestrian Intentions“. In: *2013 IEEE Intelligent Vehicles Symposium (IV)*. Gold Coast, Australien, Seiten 519–526 (siehe Seiten 33, 34, 114, 116).

Kooij, J. F. P., N. Schneider, F. Flohr und D. M. Gavrila (2014). „Context-Based Pedestrian Path Prediction“. In: *Proceedings of European Conference on Computer Vision (ECCV)*. Zürich, Schweiz, Seiten 618–633 (siehe Seiten 36, 53).

Kotte, J. und A. Pütz (2016). „Analyse der Interaktion zwischen Fußgängern und Fahrzeugen im Realverkehr und kontrollierten Feld“. In: *UR:BAN-Konferenz*. Garching (siehe Seite 22).

Lavrenko, V. und N. Goddard (2014). *Introductory Applied Machine Learning*. University of Edinburgh, School of Informatics (siehe Seite 59).

Lee, D. N. (1976). „A theory of visual control of braking based on information about time-to-collision“. In: *Perception*, Band 5, Seiten 437–459 (siehe Seite 2).

Li, P., Y. Bian, J. Rong, L. Zhao und S. Shu (2013). „Pedestrian Crossing Behavior at Unsignalized Mid-block Crosswalks Around the Primary School“. In: *Procedia - Social and Behavioral Sciences*, Band 96, Seiten 442–450 (siehe Seite 23).

Li, Y. und G. Fernie (2010). „Pedestrian behavior and safety on a two-stage crossing with a center refuge island and the effect of winter weather on pedestrian compliance rate“. In: *Accident Analysis and Prevention*, Band 42, Nr. 4, Seiten 1156–1163 (siehe Seite 16).

Limbourg, M. (2011). „Mobilitäts-/Verkehrserziehung als Beitrag zur Sozialerziehung“. In: *Sozialerziehung in der Schule*. Herausgegeben von M. Limbourg und G. Steins. 1. Auflage. Wiesbaden: VS Verlag für Sozialwissenschaften. Kapitel 19, Seiten 399–424 (siehe Seite 22).

Liu, L., S. Lao, P. W. Fieguth, Y. Guo, X. Wang und M. Pietikäinen (2016). „Median Robust Extended Local Binary Pattern for Texture Classification“. In: *IEEE Transactions on Image Processing*, Band 25, Nr. 3, Seiten 1368–1381 (siehe Seiten 79, 80).

Manning, C. D., P. Raghavan und H. Schütze (2008). *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, Seite 506 (siehe Seite 71).

Meinshausen, N. (2006). „Quantile Regression Forests“. In: *Journal of Machine Learning Research*, Band 7, Seiten 983–999 (siehe Seite 37).

Minitab Inc. (2016). *Muster in Residuendiagrammen*. URL: `http://support.minitab.com/de-de/minitab/17/topic-library/modeling-statistics/regression-and-correlation/residuals-and-residual-plots/patterns-in-residual-plots/` (besucht am 01. 11. 2017) (siehe Seite 75).

Montel, M. C., T. Brenac, M.-A. Granie, M. Millot und C. Coquelet (2013). „Urban environments, pedestrian-friendliness and crossing decisions“. In: *Transportation Research Board 92nd Annual Meeting*. Frankreich, Seite 13 (siehe Seiten 18, 19).

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: The MIT Press (siehe Seiten 56, 64, 65).

Nee, J. und M. E. Hallenbeck (2003). *A Motorist and Pedestrian Behavioral Analysis Relating To Pedestrian Safety Improvements*. Technischer Bericht. Washington State Transportation Center (TRAC) (siehe Seiten 18, 23).

Nguyen, Q., H. Valizadegan und M. Hauskrecht (2014). „Learning classification models with soft-label information.“ In: *Journal of the American Medical Informatics Association (JAMIA)*, Band 21, Nr. 3, Seiten 501–508 (siehe Seiten 57, 58).

OECD/ITF (2015). „Germany“. In: *Road Safety Annual Report 2015*. Paris, Frankreich: OECD Publishing. Kapitel 13, Seiten 173–185 (siehe Seite 4).

OpenCV (2017). *Open Source Computer Vision Library*. URL: `https://opencv.org` (besucht am 01. 11. 2017) (siehe Seite 101).

OriginLab Corporation (2012). *Grafische Residuenanalyse*. URL: `http://www.originlab.de/doc/Origin-Help/Residual-Plot-Analysis` (besucht am 01. 11. 2017) (siehe Seiten 74, 75).

Oxley, J. A., E. Ihsen, B. N. Fildes, J. L. Charlton und R. H. Day (2005). „Crossing roads safely: An experimental study of age differences in gap selection by pedestrians“. In: *Accident Analysis and Prevention*, Band 37, Nr. 5, Seiten 962–971 (siehe Seite 22).

Papadimitriou, E., G. Yannis und J. Golias (2009). „A critical assessment of pedestrian behaviour models“. In: *Transportation Research Part F: Traffic Psychology and Behaviour*, Band 12, Nr. 3, Seiten 242–255 (siehe Seiten 15, 16).

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, R. Weiss und M. Brucher (2011). „Scikit-learn: Machine Learning in Python“. In: *Journal of Machine Learning Research (JMLR)*, Band 12, Seiten 2825–2830 (siehe Seiten 62, 64, 69, 117, 139).

Peng, P., R. C.-W. Wong und P. S. Yu (2014). „Learning on Probabilistic Labels“. In: *Proceedings of the 2014 SIAM International Conference on Data Mining*. Philadelphia, PA, USA: Society for Industrial und Applied Mathematics, Seiten 307–315 (siehe Seite 57).

Petersen, J. (2003). *Verkehrsunfallaufnahme*. URL: `www.unfallaufnahme.info` (besucht am 01. 11. 2017) (siehe Seite 16).

Pietikainen, M. und T. Maenpaa (2002). *Local Binary Patterns for Still Images* (siehe Seite 79).

Platt, J. C. (1999). „Probabilistc Outputs for Support Vector Machines and Comparisons to Regularized Likelihodd Methods“. In: *Advances in Large Margin Classifiers*, Band 10, Nr. 3, Seiten 61–74 (siehe Seite 64).

Powers, D. M. W. (2007). *Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation*. Technischer Bericht SIE-07-001. Adelaide, Australien: Flinders University (siehe Seiten 70, 73).

Puca, R. M. (2014). „Intention“. In: *Dorsch - Lexikon der Psychologie*. Herausgegeben von M. A. Wirtz. 17. Auflage. Göttingen: Verlag Hans Huber, Seite 801 (siehe Seite 10).

Quintero, R., J. Almeida, D. F. Llorca und M. A. Sotelo (2014). „Pedestrian Path Prediction using Body Language Traits“. In: *2014 IEEE Intelligent Vehicle Symposium (IV)*. Dearborn, MI, USA, Seiten 317–323 (siehe Seite 32).

