

Extraktion und Klassifikation realer Verkehrsszenarien zur Absicherung autonomer Fahrzeuge

Extraction and classification of real traffic scenarios for validation of autonomous vehicles

Semesterarbeit

an der Fakultät für Maschinenwesen der Technischen Universität München

Betreut von Prof. Dr.-Ing. Markus Lienkamp
Thomas Ponn, M.Sc.
Lehrstuhl für Fahrzeugtechnik

Eingereicht von Matthias Breitfuß, B.Sc.
Urslaustraße 8
5761 Maria Alm

Eingereicht am 28.04.2020

Aufgabenstellung

Extraktion und Klassifikation realer Verkehrsszenarien zur Absicherung autonomer Fahrzeuge

Die Bewertung der Sicherheit von autonomen Fahrzeugen ist aufgrund der unendlichen Zahl an möglichen Verkehrsszenarien äußerst aufwändig und komplex. Deshalb wird aktuell an verschiedenen Verfahren geforscht, mit denen eine begründete Reduzierung der relevanten Testfälle möglich ist. Um dennoch eine hohe Aussagekraft über das Systemverhalten zu bekommen, muss die ausgewählte Teilmenge der Szenarien möglichst herausfordernd für das zu testende System sein. Eine Möglichkeit einen Katalog an derartigen Fällen zu generieren, besteht in der Extraktion der Testsituationen aus Datensätzen realer Verkehrsaufzeichnungen. Das Potential dieser Variante bezüglich einer möglichst breiten Abdeckung des wirklichen Verkehrsgeschehens durch die ermittelten Testsituationen, verdeutlicht die Relevanz dieser Herangehensweise in Verbindung mit der Absicherung autonomer Fahrzeuge unter möglichst realitätsnahen Bedingungen.

In einer theoretischen Semesterarbeit soll ein Verfahren zur Extraktion relevanter Verkehrssituationen aus Realdaten konzipiert und implementiert werden. Zudem sollen die erfassten Testfälle in definierte Szenarioklassen zur Generierung eingeordnet werden können.

Folgende Punkte sind durch Herrn Matthias Breitfuß zu bearbeiten:

- Einarbeitung in die Absicherung automatisierter Fahrzeuge
- Einarbeitung in die Themenfelder Clusteranalyseverfahren und Klassifikationsmethoden
- Entwurf einer geeigneten Verarbeitungsstrategie zur Erzeugung eines Katalogs an relevanten Testfällen auf Basis einer Datenbank realer Verkehrsaufzeichnungen
- Implementierung des festgelegten Verfahrens in Matlab
- Validierung und Diskussion der Methode
- Dokumentation der Vorgehensweise und der erzielten Ergebnisse

Die Ausarbeitung soll die einzelnen Arbeitsschritte in übersichtlicher Form dokumentieren. Der Kandidat/Die Kandidatin verpflichtet sich, die Semesterarbeit selbständig durchzuführen und die von ihm verwendeten wissenschaftlichen Hilfsmittel anzugeben.

Die eingereichte Arbeit verbleibt als Prüfungsunterlage im Eigentum des Lehrstuhls

Ausgabe: 30.10.2019

Abgabe: 28.04.2020

Prof. Dr.-Ing. M. Lienkamp

Betreuer: Thomas Ponn, M. Sc.

Geheimhaltungsverpflichtung

Herr/Frau: **Breitfuß, Matthias**

Gegenstand der Geheimhaltungsverpflichtung sind alle mündlichen, schriftlichen und digitalen Informationen und Materialien, die der Unterzeichner vom Lehrstuhl oder von Dritten im Rahmen seiner Tätigkeit am Lehrstuhl erhält. Dazu zählen vor allem Daten, Simulationswerkzeuge und Programmcode sowie Informationen zu Projekten, Prototypen und Produkten.

Der Unterzeichner verpflichtet sich, alle derartigen Informationen und Unterlagen, die ihm während seiner Tätigkeit am Lehrstuhl für Fahrzeugtechnik zugänglich werden, strikt vertraulich zu behandeln.

Er verpflichtet sich insbesondere:

- derartige Informationen betriebsintern zum Zwecke der Diskussion nur dann zu verwenden, wenn ein ihm erteilter Auftrag dies erfordert,
- keine derartigen Informationen ohne die vorherige schriftliche Zustimmung des Betreuers an Dritte weiterzuleiten,
- ohne Zustimmung eines Mitarbeiters keine Fotografien, Zeichnungen oder sonstige Darstellungen von Prototypen oder technischen Unterlagen hierzu anzufertigen,
- auf Anforderung des Lehrstuhls für Fahrzeugtechnik oder unaufgefordert spätestens bei seinem Ausscheiden aus dem Lehrstuhl für Fahrzeugtechnik alle Dokumente und Datenträger, die derartige Informationen enthalten, an den Lehrstuhl für Fahrzeugtechnik zurückzugeben.

Besondere Sorgfalt gilt im Umgang mit digitalen Daten:

- Für den Dateiaustausch dürfen keine Dienste verwendet werden, bei denen die Daten über einen Server im Ausland geleitet oder gespeichert werden (Es dürfen nur Dienste des LRZ genutzt werden (Lehrstuhlaufwerke, Sync&Share, GigaMove).
- Vertrauliche Informationen dürfen nur in verschlüsselter Form per E-Mail versendet werden.
- Nachrichten des geschäftlichen E-Mail Kontos, die vertrauliche Informationen enthalten, dürfen nicht an einen externen E-Mail Anbieter weitergeleitet werden.
- Die Kommunikation sollte nach Möglichkeit über die (my)TUM-Mailadresse erfolgen.

Die Verpflichtung zur Geheimhaltung endet nicht mit dem Ausscheiden aus dem Lehrstuhl für Fahrzeugtechnik, sondern bleibt 5 Jahre nach dem Zeitpunkt des Ausscheidens in vollem Umfang bestehen. Die eingereichte schriftliche Ausarbeitung darf der Unterzeichner nach Bekanntgabe der Note frei veröffentlichen.

Der Unterzeichner willigt ein, dass die Inhalte seiner Studienarbeit in darauf aufbauenden Studienarbeiten und Dissertationen mit der nötigen Kennzeichnung verwendet werden dürfen.

Datum: 28.04.2020

Unterschrift: _____

Erklärung

Ich versichere hiermit, dass ich die von mir eingereichte Abschlussarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Garching, den 28.04.2020

Matthias Breitfuß, B. Sc.

Declaration of Consent, Open Source

Hereby I, **Breitfuß, Matthias**, born on 19.09.1994, make the software I developed during my semester thesis available to the Institute of Automotive Technology under the terms of the license below.

Garching, 28.04.2020

Matthias Breitfuß, B. Sc.

Copyright **2020 Breitfuß, Matthias**

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
Formelzeichen.....	V
1 Einleitung	1
1.1 Motivation.....	1
1.2 Aufgabenstellung und Zielsetzung.....	2
1.3 Gliederung der Arbeit.....	3
2 Stand der Technik	5
2.1 Grade der Automatisierung.....	5
2.2 Absicherung autonomer Fahrfunktionen.....	7
2.3 HighD-Datensatz.....	11
2.3.1 Verfahren der Datenerfassung.....	11
2.3.2 Aufbau des Datensatzes.....	12
2.4 Clusteranalysemethoden.....	13
2.4.1 K-means Clustering	15
2.4.2 Hierarchisch-Agglomeratives Clustering (HAC).....	17
2.5 Prädiktive Methoden zur Kollisionsvorhersage	22
2.5.1 Grundlagen der Kollisionserkennung.....	22
2.5.2 Grundlagen zur Fahrzustands-Prädiktion	23
2.6 Klassifikation von Verkehrsszenarien.....	25
2.7 Forschungsfrage	31
3 Methode.....	33
3.1 Überblick	33
3.2 Subfunktionen der Toolkette.....	34
3.2.1 Preprocessing der Daten	34
3.2.2 Clusteranalyse	38
3.2.1 Challenger-Prädiktion	41
3.2.2 Regelbasiert-semantische Klassifikation der Szenarien.....	45
3.2.3 Aufbau des Szenarienkatalogs über Fusion von Challenger-Prädiktion und Manöveridentifikation	52

3.3	Zusammenfassung der Methodik	54
4	Ergebnisse	55
4.1	Parameterstudie	55
4.1.1	Parameterstudie zur Challenger-Prädiktion	57
4.1.2	Parameterstudie zur Klassifikation.....	67
4.1.3	Resultat Parameterstudien	70
4.2	Stichprobenanalyse zur Überprüfung der Güte des Klassifikators.....	71
4.3	Gesamtauswertung highD-Datensatz.....	75
5	Diskussion	81
6	Zusammenfassung.....	83
	Abbildungsverzeichnis	i
	Tabellenverzeichnis.....	v
	Literaturverzeichnis.....	vii
	Anhang	xiii

Abkürzungsverzeichnis

FAS	Fahrerassistenzsysteme
HAF	Hochautomatisiertes Fahren
ISO	International Organization for Standardization
PAS	Publicly Available Specification
SOTIF	Safety Of The Intended Functionality
BAST	Bundesanstalt für Straßenwesen
ACC	Adaptive Cruise Control
LCA	Lane Change Assist
SAE	Society of Automotive Engineers
EBA	Emergency Break Assist
PEGASUS	Projekt zur Etablierung von generell akzeptierten Gütekriterien, Werkzeugen und Methoden sowie Szenarien und Situationen zur Freigabe hochautomatisierter Fahrfunktionen
GIDAS	German In-Depth Accident Study
FPS	Frames per Second
HAC	Hierarchisch-agglomeratives Clustering
PAM	Partitioning Around Medoids
HMM	Hidden Markov Model
SV	Surrounding Vehicle
OEM	Original Equipment Manufacturer
ROI	Region of Interest
SA	Sicherheitsabstand

EKA	Entwurfsklasse
FBA	Follow Drive Behind Approach
HIGHD	Highway Drone
FSW	Fahstreifenwechsel

Formelzeichen

Formelzeichen	Einheit	Beschreibung
k	-	Anzahl Cluster k-means
$Cut - off$	-	Schnitthöhe Dendrogramm (HAC)
SA	[m]	Sicherheitsabstand
t_{SD}	[s]	Zeit für Berechnung des Sicherheitsabstands
d	-	Index Fahrtrichtung [ew, we, all]
ew	-	Fahrtrichtung east-west
we	-	Fahrtrichtung west-east
all	-	Beide Fahrtrichtungen
i	-	Index Fahrzeug
i_d	-	Fahrtrichtungsspezifischer Index Fahrzeug
$N_{veh,d}$	-	Fahrtrichtungsspezifische Anzahl Fahrzeuge im Datensatz
j	-	Index Frame
N_{frame}	-	Anzahl an Frames der Aufzeichnung
$N_{frames,act,i}$	-	Fahrzeugspezifische Anzahl aktive Frames
$x_{min,i,j}$	[m]	Fahrzeugheck minimale x-Position (Bounding-Box)
$y_{min,i,j}$	[m]	Fahrzeugoberkante minimale y-Position (Bounding-Box)
l_i	[m]	Fahrzeuglänge (Bounding-Box)
b_i	[m]	Fahrzeugbreite (Bounding-Box)
$x_{i,j}$	[m]	Fahrzeugposition x-Richtung (Fahrzeugmitte)
$y_{i,j}$	[m]	Fahrzeugposition y-Richtung (Fahrzeugmitte)
x_{max}	[m]	Maximale x-Position aller Fahrzeuge der Aufzeichnung
x_{trans}	[m]	x-Position der Transformationsachse zur Fahrtrichtungsumkehr
$y_{trans,d}$	[m]	Fahrtrichtungsspezifische y-Position der Transformationsachse zur y-Adaption der Aufzeichnungen

Formelzeichen

$v_{x,i,j}$	[m/s]	Fahrzeuggeschwindigkeit x-Richtung
$v_{y,i,j}$	[m/s]	Fahrzeuggeschwindigkeit y-Richtung
$a_{x,i,j}$	[m/s ²]	Fahrzeugbeschleunigung x-Richtung
$a_{y,i,j}$	[m/s ²]	Fahrzeugbeschleunigung y-Richtung
$x_{pred,i,j}$	[m]	Fahrzeugposition x-Richtung Prädiktion (Fahrzeugmitte)
$y_{pred,i,j}$	[m]	Fahrzeugposition y-Richtung Prädiktion (Fahrzeugmitte)
$v_{x,pred,i,j}$	[m/s]	Fahrzeuggeschwindigkeit x-Richtung Prädiktion
$v_{y,pred,i,j}$	[m/s]	Fahrzeuggeschwindigkeit y-Richtung Prädiktion
t_{pred}	[s]	Vektor der Prädiktionszeitpunkte
$n_{frame,pred}$	-	Anzahl an prädizierten Frames
k_{SD}	-	Faktor Berechnung Region-of-Interest
$l_{ROI,front,i,j}$	[m]	Länge Region-of-Interest vor dem Ego-Fahrzeug
$l_{ROI,back,i,j}$	[m]	Länge Region-of-Interest hinter dem Ego-Fahrzeug
b_{coll}	[m]	Breite Kollisionsbereich
l_{coll}	[m]	Länge Kollisionsbereich
$l_{coll,fact}$	-	Faktor zur Berechnung von l_{coll}
$v_{y,lim,LC}$	[m/s]	Grenzwert Quergeschwindigkeit für Spurwechsel
N	-	Populationsgröße der Stichprobe
z	-	Konfidenzniveau der Stichprobe
e	-	Fehlermarge der Stichprobe
p	-	Erwartete Standardabweichung der Stichprobe
S	-	Anzahl an Datenpunkten für k-means
s_n	-	Eintrag n der Menge S
R^d	-	d-dimensionaler Raum
N_k	-	Anzahl an Datenpunkten in Cluster k
C_k	-	Cluster k
c_k	-	Clusterzentrum Cluster k
$C_{\bar{k}}$	-	Nächstgelegenes Cluster zu einem Datenpunkt (euklidische Distanz)

1 Einleitung

Leistungsstarke Rechner, optimierte Sensortechnik und kontinuierlich steigendes Know-how sind vorrangige Gründe für die rasanten Entwicklungen im Feld der Fahrerassistenzsysteme (FAS), sowie den Realisierungsansätzen im Bereich der Hoch- bis Vollautomatisierung von Fahrzeugen [1]. Trotz unzähliger Maßnahmen zur Absicherung der automatisierten bzw. autonomen Fahrfunktionen von Seiten der Entwickler, obliegt es am Ende immer noch externen technischen Prüfeinrichtungen, die Sicherheit der Systeme zu untersuchen und die Verwendung der neuen Technologien im alltäglichen Leben freizugeben. Entsprechend der rapiden Fortschritte in Zusammenhang mit autonomen Fahrfunktionen müssen zertifizierte Prüfinstitutionen ebenso zügig qualitativ sichere Verfahren und Methodiken zur Funktionsanalyse und -absicherung derartiger Systeme aufbauen.

1.1 Motivation

Da es nicht möglich sein wird, jedes erdenkliche Verkehrsszenario im Zuge realer Versuchsfahrten zu überprüfen, werden Ansätze verfolgt, die Absicherung auf wenige, jedoch besonders komplexe Situationen zu beschränken. Dabei wird die Annahme getroffen, dass einfachere Szenarien einer definierten Klasse keine Schwierigkeiten für ein autonomes Fahrzeug darstellen, sofern jener Testfall der Klasse, der mit der höchsten Komplexität eingestuft wird, vom autonomen System bewältigt wird.

Somit werden Absicherungsmethoden autonomer Fahrzeuge in Zukunft neben bereits bestehenden und etablierten Überprüfungen der funktionalen Sicherheit (ISO 26262) [2, S. 91-99], einen immer bedeutender werdenden Anteil an Tätigkeiten zur Validierung der Sollfunktionssicherheit, der SOTIF (engl. Safety Of The Intended Functionality), einbeziehen (ISO PAS 21448) [3, S. 8-9].

Übergeordnetes Ziel ist dabei stets, das sicherere Verhalten der Systeme unter allen erdenklichen Umständen, beziehungsweise potentiellen Einsatzsituationen zu gewährleisten. In anderen Worten – Maximierung des Anteils an garantiert sicher bewältigbaren Situationen und Minimierung des Bereichs an unbekanntem und somit risikobehafteten Verkehrslagen bezogen auf die Gesamtheit aller potentiell eintretenden Szenarien [4, S. 16].

Abbildung 1.1 verdeutlicht diesen Zusammenhang und repräsentiert anschaulich, welche Bedeutung dem Aufbau eines validen Testkatalogs mit möglichst hohem Abdeckungsgrad des breiten Spektrums an potentiellen Verkehrssituationen daher zugeschrieben wird.

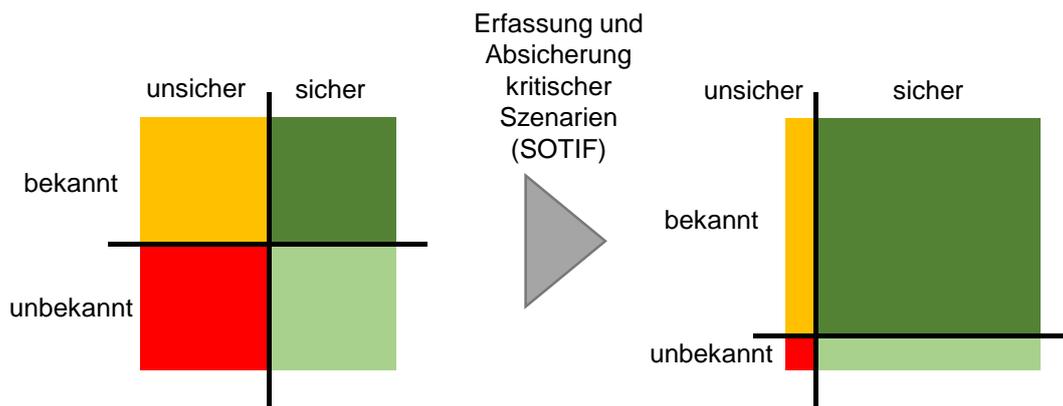


Abbildung 1.1: Reduktion unbekannter und damit risikobehafteter Szenarien durch Integration von SOTIF-Aktivitäten [4, S. 16].

1.2 Aufgabenstellung und Zielsetzung

Aufgabe der vorliegenden Arbeit ist es, das Fundament für ein stabiles Absicherungsprogramm zu schaffen, mit dessen Hilfe automatisierte Fahrfunktionen in potentiell eintretenden Verkehrssituationen getestet und deren Reaktionen analysiert werden können.

Neben dem Fokus einiger Hersteller auf der Entwicklung autonomer Transportsysteme für den urbanen Bereich, spezialisieren sich andere in erster Linie auf die Realisierung eines vollständig autonomen Autobahnpiloten. Die vorliegende Arbeit befasst sich primär mit letztgenanntem Einsatzgebiet, weswegen als Grundlage für das zu entwickelnde Absicherungswerkzeug ein Datensatz aus Aufzeichnungen des realen Verkehrsgeschehens auf deutschen Autobahnen herangezogen wird.

Diesen gilt es zu verarbeiten, wobei sich ein wesentlicher Teil der Arbeit auf die Aufbereitung der Rohdaten in eine einheitliche und zum weiteren Gebrauch geeignete Darstellungsform bezieht. Des Weiteren soll genannter Datensatz anhand expliziter Kriterien untersucht werden, um zeitlich terminierte, besonders relevante und möglichst komplexe Situationen im Sinne einzelner Verkehrsszenarien zu extrahieren. Dabei werden teilweise Methodiken aus dem Feld der Clusteranalyseverfahren, sowie partiell aus dem Bereich des maschinellen Lernens angewandt, um beschriebenes Ziel zu erreichen. Die Szenarien sollen anschließend in einem regelbasierten Klassifikationsschritt logischen Gruppen zugeordnet werden, wodurch als finale Ausgabe des zu entwickelnden Werkzeugs ein intuitiver und einfach zu handhabender Testkatalog für folgende Absicherungsverfahren resultieren soll.

Die verwendeten Methoden und eingesetzten Parameter sollen im Zuge der Ergebnisanalyse auf Korrektheit untersucht und deren jeweiliger Einfluss auf die Zusammenstellung des Testkatalogs eruiert werden. Demnach steht die Reduktion des Umfangs des vorgelegten Datensatzes bei gleichzeitiger Erhöhung des Wissens über inhärente Merkmale und Muster innerhalb der Daten im Mittelpunkt der vorliegenden Arbeit.

1.3 Gliederung der Arbeit

Eine ausführliche Untersuchung des aktuellen Stands der Technik leitet die vorliegende Arbeit im folgenden Abschnitt ein. Dabei werden insbesondere die Grundlagen der verwendeten Verfahren und Methoden zur Datenverarbeitung aufgezeigt und zusätzlich aktuelle Anwendungen der vorgestellten Techniken, vorrangig im Bereich der Analyse von Verkehrsdaten, erörtert.

Im darauf anschließenden Kapitel wird die in der vorliegenden Arbeit umgesetzte Methodik präsentiert. Kern dieses Abschnitts ist die Vermittlung der wesentlichen Prozessschritte und deren Einfluss auf die Resultate des Tools.

Kapitel 4 befasst sich mit der Auswertung und Analyse der gewonnenen Ergebnisse. Zum einen soll dabei auf die Untersuchung des Testkatalogs bei Verarbeitung der gesamten, zur Verfügung stehenden Datenbank eingegangen werden. Zum anderen soll anhand einer repräsentativ gewählten Teilmenge der Rohdaten der Einfluss der verwendeten Parameter auf das Endergebnis untersucht werden.

Im Abschnitt 5 erfolgt die kritische Diskussion des implementierten Verfahrens und eine objektive Beurteilung der Ergebnisse und deren Nutzen.

Abgeschlossen wird die vorliegende Arbeit mit einer kurzen Zusammenfassung und Anregungen zur möglichen Weiterentwicklung des Werkzeugs, sowie einem Ausblick bezüglich des Einsatzpotentials des entwickelten Katalogs an absicherungsrelevanten Verkehrsszenarien.

2 Stand der Technik

Kapitel 2 der Arbeit befasst sich mit dem aktuellen Stand der Technik im Feld der Absicherung autonomer Fahrzeuge, wobei der Fokus der folgenden Ausführungen auf Methoden und Ansätzen zur Erzeugung relevanter Testfälle für die Evaluierung genannter Systeme gelegt wird. Zu Beginn dieses Abschnitts werden die Automatisierungsgrade nach der Definition der Bundesanstalt für Straßenwesen (BASt) vorgestellt und erörtert, in welchem Bereich daraus sich die vorliegende Arbeit einordnen lässt. Darauf folgt ein Einblick in aktuell gängige Verfahren zur Validierung von FAS, wobei in diesem Unterkapitel die wesentlichen Unterschiede zur Absicherung von Systemen des hochautomatisierten Fahrens (HAF) aufgezeigt werden. Der sich daraus ergebende Vorteil szenarienbasierter Testansätze auf Realdatenbasis leitet zum anschließenden Unterkapitel, der Vorstellung des highD-Datensatzes, über. Das Kapitel 2 wird durch die Erörterung von Clusteranalyse- und Klassifikationsmethoden zur Verarbeitung umfangreicher Datenbanken komplementiert, wobei im Speziellen der Bezug zu Anwendungen dieser Verfahren in Verbindung mit Verkehrsdaten hergestellt wird. Darauf basierend wird der Stand der Technik mit der Ableitung des zentralen Forschungsschwerpunkts der vorliegenden Arbeit abgeschlossen.

2.1 Grade der Automatisierung

Die Notwendigkeit nach mehr Sicherheit im Straßenverkehr und die immer präsenter werdende Kundenanforderung nach erhöhtem Komfort und Connectivity haben insbesondere in den vergangenen Jahren zu einer deutlichen Zunahme an Assistenzsystemen im Kraftfahrzeug geführt [5, S. 1]. Neben mittlerweile standardmäßig verbauten passiven Sicherheitssystemen wie Airbags, verbreiten sich zunehmend aktive Systeme wie die ACC (Adaptive Cruise Control) und der LCA (Lane Change Assist) vorrangig im Bereich der Mittel- und Oberklasse [6].

Die Funktionen aktiver Assistenzsysteme werden dabei über das vernetzte Zusammenspiel diverser Sensoren und Aktoren ermöglicht. Sensoren übernehmen für ein System in diesem Zusammenhang die Aufgabe der Wahrnehmung von Fahrzeugzuständen und Umgebungseinflüssen, während über die Aktorik ein gezieltes Beeinflussen von Zustandsgrößen ermöglicht wird [7, S. 206]. Obwohl moderne Assistenzsysteme und insbesondere elektronischen Komponenten derselben mit Ausfallraten zwischen $2 \cdot 10^{-9}/h$ und $5 \cdot 10^{-9}/h$ [8, S. 263-264] höchsten Sicherheitsanforderungen genügen, ist der menschliche Fahrer bis dato immer noch gesetzlich dazu verpflichtet die Systeme permanent zu überwachen und als Rückfallebene zur Verfügung zu stehen.

Die Tendenz aktueller Entwicklungen zielt jedoch verstärkt darauf ab, den Menschen zunehmend aus der Verantwortung zu nehmen, beziehungsweise die Fahraufgabe zu einem stetig wachsenden Anteil an das technische System zu delegieren. Unter Betrachtung der Verteilung aller für die sichere Erfüllung der Fahraufgabe notwendigen Tätigkeiten zwischen dem

Assistenzsystem und dem Menschen, lassen sich nach der Bundesanstalt für Straßenwesen (BASt) [1] 5 Grade der Automatisierung festlegen. Die genaue Einteilung der Automatisierungslevel und die exakte Beschreibung des notwendigen Tätigkeitsbereichs des Menschen ist Tabelle 2.1 zu entnehmen.

Tabelle 2.1: Grade der Automatisierung nach BASt [1]

Level	Nomenklatur	Fahraufgabe des Fahrers nach Automatisierungsgrad
4	Vollautomatisiert	<p>Das System übernimmt Quer- und Längsführung vollständig in einem definierten Anwendungsfall</p> <ul style="list-style-type: none"> • Der Fahrer muss das System dabei nicht überwachen • Vor dem Verlassen des Anwendungsfalles fordert das System den Fahrer mit ausreichender Zeitreserve zur Übernahme der Fahraufgabe auf • Erfolgt dies nicht, wird in den risikominimalen Systemzustand zurückgeführt • Systemgrenzen werden alle vom System erkannt; das System ist in allen Situationen in der Lage, in den risikominimalen Systemzustand überzugehen
3	Hochautomatisiert	<p>Das System übernimmt Quer- und Längsführung für einen gewissen Zeitraum in spezifischen Situationen</p> <ul style="list-style-type: none"> • Der Fahrer muss das System dabei nicht überwachen • Bei Bedarf wird der Fahrer zur Übernahme der Fahraufgabe mit ausreichender Zeitreserve aufgefordert • Systemgrenzen werden alle vom System erkannt. Das System ist nicht in der Lage, aus jeder Ausgangssituation den risikominimalen Zustand herbeizuführen
2	Teilautomatisiert	<p>Das System übernimmt Quer- und Längsführung (für einen gewissen Zeitraum und/oder in spezifischen Situationen)</p> <ul style="list-style-type: none"> • Der Fahrer muss das System dauerhaft überwachen • Der Fahrer muss jederzeit zur vollständigen Übernahme der Fahrzeugführung bereit sein
1	Assistiert	<p>Fahrer führt dauerhaft entweder die Quer- oder die Längsführung aus. Die jeweils andere Fahraufgabe wird in gewissen Grenzen vom System ausgeführt</p> <ul style="list-style-type: none"> • Der Fahrer muss das System dauerhaft überwachen • Der Fahrer muss jederzeit zur vollständigen Übernahme der Fahrzeugführung bereit sein
0	Driver only	<p>Fahrer führt dauerhaft (während der gesamten Fahrt) die Längsführung (Beschleunigen/Verzögern) und die Querführung (Lenken) aus</p>

Bis dato serienreife und zugelassene Assistenzsysteme finden sich ausschließlich in den Automatisierungsstufen 1 und 2 wieder, wobei der Trend diverser Hersteller immer weiter in Richtung Stufe 3 und höher angesetzt wird. Zukunftsvisionen reichen zudem über die hier dargestellte Einteilung hinaus. So erweitert die SAE – Society of Automotive Engineers – die dargestellten Automatisierungsstufen um Level 5, dem vollständig und unter allen Umständen autonom fahrenden Vehikel [9].

Insbesondere der Sprung von Stufe 2 auf 3 ist mit erschwerenden Einflüssen verbunden. Der Fahrer darf ab Level 3 während des Betriebs des Automatisierungssystems seine Aufmerksamkeit abwenden, muss im Bedarfsfall jedoch übernahmefähig sei. Dies macht die technische Realisierung und rechtliche Absicherung deutlich komplexer, als es bei Systemen niedrigerer Stufen der Fall ist [10, S. 34]. Dementsprechend sind zuverlässige und vor allem allgemeingültige Absicherungsmethoden essenziell für die Freigabe von Systemen ab Stufe 3.

2.2 Absicherung autonomer Fahrfunktionen

Die wesentliche Herausforderung bezüglich einer Zertifizierung und Freigabe von Automatisierungssystemen ab Stufe 3 (Tabelle 2.1), sowie vollständig autonomen Fahrfunktionen (Stufe 5 nach [9]) im Automotive-Bereich liegt insbesondere in einer zurzeit unausgereiften Absicherungsstrategie und fehlenden Prüfverfahren. Diese Problematik kann mitunter dadurch begründet werden, dass die vorgesehene und sichere Funktionalität von Automatisierungssystemen höherer Grade nicht mehr allein auf der einwandfreien Funktion der verwendeten Hardware- und Softwarekomponenten beruht, sondern vielmehr eine ausgeprägte Abhängigkeit von potentiell stark variierenden Umgebungseinflüssen aufweist.

Dementsprechend muss zur Beurteilung der Sicherheit komplexer Automatisierungssysteme ab Level 3 sowohl ein Nachweis der einwandfreien Funktion auf Komponentenebene erfolgen (ISO 26262 [2, S. 91-99]), als auch der situative Sicherheitsaspekt bestätigt werden. Dieser Umstand hat mitunter zur Ausarbeitung und Einführung der ISO PAS 21448 geführt. Der Fokus dieser Norm liegt auf der Erkennung von Risiken durch das Abweichen eines Systems von einer gewünschten Sollfunktion (SOTIF) aufgrund äußerer, nicht spezifizierter Einflüsse oder einem inkorrekten Gebrauch [3, S. 8-9]. Insbesondere im Bereich der Fahrzeugtechnik ist die Einhaltung der SOTIF eines automatisierten Systems von größter Wichtigkeit, um Fehlfunktionen und damit verbundene Risiken für den Fahrer, das Fahrzeug und das Umfeld zu vermeiden.

Der Nachweis der Gebrauchssicherheit ist somit ein zentrales Kernelement eines Absicherungsprozesses für hochautomatisierte und autonome Fahrzeuge. Bis dato wurde in Zusammenhang mit FAS genannte Bestätigung der korrekten Funktionalität der Systeme zu einem großen Anteil durch Fahrzeugerprobungen am Prüfgelände und abschließenden Realfahrten im Straßenverkehr erreicht [11, S. 58].

Für Systeme ab Level 3 ist dieser Ansatz jedoch aus zeitlichen und wirtschaftlichen Gründen nur bedingt einsetzbar, weshalb die Notwendigkeit besteht, neuartige Freigabekonzepte zu entwickeln. Im Folgenden wird dahingehend auf die Nachteile etablierter Validierungsmethoden in der Automobilbranche in Zusammenhang mit der Testung von HAF eingegangen und ein vielversprechender Ansatz vorgestellt, der gegenüber den gängigen Methoden wesentliche Vorteile bietet.

Validierungsmethoden

Repräsentativ und beobachtbar, ökonomisch und reproduzierbar – das sind nur einige der Stichworte die mit Freigabekonzepten von Automatisierungssystemen im Fahrzeugsektor in Verbindung gebracht werden sollen. In oberster Ebene müssen Validierungsmethoden demnach sowohl Kriterien bezüglich eines stabilen Verfahrens, als auch bezüglich einer möglichst hohen Effizienz erfüllen [12, S. 447-448].

Zwei der gebräuchlichsten Ansätze in Verbindung mit der Systementwicklung und -absicherung von FAS sind das V-Modell und der Dauerlauf.

V-Modell

Das V-Modell repräsentiert ein etabliertes Vorgehensmodell zur Planung und Durchführung von Entwicklungsprozessen vorrangig im Feld elektrischer/elektronischer Komponenten und Softwareprodukten [13]. Das Modell illustriert eine V-förmige Anordnung definierter, sequenziell abzuarbeitender Phasen während der Produktentwicklung, wobei sich der linke Ast auf die Entwurfs- und Entwicklungsaktivitäten fokussiert, während auf der gegenüberliegenden Seite des Modells korrespondierende Test- und Validierungstätigkeiten eingetragen werden. Entsprechend Abbildung 2.1 wird somit eine Phase links im V-Modell über die Definition geeigneter Testfälle, beziehungsweise Testszenarien mit den zugehörigen Absicherungstätigkeiten verknüpft. Die vorliegende Arbeit lässt sich auf der Validierungsebene des Modells eingliedern. Nach FIRESMITH [14] ist der strukturierte Aufbau und die logische Abarbeitungsreihenfolge der einzelnen Phasen des V-Modells vorteilhaft im Sinne einer kontrollierbaren systematischen Vorgehensweise. Jedoch sieht der Autor in dieser strikt vorgegebene und einzuhaltende Struktur auch einen negativen Aspekt des Vorgehensmodells, weil diese Eigenschaft mit einer hohen Inflexibilität und hohen Kosten bei späten Adaptionen von Elementen früher Entwurfsphasen einhergeht.

Abbildung 2.1 verdeutlicht dahingehen, dass bereits während der frühen Entwicklungsphase eines Systems möglichst alle Anforderungen an dieses festzulegen sind, um darauf basierend geeignete Szenarien abzuleiten, die am Ende der Prozesskette zur Validierung überprüft werden sollen.

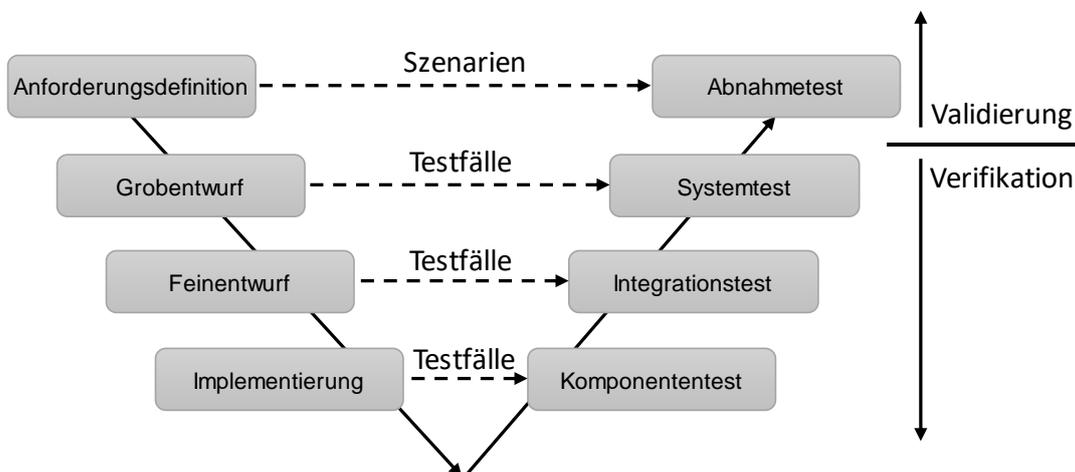


Abbildung 2.1: Vereinfachte Darstellung des V-Modells nach [15]

Nach WILHELM ET AL. [10, S. 98-99] stellt diese Notwendigkeit an eine vollständige Anforderungsdefinition jedoch das größte Problem bezüglich der validen Absicherung eines Systems dar. Die Autoren begründen das damit, dass bereits 55 % aller Fehler eines Systems auf unzureichende und falsche Anforderungen in der frühen Entwurfsphase zurückzuführen sind. Die Tatsache, dass es für immer komplexer werdende FAS und vor allem autonome Fahrzeugsysteme nahezu unmöglich ist, bereits zu Beginn des Produktentwicklungsprozesses eine vollständige Systemanforderung und korrespondierende Absicherungsszenarien definieren zu können, verdeutlicht den Bedarf an einer agileren und dennoch strukturierten Validierungsmethode in Verbindung mit der Entwicklung komplexer FAS und Systemen aus dem Feld des HAF.

Dauerlauf

Der Dauerlauf im realen Straßenverkehr oder auf einem abgesicherten Testgelände stellt eine weitere im Automotive-Sektor verbreitete Test- und Validierungsmethode dar. Im Vergleich zu vorangegangenen Verfahren, in dem die Validierungstätigkeit auf explizit definierten Testszenarien auf Basis von Anforderungsdefinitionen beruht, setzt das Prinzip des Dauerlaufs auf das Argument, über die Quantität an Versuchen und Testfahrten den Sicherheitsnachweis unter allen möglichen Einsatzbedingungen des Systems erbringen zu können. Es ist naheliegend, dass diese Form des Funktionalitätsnachweises insbesondere durch einen hohen zeitlichen und finanziellen Aufwand gekennzeichnet ist.

Trotz dieser Nachteile ist nach MAUERER [12, S. 446-447] die Realfahrt ein unerlässlicher Bestandteil des Absicherungsprozesses von FAS und in weiterer Folge autonomen Fahrzeugen. Dennoch weist der Autor explizit darauf hin, dass die alleinige Nutzung dieser Form der Fahrzeugerprobung in Verbindung mit der steigenden Komplexität der Systeme zunehmend an ihre Grenzen stößt. So führt MAURER [12, S. 446] exemplarisch den Umfang zur Absicherung eines automatischen Notbremssystems (EBA-Emergency Break Assist) an, der mit bis zu zwei Millionen Testkilometer verdeutlicht, dass bereits die Freigabe aktueller FAS mit einem erheblichen Zeit- und Kosteneinsatz verbunden sind. Übertragen auf die Absicherung der wesentlich komplexeren HAF-Systeme steigt dieser Anteil an durchzuführenden Testkilometern auf ein in angemessener Zeit nicht zu realisierendes Vielfaches des genannten Umfangs an, weshalb es unumgänglich ist, alternative Methoden zur Validierung dieser Systeme zu entwickeln.

Die vorangegangenen Ausführungen veranschaulichen die Schwächen der gängigen Validierungsverfahren im Bereich der Fahrzeugabsicherung. Die Notwendigkeit zur Entwicklung einer alternativen Methode, die diese Nachteile überwindet und in weiterer Folge als Standard zur Absicherung von autonomen System herangezogen werden soll, hat zur Initiierung des Projekts PEGASUS geführt, das im Folgenden vorgestellt wird.

PEGASUS-Projekt

Das PEGASUS-Projekt (Projekt zur Etablierung von generell akzeptierten Gütekriterien, Werkzeugen und Methoden sowie Szenarien und Situationen zur Freigabe hochautomatisierter Fahrfunktionen) beschreibt der Bedeutung seines Namens entsprechend ein Forschungsprojekt, dessen übergeordnetes Ziel die Entwicklung einer Methodik inklusive der notwendigen Werkzeuge für die Testung automatisierter Fahrzeugfunktionen war [16, S. 5]. Die im Projekt erarbeitete Methode kann anhand Abbildung 2.2 erörtert werden.

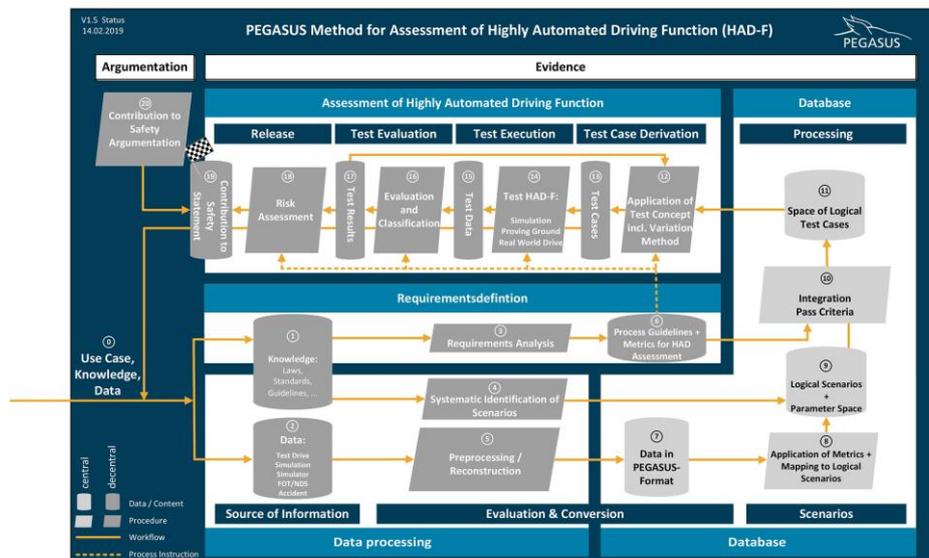


Abbildung 2.2: Schematischer Ablauf der im Zuge des PEGASUS-Projekts entwickelten Methode [17]

Die Methode hinter dem PEGASUS-Projekt setzt sich aus 20 Einzelschritten zusammen. Diese repräsentieren die Elemente der vier zentralen Bereiche:

- Datenverarbeitung (Data processing)
- Datenbankaufbau aus logischen Testszenarien (Database)
- Anforderungsdefinition (Requirementsdefinition)
- Systembewertung der HAF (Assessment of Highly Automated Driving Functions)

Die vorliegende Arbeit lässt sich in die beiden erstgenannten Bereiche einordnen, wobei insbesondere die Punkte 2, 4 und 5, sowie 8 und 9 aus Abbildung 2.2 von zentraler Bedeutung sind. Genannte Teilschritte des Projekts zielen auf die geeignete Verarbeitung von Rohdaten zur Erzeugung einer Datenbank an Testszenarien für die folgenden Absicherungsprozesse im Bereich der Systembewertung ab.