Rehder, E. und H. Kloeden (2015). „Goal-Directed Pedestrian Prediction“. In: *2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*. Santiago, Chile, Seiten 139–147 (siehe Seite 29).

Rehder, E., H. Kloeden und C. Stiller (2014). „Head Detection and Orientation Estimation for Pedestrian Safety“. In: *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. Qingdao, China, Seiten 2292–2297 (siehe Seiten 76, 80, 111, 125).

Rehder, E., H. Kloeden und C. Stiller (2015). „Planungsbasierte Fußgängerprädiktion“. In: *10. Workshop Fahrerassistenzsysteme*. Waltingen: Uni-DAS e.V., Seiten 129–138 (siehe Seiten 3, 4, 25–30).

Roehder, M. (2011). *Ein situativ entscheidendes Fahrzeugsystem für den vorausschauenden Fußgängerschutz*. Dissertation. TU Kaiserslautern. Cuvillier Verlag (siehe Seiten 18, 26, 38).

Rohrlack, C. (2009). „Logistische und Ordinale Regression“. In: *Methodik der empirischen Forschung*. Herausgegeben von S. Albers, D. Klapper, U. Konradt, A. Walter und J. Wolf. 3. Auflage. Wiesbaden: Gabler Verlag. Kapitel 18, Seite 16 (siehe Seite 64).

Rohrmann, B. (1978). „Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung“. In: *Zeitschrift für Sozialpsychologie*, Band 9, Seiten 222–245 (siehe Seite 86).

Rosbach, S. (2016). „Pedestrian Orientation Estimation in Autonomous Driving“. Masterarbeit. FAU Erlangen (siehe Seite 125).

Rosenbloom, T. (2009). „Crossing at a red light: Behaviour of individuals and groups“. In: *Transportation Research Part F: Traffic Psychology and Behaviour*, Band 12, Seiten 389–394 (siehe Seite 23).

RoSPA (2014). *The Green Cross Code*. Edgbaston, UK: Royal Society for the Prevention of Accidents, Seiten 1–2 (siehe Seite 20).

Rossant, C. (2014). *IPython Interactive Computing and Visualization Cookbook*. Birmingham, UK: Packt Publishing (siehe Seiten 58, 60).

Scherf, O. und S. Zecha (2009). „Verfahren zum Bestimmen eines wahrscheinlichen Bewegungs-Aufenthaltsbereichs eines Lebewesens“. Patentschrift DE102007037610 A1 (siehe Seite 3).

Schilde, M. (2007). „Erfassung des Querungsverhaltens von mobilitätseingeschränkten, nicht motorisierten Verkehrsteilnehmern“. Studienarbeit. TU Dresden (siehe Seite 19).

Schmidt, S. und B. Färber (2009). „Pedestrians at the kerb - Recognising the action intentions of humans“. In: *Transportation Research Part F: Traffic Psychology and Behaviour*, Band 12, Nr. 4, Seiten 300–310 (siehe Seiten 19, 22).

Schmidt, S., B. Färber und A. Pèrez Grassi (2008). „Geht er oder geht er nicht? - Ein FAS zur Vorhersage von Fußgängerabsichten“. In: *Workshop Fahrerassistenzsysteme*. Herausgegeben von M. Maurer und C. Stiller. Walting: Freundeskreis Mess- und Regelungstechnik Karlsruhe e.V., Seiten 176–184 (siehe Seiten 4, 21, 40, 41).

Schnabel, W. und D. Lohse (2011). *Straßenverkehrstechnik*. 3. Auflage. Berlin: Beuth Verlag GmbH (siehe Seite 16).

Schneemann, F. und I. Gohl (2016). „Analyzing Driver-Pedestrian Interaction at Crosswalks: A Contribution to Autonomous Driving in Urban Environments“. In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. Göteborg, Schweden, Seiten 38–43 (siehe Seiten 2, 4, 5, 21, 23).

Schneemann, F. und P. Heinemann (2016). „Context-based Detection of Pedestrian Crossing Intention for Autonomous Driving in Urban Environments“. In: *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, Süd Korea, Seiten 2243–2248 (siehe Seiten 82, 108).

Schneider, N. und D. M. Gavrila (2013). „Pedestrian path prediction with recursive Bayesian filters: A comparative study“. In: *Proceedings ot the German Conference on Pattern Recognition (GCPR)*. Saarbrücken, Seiten 174–183 (siehe Seiten 25, 38).

Schoon, J. G. (2006). „Pedestrian Behaviour at Uncontrolled Crossings“. In: *Traffic Engineering and Control*, Band 47, Nr. 6, Seiten 229–235 (siehe Seite 20).

Schrempf, O. (2008). *Stochastische Behandlung von Unsicherheiten in kaskadierten dynamischen Systemen*. Dissertation. Universität Karlsruhe (TH). Universitätsverlag Karlsruhe (siehe Seiten 13, 49).

Schulz, A., N. Damer, M. Fischer und R. Stiefelhagen (2011). „Combined Head Localization and Head Pose Estimation for Video-Based Advanced Driver Assistance

Systems“. In: *Pattern Recognition - 33rd DAGM Symposium*. Frankfurt, Seiten 51–60 (siehe Seite 125).

Schweizer, T., C. Thomas und P. Regli (2009). *Verhalten am Fussgängerstreifen*. Technischer Bericht. Zürich, Schweiz: Fussverkehr Schweiz (siehe Seiten 18, 21–23).

Sisiopiku, V. P. und D. Akin (2003). „Pedestrian behaviors at and perceptions towards various pedestrian facilities: An examination based on observation and survey data“. In: *Transportation Research Part F: Traffic Psychology and Behaviour*, Band 6, Seiten 249–274 (siehe Seite 19).

Smith, S. W. (1997). *The Scientist & Engineer's Guide to Digital Signal Processing*. San Diego, CA, USA: California Technical Publications (siehe Seite 186).

Sonka, M., V. Hlavac und R. Boyle (2013). *Image Processing, Analysis, and Machine Vision*. 4. Auflage. Boston, MA, USA: Cengage Learning (siehe Seite 77).

Sullman, M. J. M., M. E. Gras, S. Font-Mayolas, L. Masferrer, M. Cunill und M. Planes (2011). „The Pedestrian Behaviour of Spanish Adolescents“. In: *Journal of Adolescence*, Band 34, Nr. 3, Seiten 531–539 (siehe Seiten 21, 24).

Thornton, C. (2008). *Support Vector Machines* (siehe Seite 63).

Tiemann, N. (2012). *Ein Beitrag zur Situationsanalyse im vorausschauenden Fußgängerschutz*. Dissertation. Universität Duisburg-Essen (siehe Seiten 17, 25).