Damit bildet der beschriebene Ausschnitt aus der PEGASUS-Methode das Fundament für ein Validierungsverfahren von FAS und HAF, indem auf Basis definierter Rohdatensätze der notwendige Input für den Bereich der Systembewertung geschaffen wird. Ein entscheidender Aspekt bezüglich der Qualität der erzeugten Datenbank an Testfällen ist mitunter in Punkt 2 aus Abbildung 2.2 dargestellt, weil die Wahl der Rohdaten, beziehungsweise das Verfahren zur Generierung selbiger einen grundlegenden Einfluss auf die charakteristischen Eigenschaften der resultierenden Testfälle hat. Im Wesentlichen besteht dahingehend die Möglichkeit Rohdaten für die Weiterverarbeitung aus Simulationen zu extrahieren, oder auf Datenbanken realer Verkehrsaufzeichnungen zurückzugreifen.

Die vorliegende Arbeit befasst sich ausschließlich mit Daten der zweitgenannten Quelle in Form des im Folgenden vorgestellten „highD“ Datensatzes (highway drone dataset) [18, S. 2118-2125]. Diese Wahl kann derart argumentiert werden, dass der zu erzeugenden Testkatalog einen

möglichst hohen Abdeckungsgrad an potentiell eintretenden Verkehrssituationen aufweisen soll, wobei diese Anforderung bei der synthetischen Erzeugung von Testfällen zumeist nur näherungsweise und unter erheblichem Aufwand erfüllt werden kann.

2.3 HighD-Datensatz

Institutionen und Forschungseinrichtungen, wie beispielsweise die Bundesanstalt für Straßenwesen im Projekt GIDAS (German In-Depth Accident Study) [19], oder das Deutsche Zentrum für Luft- und Raumfahrt [20] befassen sich mit der Erfassung und Aufbereitung von umfangreichen Datensätzen realer Unfallsituationen und allgemeiner Verkehrsgeschehnissen. Aufgrund ihrer Realitätsnähe bilden die Datenbanken dieser Projekte eine besonders geeignete Basis für Absicherungswerkzeuge autonomer Systeme im Automotive-Bereich

Neben Methoden, wie etwa der Verlegung von Induktionsschleifen [21, S. 13-15], stellen insbesondere Drohnenaufnahmen eine effiziente und aussagekräftige Varianten der Datenerhebung des realen Verkehrsgeschehens im Bereich mehrspuriger Autobahnen dar.

2.3.1 Verfahren der Datenerfassung

KRAJEWSKI ET AL. [18, S. 2118-2125] haben sich auf eine solche Datenerhebung spezialisiert. Dabei wird das Verkehrsgeschehen (vorrangig Autobahnabschnitte, mittlerweile auch innerstädtischer Verkehr [22]) mit Hilfe hochauflösender Kameras an einer Drohne in Vogelperspektive aufgezeichnet (Abbildung 2.3). Anschließend werden Bildverarbeitungsalgorithmen eingesetzt, um die Daten zu glätten, die einzelnen Verkehrsteilnehmer im Bildfluss zu lokalisieren und darüber deren Positionen, Geschwindigkeiten, Beschleunigungen und weitere Kenngrößen zu extrahieren [18, S. 2122].

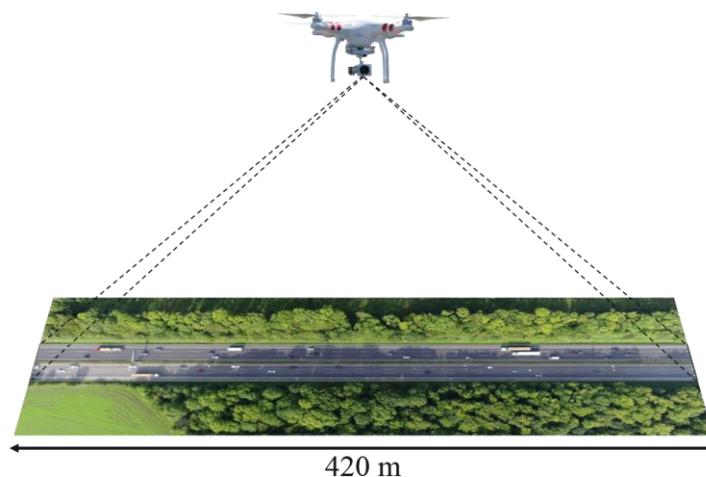


Abbildung 2.3: Prinzip der Datenerfassung mittels Drohne [18]

Neben der kostengünstigen Erhebungsmethode zeichnet sich dieses Verfahren vor allem durch die kurze Dauer zur Erstellung umfangreicher Datensätze aus. Zudem ist die Genauigkeit der Aufzeichnungen besonders hervorzuheben, die mit 25 Frames per Second (fps) in einer angemessenen Frequenz und zudem mit einer Auflösung von 10x10 cm pro Pixel ausreichend exakt dargestellt werden. Dementsprechend repräsentiert der hier beschriebene „highD-Datensatz“ ein geeignetes Ausgangsprodukt für ein auf Realdaten basierendes Absicherungstool [18, S. 2121].

2.3.2 Aufbau des Datensatzes

Der HighD-Datensatz setzt sich aus 60 Aufnahmen von durchschnittlich 17 Minuten Dauer zusammen, wobei die einzelnen Aufnahmen an jeweils einem von sechs Autobahnteilstücken in der näheren Umgebung von Köln durchgeführt wurden [18, S. 2121]. Jede der 60 Dateien beinhaltet die ermittelten Werte aller Fahrzeuge für beide Fahrrichtungen, wobei sowohl 2-, als auch 3-spurige Autobahnabschnitte und der Einbezug unterschiedlicher Verkehrsflusssituationen (Stau, Freifahrten, etc.) eine hohe Variabilität der Daten verspricht. Diese Eigenschaften deuten a priori auf einen großen Nutzen für die Verwendung in realdaten-basierten Absicherungswerkzeugen hin. Die Aufzeichnungen beinhalten aufgrund der variierenden Rahmenbedingungen mit hoher Wahrscheinlichkeit eine Vielzahl an Testfällen, die über ein synthetisches Erstellungsverfahren womöglich nicht in Betracht gezogen werden. Repräsentativ für den Aufbau der Dateien sind in Tabelle 2.2 die Einträge der ersten 5 Frames eines Fahrzeugs dargestellt [23].

Die Identifikation der einzelnen Fahrzeuge geschieht über eine fortlaufende, von der Fahrtrichtung unabhängige Nummerierung (ID) auf Basis des Erscheinungszeitpunkts am Autobahnabschnitt (Abbildung 2.4). Neben den in Tabelle 2.2 aufgelisteten Einträgen beinhalten die Datensätze weitere, für die vorliegende Arbeit zum größten Teil jedoch irrelevante Informationen, auf die deshalb nicht näher eingegangen wird. KRAJEWSKI ET AL. [18, S. 2121] stellen neben den Daten zusätzlich geeignete Werkzeuge für das Einlesen derselben in MATLAB, sowie zur Visualisierung der Fahrzeuge und Trajektorien zur Verfügung.

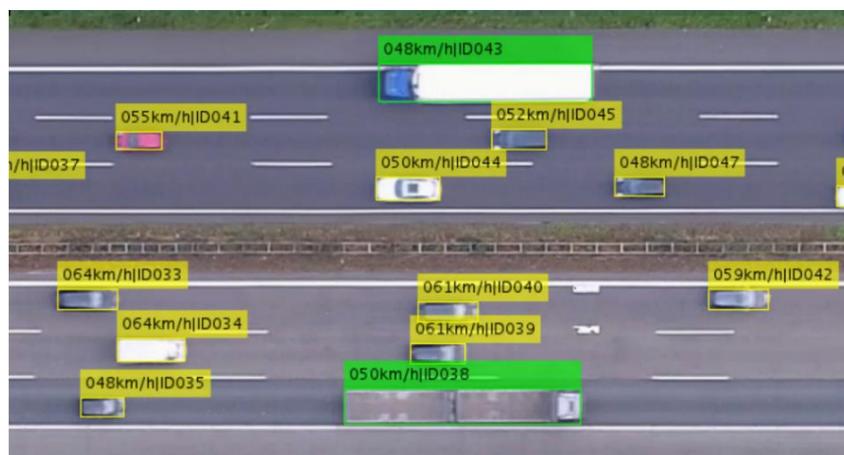


Abbildung 2.4: Visualisierung der Aufzeichnungen. Mittels Bildverarbeitungsalgorithmen werden die Verkehrsteilnehmer extrahiert [18, S. 2123].

Tabelle 2.2: Ausschnitt aus dem highD Datensatz in Form der ersten 5 Frames des Fahrzeugs mit ID 1 [23].

Frame	ID	x [m]	y [m]	Fzg-Länge [m]	Fzg-Breite [m]	x-Geschwindigkeit [m/s]	y-Geschwindigkeit [m/s]	x-Beschleunigung [m/s ²]	y-Beschleunigung [m/s ²]
1	1	362.26	21.68	4.95	2.22	40.85	0.00	0.30	0.00
2	1	363.73	21.68	4.95	2.22	40.87	0.00	0.30	0.00
3	1	365.27	21.68	4.95	2.22	40.88	0.00	0.31	0.00
4	1	366.83	21.68	4.95	2.22	40.89	0.00	0.32	0.00
5	1	368.42	21.68	4.95	2.22	40.90	0.00	0.32	0.00

Nachteilig ist in Zusammenhang mit der Aufzeichnungsmethode die mit durchschnittlich 420m stark beschränkte Länge des erfassten Autobahnabschnitts zu erwähnen. Der aus den Daten gewonnenen Median der Aufenthaltsdauer aller Fahrzeuge auf den Abschnitten liegt bei 13.6s [18, S. 2121]. Folglich ist das Erfassen geeigneter Testfälle auf diese kurze Zeitspanne begrenzt und stellt daher eine der wesentlichsten Herausforderungen im Umgang mit dem vorliegenden Datensatz dar.

Trotz dieses Mankos spricht besonders ein Argument für die vorgestellte Form der Datenerhebung. Der Aufnahmeprozess selbst hat keinen Einfluss auf das Verkehrsgeschehen, wodurch Verhaltensmuster des realen Verkehrs in unverfälschter Form erfasst werden können. Wird dazu beispielsweise die Rekonstruktion von Verkehrssituationen über die Auswertung von Aufzeichnungen aus Testfahrten verglichen, so sind die dadurch ermittelbaren Situationen stets durch die Wirkung des eigenen Fahrverhaltens verzerrt.

Der Umfang und die inhaltliche Struktur der highD Daten gelten in Verbindung mit der vorliegenden Arbeit als besonders geeignet zur Erstellung eines validen, realdatenbasierten Testkatalogs. Dahingehend bedarf es der Anwendung geeigneter Verfahren zur Bildung zeitlich terminierter, zusammengehöriger Cluster, um aus der ursprünglichen Form der Rohdaten als minutenlange Aufzeichnungen der Autobahnabschnitte kurze Verkehrsszenarien extrahieren zu können. Im folgenden Abschnitt werden dahingehend gebräuchliche Methoden vorgestellt.

2.4 Clusteranalysemethoden

Clusteranalysemethoden stellen einen Kernbereich des Data Minings dar und fassen eine spezielle Klasse von Algorithmen zusammen, die in erster Linie zur merkmalsbasierten Gruppierung von Informationen herangezogen werden. Die Algorithmen bilden die Gruppen auf Basis von Gemeinsamkeiten innerhalb der Daten und teilen sie im Gegenzug bei zu starken Abweichungen charakteristischer Größen. Dies lässt eine Einteilung der Clusteranalyseverfahren in den

Bereich des unüberwachten (unsupervised) Lernens zu, weil die Ergebnisse nicht auf externen Vorgaben wie beispielsweise Klassenbezeichnungen beruhen [24, S. 2].

In der Wissenschaft und Literatur finden sich zahlreiche Algorithmen aus dem Feld der Clusteranalysemethoden wieder, wobei die Wahl des geeignetsten Verfahrens auf die eigentliche Problemstellung bezogen werden muss, um sinnvolle Ergebnisse generieren zu können.

Wu [24, S. 3-4] führt dahingehend drei Punkte als vorrangige Begründung der Vielzahl an Algorithmen an:

1. Clustering stellt ein unerschöpfliches kombinatorisches Problem dar
2. Es gibt keine allgemeingültige Theorie in Zusammenhang mit Clusteranalysen
3. Clusteranalyse birgt ein gewisses Maß an Willkür bezüglich der Definition eines Clusters und dem Verständnis der charakteristischen Eigenschaften der Daten.

Trotz der Menge an unterschiedlichen Verfahren lassen sich gewisse Methoden aufgrund ähnlicher Arbeitsweisen zu Klassen zusammenfassen. Ein Überblick über die möglichen Prinzipien der Clusteranalyse ist nach Wu [24, S. 3-5] in Tabelle 2.3 aufgeführt, wobei im Folgenden repräsentativ zwei der gängigsten Verfahren vorgestellt werden.

Tabelle 2.3: Klassen der Clusteranalysemethoden nach Wu [24, S. 3-5]

Klasse	Beschreibung
Prototypen-basierter Algorithmus	<ul style="list-style-type: none">• Separate Cluster über Prototypen repräsentiert• Inputdaten bauen tatsächliches Cluster um Prototypen herum auf• Vertreter: K-means, Self-Organizing Maps (SOM)
Dichte-basierter Algorithmus	<ul style="list-style-type: none">• Gebiete mit hoher Dichte an Datenpunkten bilden Cluster• Gebiete mit niedriger Dichte an Datenpunkten trennen separate Cluster voneinander• Verwendung bei Vorhandensein von Ausreißern und verrauschten Daten• Vertreter: DBSCAN, DENCLUE
Graphen-basierter Algorithmus	<ul style="list-style-type: none">• Datenpunkte als Knoten in Graph dargestellt• Abstand zwischen Knoten als Gewicht der Verbindungslinie erfasst• Cluster werden über die Ermittlung verknüpfter Subgraphen gebildet• Vertreter: Hierarchisch-agglomeratives Clustering (HAC)
Hybride Algorithmen	<ul style="list-style-type: none">• Kombination von zwei oder mehr Clustering-Algorithmen• Vertreter: Chameleon

2.4.1 K-means Clustering

Ein besonderes Augenmerk gilt dem speziellen Clusteranalyseverfahren „k-means“, das einen des ältesten, jedoch bis heute vielfach verwendeten Algorithmus zur Gruppierung von Daten repräsentiert [24, S. 7]. Wie aus Tabelle 2.3 zu entnehmen wird das k-means Verfahren den prototypen-basierten Algorithmen zugeordnet. Vorrangiges Ziel der Methode ist das Auffinden einer definierten Anzahl k Cluster innerhalb einer Menge an Datenpunkten mit der Voraussetzung der überlappungsfreien Gruppierung.

Zu Beginn der k-means Methode wird vom Anwender eine feste Anzahl an zu findenden Clustern angegeben. Die geeignete Einstellung dieses Parameters k ist von großer Bedeutung bezüglich der zu erwartenden Ergebnisse des Algorithmus. Am Beispiel in Abbildung 2.5 ist zu erkennen, dass die Wahl von $k = 2$ ein sinnvolles Resultat liefert. Wird hingegen ein größeres k gewählt, so werden die dargestellten Cluster weiter unterteilt, womit die intuitive und im dargestellten Beispiel naheliegende Lösung verloren geht. In vielen Anwendungen des k-means Verfahrens ist die Wahl der Clusteranzahl jedoch nicht anhand einer subjektiven Abschätzung möglich. In Folge dessen haben sich Methoden etabliert, die dahingehend eine Hilfestellung anbieten. Der Vollständigkeit halber wird in diesem Zusammenhang auf die Ellbogen-Methode und das Silhouetten-Verfahren hingewiesen [25].

Der in Abbildung 2.5 dargestellte Algorithmus weist in einem iterativen Prozess alle Datenpunkte (1) k Clustern auf Grundlage des Abstands zu deren Zentren zu (2) und positioniert danach die Clusterzentren auf Basis der zugewiesenen Punkte neu (3). Sobald alle Punkte einem Cluster zugehörig sind und weitere Iterationen des Verfahrens keine Änderungen der Zuweisungen hervorrufen, ist das Clustering beendet (4) [24, S. 7]. Die exakten Schritte der Methode werden basierend auf [26, S. 102-103] im Folgenden erörtert.

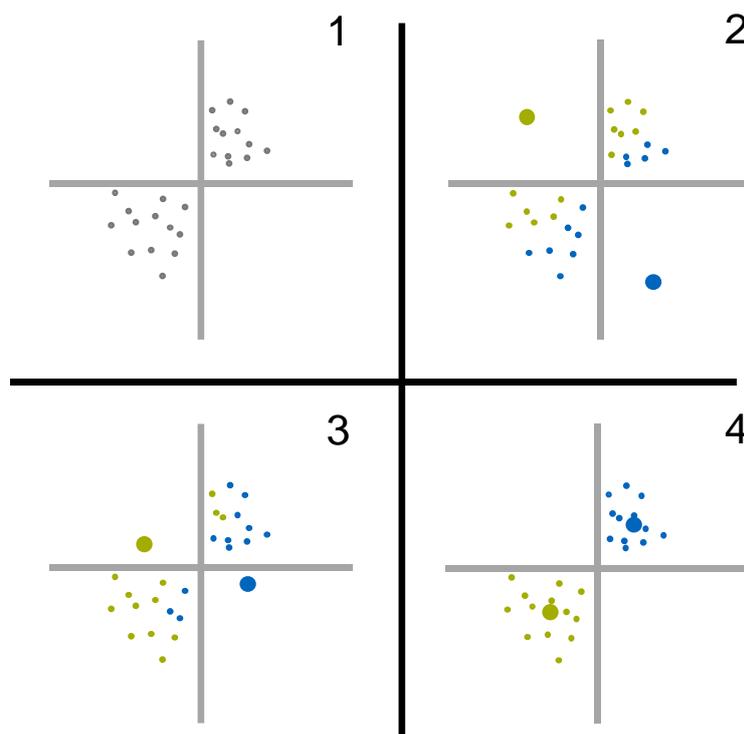


Abbildung 2.5: Beispiel für den Ablauf des *k-means* Clustering nach [27, S. 34]

Ausgangspunkt ist eine Menge S aus N Vektoren s_n , mit $S = \{s_n : s_n \in \mathbb{R}^d \quad n = 1, \dots, N\}$ im d -dimensionalen Raum. Abhängig von der Anwendung stellen die Vektoren Werttupel des Datensatzes dar, wobei deren Einträge charakteristische Größen des zu untersuchenden Problems repräsentieren. So können die Größen *Länge* und *Masse* von Fahrzeugen in einem 2-dimensionalen Koordinatensystem, wie in Abbildung 2.5 gezeigt, eingetragen werden, wodurch mittels k-means Clustering beispielsweise eine Gruppierung in 2 Cluster (Pkw und Lkw) denkbar ist.

Schritt 1:

Im ersten Schritt des Verfahrens werden die N Elemente $s_n \in S$ jeweils dem nächstgelegenen der k Clusterzentren zugewiesen. Die initialen Positionen dieser Zentren werden unter der Bedingung der Überlappungsfreiheit entweder bewusst oder zufällig vorgegeben. Jedes Zentrum gruppiert somit eine Menge an N_k Vektoren in der Untergruppe C_k (Abbildung 2.5, Schritt 2)

Schritt 2:

Die Positionen der Zentren c_k der gebildeten Subgruppen C_k werden neu berechnet:

$$c_k = \frac{1}{N_k} \sum_{s_n \in C_k} s_n \quad k = 1, \dots, K \quad (2.1)$$

Es handelt sich offensichtlich um die Bildung des Mittelwerts aller dem betrachteten Cluster zugeordneten Datenpunkte (Abbildung 2.5, Schritt 3).

Schritt 3:

Im nächsten Schritt erfolgt die Korrektur der Zuweisung der Datenpunkte zu den Clusterzentren (Abbildung 2.5, Schritt 3). $\forall s_n \in S$ wird jenes Cluster $C_{\bar{k}}$ ermittelt, das die Euklidische Distanz zum Clusterzentrum minimiert. Anstelle der euklidischen Norm kann die Ähnlichkeit zweier Datenpunkte auch über andere Metriken berechnet werden:

$$s_n \in C_{\bar{k}} \Leftrightarrow \|s_n - c_{\bar{k}}\|_2 = \min_{k=1, \dots, K} \|s_n - c_k\|_2 \quad (2.2)$$

Schritt 4:

Die einzelnen Datenpunkte s_n werden den neuen Clustern $C_{\bar{k}}$ zugeordnet (Abbildung 2.5, Schritt 3).

Schritt 5:

Die Schritte 2-4 werden so lange iteriert, bis sich die Zuordnung der Datenpunkte s_n zu den Clustern nicht mehr ändert (Abbildung 2.5, Schritt 4).

Das k-means Clustering findet aufgrund der universellen Einsetzbarkeit in verschiedensten Wissenschaftszweigen Anwendung, weil es mit wenig Aufwand auf diverse Problemstellungen übertragbar ist. Beispiele für den konkreten Gebrauch des Verfahrens, insbesondere im Bereich der Analyse von Verkehrsdaten, werden am Ende des Kapitels 2.4 angeführt.

2.4.2 Hierarchisch-Agglomeratives Clustering (HAC)

Eine alternative Herangehensweise an ein Clustering von unbekanntem Daten, das explizit keine initiale Vorgabe der Clusteranzahl verlangt, ist das hierarchisch-agglomerative Verfahren [28, S. 71]. Dabei wird zu Beginn der Methode jedes Datenelement als separates Cluster betrachtet. In einem iterativen Prozess werden einzelne, graphisch als Knotenpunkte dargestellte Cluster, die sich bezüglich einer explizit gewählten Eigenschaft der Daten am meisten ähneln, fusioniert. Demensprechend verringert sich die Anzahl an Cluster von Iteration zu Iteration, bis alle Datenpunkte zu einer übergeordneten Gruppe vereint sind. Die Ähnlichkeit, oder auch Unähnlichkeit der Daten bezüglich der betrachteten Eigenschaft kann dabei als Abstand der Datenpunkte zueinander interpretiert werden. Die Fusion der Datenpunkte oder -gruppen wird über die Anwendung einer speziellen Verbindungsmethode, beispielsweise einem der beiden Verfahren single-linkage oder average-linkage realisiert [29, S. 148]. Das Ergebnis der Methode kann graphisch in Form eines Dendrogramms dargestellt werden, wobei mit Hilfe eines einstellbaren Wertes, in der Literatur wiederholt als Schwellwert oder Cut-off Wert bezeichnet [30, S. 241], unähnliche Daten aufgetrennt und ähnliche zu gemeinsamen Clustern gruppiert werden können [28, S. 73-77].

Beispiel Clusteranalyse

Im Folgenden werden anhand eines exemplarischen Beispiels die Prinzipien der beiden vorgestellten Clusterverfahren und der Einfluss der Wahl der Clusteranzahl, sowie des Cut-off Werts in Kombination mit dem hierarchisch-agglomerativen Clustering (HAC) gezeigt. Repräsentativ für eine konkrete Anwendung wird die Gruppierung von Fahrzeugdaten (Länge und Masse) untersucht. Tabelle 2.4 stellt den im Beispiel verwendeten, fiktiven Datensatz dar. Dieser beinhaltet 9 Fahrzeuge unterschiedlicher Masse und Länge, wobei zur Veranschaulichung Werte aus den vorab bekannten Gruppen Oberklasse, Mittelklasse und Kleinwagen angenommen werden.

Tabelle 2.4: Datenpunkte Clusterbeispiel

Fahrzeug-ID	Fahrzeugklasse	Fahrzeuglänge [m]	Fahrzeugmasse [kg]
1	Kleinwagen	3.74	1150
2	Oberklasse	4.81	1580
3	Mittelklasse	4.50	1390
4	Oberklasse	4.75	1595
5	Kleinwagen	4.02	1170
6	Mittelklasse	4.60	1420
7	Kleinwagen	3.92	1200
8	Oberklasse	4.76	1620
9	Mittelklasse	4.68	1400

Abbildung 2.6 stellt die Resultate der k-means Clusteranalyse graphisch dar, wobei bereits in dieser visuellen Aufbereitung die intuitive Einteilung der Datenpunkte in drei Ähnlichkeitsklassen (Cluster) ersichtlich ist. Diese Einschätzung deckt sich auch mit den Ergebnissen des Verfahrens bei einer vorab definierten Clusteranzahl von $k = 3$ (Tabelle 2.5).

Die untere Darstellung in Abbildung 2.6 verdeutlicht die negative Auswirkung einer ungünstigen Wahl der zu findenden Clusteranzahl k . Im vorliegenden Fall wird abhängig von der zufällig gewählten Initialposition der Clusterzentren eine der drei intuitiv erfassten Gruppen zusätzlich unterteilt. Obwohl aufgrund der Trivialität des gewählten Beispiels der geeignetste Wert von k offensichtlich erscheint, kann im Fall umfangreicherer Datensätze diese Wahl deutlich komplexer und die Güte und Aussagekraft der resultierenden Ergebnisse damit maßgeblich beeinflusst werden.

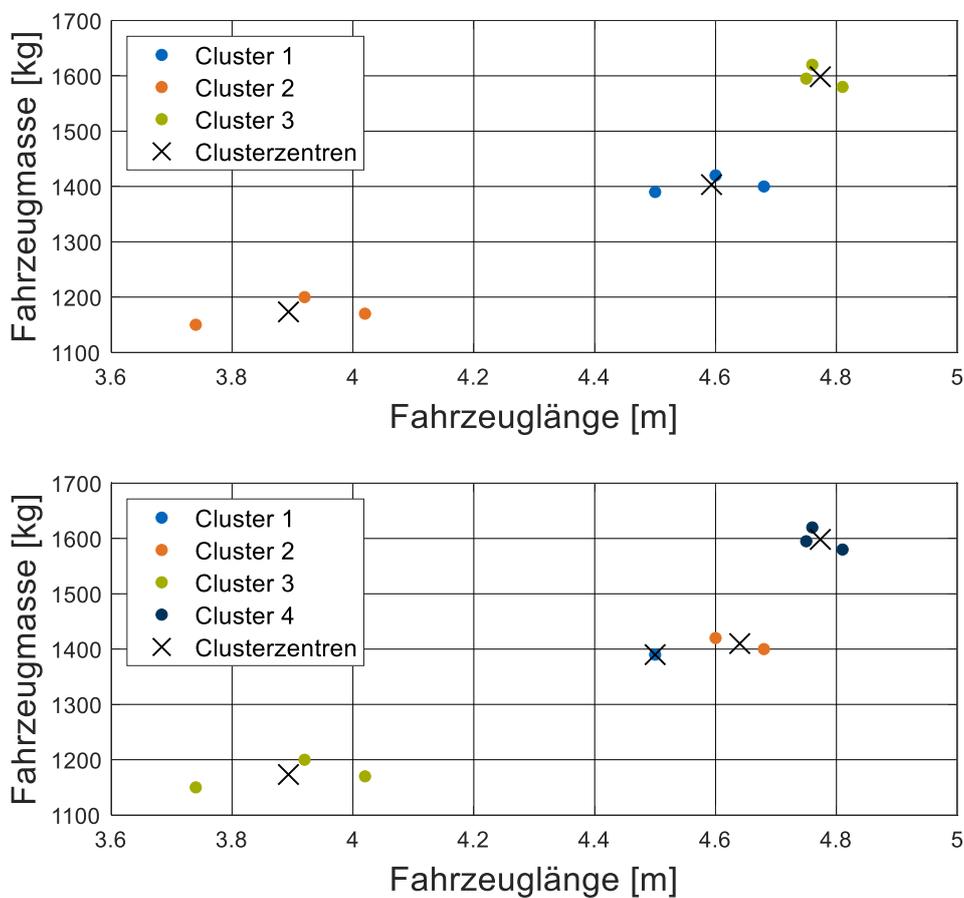


Abbildung 2.6: Ergebnisse k-means mit $k = 3$ (oben) und $k = 4$ (unten)

Tabelle 2.5: Ergebnis k-means bei Wahl $k = 3$

Cluster	Fahrzeugklasse	Clusterzentrum	Fahrzeug-ID
1	Mittelklasse	$x = 4.593\text{m}, y = 1403\text{kg}$	[3, 6, 9]
2	Kleinwagen	$x = 3.893\text{m}, y = 1173\text{kg}$	[1, 5, 7]
3	Oberklasse	$x = 4.773\text{m}, y = 1598\text{kg}$	[2, 4, 8]

Ein ähnliches Problem ergibt sich bei Anwendung des hierarchisch-agglomerativen Clusterings (HAC). Im Vergleich zum k-means Verfahren ist die Anzahl an Cluster nicht vorab zu definieren, jedoch muss zur Bildung der Gruppen im Nachhinein ein konkreter Cut-off Wert angegeben werden. Visuell lässt sich die Problematik anhand Abbildung 2.7 nachvollziehen. Darin sind die Resultate des HAC in Form von Dendrogrammen dargestellt. Diese repräsentieren die hierarchische Aufteilung der Datenpunkte, die sich im Zuge des Clusteringprozesses einstellt.

Die oberste horizontale Linie, die Wurzel des Dendrogramms, stellt ein einziges Cluster aus allen Datenpunkten dar. Im Vergleich dazu entsprechen die nummerierten Einträge auf der Abszisse Clustern aus jeweils einem einzelnen Datenpunkt (Blätter). Die horizontalen Kanten zwischen den vertikalen Linien repräsentieren Fusionen einzelner Datenpunkte oder Subcluster, wobei der Ordinatenwert dieser Verbindungslinien dem Abstand zwischen den fusionierten Elementen entspricht [31, S. 77-78].

Der Schnitt durch die Dendrogramme an konkreten Ordinatenwerten (Cut-off) ermöglicht die Einteilung der darunterliegenden Elemente aufgrund besonderer Ähnlichkeiten bezüglich des untersuchten Merkmals zu gemeinsamen Clustern. In Zusammenhang mit dem exemplarisch gewählten Problem repräsentiert der Massenunterschiede zwischen den Fahrzeugen die charakteristische Größe (Abbildung 2.7). Analog der erwähnten Problematik bezogen auf die Wahl von k für das k-means Verfahren, hat auch die subjektive Festlegung des Cut-off Werts einen entscheidenden Einfluss auf die Güte der Resultate des HAC.

Wird für das vorliegende Problem beispielsweise der Wert 250 kg gewählt (Abbildung 2.7 unten), so reduziert sich die Anzahl der extrahierbaren Cluster auf 2. Wird der Cut-off Wert hingegen im Bereich von 150 kg angesetzt (Abbildung 2.7 oben), so ergibt sich die erwartete Gruppierung der Daten in 3 getrennte Cluster. Der Vergleich mit den Resultaten der k-means Methode bestätigt die Übereinstimmung der Ergebnisse beider Verfahren und damit die Gültigkeit des intuitiv festgelegten Cut-offs (Tabelle 2.6).

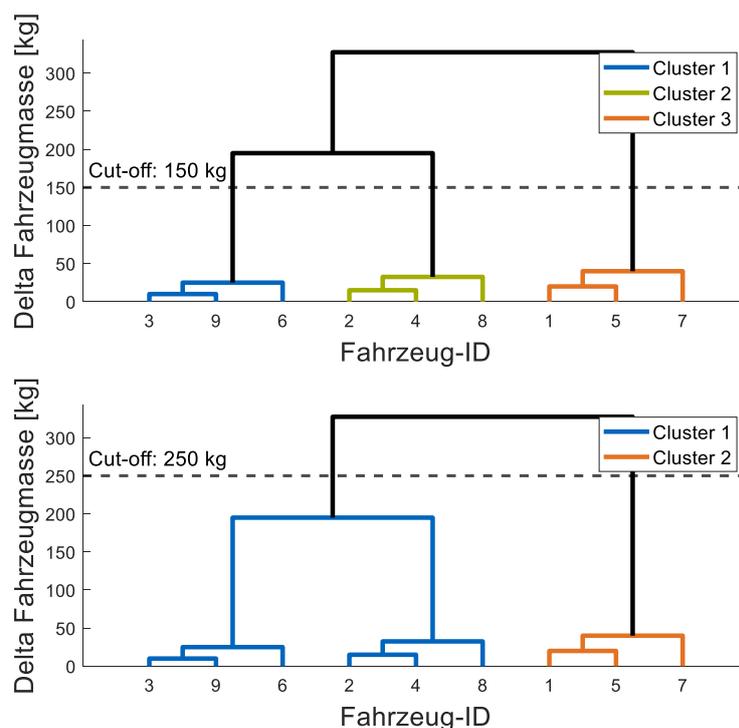


Abbildung 2.7: Dendrogramme des HAC auf Basis der Fahrzeugmassen. Die subjektive Wahl des Cut-off Werts zu 150 liefert das intuitiv erwartete Ergebnis von 3 Clustern.

Tabelle 2.6: Vergleich der Ergebnisse beider Clusterverfahren

Cluster	Fahrzeugklasse	Elemente k-means	Elemente HAC
1	Mittelklasse	[3, 6, 9]	[3, 6, 9]
2	Kleiwagen	[1, 5, 7]	[1, 5, 7]
3	Oberklasse	[2, 4, 8]	[2, 4, 8]

Anwendung der Clusteranalyse im Verkehrswesen

Clusteranalysemethoden finden in den unterschiedlichsten Wirtschafts- und Wissenschaftszweigen Anwendung. So werden beispielsweise Marketingstrategien über die Auswertung von Daten des spezifischen Kaufverhaltens einzelner Kundenschichten mit Hilfe diverser Clusteringverfahren entwickelt. Ebenso werden genannte Methoden in der Biologie und Medizin zur Differenzierung von Bakterienstämmen, Krankheitsbildern oder psychologischen Verhaltensmustern eingesetzt [32, S. 4].

Von besonderem Interesse für die vorliegende Arbeit sind spezielle Anwendungen der Clusteranalyseverfahren zur Auswertung von Verkehrsaufzeichnungen und Unfalldaten. Vermehrt liegen erwähnte Informationen in Form umfangreicher und unverarbeiteter Datenbanken vor (Beispiel GIDAS [19]), wodurch sich in vielen Fällen kein unmittelbarer Nutzen durch die Daten ergibt, beziehungsweise kein direkter Erkenntnisgewinn möglich ist. Erst eine gezielte und sinnvolle Verarbeitung erlaubt die Extraktion verwertbaren Wissens aus den Datensätzen. Folglich wird in vielen Wissenschaftsbereichen bei Anwendung von Clusteranalyseverfahren das in Abbildung 2.8 dargestellte Prinzip verfolgt. Ziel ist stets die Reduktion an Datenumfang bei gleichzeitiger Steigerung des Wissens über verborgene Strukturen und Muster innerhalb der Daten [33, S. 41].

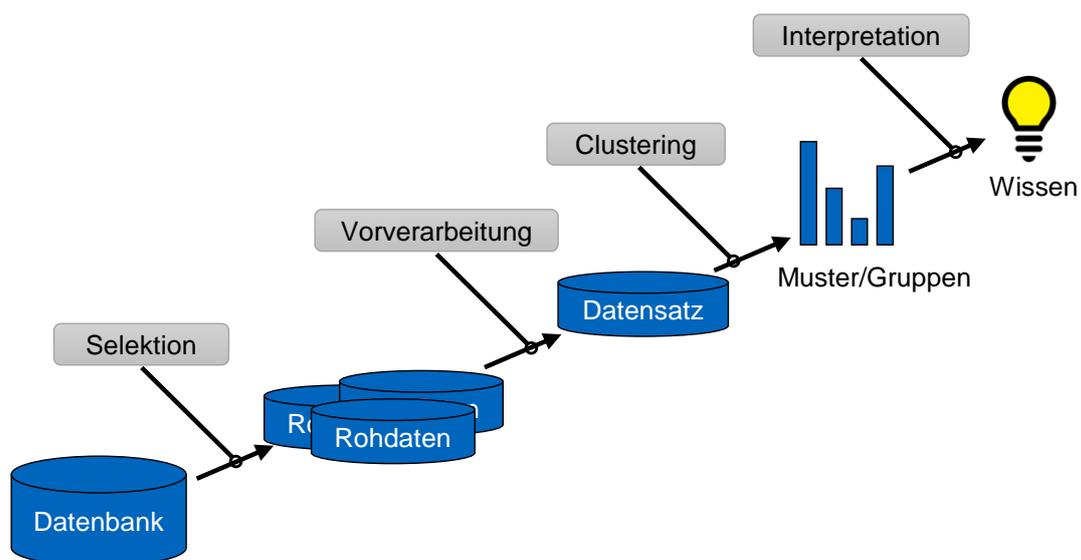


Abbildung 2.8: Prozesskette Clusterverfahren in Anlehnung an [33, S. 41]

WATANABE ET AL. [34] setzen in ihrem Verarbeitungsprozess von realen Verkehrsdaten eine abgewandelte Form des k-means Algorithmus, sowie ein hierarchisch-agglomeratives Clustering zur Extraktion von Testszenarien für die Entwicklung sicherheitsrelevanter Funktionen im Automotive-Bereich ein.

VON BUXHOEVEDEN [35, S. 80] ermittelt ortsbezogen Schwerpunkte vermehrter Unfallgeschehnisse über Clustering von Ereignisdaten, wodurch gezielte Maßnahmen zur lokalen Verbesserung des Verkehrskonzepts abgeleitet und damit eine Reduktion des Unfallpotentials an entsprechenden Orten ermöglicht werden soll.

NECULA [36, S. 279-282] wertet unter Zuhilfenahme der vorgestellten Gruppierungsverfahren GPS-Daten von Fahrzeugen aus, um darin sich wiederholende Muster innerhalb der Entwicklungen des täglichen Verkehrsflusses zu erkennen.

Ein ähnliches Ziel verfolgen KIANFAR ET AL. [37, S. 430-439] in ihrem Ansatz. Unter Kombination der fundamentalen Variablen Verkehrsfluss, Verkehrsdichte und Geschwindigkeit mit unterschiedlichen Clusteringverfahren wird versucht eine geeignete Methode zu etablieren, das Auftreten von unterteilbaren Verkehrssituationen zu erfassen und Datenpunkte auf sinnvolle Weise in die Klassen freier Fluss und Stau zu aggregieren.

LENARD ET AL. [38, S. 1] extrahieren explizit Frontalzusammenstöße von Fahrzeugen mit Fußgängern aus einer Unfalldatenbank und benutzen hierarchische Clusteringverfahren, um auf Basis von Mustern in den Unfallverläufen die Einzelfälle in ähnliche Szenarioklassen zusammenzufassen.

SANDER ET AL. [39, S. 10] beschäftigen sich in ihrer Arbeit mit der Auswertung von Kreuzungsunfällen und versuchen über den Einsatz von Clusteringmethoden homogene Gruppen ähnlicher Unfallmuster zu finden. Die Daten werden insbesondere dahingehend untersucht, ob die Unfälle durch den Eingriff eines automatischen Notbremssystems verhindert hätten werden können. Die Autoren stellen jedoch fest, dass aufgrund einer hohen Variabilität der charakteristischen Unfallgrößen dahingehend keine eindeutige Unterteilung der Datenbank mittels Clusteranalyseverfahren möglich ist.

Ebenso befassen sich NITSCHKE ET AL. [40, S. 140-148] mit der Auswertung von Unfalldaten an Kreuzungen. Über die Resultate des von ihnen verwendeten Clusteringverfahrens PAM (partitioning around medoids) identifizieren sie basierend auf Attributen des Unfalls wie beispielsweise dem maximalen Verletzungsgrad der beteiligten Personen und der geometrischen Ausführung der Kreuzung 13 Unfallklassen für T-förmige Kreuzungen und 6 Klassen für Kreuzungen mit 4 zusammenlaufenden Straßen. Ein Auszug dieser Unterteilungen ist in Abbildung 2.9 zu sehen.