Tsimhoni, O., A. S. Kandt und M. J. Flannagan (2008). *Driver Perception of Potential Pedestrian Conflict*. Technischer Bericht UMTRI-2008-46. Ann Arbor, MI, USA: University of Michigan, Transportation Research Institute (siehe Seiten 4, 40).

Tukey, J. (1949). „One degree of freedom for nonadditivity“. In: *Biometrics*, Band 5, Nr. 3, Seiten 232–242 (siehe Seite 45).

Velodyne (2016). *Velodyne LiDAR*. URL: `http://velodynelidar.com` (besucht am 01. 11. 2017) (siehe Seite 37).

Voelz, B., H. Mielenz, G. Agamennoni und R. Siegwart (2015). „Feature Relevance Estimation for Learning Pedestrian Behavior at Crosswalks“. In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*. Las Palmas, Spanien, Seiten 854–860 (siehe Seiten 36, 37, 39, 53, 184, 185).

Voelz, B., H. Mielenz, R. Siegwart und J. Nieto (2016). „Predicting Pedestrian Crossing using Quantile Regression Forests“. In: *2016 IEEE Intelligent Vehicles Symposium (IV)*. Göteborg, Schweden, Seiten 426–432 (siehe Seite 37).

Vollrath, M., S. Briest und C. Schießl (2006). „Ableitung von Anforderungen an Fahrerassistenzsysteme aus Sicht der Verkehrssicherheit“. In: *Berichte der Bundesanstalt für Straßenwesen*, Band F60 (siehe Seite 4).

Wakim, C. F., S. Capperon und J. Oksman (2004). „A Markovian model of pedestrian behavior“. In: *2004 IEEE International Conference on Systems, Man and Cybernetics*. Band 4. Den Haag, Niederlande, Seiten 4028–4033 (siehe Seite 30).

Wang, J. M., D. J. Fleet und A. Hertzmann (2008). „Gaussian Process Dynamical Models for Human Motion“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Band 30, Nr. 2, Seiten 283–298 (siehe Seite 31).

Wang, L. (2010). *Support Vector Machines: Theory and Applications*. New York, NY, USA: Springer (siehe Seite 63).

Wedel, A., T. Pock, J. Braun, U. Franke und D. Cremers (2008). „Duality TV-L1 Flow with Fundamental Matrix Prior“. In: *2008 23rd International Conference Image and Vision Computing New Zealand (IVCNZ)*. Christchurch, Neuseeland (siehe Seite 31).

Weidmann, U. (1992). „Transporttechnische Eigenschaften des Fussgängerverkehrs Literaturauswertung“. In: *Transporttechnik der Fußgänger*. Schriftenreihe des IVT Nr. 90. Insitut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT), ETH Zürich (siehe Seite 16).

Wilson, D. G. und G. B. Grayson (1980). *Age-Related Differences in the Road Crossing Behaviour of Adult Pedestrians*. Technischer Bericht. Crowthorne, England: Transport und Road Research Laboratory Report LR 933 (siehe Seiten 20, 24).

Winner, H., S. Hakuli, F. Lotz und C. Singer (2015). *Handbuch Fahrerassistenzsysteme*. Herausgegeben von H. Winner, S. Hakuli, F. Lotz und C. Singer. 3. Auflage. Wiesbaden: Springer (siehe Seite 3).

Wirtz, M. und F. Caspar (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe (siehe Seiten 38, 39, 41–45, 233).

Yanqing, W., Z. Xiaoqing und L. Yang (2014). „Similar Normal Distribution of Pedestrian Speeds at Signalized Intersection Crosswalks“. In: *2014 Fifth International Conference on Intelligent Systems Design and Engineering Applications (ISDEA)*. Hunan, China (siehe Seite 16).

Yi, S., H. Li und X. Wang (2015). „Understanding Pedestrian Behaviors from Stationary Crowd Groups“. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA, USA, Seiten 3488–3496 (siehe Seite 26).

Zhang, T. Y. und C. Y. Suen (1984). „A Fast Parallel Algorithm for Thinning Digital Patterns“. In: *Communications of the ACM*, Band 27, Nr. 3, Seiten 236–239 (siehe Seite 114).

Abbildungsverzeichnis

Tabellenverzeichnis

Abkürzungsverzeichnis

ADTF	Automotive Data and Time-Triggered Framework
AC	Accuracy
AUROC	Area Under the ROC-Curve
CHOG	Context-based Histograms of Oriented Gradients
CMHI	Context-based Movement History Image
CO	Crosswalk Occupancy
DBN	Dynamic Bayesian Network
FAS	Fahrerassistenzsystem
FGS	Fußgängerschutzsystem
FN	False Negative
FOV	Field of view
FP	False Positive
FPR	False Positive Rate
fps	Frames Per Second
GPDM	Gaussian Process Dynamical Model
GUI	Grafische Benutzeroberfläche

HM	Harmonische Mittel
HMM	Hidden Markov Model
HOG	Histograms of Oriented Gradients
HOM	Histograms of Orientation Motion
ICC	Intraklassenkorrelation
LBP	Local Binary Pattern
LoPL	Learning on Probabilistic Labels
LSFF	Lateral Scene Flow Features
MAE	Mean Absolute Error
MCHOG	Motion Contour Histograms of Oriented Gradients
MDP	Markov Decision Process
MHI	Motion History Image
MLC	Maximum Likelihood Classifier
MMH	Maximum Margin Hyperebene
MOMDP	Mixed Observable Markov Decision Process
MoG-HMM	Mixture of Gaussians Hidden Markov Model
MW	Mittelwert
Min	Minimaler Wert
Max	Maximaler Wert
OF	Orientation Feature
PAF	Posture Appearance Feature

PCA	Principal Component Analysis
PN	Predicted Negative
PP	Predicted Positive
PR	Precision
PÜ	Prozentuale Übereinstimmung
px	Pixel
QI	Querungsintention
RBF	Radiale Basisfunktion
RC	Recall
RMSE	Root Mean Squared Error
RN	Real Negative
ROC	Receiver Operation Characteristic
RP	Real Positive
SD	Standardabweichung
SLDS	Switching Linear Dynamical System
SLP	Single Layer Perceptron
s.t.	subject to
StVO	Straßenverkehrs-Ordnung
SVM	Support Vector Machine
SVR	Support Vector Regression
TN	True Negative

TP True Positive

TPR True Positive Rate

TNR True Negative Rate

TTC Time-to-Collision

WAO Waiting Area Occupancy

Anhang A

Details zur Berechnung der Beobachterreliabilität

Ergänzend zum Abschnitt 2.4.2 werden im Folgenden die zur Berechnung der Beobachterreliabilität nötigen Berechnugngsvorschriften erläutert. Die Darstellung basiert auf (Wirtz und Caspar, 2002)

A.1 Berechnung der Varianzbestandteile der *ICC*

Um über die Gleichungen 2.7 die Intraklassenkorrelation ICC_{unjust} und über die Gleichung 2.8 die ICC_{just} zu bestimmen, müssen zuvor die einzelnen Varianzbestandteile berechnet werden. Wie in Tabelle A.1 dargestellt, wird hierzu zunächst für jedes der N beurteilten Objekte der Mittelwert $\bar{o}_i$ der Urteile u_{ij} aller k Beobachter mit

$$\bar{o}_i = \frac{\sum_{j=1}^{k} u_{ij}}{k} \tag{A.1}$$

bestimmt. Auf gleiche Weise wird für jeden Beobachter der Mittelwert $\bar{b}_j$ seiner Urteile berechnet. Daraus ergibt sich der Mittelwert der gesamten Stichprobe $\bar{u}_{Ges}$ mit

$$\bar{u}_{Ges} = \frac{\sum_{i=1}^{N} \bar{o}_i}{N} = \frac{\sum_{j=1}^{k} \bar{b}_j}{k}. \tag{A.2}$$

Zur Bestimmung der Varianz zwischen den Objekten (MS_{obj}) wird anschließend die k-

Tabelle A.1: Ergebnismatrix für k Beobachter, die N Objekte beurteilt haben

	Beobachter 1	Beobachter 2	...	Beobachter k	$\bar{o}_i$
Objekt 1	u_{11}	u_{12}	...	u_{1k}	$\bar{o}_1$
Objekt 2	u_{21}	u_{22}	...	u_{2k}	$\bar{o}_2$
...	...	...	...	...	...
Objekt N	u_{N1}	u_{N2}	...	u_{Nk}	$\bar{o}_N$
$\bar{b}_j$	$\bar{b}_1$	$\bar{b}_2$	...	$\bar{b}_k$	$\bar{u}_{Ges}$

fache Summe der quadrierten Abweichungen der individuellen Objektmittelwerte vom Gesamtmittelwert (QS_{obj}) ermittelt. Es gilt

$$QS_{obj} = k \sum_{i=1}^{N} \left(\bar{o}_i - \bar{u}_{Ges}\right)^2 . \tag{A.3}$$

Mit der Anzahl der Freiheitsgrade (df) als Divisor ergibt sich

$$MS_{obj} = \frac{QS_{obj}}{df_{obj}} = \frac{QS_{obj}}{N-1}. \tag{A.4}$$

Analog wird die Varianz zwischen den Beobachtern (MS_{rat}) mit

$$QS_{rat} = N \sum_{j=1}^{k} \left(\bar{b}_j - \bar{u}_{Ges}\right)^2 \tag{A.5}$$

und

$$MS_{rat} = \frac{QS_{rat}}{df_{rat}} = \frac{QS_{rat}}{k-1} \tag{A.6}$$

bestimmt. Die für die Berechnung der Restvarianz MS_{err} benötigte Quadratsumme QS_{err} lässt sich indirekt über

$$QS_{Ges} = QS_{obj} + QS_{rat} + QS_{err} \tag{A.7}$$

bestimmen, wobei die Gesamtquadratsumme QS_{Ges} durch die Aufsummierung der quadrierten Differenzen der individuellen Urteile zum Gesamtmittelwert mit

$$QS_{Ges} = \sum_{i=1}^{N} \sum_{j=1}^{k} \left(u_{ij} - \bar{u}_{Ges}\right)^2 \tag{A.8}$$

bestimmt wird. Für MS_{err} gilt schließlich

$$MS_{err} = \frac{QS_{err}}{df_{err}} = \frac{QS_{err}}{(N-1)(k-1)}. \tag{A.9}$$

A.2 Berechnung des Konfidenzintervalls der *ICC*

Im Folgenden wird angenommen, dass ein Konfidenzintervall mit der Sicherheit $(1-\alpha)\cdot 100\%$ für eine Stichprobe, bei der k Beobachter N Objekte beurteilt haben, bestimmt wird.

A.2.1 Konfidenzintervall für die ICC_{unjust}

Zur Bestimmung des Konfidenzintervalls für die ICC_{unjust} im zweifaktoriellen Modell muss zunächst

$$\nu = \frac{(k-1)(N-1)(k\cdot r\cdot F_j + N(1+(k-1)\cdot r) - k\cdot r)^2}{(N-1)\cdot k^2 r^2 F_j^2 + (N(1+(k-1)\cdot r) - k\cdot r)^2} \tag{A.10}$$

mit

$$F_j = \frac{MS_{rat}}{MS_{err}} \tag{A.11}$$

und

$$r = ICC_{unjust} \tag{A.12}$$

bestimmt werden. Damit können die F-Werte

$$F_u = F_{(1-0.5\alpha,\, N-1,\, \nu)} \tag{A.13}$$

und

$$F_o = F_{(1-0.5\alpha,\, \nu,\, N-1)} \tag{A.14}$$

in Tabellen zu kritischen Werten der F-Verteilung nachgeschlagen werden.

Für das Konfidenzintervall

$$P\left(L_u < \rho < L_o\right) = 1-\alpha \tag{A.15}$$

gilt schließlich

$$L_u = \frac{N\left(MS_{obj} - F_u\cdot MS_{err}\right)}{F_u\left(k\cdot MS_{rat} + (k\cdot N - k - N)\cdot MS_{err}\right) + N\cdot MS_{obj}} \tag{A.16}$$

$$L_o = \frac{N(MS_{obj}\cdot F_o - MS_{err})}{k\cdot MS_{rat} + (k\cdot N - k - N)\cdot MS_{err} + N\cdot MS_{obj}\cdot F_o}. \tag{A.17}$$

A.2.2 Konfidenzintervall für die ICC_{just}

Für das Konfidenzintervall der ICC_{just} im zweifaktoriellen Modell müssen zunächst die F-Werte

$$F_u = \frac{F_0}{F_{(1-0.5\alpha,\, N-1,\, (N-1)(k-1))}} \tag{A.18}$$

und

$$F_o = F_0 \cdot F_{(1-0.5\alpha,\, (N-1)(k-1),\, N-1)} \tag{A.19}$$

mit

$$F_0 = \frac{MS_{obj}}{MS_{err}} \tag{A.20}$$

bestimmt werden. Das Konfidenzintervall ergibt sich dann über

$$P\left(\frac{F_u - 1}{F_u + (k-1)} < \rho < \frac{F_o - 1}{F_o + (k-1)}\right) = 1 - \alpha. \tag{A.21}$$

Anhang B

Beobachterschulung und Beobachterbefragung

Bei der Kapitel 4 beschriebenen Videoannotationsstudie wurden die Beobachter vor der Durchführung der Studie mit der in Abschnitt B.1 gezeigten Anleitung geschult. Im Anschluss wurden die Beobachter aufgefordert den in Abschnitt B.2 gezeigten Fragebogen zur Erfassung der demographischen Daten der Beobachter auszufüllen.