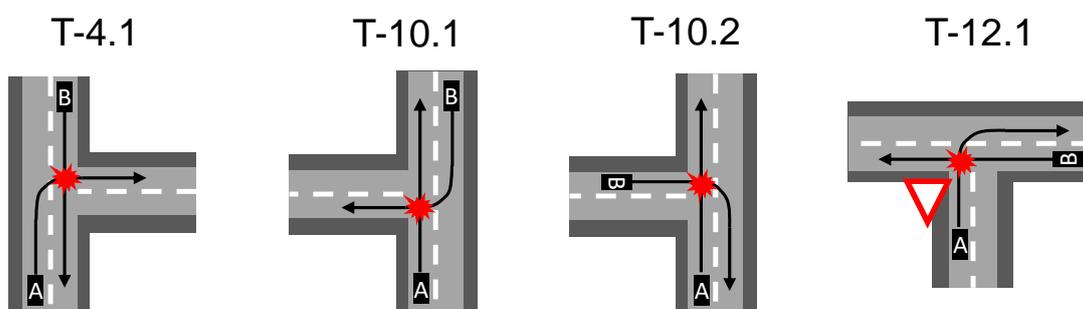


Abbildung 2.9: Auszug der identifizierten Unfallklassen für T-förmige Kreuzungen nach [40, S. 148]

2.5 Prädiktive Methoden zur Kollisionsvorhersage

Aus dem vorangegangenen Abschnitt lässt sich schlussfolgern, dass Clusteranalysemethoden geeignete Verfahren darstellen, aus unübersichtlich großen Datensätzen wertvolle Informationen extrahieren zu können. Neben diesen Analysen werden in der vorliegenden Arbeit weitere Ansätze benötigt, um die Datenmenge zusätzlich zu reduzieren und damit schrittweise näher an das Ziel eines aussagekräftigen Szenarienkatalogs zu gelangen. Eine wesentliche Eigenschaft der Elemente eines derartigen Katalogs zur Absicherung automatisierter Fahrfunktionen ist definitiv, dass die Situationen besonders herausfordernd für ein System sein müssen, um als absicherungsrelevant zu gelten. In anderen Worten ist ein Verkehrsszenario mit Beteiligung eines automatisiert geregelten Fahrzeugs genau dann relevant und damit in den Katalog aufzunehmen, wenn zu einem konkreten Zeitpunkt der Fahrt eine Reaktion des autonomen Systems notwendig wird, um eine Kollision mit anderen Verkehrsteilnehmern zu vermeiden.

In der vorliegenden Arbeit wird ein prädiktives Modell angesetzt, um potentielle Kollisionen zwischen den Verkehrsteilnehmern auf Basis der tatsächlichen Beschleunigungswerte zu erkennen und damit die Einteilung in relevante und irrelevante Situationen zu ermöglichen. Der folgende Abschnitt befasst sich folglich mit den Grundlagen prädiktiver Methoden zur Vorhersage von Fahrzeugkollisionen und führt diverse Einsatzgebiete dieser Verfahren im Automotive-Sektor an.

2.5.1 Grundlagen der Kollisionserkennung

Als Kollision wird im allgemeinen Sprachgebrauch das Zusammenstoßen mindestens zweier Objekte verstanden. Der Begriff findet in den unterschiedlichsten Wissenschaftsbereichen Verwendung. So bezeichnen Astronomen beispielsweise das Zusammentreffen zweier Sterne als Sternenkollision und Chemiker verwenden den Begriff synonym für das Aufeinanderprallen von Teilchen während chemischer Reaktionen.

Mathematisch betrachtet kann eine Kollision als geometrische Überlappung polygonförmiger Objekte verstanden werden [41, S. 274-275]. Selbstverständlich ist eine solche Herangehensweise in Verbindung mit starren oder deformierbaren Körpern nicht zweckmäßig (Verkehrsunfälle). Die vorliegende Arbeit greift jedoch auf eine rein mathematische Kollisionsuntersuchung zurück (fiktive, prädizierte Kollisionen), weshalb ein besonderes Augenmerk auf genannte Definition gelegt wird. In Abhängigkeit der überlappenden Fläche (2-dimensional), beziehungsweise des Schnittvolumens bei Behandlung einer 3-dimensionalen Problemstellung, lässt sich für konkrete Anwendungsfälle der Grad einer Kollision definieren (Abbildung 2.10) [42, S. 50].

Analog kann neben dem Grad einer Kollision auch die Gefahr eines potenziellen Zusammenstoßes während der Annäherung zweier Objekte ermittelt werden. WOENCKHAUS [42, S. 51] umhüllt dazu die Objekte mit eindeutig parametrisierten Volumina (beispielsweise Sphären mit den Radien R_1 , R_2), die auf idente Weise wie die Körper selbst auf Überlappung überprüfbar sind. Dadurch kann die Höhe der Kollisionsgefahr bei Überschneidung zweier Hüllkurven unter denselben argumentativen Gesichtspunkten wie bei der eigentlichen Kollisionsüberprüfung festgestellt werden (Abbildung 2.11).

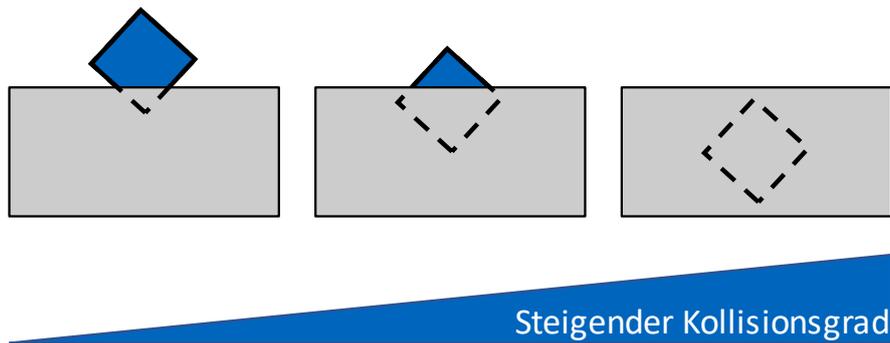


Abbildung 2.10: Polygonüberschneidung zweier Objekte und Grad der Kollision nach [42, S. 50]

Im Bereich der Sicherheitsbewertung von Fahrzeugen wird in der Literatur das Kollisionsrisiko vielfach als Gefahrenpotential einer Verkehrssituation herangezogen. So wird in vielen Fällen eine Kollisionswahrscheinlichkeit berechnet [43], oder ähnlich dem Prinzip in Abbildung 2.11 der minimale Abstand zweier Fahrzeuge als relevante Größe zur Beschreibung der Kritikalität herangezogen [44, S. 907].

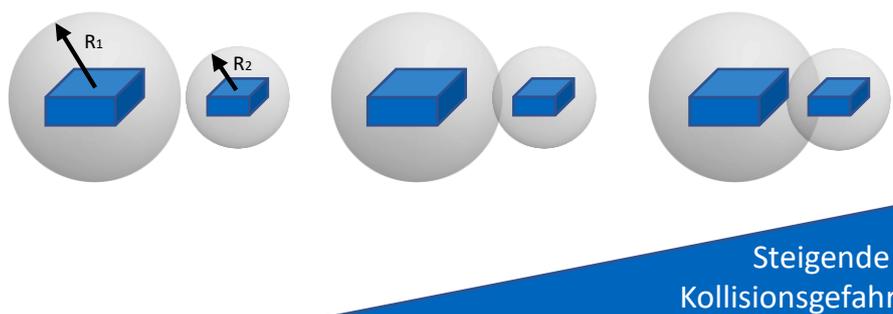


Abbildung 2.11: Erhöhung der Kollisionsgefahr durch zunehmende Überlappung der Hüllvolumen nach [42, S. 51], beispielhaft ausgeführt als Kugelvolumen mit Radien R_1 und R_2 .

2.5.2 Grundlagen zur Fahrzustands-Prädiktion

Eine der wesentlichsten Charakteristika einer testbasierten Absicherungsstrategie autonomer Fahrzeuge ist die Verwendung möglichst relevanter, also sicherheitskritischer Testfälle. Es besteht die Möglichkeit diese herausfordernden Testsituationen unmittelbar aus Unfalldatenbanken wie beispielsweise jener der Studie GIDAS [19] abzuleiten. Diese Daten repräsentieren jedoch in erster Linie kritische Situationen für einen menschlichen Fahrer, weshalb nicht gewährleistet ist, dass sie ebenso relevant für autonome Fahrzeuge sind. In Folge dessen werden in Verbindung mit der Absicherung autonomer Systeme alternative Methoden zur Extraktion sicherheitskritischer Testfälle aus realen Verkehrsdaten bevorzugt.

Eine Möglichkeit dies zu tun ist die Verwendung prädiktiver Verfahren. Dabei werden zukünftige Positionen und Dynamikwerte eines Fahrzeugs (Geschwindigkeit, Beschleunigung) über ein

Vorhersagemodell ermittelt und darauf beruhend eine Abschätzung des Gefahrenpotentials der prädizierten Zustände getätigt. Darüber lassen sich Verkehrssituationen bezüglich ihrer Relevanz für die Aufnahme in einen Testkatalog eines Absicherungswerkzeugs beurteilen, wodurch auch unfallfreie Realdata in die Sicherheitsüberprüfung aufgenommen werden können.

In der Literatur finden sich diverse Beispiele zur Anwendung des dargestellten Ansatzes. So verweist FRESE [45, S. 11] in seiner Arbeit auf einen vermehrten Einsatz diverser Varianten des Kalman-Filters zur Fahrzeug-Prädiktion und erwähnt Markov-Ketten, die eine Analyse der Erreichbarkeit zukünftiger Zustände erlauben.

Im Vergleich dazu nutzen BLASCHKE ET AL. [46, S. 4] fahrzeuginterne Zustandsgrößen wie den Fahrpedalwert und externe Indikatoren wie die Differenzgeschwindigkeit zwischen zwei Fahrzeugen, um auf Grundlage eines Fuzzy-Systems Überholmanöver vorherzusagen.

Trotz der Unterschiede zwischen den Prädiktionsmethoden haben alle Verfahren eine wesentliche Charakteristik gemein. Jede Methode nutzt zur Vorhersage eines gewissen Zeithorizonts unterschiedliche Fahrzeugparameter und Zustandsgrößen, jedoch können die Verfahren die Realität nur näherungsweise prädizieren. Um dennoch möglichst genaue Vorhersagen treffen zu können, werden in Abhängigkeit der gewünschten Prädiktionsdauer mehr oder weniger komplexe Modelle angesetzt. Nach FRESE [45, S. 11] liefert für kurzfristigen Prädiktionen ein auf konstanten Handlungen basierendes Modell ausreichend exakte Ergebnisse. Zeitlich darüber hinausgehende Schätzungen benötigen für zuverlässige Resultate deutlich mehr Informationen, wie beispielsweise den zukünftigen Straßenverlauf.

Unabhängig vom gewählten Prädiktionsmodell und der Vorhersagedauer können die vorgestellten Verfahren mitunter zur Abschätzung des zukünftigen Risikopotentials einer Verkehrssituation genutzt werden. Gerade in Zusammenhang mit der Trajektorienplanung eines autonomen Fahrzeugs ist dies von besonderer Bedeutung, um sich bei der Generierung beziehungsweise Wahl der auszuführenden Trajektorie für die bestmögliche und somit sicherste Option zu entscheiden. Dazu entwerfen DAMEROV ET AL. [47, S. 703-710] in ihrer Arbeit prädiktive Risikokarten (Abbildung 2.12), mit deren Hilfe eine Risikoanalyse bezüglich variabler Reaktionsmöglichkeiten des betrachteten autonomen Fahrzeugs durchgeführt werden kann.

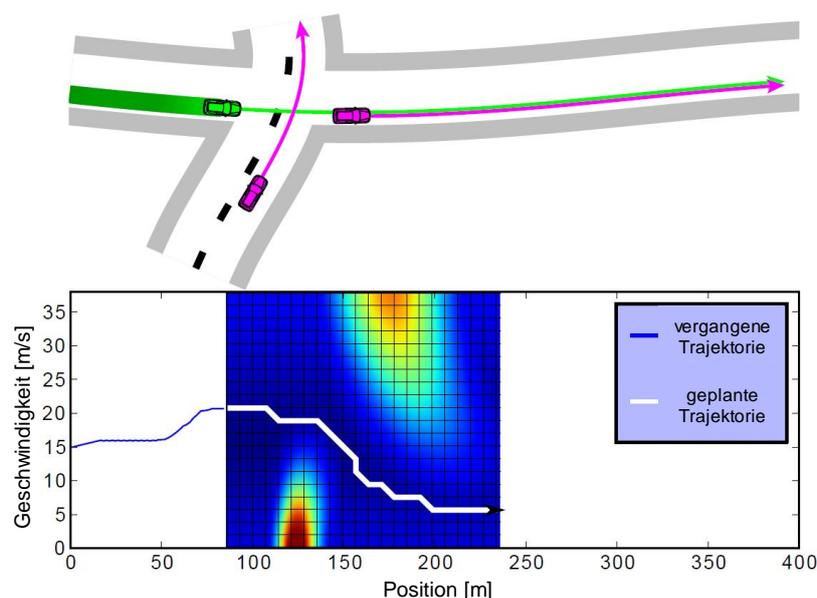


Abbildung 2.12: Risk-Map einer Kreuzungssituation aus [47, S. 706].

Beispielsweise lässt sich durch Anpassung der zukünftigen Trajektorie (weiße Linie) das Kreuzungsszenario in Abbildung 2.12 möglichst sicher bewältigen. Hierzu wird die Planung der Trajektorie des betrachteten autonomen Fahrzeugs (grün) dahingehend optimiert, dass sie in einem Bereich mit möglichst geringem prädiziertem Risiko (blauer Bereich der Risikokarte) verläuft.

Ein vergleichbares Verfahren wenden ALTHOFF ET AL. [48, S. 301-307] in ihrem Konzept zur modellbasierten Kollisionserkennung an. In dieser Arbeit werden ebenfalls geplante Trajektorien eines zentralen Fahrzeugs mit der Vorhersage des Verkehrsgeschehens verknüpft, wodurch eine Aussage über das Kollisionspotential getroffen werden kann. Ähnlich dem vorangegangenen Prinzip, beziehen die Autoren Unsicherheiten der Messungen, sowie die Unschärfe bei der Vorhersage von Handlungen und Reaktionen beteiligter Verkehrsteilnehmer in ihre Berechnungen mit ein, wodurch eine wahrscheinlichkeitsbasierte Aussage über mögliche Zusammenstöße getroffen werden kann [48, S. 301-307].

Die vorgestellten Ansätzen liefern die methodischen Grundlagen für das in der vorliegenden Arbeit entwickelte Prädiktionsmodell zur Einschätzung des zukünftigen Gefahrenpotentials von Verkehrsszenarien. Darüber wird eine Identifikation sicherheitskritischer und somit absicherungsrelevanter Verkehrsszenarien ermöglicht.

2.6 Klassifikation von Verkehrsszenarien

Vorangegangene Erörterungen haben verdeutlicht, dass es mit Hilfe prädiktiver Methoden und anschließenden Überprüfungen von Kollisionswahrscheinlichkeiten in Kombination mit Verfahren zur Bildung ähnlichkeitsbasierter Cluster möglich ist, aus einer unübersichtlichen und umfangreichen Datenbank an realen Verkehrssituationen besonders herausfordernde Szenarien zu extrahieren. Die Resultate dieser Verarbeitungsschritte bilden somit die Kernelemente des gewünschten Szenarienkatalogs. Die Fertigstellung desselben hängt jedoch von einer weiteren Untersuchung der extrahierten Testfälle ab, weil bis zu diesem Punkt der Datenverarbeitung eine konkrete Beschreibung der Szenarien innerhalb der Testfälle fehlt. Anders formuliert repräsentiert ein Katalog ein nach einem bestimmten System geordnetes Verzeichnis von Gegenständen [49]. Folglich lassen sich die Testfälle nur dann systematisch in einem spezifischen Verzeichnis verteilen, wenn einerseits die Klassen des Katalogs definiert sind und andererseits die Szenarien diesen Klassen zugeordnet werden können. Eine derartige Gruppierung ermöglicht es, den Katalog für Absicherungsuntersuchungen autonomer Fahrzeuge zu verwenden und dabei gezielt Testfälle aus einzelnen Klassen, die besondere Herausforderungen für die Algorithmen der Automatisierungssysteme darstellen, zu überprüfen.

So ist beispielsweise die Fahrt auf einer freien Spur für einen Abstandsregelautomaten (ACC) wenig problematisch, während ein plötzlich vor dem betrachteten Vehikel einscherendes Umgebungsfahrzeug eine besondere Schwierigkeit für genanntes Assistenzsystem darstellt. Bereits für die Absicherung der Reaktionen des hier repräsentativ gewählten Beispiels zeigt sich der große Nutzen eines passend strukturierten Szenarienkatalogs, um eine einfache, schnelle und effektive Untersuchung des Systems zu ermöglichen.

In der Literatur werden unterschiedliche Ansätze verfolgt, um die erwähnte Klassifikation der Verkehrssituationen zu realisieren. Ein vielversprechendes Verfahren wird von DEO ET AL. [50, S. 132] vorgeschlagen, wobei in dieser Ausführung verborgene Markovmodelle (Hidden Markov

Model - HMM) eingesetzt werden, um Manöver auf Basis der Trajektorien der beteiligten Fahrzeuge zu erkennen. Dazu fusionieren die Autoren drei Module, die zum einen die aktuellen Fahrmanöver der Fahrzeuge im Umfeld des Ego-Vehikels auf Basis der vergangenen Trajektorien und der momentanen Fahrzustände erkennen und zum anderen darauf aufbauend eine Prädiktion der Positions- und Fahrzustandsverläufe realisieren. Über Muster in den Trajektorienverläufen und -prädiktionen leiten die Autoren 10 Klassen von Verkehrsszenarien einer Autobahnfahrt ab.

Die von DEO ET AL. [50, S. 132] verwendeten verborgenen Markov Modelle ermöglichen allgemein die Darstellung dynamischer Systeme über die Modellierung endlich viele Zustände und der Transitionen zwischen ihnen [51, S. 81-83]. In einem einfachen Markov Modell (Abbildung 2.13) sind in jedem Zeitschritt Transitionen in benachbarte Zustände (bspw. V1 zu V2 via a_{12}) oder der Verbleib im aktuellen Zustand möglich (bspw. V2 via a_{22}).

Der Unterschied eines einfachen Markov Modells zu den erwähnten HMM zeigt sich darin, dass der relevante Bereich eines betrachteten Systems nicht unmittelbar mess- oder beobachtbar ist und daher in Form verborgener Zustände (V_i) im Modell vorliegt. Ein weiterer Anteil des Systems, der mit erstgenanntem Bereich verkoppelt ist, weist diese beschränkende Eigenschaft hingegen nicht auf und kann daher im Modell als beobachtbare Zustände (B_k) integriert werden. Über die Betrachtung der bekannten Zustände B_k kann in weiterer Folge auf die verborgenen Zustände zurückgeschlossen werden. Dies ist möglich, weil die Zustände B_k mit definierten Wahrscheinlichkeiten b_{ik} von den verborgenen Zuständen V_i beeinflusst werden (Verkopplung). Somit kann über die Beobachtung messbarer Zustände eine Aussage über die systeminternen, verborgenen Zustände getroffen werden (Abbildung 2.13).

Die wahrscheinlichkeitsbasierte Struktur des Modells ermöglicht bei günstiger Wahl der Transitionparameter eine Vorhersage zukünftiger Zustände. Im Kontext mit Verkehrsszenarien lassen sich so Fahrmanöver frühzeitig erkennen. Repräsentativ für das Prinzip hinter einem verborgenen Markov Modell kann die Erkennung der Intention eines Spurwechsel- oder Überholmanövers (verborgene Information) durch Beobachtung der tatsächlich messbaren Zustandsgrößen Beschleunigung und Fahrzeugausrichtung erwähnt werden [51, S. 81-83].