B.1 Anleitung

Lieber Teilnehmer,

vielen Dank für die Teilnahme an meiner Beobachtungsstudie. In dieser Studie interessiere ich mich für die Beobachtbarkeit verschiedener Verhaltensweisen von Fußgängern. In diesem Rahmen werden Ihnen Video-Aufnahmen gezeigt, die im realen Straßenverkehr aus Sicht eines Fahrzeugs aufgezeichnet wurden. Ihre Aufgabe ist es, eine Einschätzung bezüglich der **Querungsintention** der abgebildeten Fußgänger abzugeben.

1. **Konzept**

Der Begriff Querungsintention beschreibt die **prinzipielle Absicht** des Fußgängers, den **Fahrstreifen** des filmenden Fahrzeugs queren zu wollen. Hierbei ist es irrelevant, ob der Fußgänger die Querung noch vor dem sich nähernden Fahrzeug z.B. durch das Betreten der Straße initiiert, oder erst nach dessen Vorbeifahrt. Sie sollen also NICHT einschätzen, ob der Fußgänger noch vor dem filmenden Fahrzeug auf die Straße geht oder nicht, sondern ob der Fußgänger prinzipiell die Absicht hat, die Straße queren zu wollen.

Die zwei nachfolgenden Bilder zeigen beispielhaft eine Situation mit einem Fußgänger mit Querungsintention und eine mit einem Fußgänger ohne Querungsintention:

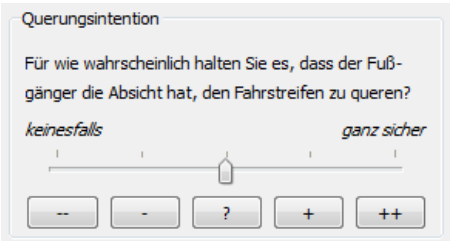
Bild 1:

Ein Fußgänger steht an einer Fußgängerampel und wartet darauf, dass diese auf grün umschaltet, um die Straße sicher queren zu können. Der Fußgänger hat somit eine Querungsintention, unabhängig davon, dass er die Straße nicht vor dem sich nähernden Fahrzeug betreten wird.

Bild 2:

Ein Fußgänger steht an einer Bushaltestelle und wartet auf einen Bus. Er hat keine Absicht, den Fahrstreifen des sich nähernden Fahrzeugs zu queren und somit keine Querungsintention.

Um die Querungsintention eines Fußgänger einschätzen zu können, steht Ihnen eine **5-stufige Skala** zur Verfügung, bei der die Abstände zwischen den einzelnen Stufen gleich groß sind:



Die einzelnen Skalenwerte haben folgende Bedeutung:

- - Der Fußgänger hat *keinesfalls* die Absicht den Fahrstreifen zu queren
- Der Fußgänger hat *wahrscheinlich nicht* die Absicht den Fahrstreifen zu queren
? Der Fußgänger hat *vielleicht* die Absicht den Fahrstreifen zu queren
+ Der Fußgänger hat *ziemlich wahrscheinlich* die Absicht den Fahrstreifen zu queren
+ + Der Fußgänger hat *ganz sicher* die Absicht den Fahrstreifen zu queren

Die Beurteilung geschieht immer bei pausiertem Video und bezieht sich auf den Fußgänger und sein Verhalten **bis** zum **aktuellen Zeitpunkt**. Je nach Situation kann sich Ihre

Einschätzung mit fortschreitender Beobachtungszeit ändern. Dieses sollen Sie durch die Abgabe einer neuen Einschätzung anzeigen. D.h. eine Einschätzung ist so lange gültig, bis Sie eine neue, geänderte Einschätzung abgeben.

Die zwei nachfolgenden Skizzen zeigen zwei beispielhafte Situationen, in denen sich die Querungsintention eines Fußgängers ändert:

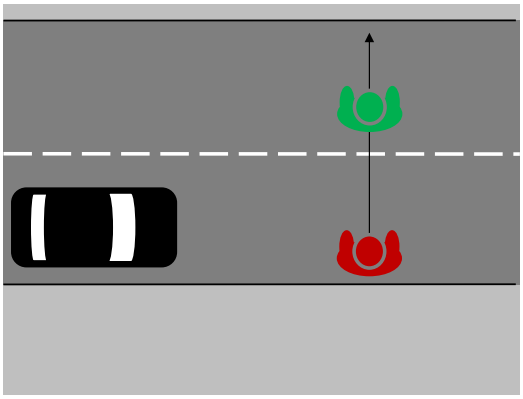


Bild 3:

Ein Fußgänger quert den Fahrstreifen des sich nähernden Fahrzeugs und hat somit eine Querungsintention (rot). Nach dem Verlassen des Fahrstreifen des Ego-Fahrzeugs zeigt der Fußgänger keine Absicht, diesen nochmal zu queren. Der Fußgänger hat somit keine Querungsintention mehr (grün), auch wenn er die Straße auf der Gegenfahrbahn noch quert.

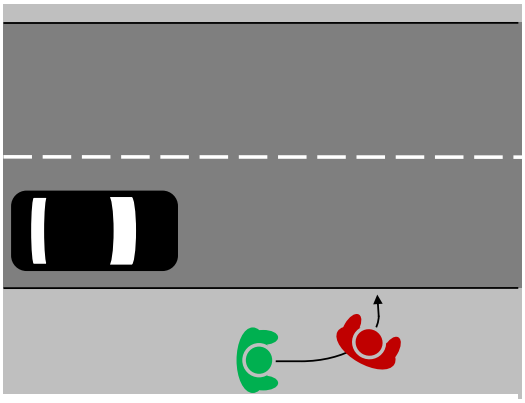


Bild 4:

Der Fußgänger läuft parallel zum Fahrbahnrand und zeigt kein Verhalten, das auf eine Querungsintention schließen lässt (grün). Nach einigen Metern bewegt sich der Fußgänger Richtung Fahrbahnkante und zeigt ein deutliches Absicherungsverhalten, dass auf eine Querungsintention schließen lässt (rot).

Bei Unklarheiten und Fragen bezüglich der zu bewertenden Querungsintention wenden Sie sich bitte an den Versuchsleiter.

2. Tooling

Zur Durchführung der Studie steht Ihnen das **PCI_Labeltool** zur Verfügung. Dieses wird durch einen Doppelklick auf die Datei *PCI_MLBevo_Labeling.bat* gestartet.

Machen Sie sich zunächst anhand eines Beispielvideos mit dem Labeltool vertraut. Das Video wird durch einen Doppelklick auf den Dateinamen *Beispiel.dat* im linken Bereich des Tools (Bild 5) geladen. Sollte sich das Fenster des Labeltools nicht automatisch öffnen, gehen Sie analog den in Bild 6 beschriebenen Schritten vor.