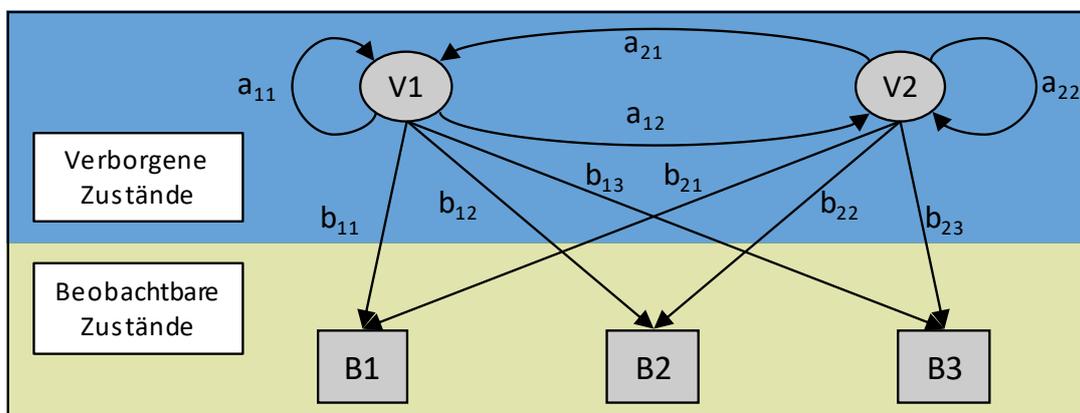


Abbildung 2.13: Beispiel eines verborgenen Markov Modells in Anlehnung an [51, S. 83]. Zwischen den Zuständen V und B kann mit den Wahrscheinlichkeiten a/b gewechselt werden.

SCHNEIDER [52, S. 41-44] erwähnt im Vergleich zum wahrscheinlichkeitsbasierten Ansatz ein regelbasiertes Prinzip zur Bestimmung von Fahrmanövern, wobei diese Idee als Basis für die vorliegende Arbeit genutzt wird. Grundlage für das Manöveridentifikationssystem ist die exakte Kenntnis der Manöverklassen und der jeweiligen Zustandsverläufe. Ähnlich einem Markov-Modell aus Abbildung 2.13 beschreibt SCHNEIDER [52, S. 41] ein Manöver als eine Aneinanderreihung von wohldefinierten Zuständen, die während eines Manövers sequenziell durchlaufen werden. Dabei ist ein Übergang zwischen zwei benachbarten Zuständen nur bei Erfüllung der zugehörigen Transitions- oder Übergangsbedingung möglich.

Hier zeigt sich der wesentlichste Unterschied zu vorheriger Methode, weil im Vergleich zu probabilistischen Werten als Übergangscharakteristik nun deterministische und eindeutig definierte Transitionsbedingungen vorliegen. Dies ist einerseits vorteilhaft, weil die Ausgaben des Erkennungsmoduls nachvollzogen werden können, andererseits ist der Aufwand zur adäquaten Erfassung aller potentiellen Manöver entsprechend besonders umfangreich.

SCHNEIDER [52, S. 42] führt zur Veranschaulichung des Prinzips den Fall eines Überholmanövers an, der zur Verdeutlichung des Ansatzes einer Zustandskette mit expliziten Übergangsbedingungen vorgestellt wird (Abbildung 2.14 und Tabelle 2.7). Über die Erkennung derartiger Zustandslisten können in weiterer Folge Verkehrssituationen zu definierten Manöverklassen zugeordnet werden.

Tabelle 2.7: Zustandsabfolge des Überholmanövers nach [52, S. 42]

Zustand	Beschreibung
S1	Warten auf Überholen
S2	Beginn des Spurwechsels
S3	Überfahren der Spurmarkierung
S4	Ende des Spurwechsels
S5	Vorbeifahren
S6	Ende Vorbeifahren
S7	Beginn des Spurwechsels
S8	Überfahren der Spurmarkierung
S9	Ende des Spurwechsels

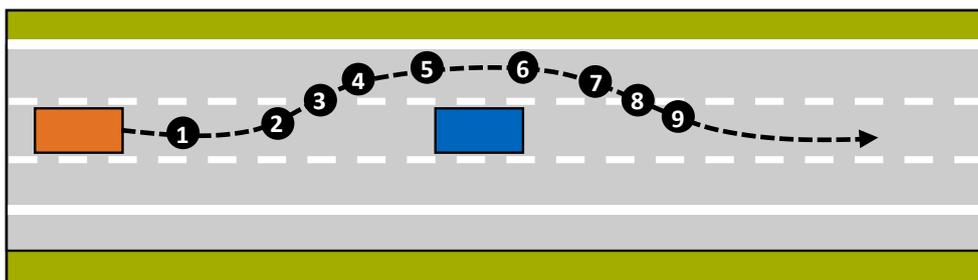


Abbildung 2.14: Zustandsabfolge des Überholmanövers nach [52, S. 42]. Das betrachtete Manöver ist anhand des definierten Zustandsverlaufs (Tabelle 2.7) eindeutig identifizierbar.

Im Vergleich zu den vorgestellten wahrscheinlichkeits- und zustandsbasierten Herangehensweisen wird im Projekt PEGASUS [16, S. 62-64] die Erkennung definierter Verkehrsszenarien über die Kombination aus zwei situationsspezifischen Größen realisiert. Diese Kriterien sind einerseits der Aufprallort und andererseits die initiale Position der Umgebungsfahrzeuge (engl. Surrounding Vehicles (SV)) zu Beginn eines sicherheitskritischen Szenarios (Abbildung 2.15) [16, S. 63]. Unter dem Begriff sicherheitskritisches Szenario wird in diesem Zusammenhang eine Verkehrssituation verstanden, in der sich ein Kollisionspotential zwischen dem betrachteten Ego-Vehikel und einem der Umgebungsfahrzeuge ausbildet. Dieses Potential kennzeichnet sich dadurch aus, dass das Ego-Fahrzeug zur Vermeidung eines Zusammenstoßes mit dem Umgebungsvehikel eine angemessene Maßnahme, wie beispielsweise einer Geschwindigkeitsreduktion mittels Bremsbetätigung, ergreifen muss. Ist eine solche Reaktion notwendig, so wird das beteiligte SV als Herausforderer (engl. Challenger) bezeichnet und das korrespondierende Szenario als relevant erachtet.

Als Aufprallort wird jener Bereich am Ego-Fahrzeug beschrieben, in welchem der erste potenzielle Zusammenstoß mit einem solchen Challenger stattfindet, sofern keine adäquate Reaktion zur Unfallvermeidung umgesetzt wird. In erster Linie kann eine Kollision im vorderen (Front), hinteren (Heck), oder seitlichen Bereich des Ego-Fahrzeugs erfolgen. In Kombination mit der Initialposition des Challengers (Abbildung 2.15 1-7) lassen sich 9 spezifische Kollisionspfade A bis I definieren, die beschreiben, welche Trajektorie das kritische Umgebungsfahrzeug vom Beginn des Manövers bis zur potenziellen Kollision zurücklegt. Diese Trajektorien stehen dabei gleichermaßen für die 9 definierten Szenarioklassen des PEGASUS-Projekts, die anhand der in Tabelle 2.8 aufgelisteten Semantik beschrieben werden [16, S. 62-64]. Über die erörterte Untersuchung der genannten Kriterien und der darauf basierenden Einteilung der Situationen in Szenarioklassen lässt sich ein strukturierter Katalog an relevanten Testfällen für die Absicherung autonomer Fahrzeuge ableiten.

Tabelle 2.8: Logische Szenarien nach [16, S. 64]

Aufprallort	Ausgangsposition	Pfad	Beschreibung
Heck	1	B	Einscherender Auffahrer
	2	A	Auffahrer
	6	F	Zurückfallender Auffahrer
Seite	1	C	Überholender Abdränger
	4	E	Abdränger
	6	G	Langsamer Abdränger
Front	1	D	Überholender Einscherer
	6	H	Langsamer Einscherer
	7	I	Langsamer Vorfahrer
Sonstige Kombinationen obiger Größen		ND	Nicht definiert

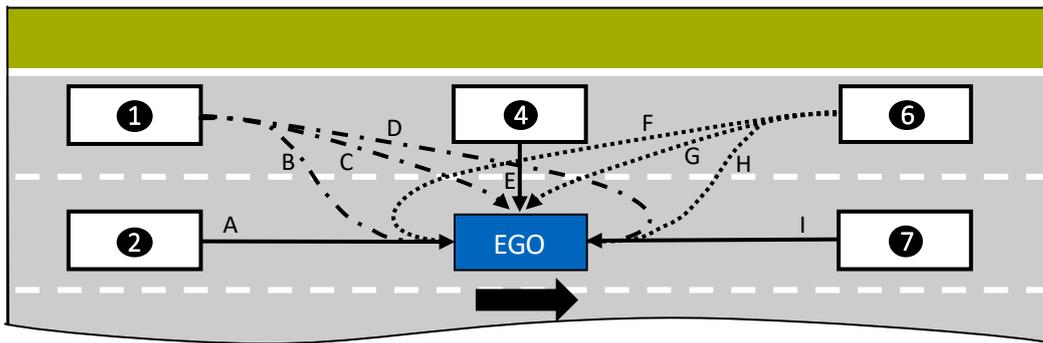


Abbildung 2.15: Trajektorien der Challenger-Fahrzeuge zur Klassifikation der Verkehrsszenarien (A-I) basierend auf [16, S. 63]. Aus Gründen der Symmetrie ist nur die linke Seite bezogen auf das Ego-Fahrzeug dargestellt.

Neben den angeführten Ansätzen zur Erkennung von Verkehrsmanövern und -szenarien ist die Arbeit von ERDOGAN ET AL. [53, S. 267-272] von besonderem Interesse für die vorliegende Arbeit. Die Autoren vergleichen darin verschiedene Methoden zur Extraktion und Identifikation absicherungsrelevanter Szenarien aus einer Datenbank realer Verkehrsaufzeichnungen. Die Untersuchungen behandeln einen regelbasierten Ansatz und überwachte, beziehungsweise unüberwachte Lernmethoden.

- Regelbasierter Ansatz:
 - Ein Szenario ist im Sinne der Autoren mit einer definierten Abfolge an bekannten Start- und Endzuständen gekoppelt, die von einem Algorithmus zur Identifikation zugrundeliegender Manöver in jedem Zeitschritt überprüft werden. Somit ähnelt dieser regelbasierte Ansatz dem von SCHNEIDER [52, S. 41-44] vorgestellten Verfahren.
- Unüberwachte Lernmethode:
 - Als unüberwachte Lernmethode nutzen ERDOGAN ET AL. [53, S. 267-272] das in Kapitel 2.4.1 vorgestellte k-means Verfahren. Sie bilden dazu mit Hilfe des genannten Verfahrens ähnlichkeitsbasierte Gruppen innerhalb der Verkehrsdaten, die sie anschließend im Sinne unterschiedlicher Manöver klassifizieren.
- Überwachte Lernmethode:
 - Die Autoren trainieren ein Long Short-Term Memory (LSTM) Netzwerk, das auf Basis zeitlicher Verläufe von Fahrzeugzustandsgrößen wie der Ego-Geschwindigkeit eine Zuordnung der Fahrsituation zu definierten Manöverklassen ermöglicht.

ERDOGAN ET AL. [53, S. 267-272] kommen in ihrer Arbeit auf das Ergebnis, dass alle vorgestellten Methoden zur Findung relevanter Verkehrsszenarien genutzt werden können. Als Nachteile der Verfahren gelten in Verbindung mit dem regelbasierten Ansatz das notwendige Expertenwissen, während das neuronale Netzwerk (überwachtes Lernen) nur auf simulierten Daten eine adäquate Genauigkeit aufweist. In Folge dessen empfehlen die Autoren eine Kombination der Methoden zur Kompensation der jeweiligen Schwachpunkte.

Künstliche Intelligenz im Feld der Szenarioklassifikation

Ähnlich der unüberwachten Lernmethode von ERDOGAN ET AL. [53, S. 267-272], finden sich weitere Ansätze zur Identifikation und Klassifikation von Verkehrsszenarien in der Literatur, die auf Prinzipien aus dem Bereich des maschinellen Lernens und der künstlichen Intelligenz zurückgreifen.

Das grundlegende Prinzip hinter dem maschinellen Lernen kann anhand Abbildung 2.16 erörtert werden. Modelle in Form neuronaler Netze werden mit umfangreichen Trainingsdaten gespeist. Diese Modelle untersuchen die Inhalte dieser Daten und liefern darauf beruhende Ausgaben. Dabei kann es sich beispielsweise um das Label eines Objektes in einem Bild handeln. Spezielle algorithmische Verfahren vergleichen die Outputs des neuronalen Netzes bei Durcharbeitung der Trainingsdaten mit den bekannten, zu erwartenden Ausgaben. Die darüber identifizierbaren Abweichungen dienen in weiterer Folge als Grundlage zur Verbesserung der modellinternen Struktur und Parameter. Dieser Vorgang wird im Sinne eines iterativen Lernprozesses vielfach wiederholt, bis die Genauigkeit des Modells, in Form geringen Abweichungen der Ausgaben von den Zielwerten, ein gewünschtes Maß erreicht hat. Ein angemessen trainiertes und damit genaues Modell kann in weiterer Folge zur Verarbeitung von neuen Inputdaten genutzt werden, die den Trainingsdaten ähnlich, jedoch a priori unbekannt sind [54, S. 4-5].

BEGLEROVIC ET AL. [55, S. 3-8] nutzen 2 Formen von neuronalen Netzen zur Klassifikation von Fahrmanövern. Zu Beginn verwenden sie ein CNN (engl. convolutional neural network), um aus rohen Messdaten von Fahrzeugsensoren definierte Eigenschaften zu extrahieren. Darauf folgt der Einsatz eines mehrschichtigen Perzeptrons, das die Aufgabe der Klassifikation von Szenarien auf Basis der zuvor ermittelten Kennwerte hat.

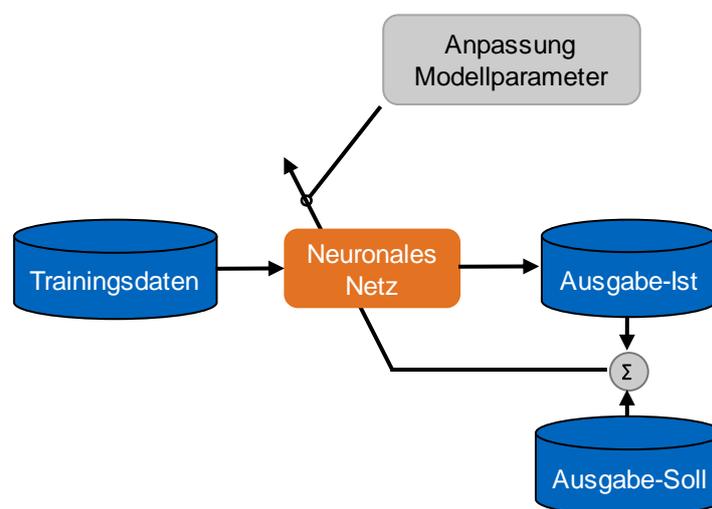


Abbildung 2.16: Prinzip des maschinellen Lernens nach [54, S. 5]. Der Prozess des Ausgabenvergleichs und der Parameteranpassung wird iterativ bis zu einem gewünschten Grad der Modellgenauigkeit ausgeführt.

2.7 Forschungsfrage

Der Stand der Technik liefert zahlreiche Methoden zur Absicherung autonomer Fahrfunktionen, wobei unterschiedliche Realisierungsansätze verfolgt werden. Diese Vielfalt geht mit Vor- und Nachteilen der einzelnen Verfahren einher, wobei bis dato kein allgemeingültiger Standard bezüglich eines validen Vorgehens für den (soll-)funktionalen Sicherheitsnachweis vorgestellt werden konnte. Insbesondere fehlen in diesem Zusammenhang geeignete Verfahren zur Erzeugung und Bereitstellung sinnvoller Testfälle für eine valide Absicherung von HAF. Diese Arbeit soll dahingehend einen Beitrag leisten und vor allem den Vorzug der Verwendung realer Verkehrsdaten für die Erstellung eines Testkatalogs gegenüber der synthetischen Erzeugung kritischer Szenarien verdeutlichen.

Der zentrale Forschungsschwerpunkt der vorliegenden Arbeit besteht folglich in der Identifikation einer passenden Prozesskette aus Clustering- und Klassifikationsverfahren zur Erzeugung eines szenariobasierten Testkatalogs. Diese Verarbeitungskette muss dabei insbesondere an die Form der verwendeten Daten angepasst werden. Dies ist notwendig, weil sich der HighD-Datensatz aufgrund der neuartigen Erhebungsmethode mittels Drohne maßgeblich von Verkehrsaufzeichnungen aus Fahrzeug-Perspektive, wie sie vielfach von OEMs erstellt werden, unterscheidet.

3 Methode

In Kapitel 3 der vorliegenden Arbeit wird die Verarbeitungskette von den Rohdaten bis hin zum fertigen Szenarienkatalog beschrieben. Es wird dazu im folgenden Abschnitt zuerst ein Überblick über den gesamten Prozess geliefert, woraufhin auf die wichtigsten Unterfunktionen des Tools genauer eingegangen, deren Aufgaben vorgestellt und die in MATLAB implementierten Ansätze präsentiert werden.

3.1 Überblick

Das Kernziel der zu implementierenden Werkzeugkette (Abbildung 3.1) ist es, aus einem umfangreichen Datensatz realer Verkehrsaufzeichnungen Situationen zu extrahieren, welchen eine besondere Relevanz im Zuge der Absicherung autonomer Fahrzeuge zugeschrieben wird.

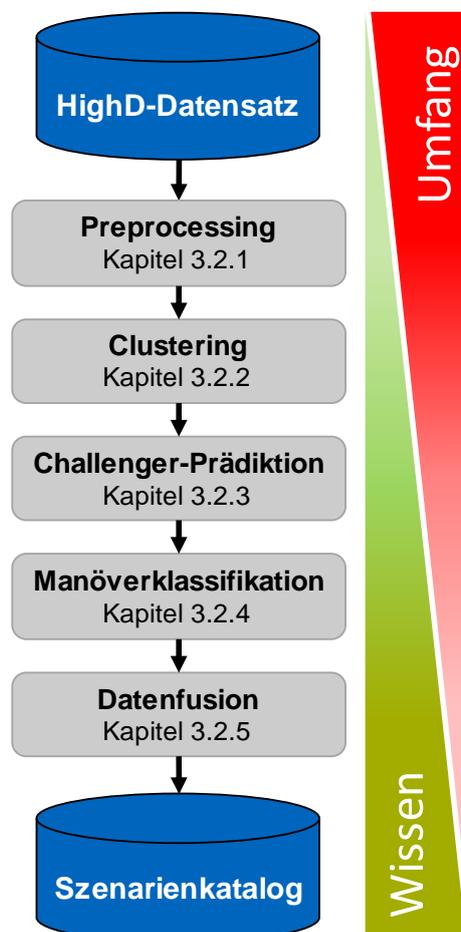


Abbildung 3.1: Überblick über die gesamte Prozesskette des Tools.

Dazu besteht im Anschluss an eine notwendige Vorverarbeitung der Daten die erste zentrale Aufgabe in der Extraktion kontextuell zusammengehöriger und auf kurze Zeitspannen beschränkter Situationen aus den Rohdaten. Dies ist Aufgabe der Clusteranalyse. Der Umfang der gefundenen Situationen nach diesem Schritt soll weiter reduziert werden, indem aus der Menge an separierten Clustern all jene hinausgeworfen werden, die als irrelevant für die Untersuchung eines autonomen Fahrzeugs eingestuft werden. Der dafür notwendige Prozessschritt ist in Form einer Challenger-Prädiktion realisiert. Die verbliebenen Situationen werden schließlich detaillierter untersucht, wodurch eine Klassifikation der kurzen Verkehrsgeschehnisse in logische Manöver ermöglicht wird. Über einen Fusionsvorgang der gewonnenen Informationen wird am Ende der Prozesskette der angestrebte Output eines klar strukturierten Szenarienkatalogs realisiert.

Abbildung 3.1 zeigt anschaulich in welcher Reihenfolge die einzelnen Subfunktionen der Prozesskette aufgerufen werden. Insbesondere das Clustering und die Challenger-Prädiktion tragen einen maßgeblichen Teil dazu bei, den angestrebten Verlauf (rot) bezüglich der Verringerung des Datenumfangs zu realisieren. Im Vergleich dazu sind es die beiden letzten Schritte der Kette, die eigentliche Manöverklassifikation sowie der finale Fusionsvorgang, die primär zur Erhöhung des Wissens (grün) über inhärente Muster und Vorgänge innerhalb der Daten beisteuern.