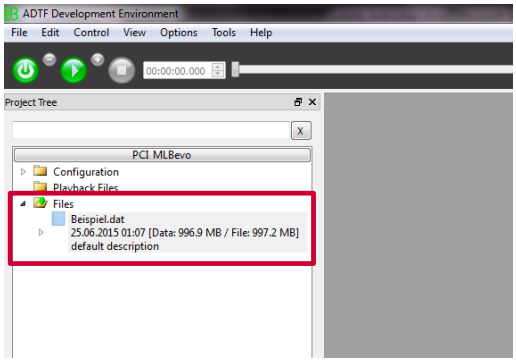


Bild 5:

Ein Doppelklick auf die Datei Beispiel.dat lädt das Video.

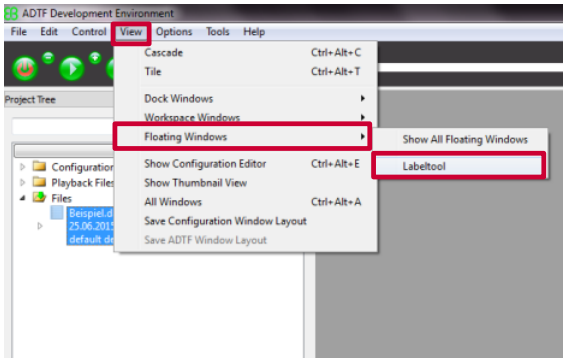


Bild 6:

Klicken Sie zum Öffnen des Labeltools auf: View → Floating Windows → Labeltool.

Das Labeltool besteht aus den vier in Bild 7 markierten Bereichen:

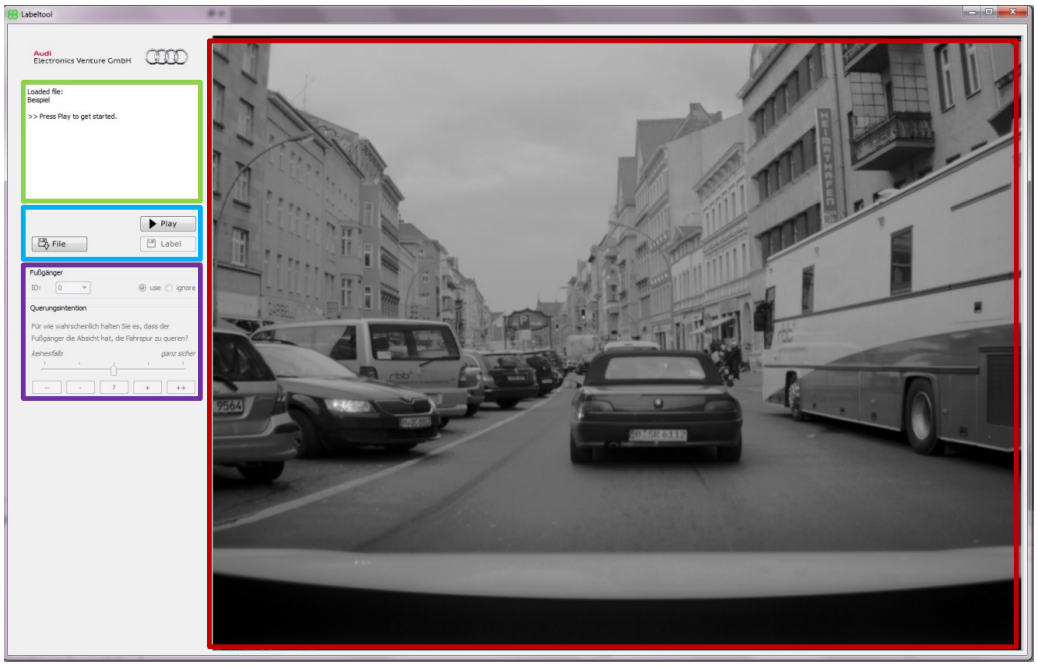


Bild 7:

Das Labeltool mit **Video**- , **Info**-, **Steuer**- und **Bewertungs**bereich.

Starten Sie das Video, in dem Sie im Steuerbereich auf den Play – Button klicken. Bitte beobachten Sie das im Videobereich gezeigte Verkehrsgeschehen aufmerksam.

Das Video hält automatisch an, sobald sie einen **Fußgänger bewerten** sollen. Der zu bewertende Fußgänger wird mit einer roten Box markiert (Bild 8: roter Pfeil) und der Bewertungsbereich auf der linken Seite ist freigeschaltet. Zudem wird der Fahrstreifen, auf den sich die zu bewertende Querungsintention bezieht, durch zwei weiße, gestrichelte Linien angezeigt (Bild 8: weiße Pfeile).

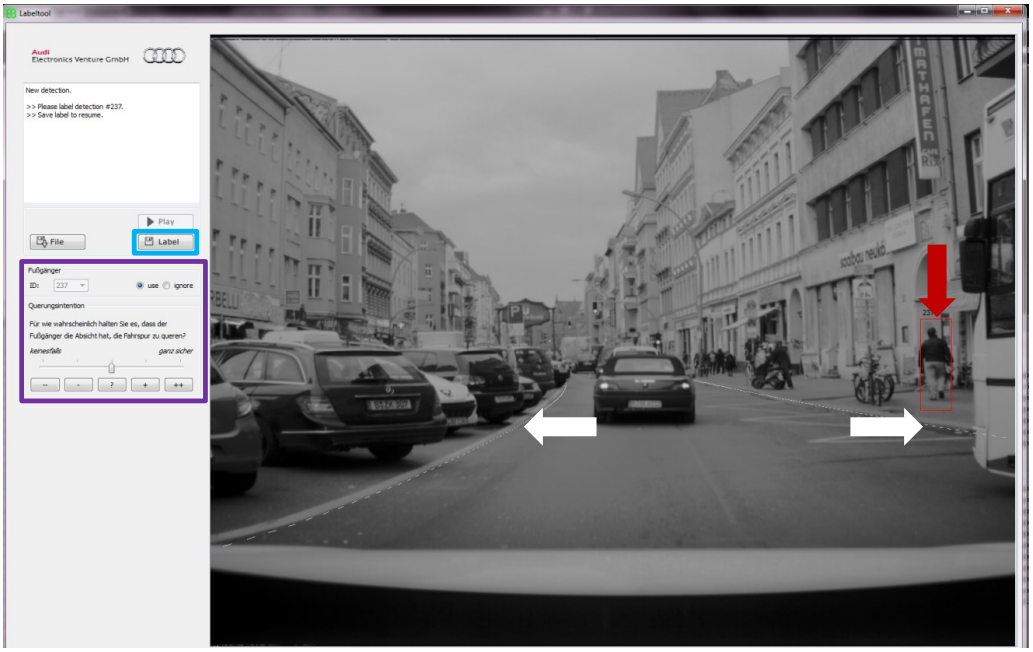


Bild 8:

Ansicht des Labeltools, wenn Sie zur Bewertung eines Fußgängers aufgefordert werden.