3.2 Subfunktionen der Toolkette

Die Prozesskette setzt sich also, wie Abbildung 3.1 zu entnehmen, aus 5 Subfunktionen zusammen, wobei die Ausgabe einer Funktionen gleichermaßen als Input für den darauffolgenden Verarbeitungsschritt dient. Auf diese Art und Weise wird ein sequenzieller Ablauf der einzelnen Prozessschritte umgesetzt, auf den in den folgenden Unterpunkten Funktion für Funktion genauer eingegangen wird.

3.2.1 Preprocessing der Daten

Wie bereits unter Punkt 2.3 eingeleitet, baut die vorliegende Toolkette auf dem highD Datensatz auf [18, S. 2118-2125]. Es handelt sich dabei um eine 60 Dateien umfassende Datenbank realer Verkehrsaufzeichnungen deutscher Autobahnen, die in Form von CSV-Files (engl. comma-separated values) vorliegen.

Die Dateien beinhalten jeweils Positions- und Geschwindigkeitswerte, Beschleunigungsverläufe und einige weitere Informationen bezüglich der einzelnen Fahrzeuge und Rahmenbedingungen der Aufzeichnungen. Die Verarbeitung der Daten in dieser ursprünglichen Form ist besonders umständlich, weswegen sich der erste Schritt der Methode dem Preprocessing (Vorverarbeitung) der Rohdaten widmet.

Neben dem Einlesen der Daten und der Aufbereitung derselben in eine einfach zu handhabende Form, sind die wesentlichen Aufgaben dieser ersten Subfunktion die Vereinheitlichung der Informationen, um für alle weiteren Verarbeitungsvorgänge eine universelle Ausgangslage zu schaffen. Dabei werden die Schritte Fahrtrichtungskorrektur, Zuschnitt, Koordinatenadaptation und Korrektur Fahrstreifennummerierung sequenziell entsprechend Abbildung 3.2 abgearbeitet.

Der Ursprung des für die Beschreibung der Kenngrößen verwendeten Koordinatensystems sitzt definitionsgemäß in der linken oberen Ecke des visualisierten Autobahnabschnitts (Abbildung 3.2, Originale Darstellung). Es werden dabei die Koordinatenwerte in x-Richtung von links nach rechts und in y-Richtung von oben nach unten positiv angenommen. Eine für die folgenden Verarbeitungsschritte besonders vorteilhafte Eigenschaft der Rohdaten ist die vorhandene Aufbereitung der Einträge in SI-Einheiten. Dies ermöglicht beispielsweise die Behandlung der x- und y-Koordinaten als Längenangaben in Meter.

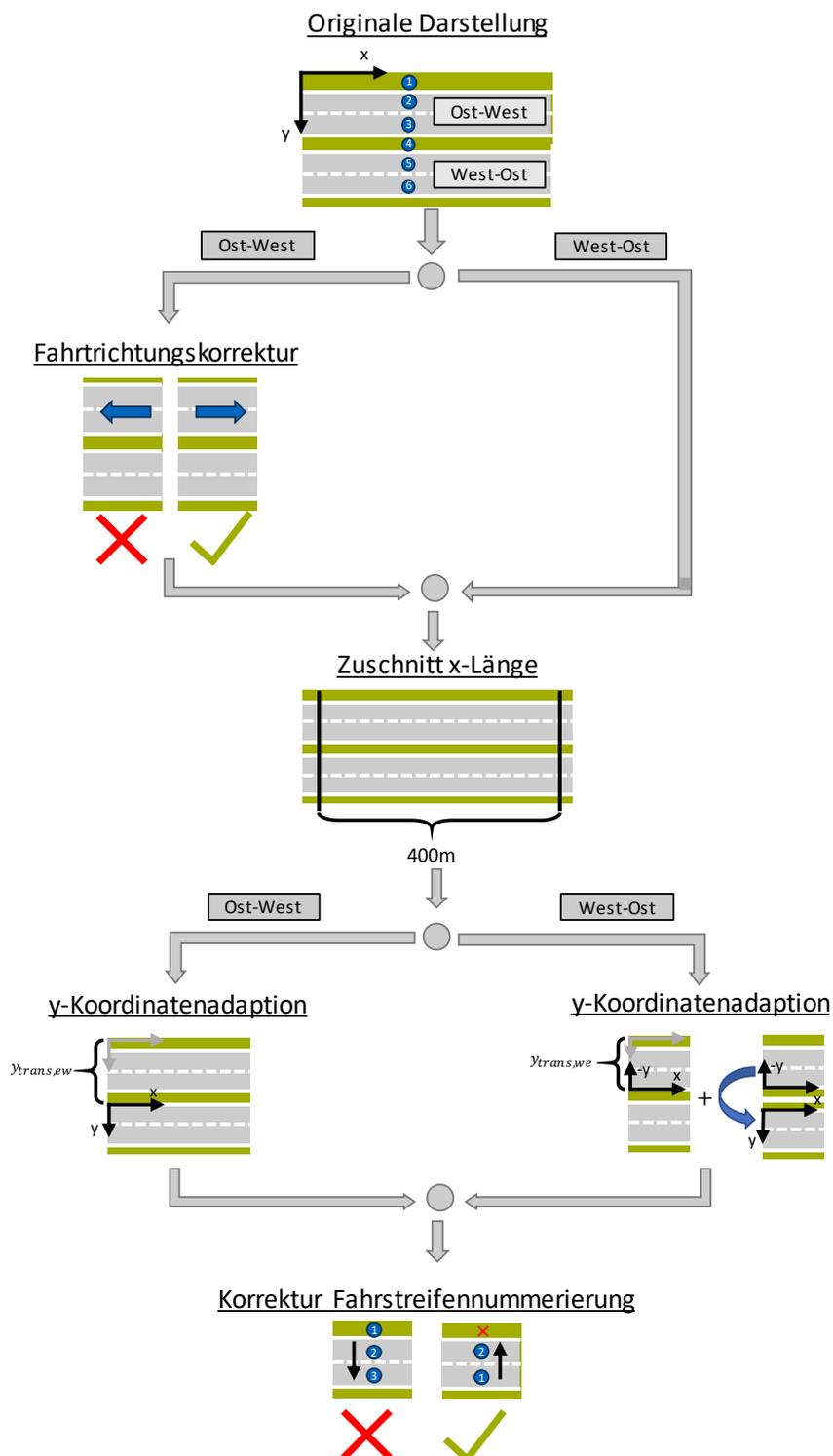


Abbildung 3.2: Darstellung der einzelnen Verarbeitungsschritte des Preprocessing

Fahrtrichtungskorrektur

Jede der 60 Aufzeichnung des highD Datensatzes beinhaltet die Fahrstreifen beider Fahrtrichtungen des Autobahnabschnitts mit horizontal angeordneten Fahrstreifenmarkierungen. Zur Homogenisierung der Daten kehrt der beschriebene Korrekturschritt die Orientierung aller Fahrzeuge mit ursprünglicher Ost-West Fahrtrichtung um, sodass alle Vehikel eine einheitliche Ausrichtung vorweisen (Abbildung 3.2, Fahrtrichtungskorrektur). Die dafür notwendige Transformation wird wie folgt beschrieben (Gleichungen (3.1) und (3.2)).

$$x_{\max} = \max_{ij} x_{ij} \quad \forall i \in N_{\text{veh,ew}}, j \in N_{\text{frame}} \quad (3.1)$$

$$x_{\text{trans}} = \frac{x_{\max}}{2} \quad (3.2)$$

Darin ist $x_{i,j}$ die x-Position (Fahrzeugmitte) von Fahrzeug i in Frame j mit $i \in N_{\text{veh,ew}}$ (Menge aller Vehikel der Ost-West Richtung) und $j \in N_{\text{frame}}$ (Anzahl der aufgezeichneten Frames des betrachteten Datensatzes). Die x-Position x_{trans} der gesuchten Transformationsachse berechnet sich über die Halbierung der unter allen Fahrzeugen i maximal vorkommenden x-Position x_{\max} (3.2). Die resultierende Transformationslänge gewährleistet eine korrekte Richtungsumkehr ohne Verfälschung der relativen Abstände zwischen den Fahrzeugen.

Die Positionswerte der Fahrzeuge i werden in jedem Frame j an x_{trans} gespiegelt. Dabei ist eine Fallunterscheidung zur fehlerfreien Transformation der x-Koordinate auszuführen, weil die Positionswerte der Vehikel in Form von Bounding-Boxen mit den Einträgen $x_{\min,i,j}$, $y_{\min,i,j}$, Fahrzeuglänge l_i und Fahrzeugbreite b_i (bezogen auf das in Abbildung 3.2 (Originale Darstellung) angeführte Koordinatensystem) vorliegen. Zur korrekten Transformation dieser Bounding-Boxen muss neben der trivialen Spiegelung von $x_{\min,i,j}$ an der Achse bei x_{trans} , zusätzlich eine Drehung des Vehikels um 180° realisiert werden. Das lässt sich für Fahrzeug i mit Fahrzeuglänge l_i in jedem Frame j über folgende Fallunterscheidung (3.3) berechnen. Die y-Position des Vehikels wird in einem späteren Verarbeitungsschritt des Preprocessing angepasst, weshalb vorerst nur der x-Wert der Bounding-Box neu berechnet werden muss.

$$x_{\min,\text{neu},i,j} = \begin{cases} x_{\text{trans}} + |x_{\text{trans}} - x_{\min,i,j}| - l_i, & \text{falls } x_{\min,i,j} < x_{\text{trans}} \\ x_{\text{trans}} - |x_{\text{trans}} - x_{\min,i,j}| - l_i, & \text{falls } x_{\min,i,j} > x_{\text{trans}} \end{cases} \quad (3.3)$$

mit $i \in N_{\text{veh,ew}}$ und $j \in N_{\text{frame}}$.

Aufgrund der ursprünglichen Fahrt der betroffenen Fahrzeuge entgegen der positiven x-Richtung (Ost-West) sind deren Geschwindigkeits- und Beschleunigungswerte in Längsrichtung negativ. Zur Homogenisierung der Daten werden daher neben der Transformation der x-Positionen (Formel (3.3)), genannte Dynamikwerte in ihrem Vorzeichen umgekehrt.

Zuschnitt X-Länge

Aufgrund von Schwankungen innerhalb der Längen der aufgezeichneten Autobahnabschnitte, wird in diesem Funktionsteil ein Zuschnitt beider Fahrrichtungen auf eine einheitliche Abmessung von 400m umgesetzt. Dazu werden symmetrisch um den zuvor berechneten Transformationspunkt x_{trans} Bereiche über 200m Abstand zu demselben aus den Daten entfernt. Der Koordinatenursprung wird auf die neu entstandene westliche Außenkante des Autobahnabschnitts verschoben, wodurch die Beschränkung der Positionswerte in x-Richtung auf den positiven Bereich aufrechterhalten bleibt.

Koordinatenadaption y-Richtung

Vorangegangene Preprocessing-Schritte passen ausschließlich die x-Werte der Fahrzeugdaten an. Zur weiteren Vereinheitlichung müssen hingegen zusätzlich alle y-Größen adaptiert werden. Der aktuelle Verarbeitungsschritt referenziert dazu die betroffenen Größen auf das in Abbildung 3.2 (y-Koordinatenadaption) dargestellte Koordinatensystem mit Ursprung auf der westlichen Außenkante des Autobahnabschnitts auf Höhe der äußeren, durchgezogenen Fahrstreifenmarkierung der jeweiligen Überholspur. Es ist zu beachten, dass für eine korrekte y-Adaption der ursprünglichen Ost-West-Richtung neben der Festlegung des neuen Koordinatensystems zusätzlich eine Spiegelung der gesamten oberen Autobahnhälfte an der neuen x-Achse notwendig ist.

Die notwendige Transformationsvorschrift kann als Gleichung (3.4) angegeben werden. Die benötigte Transformationslänge $y_{\text{trans},d}$ entspricht dem y-Abstand zwischen dem alten und dem neu definierten Koordinatensystem, wobei diese Größe wie in Abbildung 3.2 dargestellt von der ursprünglichen Fahrtrichtung abhängt. Somit werden für die y-Anpassung zwei unterschiedliche Längen $y_{\text{trans},d}$ mit $d \in [ew, we]$ verwendet, wobei für jede Richtung nur die Menge $N_{veh,d}$ der zugehörigen Fahrzeuge i_d , mit $i_d \in N_{veh,d}$, transformiert werden. Die Transformation wird jedem Frame $j \in N_{\text{frame}}$ der Aufzeichnung ausgeführt.

$$y_{\text{min,neu},i_d,j} = y_{\text{min},i_d,j} - y_{\text{trans},d} \quad (3.4)$$

Der zuvor erwähnte zusätzliche Schritt bei der y-Adaption der ursprünglichen Ost-West-Richtung wird analog zur beschriebenen Anpassung der x-Größen über eine Vorzeichenumkehr der y-abhängigen Dynamikwerte Quergeschwindigkeit und Querschleunigung erreicht.

Nach diesem Prozessschritt liegt der Datensatz bezüglich Positions- und Dynamikwerten vollständig in der gewünschten Darstellungsform vor, wobei die Relativwerte zwischen den Fahrzeugen durch die angeführten Homogenisierungsschritte erhalten bleiben und die notwendigen Adaptionen folglich zulässig sind.

Korrektur Fahrstreifennummerierung

Der letzte notwendige Schritt zu einem vollständig vereinheitlichten Datensatz bezieht sich auf die Änderung der Nummerierung der einzelnen Fahrstreifen (Abbildung 3.3). Ausgangslage ist die highD-interne Definition in Form einer aufsteigenden Bezifferung von oben nach unten, wobei auch die Abschnitte außerhalb des befahrbaren Straßenbereichs miteinbezogen werden. Diese Nummerierung wird wie in Abbildung 3.3 dargestellt umgewandelt, sodass für beide Fahrrichtungen der innerste Fahrstreifen (Lkw-Streifen) mit der Ziffer 1 identifiziert wird. Je nach Ausführung der Autobahn (2-spurig, 3-spurig, ...) werden die darüberliegenden Fahrstreifen aufsteigend durchnummeriert.

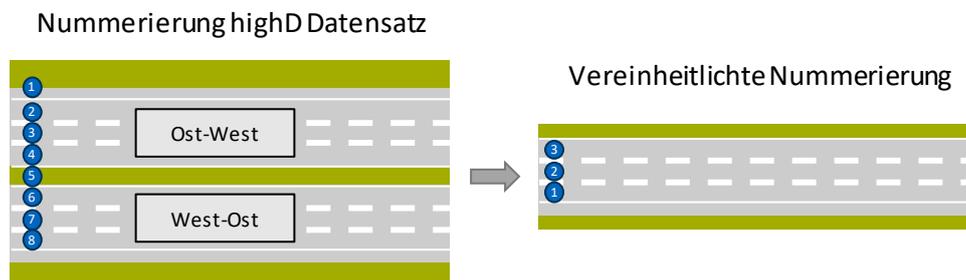


Abbildung 3.3: Anpassung der Fahrstreifennummerierung

3.2.2 Clusteranalyse

Nach erfolgreicher Ausführung der vorgegangenen Umwandelungsschritte kann der Datensatz zur weiteren Verarbeitung verwendet werden. Wie Abbildung 3.1 zu entnehmen, erfolgt als nächster Schritt der Prozesskette das Clustering der aufbereiteten Daten. Im Fokus dieses Verarbeitungsvorgangs steht die Aufteilung der einzelnen Verkehrsaufnahmen in zeitlich kurze Ausschnitte, die im Weiteren als separierte und kontextuell zusammengehörige Verkehrssituationen behandelt werden können. Die Ausführungen im Kapitel 2.4 haben gezeigt, dass sich die darin vorgestellten Verfahren besonders zur Identifikation von Mustern und Strukturen in a priori unübersichtlichen Datenmengen eignen. Dabei ist allgemein kein tieferes Wissen über die Dateninhalte nötig, um die Methoden anwenden zu können. Dennoch liefern herkömmliche Clusterverfahren, wie die vorgestellte k-means-Methode oder das hierarchisch-agglomerative Clustering nur dann sinnvolle Resultate, wenn zum einen das für den zu verarbeitenden Datensatz geeignetste Verfahren ausgewählt und zum anderen die anpassbaren Parameter günstig eingestellt werden. Beispielsweise sollen die charakteristischen Werte der Datenpunkte bei Einsatz eines der genannten Verfahren innerhalb der Gruppen möglichst homogen sein, während sich dieselben Größen zwischen den Clustern deutlich unterscheiden sollen. Die Struktur des highD Datensatzes und die Form der darin zur Verfügung stehenden Einträge weisen derartige Eigenschaften a priori nicht auf. Somit wird der unmittelbare Einsatz eines der bekannten Clusteranalyseverfahren deutlich erschwert. Aus genanntem Grund wird im Folgenden ein eigenes Konzept einer Clustermethode vorgestellt, die gezielt auf die Verarbeitung des vorliegenden Datensatzes optimiert ist.

Die Ausgangslage des angepassten Clusteringprozesses ähnelt jener des herkömmlichen hierarchisch-agglomerativen Prinzips. Jedes Fahrzeug eines Datensatzes bildet von vornherein ein

eigenes Cluster, das im Zuge des iterativen Verarbeitungsvorgangs mit weiteren Fahrzeugen zu größeren Gruppen fusioniert wird. Die für das Clustering charakteristische Größe ist dabei der Abstand der Fahrzeuge zueinander. Die grundlegende Idee dahinter ist, dass weit entfernte Vehikel keine Bedeutung für das zentrale Ego-Fahrzeug haben, während jene in der näheren Umgebung einen maßgeblichen Einfluss auf dessen Fahrverhalten ausüben. Dieser Regel entsprechend wird ein Fahrzeug genau dann dem Ego-Cluster hinzugefügt, wenn es sich zu einem beliebigen Zeitpunkt innerhalb einer definierten Umgebung um das Ego-Vehikel herum befindet. Diese Umgebung wird in der vorliegenden Arbeit ROI – Region of Interest bezeichnet und als ein geschwindigkeits- und positionsabhängiges und somit dynamisch veränderliches Areal ausgeführt (Abbildung 3.4).

Die Größe des rechteckigen Bereichs der ROI wird über zwei variable Werte beschrieben. Zum einen legt die y -Koordinate des Ego-Fahrzeugs und damit der aktuell befahrene Fahrstreifen die Breite der ROI fest, weil sich diese per Definition über den eigenen und jeweils benachbarten Fahrstreifen des Autobahnabschnitts erstreckt. Zum anderen definiert die x -Koordinate der Fahrzeugmitte den Referenzpunkt für die Ausdehnung der ROI in x -Richtung, wobei die beiden Variablen $l_{ROI,back}$ und $l_{ROI,front}$ deren momentane Länge definieren.

Genannte Variablen sind abhängig vom gesetzlich geregelten Sicherheitsabstand (engl. safety distance SD) auf deutschen Autobahnen [56] und werden über die Formeln (3.5) und (3.6) ermittelt. Vom Gesetzgeber wird vorgeschrieben zum vorausfahrenden Fahrzeug mindestens einen Abstand gleich dem halben Tacho in Meter einzuhalten. Dieser Wert entspricht exakt jener Distanz die das Fahrzeug bei Konstanthaltung der Momentangeschwindigkeit in $t_{SD} = 1.8s$ zurücklegt.

Der in Formel (3.6) verwendete Faktor k_{SD} stellt zudem einen parametrierbaren Wert dar, der zur Adaption der Länge der ROI vor dem Ego-Fahrzeug dient. In dieser Arbeit wird $k_{SD} = 2$ gesetzt. Dies ist darauf zurückzuführen, dass unter der Annahme des Einhaltens des gesetzlich vorgeschriebenen Sicherheitsabstands von allen Verkehrsteilnehmern bei einheitlicher Geschwindigkeit exakt 2 Umgebungsfahrzeuge vor dem Ego erfasst werden.

$$l_{ROI,back,i,j} = t_{SD} v_{x,i,j} \quad (3.5)$$

$$l_{ROI,front,i,j} = t_{SD} k_{SD} v_{x,i,j} \quad (3.6)$$

mit $v_{x,i,j}$ als Längsgeschwindigkeit des betrachteten Ego-Fahrzeugs $i \in N_{veh,all}$ (Menge aller Fahrzeuge des vereinheitlichten Datensatzes) im aktuellen Frame $j \in N_{frame,act,i}$. Die Anzahl der zu untersuchenden Frames $N_{frame,act,i}$ ist dabei abhängig vom aktuell gewählten Ego-Vehikel i , weil nur jene Frames betrachtet werden müssen, zu welchen sich das aktuelle Ego-Fahrzeug am Autobahnabschnitt befindet. In Verbindung mit der vorliegenden Arbeit wird in diesem Zusammenhang von den aktiven (active) Frames des Fahrzeugs i gesprochen.