Um Ihr **Urteil** zur Querungsintention des markierten Fußgängers **abzugeben**, können Sie

a) die Buttons unter der 5-stufigen Skala verwenden

oder b) den Slider entlang der Skala mit der Maus verschieben.

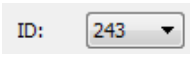
Bitte urteilen Sie immer möglichst spontan. Uns interessiert Ihre persönliche Einschätzung, d.h. es gibt keine „richtigen“ oder „falschen“ Antworten.

Klicken Sie anschließend im Steuerbereich auf den Label – Button, um Ihr Urteil zu **speichern**. Über den Play – Button können Sie anschließend mit dem Abspielen des Videos fortfahren.

Ändert sich Ihr Urteil zur Querungsintention eines Fußgängers mit fortlaufender Beobachtung, können Sie das Video zu jedem Zeitpunkt über den Pause – Button oder

die Leertaste **anhalten** und eine geänderte Bewertung abgeben und speichern. Der Slider zeigt stets ihre aktuelle Bewertung an.

Sind **mehrere** zu bewertende **Fußgänger** im Bild, können Sie den gewünschten Fußgänger durch:

a) die Auswahl seiner ID in dem Pull-Down Menü 

oder b) einen Klick in die schwarze Box, die den Fußgänger umgibt

auswählen.

Zusätzlich werden Sie nach einem festen Zeitintervall aufgefordert, Ihr bisheriges Urteil zur Querungsintention des Fußgängers zu **überprüfen**. Analog zur Erstbewertung wird das Video hierzu automatisch angehalten und der zu bewertende Fußgänger ausgewählt. Der Slider zeigt auch hier ihre vorherige Bewertung an. Diese können Sie durch einen Klick auf den 

– Button direkt bestätigen oder ggf. vorher anpassen. Alternativ können Sie auch mit der Leertaste das Label zunächst bestätigen und mit einem zweiten Tastendruck das Video weiter abspielen.

Haben Sie das **Ende** der aktuellen Video-Datei erreicht, müssen Sie Ihre abgegebenen Beurteilungen durch einen Klick auf den 

– Button speichern.

Der Infobereich unterstützt Sie bei der Durchführung der Studie mit Feedback zum aktuellen Status des Labeltools und Hinweisen auf die als nächstes auszuführenden Aktionen.

Machen Sie sich zunächst anhand des Beispiel-Videos mit der Bedienung des Labeltools vertraut. Bei Unklarheiten und Fragen wenden Sie sich bitte an den Versuchsleiter. Anschließend können Sie mit der Bewertung der für diese Studie ausgewählten Videos beginnen.

3. Studienablauf

In dieser Studie bitten wir Sie, die Fußgänger in insgesamt 41 Video-Dateien zu bewerten. Die Videos sind von file_01 bis file_41 durchlaufend nummeriert und auf vier Ordner aufgeteilt (01 bis 04). Bitte beginnen Sie mit file_01 in Ordner 01 und arbeiten sich der Reihe nach durch die vier Ordner. Bild 9 zeigt Ihnen, wie Sie die Ordner auswählen.

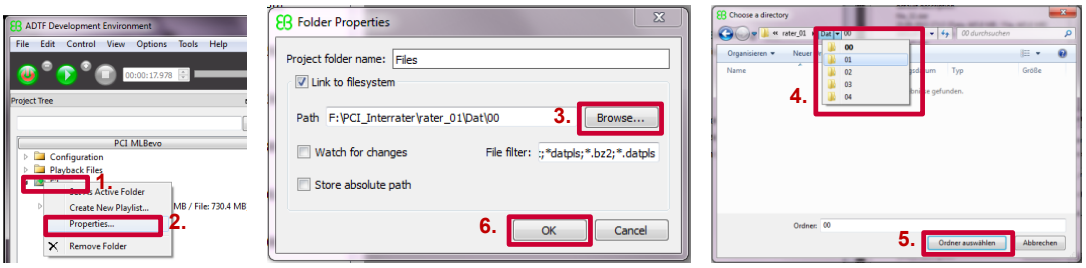


Bild 9:

Auswahl eines Ordners über: 1. Rechts-Klick auf Files, 2. Klick auf Properties..., 3. Klick auf Browse…, 4. Orderauswahl (01…04), 5. Klick auf Ordner auswählen, 6. Klick auf OK

Anschließend können Sie analog zum Beispielvideo die erste Datei file_01 über einen Doppelklick laden und mit der Bewertung beginnen. Vergessen Sie am Ende nicht Ihre Bewertungen über den File – Button zu speichern, bevor Sie die Datei file_02 laden.

Achten Sie darauf, dass Sie das im Video gezeigte Verkehrsgeschehen stets aufmerksam beobachten und konsistent bewerten. Sie können zwischen den einzelnen Videodaten beliebig lange Pausen machen. Nutzen Sie die im Anhang beigefügte Liste, um den Überblick über Ihre bereits bewerteten Videos zu behalten. Bewerten Sie die Videos in der richtigen Reihenfolge und jedes Video nur einmal.

A. Anhang: Übersicht Video-Dateien

Ordner	Dateiname	Beurteilt
01	file_01	□
01	file_02	□
01	file_03	□
01	file_04	□
01	file_05	□
01	file_06	□
01	file_07	□
01	file_08	□
01	file_09	□
01	file_10	□

Ordner	Dateiname	Beurteilt
02	file_11	□
02	file_12	□
02	file_13	□
02	file_14	□
02	file_15	□
02	file_16	□
02	file_17	□
02	file_18	□
02	file_19	□
02	file_20	□

Ordner	Dateiname	Beurteilt
03	file_21	□
03	file_22	□
03	file_23	□
03	file_24	□
03	file_25	□
03	file_26	□
03	file_27	□
03	file_28	□
03	file_29	□
03	file_30	□

Ordner	Dateiname	Beurteilt
04	file_31	□
04	file_32	□
04	file_33	□
04	file_34	□
04	file_35	□
04	file_36	□
04	file_37	□
04	file_38	□
04	file_39	□
04	file_40	□
04	file_41	□

B.2 Fragebogen

An dieser Stelle möchte ich Sie bitten, die nachfolgenden Angaben zu Ihrer Person und ihren Fahrgewohnheiten zu machen. Ihre Daten werden streng vertraulich und anonym behandelt, nur zu Forschungszwecken verwendet und in keinem Fall an Dritte weitergegeben.

1. Alter in Jahren: _____
2. Geschlecht: □ weiblich □ männlich
3. Führerscheinbesitz in Jahren: _____
4. Wie viele Kilometer haben Sie im vergangenen Jahr selbst mit dem Auto zurückgelegt?
   - □ bis 5.000 km
   - □ bis 10.000 km
   - □ bis 15.000 km
   - □ bis 20.000 km
   - □ > 20.000 km
5. Wie häufig fahren Sie pro Woche durchschnittlich mit dem Auto?
   - □ täglich
   - □ drei- bis fünfmal pro Woche
   - □ ein- bis zweimal pro Woche
   - □ seltener
6. Im Vergleich zu anderen Autofahrern fahre ich überwiegend …

schnell	○	○	○	○	○	langsam
ängstlich	○	○	○	○	○	mutig
offensiv	○	○	○	○	○	defensiv
vorsichtig	○	○	○	○	○	risikobereit
sportlich	○	○	○	○	○	gemütlich

Anhang C

Details der Implementierung

C.1 Fahrzeugkoordinatensystem

Das in dieser Arbeit verwendete Fahrzeugkoordinatensystem ist ein Rechtssystem und hat seinen Ursprung in der Mitte der Fahrzeughinterachse (s. Abb. C.1). Die x-Achse zeigt dabei in Fahrtrichtung, die y-Achse nach links und die z-Achse nach oben.

C.2 Labeltools

Neben dem in Abschnitt 4.1 beschriebenen PCI_Labeltool, bedarf es zur Erstellung der in dieser Arbeit verwendeten Datenbasis drei weiteren Labeltools. Bei zwei der drei Tools handelt es sich um Eigenentwicklungen, die analog dem PCI_Labeltool als

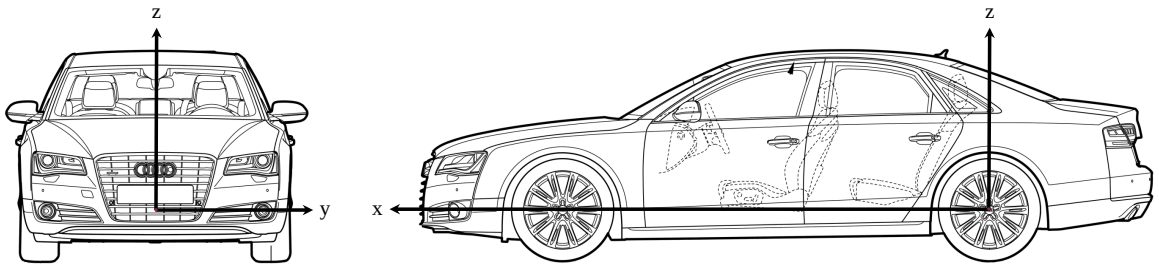


Abbildung C.1: Das in dieser Arbeit verwendete Fahrzeugkoordinatensystem.

Abbildung C.2: Die Oberfläche des zur Bestimmung der Körperorientierung entwickelten Labeltools.

eine, auf der C++–Klassenbibiliothek QT basierende und in das Framework ADTF integrierte, grafische Benutzeroberfläche (GUI) umgesetzt sind.

Abbildung C.2 zeigt die Oberfläche des ersten Tools, dass zur Bestimmung der in Abschnitt 6.1.1 beschriebenen Körperorientierung des Fußgängers verwendet wird. Im Detail handelt es sich hier um eine Anpassung des PCI_Labeltools, bei dem sich ebenfalls rechts der Videobereich mit der zu bewertenden Fußgängerdetektion und dem eingezeichneten Ego-Fahrstreifen befindet und Links der Informations-, Steuer- und Bewertungsbereich (von oben nach unten). Im Bewertungsbereich kann die Körperorientierung eines Fußgängers über die Drehskala auf ein Grad genau eingestellt werden. Die vier Buttons rechts neben der Drehskala dienen dabei einer schnellen Vorauswahl der vier am häufigsten auftretenden Körperorientierungen.

Das zweite Tool wird zur Korrektur und Ergänzung der vom verwendeten Kamerasystem ausgegebenen Fahrstreifenbegrenzungen verwendet (s. Abschn. 6.1.2). Zudem

Abbildung C.3: Die Oberfläche des zur Korrektur und Ergänzung der erkannten Fahrstreifen sowie zur Markierung der Position vorhandener Szenenelemente entwickelten Labeltools.

werden auch die Positionen der in Abschnitt 6.1.3 beschriebenen Szenenelemente, wie Fußgängerüberwege und Wartebereiche, mit Hilfe dieses Tools markiert.

Abbildung C.3 zeigt die Oberfläche der GUI. Auch hier befindet sich der Videobereich rechts im Bild. Die vom Kamerasystem erkannten Fahrstreifen werde als weiße gepunkte Linien eingezeichnet. Ebenso sind in dem gezeigten Beispiel die korrigierte linke und die ergänzte rechte Begrenzung des Ego-Fahrstreifens zu sehen (lila und pinke Linie). Die Klothoiden-Parameter der Fahrstreifenbegrenzungen können jeweils links im Bereich „Spuren“ eingestellt werden. Dabei kann über die Dropdown-Liste „ID“ der Parametersatz eines der vom Kamerasystem erkannten Fahrstreifen übernommen werden, um nur noch einzelne Größen anpassen zu müssen.

Der mittlere Bereich der GUI dient der Betrachtung der Szene aus der Vogelperspektive. Auch hier sind die Fahrstreifenbegrenzungen farbig eingezeichnet. Zudem ist die Position des detektierten Fußgängers mit einem weißen Punkt markiert und der Be-

reich des für den kontextbasierten Merkmalsvektor verwendeten Szenenausschnitts ist weiß umrandet (s. Abschn. 5.2.1, S. 105).

Die blauen Linken markieren den Fußgängerüberweg. Dieser wird über seinen Anfangs- und Endpunkt auf der jeweiligen Fahrstreifenklothoide links im Bereich „Crosswalk“ eingestellt. Die Radiobuttons dienen dabei der Auswahl der Überwegtyps.

In dem gezeigten Beispiel deaktiviert ist der Bereich „(Public Transport) Waiting Area“. Über diesen kann die Position von Wartebereichen wie Bushaltestellen markiert werden. Ein Wartebereich wird über ein Rechteck mit einem Mittelpunkt (x, y) einer Breite w, einer Tiefe d und einem Rotationswinkel a relativ zum Fahrzeugkoordinatensystem modelliert.

Bei dem dritten Labeltool handelt es sich schließlich um eine, intern bei der AUDI AG zur Verfügung stehende Applikation, die die Markierung einzelner Bildbereiche in Pixelkoordinaten sowie ein Tracking dieser Bereiche über mehrere Frames ermöglicht. Zudem können den Bildbereichen manuell definierbare Attribute zugewiesen werden. Diese Applikation wurde verwendet, um die Position der Köpfe der Fußgänger sowie die Kopforientierung zu bestimmen.