Auf Basis der Gleichungen (3.5) und (3.6) wird folglich in jedem aktiven Frame die Ausdehnung der ROI berechnet und eine Überprüfung des Eindringens von Umgebungsfahrzeugen in selbige durchgeführt. Im Wesentlichen stellt das so umgesetzte Vorgehen ein HAC in jedem aktiven Frame des betrachteten Ego-Fahrzeugs dar. Das Prinzip ist exemplarisch in Abbildung 3.4 dargestellt (Datensatz: 03tracks_we, Ego-Fahrzeug: ID 495, Frame: 190).

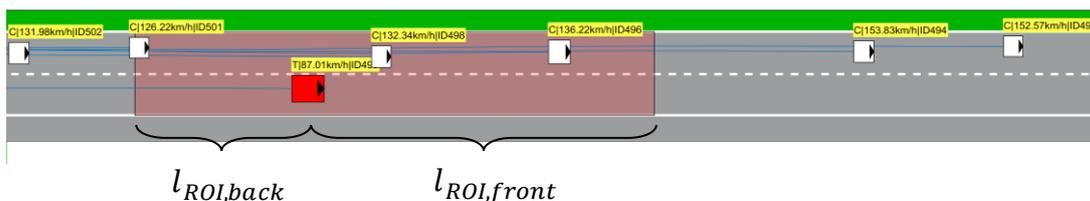


Abbildung 3.4: Darstellung der Region-of-Interest (ROI) des aktuellen Ego-Fahrzeugs (rot) auf Basis der geschwindigkeitsabhängigen Größen $l_{ROI,front}$ und $l_{ROI,back}$. Umgebungsfahrzeuge (weiß) außerhalb des rot markierten Bereichs werden nicht in das Cluster aufgenommen.

Abbildung 3.5 präsentiert das zum repräsentativ gewählten Clusteringschritt korrespondierende Dendrogramm. Auf der Ordinate des Dendrogramms ist der minimale Abstand der Umgebungsfahrzeuge relativ zum Ego-Vehikel im aktuellen Frame aufgetragen. Die vorherrschenden Geschwindigkeitswerte des Ego-Fahrzeugs führen unter Verwendung der Formeln (3.5) und (3.6) mit $k_{SD} = 2$ zu den beiden, als Cut-off Grenzen eingetragenen Werten, die gleichermaßen die Längen der ROI beschreiben. Anhand der Cut-off Werte wird festgelegt, welche Fahrzeuge im aktuellen Frame dem Cluster des Ego-Vehikels zugeschrieben werden können. Dabei ist zu beachten, dass für die Umgebungsfahrzeuge in Abhängigkeit ihrer relativen Position vor oder hinter dem Ego nur einer der beiden Cut-off Werte Gültigkeit hat (Abbildung 3.5 oben/unten). Im vorliegenden Fall werden im aktuellen Frame die Umgebungsfahrzeuge 498, 496 und 501 als Clusterelemente des Ego-Fahrzeugs mit ID 495 identifiziert.

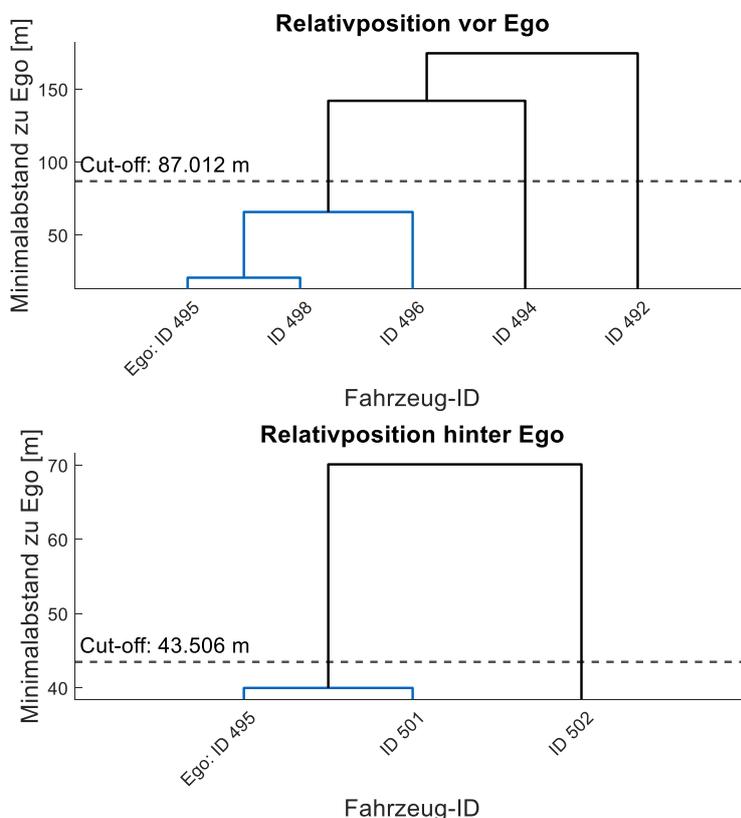


Abbildung 3.5: Dendrogramm der Minimalabstände zum Ego-Fahrzeug eines exemplarisch gewählten Frames. Die Clusterbildung beruht auf den Cut-off Werten, die über die Länge der ROI ($l_{ROI,front}$, $l_{ROI,back}$) festgelegt sind.

Die Analyse des Ego-Fahrzeugs nach beschriebenem Prinzip liefert in jedem aktiven Frame ein Cluster über die Gruppierung der im Dendrogramm blau dargestellten Kanten. Die Resultate aller Frames des Ego-Vehikels werden anschließend zusammengeführt. Darüber ergibt sich eine Teilmenge an Umgebungsfahrzeugen, die keinen Einfluss auf das Ego haben und damit ignoriert werden können. Die restlichen Fahrzeuge werden für die folgenden Verarbeitungsschritte als relevant angesehen und demnach als zugehörige Clusterelemente abgespeichert.

Die Ausführung der entwickelten Methode auf den Rohdaten setzt einerseits die gewünschte Zerteilung der Aufnahmen in zeitlich gekürzte Abschnitte um, andererseits wird durch das beschriebene Vorgehen bereits an diesem Prozessschritt ein Teil der Daten aussortiert. Dabei handelt es sich um jene Fahrzeuge, die im Zuge der Fahrt über den aufgezeichneten Autobahnabschnitt mit keinem Umgebungsvehikel wechselwirken und somit in keinem Fall ein bedeutsames Verkehrsszenario abbilden. In der vorliegenden Arbeit wird in diesem Zusammenhang von einer Freifahrt des Ego-Fahrzeugs gesprochen.

3.2.1 Challenger-Prädiktion

Nach der Durchführung des vorangegangenen Clusterings folgt der nächste Verarbeitungsschritt der Prozesskette, die Challenger-Prädiktion. Im Wesentlichen extrahiert dieser methodische Vorgang jene Verkehrssituationen aus der Gesamtheit der gefundenen Cluster, die als relevant für die Aufnahme in einen Testkatalog eingestuft werden. Das Kriterium dafür beruht auf der Existenz eines oder mehrerer Herausforderer (engl. Challenger), wobei die Erkennung solcher der Fokus des im Folgenden erörterten Prozessschritts ist.

Die verwendeten highD Daten repräsentieren reale Verkehrsgeschehnisse. Das bedeutet insbesondere, dass die Informationen und Daten hinter jedem aufgezeichneten Fahrzeug alleinige Abbildungen des Fahrverhaltens und der Reaktionen des menschlichen Lenkers darstellen. Diese Einflüsse manifestieren sich in den Aufnahmen als kontrollierte Fahrzeugführungen und allen voran im Nichtvorhandensein von Kollisionen, weil die Lenker rechtzeitig und geeignet reagieren, um Zusammenstöße mit umgebenden Vehikeln zu vermeiden.

Für die Identifikation relevanter Verkehrssituationen ist es hingegen notwendig, den Einfluss des menschlichen Fahrers eines betrachteten Ego-Vehikels weitestgehend zu eliminieren, um analysieren zu können, wie sich das Gefahrenpotential umgebender Fahrzeuge verändert, falls der Lenker keine kollisionsvermeidenden Reaktionen zeigt. In der vorliegenden Arbeit wird dies über die Verwendung eines prädiktiven Modells zur Vorhersage der zukünftigen Position des Vehikels auf Basis der momentanen Dynamikwerte umgesetzt. Grundlage hierfür bildet das Modell einer zweidimensionalen Bewegung mit konstanter Beschleunigung in x- und y-Richtung. Die dem Modell zugrundeliegenden Gleichungen sind in den Formeln (3.7) und (3.8) aufbereitet.

$$\begin{pmatrix} x_{\text{pred},i,j} \\ y_{\text{pred},i,j} \end{pmatrix} = \begin{pmatrix} x_{i,j} + a_{x,i,j}t_{\text{pred}}^2 + v_{x,i,j}t_{\text{pred}} \\ y_{i,j} + a_{y,i,j}t_{\text{pred}}^2 + v_{y,i,j}t_{\text{pred}} \end{pmatrix} \quad \forall i \in N_{\text{veh,all}}, j \in N_{\text{frame,act},i} \quad (3.7)$$

$$\begin{pmatrix} v_{x,\text{pred},i,j} \\ v_{y,\text{pred},i,j} \end{pmatrix} = \begin{pmatrix} v_{x,i,j} + a_{x,i,j}t_{\text{pred}} \\ v_{y,i,j} + a_{y,i,j}t_{\text{pred}} \end{pmatrix} \quad \forall i \in N_{\text{veh,all}}, j \in N_{\text{frame,act},i} \quad (3.8)$$

Darin beschreiben $x_{i,j}/y_{i,j}$ und $x_{\text{pred},i,j}/y_{\text{pred},i,j}$ die momentanen, beziehungsweise prädizierten Positionen, $v_{x,i,j}/v_{y,i,j}$ und $a_{x,i,j}/a_{y,i,j}$ die Geschwindigkeiten und Beschleunigungen und $v_{x,\text{pred},i,j}/v_{y,\text{pred},i,j}$ die Geschwindigkeitsprädiktionen des Ego-Vehikels $i \in N_{\text{veh,all}}$ im aktuell betrachteten aktiven Frame $j \in N_{\text{frame,act},i}$. t_{pred} beschreibt die angesetzte Prädiktionszeit, wobei diese Größe als veränderlicher Parameter vorliegt und die Festlegung der zeitlichen Distanz des Vorhersagemodells ermöglicht. Dabei ist zu beachten, dass das vorgestellte Prädiktionsmodell im Zeitbereich definiert und t_{pred} dementsprechend in Sekunden anzugeben ist. Zur Anwendung des Vorhersagemodells auf die frameweise diskretisierten highD Daten muss dementsprechend eine Transformation durchgeführt werden (Gleichung (3.9)). Diese ergibt sich aus der Aufzeichnungsrate der Daten von 25 Hz.

$$n_{\text{frame,pred}} = t_{\text{pred}} \frac{1}{25} \quad (3.9)$$

Obige Gleichungen (3.7) und (3.8) können unter definierten Umständen ungültige Ergebnisse liefern, weshalb eine gesonderte Behandlung folgender Ausgaben notwendig ist. Zum einen kann die Prädiktion der Position in x-Richtung Werte außerhalb der festgelegten Abschnittslänge von 400 m erzeugen. Dieser Fall wird abgefangen, indem die Prädiktionszeit automatisch derart beschränkt wird, dass nur Positionen unter genannter Grenze vorausgesagt werden können. Zum anderen führen gewisse Konstellationen aus der momentanen Geschwindigkeit und Beschleunigung eines Fahrzeugs bei ausreichender Prädiktionszeit zu einer Vorzeichenumkehr der Geschwindigkeitswerte. Derartige Situationen werden als andauernde Verzögerung mit anschließendem Stillstand interpretiert. In der Prädiktion werden daher die Geschwindigkeiten zu 0 gesetzt, wodurch es nie zu einer Richtungsumkehr kommen kann.

Vorgestelltes Modell zur Prädiktion der Fahrzeugzustände wird analog zum ROI-basierten Clustering in jedem aktiven Frame des Ego-Fahrzeugs angesetzt, die zukünftigen Zustandsgrößen auf Basis des eingestellten Prädiktionsparameters t_{pred} berechnet und darauf aufbauend Werte für das Kollisionspotential mit den umgebenden Fahrzeugen ermittelt. Letztgenannter Schritt beruht auf den Erörterungen aus dem Abschnitt 2.5.1, wobei insbesondere der Ansatz zur Untersuchung der Kollisionsgefahr über die Analyse der Überlappung definierter Hüllvolumina um die Objekte von Bedeutung ist. Die vorliegende Problemstellung ist ausnahmslos zweidimensional, wodurch sich die Definition ebengenannter Hüllvolumen auf Geometrien des 2D-Raums beschränkt. Abbildung 3.6 ist zu entnehmen, dass diese in der näheren Umgebung des Ego-Fahrzeugs, genauer gesagt dessen prädizierter Position, in Form trapezförmiger Kollisionsflächen vorliegen. Im Vergleich dazu werden die Umgebungsfahrzeuge zur Kollisionsüberprüfung über ihre tatsächlichen Positionen an den prädizierten Zeitpunkten und den jeweiligen Fahrzeugabmessungen dargestellt. Vorbereitend für nachfolgende Prozessschritte wird der Kollisionsbereich im und um das Ego-Fahrzeug in 5 kritische Areale eingeteilt (Abbildung 3.6), wobei die Regionen 2-5 über die beiden Parameter l_{coll} und b_{coll} festgelegt werden.

Der Wert für b_{coll} wird dabei konstant mit 0.75 m angenommen. Dies entspricht exakt dem seitlichen Abstand eines Pkw mit durchschnittlicher Breite von 1.80 m zur nächstgelegenen Fahrstreifenmarkierung, unter den Annahmen einer standardisierten minimalen Fahrstreifenbreite von 3.50 m (Entwurfsklasse EKA 1 [57, S. 70-72]) und unter Abzug der durchschnittlichen Breite der Markierung von 0.20 m [58, S. 28].

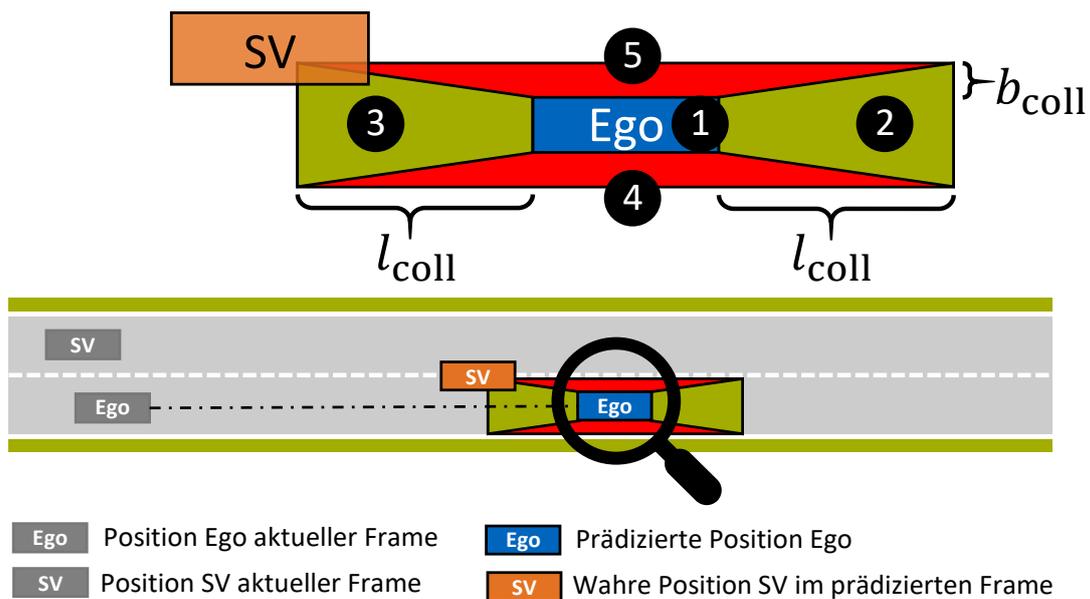


Abbildung 3.6: Darstellung des Prädiktionsprinzips und der Kollisionsbereiche (1-5) inklusive der beschreibenden Parameter.

Somit wird ein Umgebungsfahrzeug unmittelbar als potenzieller Kollisionspartner eingestuft, wenn es auf Höhe des Ego-Vehikels die Spurmarkierung in dessen Richtung überschreitet und sich letzteres dabei idealer Weise exakt in der Spurmitte befindet.

Der zweite Parameter l_{coll} wird im Vergleich dazu als geschwindigkeitsabhängige Größe angesehen. Dies ist darauf zurückzuführen, dass bei hohen Geschwindigkeiten des Ego-Fahrzeugs besonders in Kombination mit einer deutlich niedrigeren Geschwindigkeit eines vorausfahrenden Umgebungsfahrzeugs, letzteres bereits in einem größeren Abstand ein hohes Kollisionspotential aufweist.

Im Gegensatz dazu ist ein geringer Abstand zwischen den Fahrzeugen bei allgemein niedrigen Geschwindigkeiten (Stausituation) nahezu unproblematisch. Aus diesem Grund ist die geschwindigkeitsbasierte Anpassung von l_{coll} sinnvoll, wobei als zusätzliche Randbedingung die Bereiche 2 und 3 aus Abbildung 3.6 nur bei Geschwindigkeitswerten > 0 ausgebildet werden. Bei Stillstand des Ego-Fahrzeugs reduziert sich demnach der Untersuchungsaufwand auf die übrigen drei Kollisionsbereiche 1, 4 und 5. Der veränderliche Wert l_{coll} wird im Ergebnisteil der Arbeit einer Parameterstudie unterzogen, um dessen Einfluss auf die Extraktionsmenge der Challenger aufzuzeigen.

Die methodische Einschätzung der Kollisionsgefahr eines Umgebungsfahrzeugs wird anhand der überlappenden Fläche des Challengers mit den definierten Kollisionsbereichen umgesetzt. Je größer die Überdeckung, desto deutlicher repräsentiert das beteiligte Umgebungsfahrzeug einen Herausforderer in diesem Kollisionsareal. In Abbildung 3.6 ist zudem zu sehen, dass auch eine gleichzeitige Überlappung mehrerer Kollisionsbereiche (im Beispiel 3 und 5) möglich ist.

Die finale Entscheidung, in welchem der 5 Bereiche der Challenger das größte Gefährdungspotential aufweist, kann auf zwei Arten geschehen:

- **Erstkontaktprinzip:** Der Erstkontakt eines Umgebungsfahrzeugs mit den Kollisionsbereichen ist als entscheidungsrelevant zu werten. Bei initialer Überlappung mit mehreren Arealen wird zusätzlich der maximale Quotient aus überdeckter Fläche zu Größe des Umgebungsfahrzeugs in die Entscheidung mit einbezogen.
- **Maximalwertprinzip:** Der Maximalwert des Quotienten aus überdeckter Fläche zu Größe des Umgebungsfahrzeugs wird unabhängig vom Zeitpunkt des Auftretens als maßgeblich betrachtet, wodurch stets jenes Areal gewählt wird, in dem das größte Kollisionsrisiko mit dem Challenger vorliegt.

Abbildung 3.7 veranschaulicht die Verläufe des Kollisionspotentials eines repräsentativen Beispiels über die gesamte Dauer des Szenarios (Datensatz: 01tracks_ew, Cluster: veh04_nr03, $t_{\text{pred}} = 5s$). Das Ego-Fahrzeug befindet sich darin mit 3 weiteren Verkehrsteilnehmer auf dem Autobahnabschnitt, wobei zwei der Umgebungsfahrzeuge im Laufe des Szenarios ein Challenger-Potential entwickeln. Die Graphen sind dabei analog zur Bereichsdefinition aus Abbildung 3.6 angeordnet (Fahrrichtung West-Ost). Das zuvor erwähnte, optionale Kriterium bei der Wahl des Umgebungsareals mit dem maßgeblichsten Gefährdungspotential verdeutlicht in diesem Beispiel den Unterschied der beiden Ansätze. Wird der Maximalwert über der gesamten Szenariodauer betrachtet, so tritt dieser im Bereich hinter dem Ego-Vehikel in etwa bei Frame 170 auf. Grund dafür ist das sich schnell annähernde Vehikel mit der ID41.

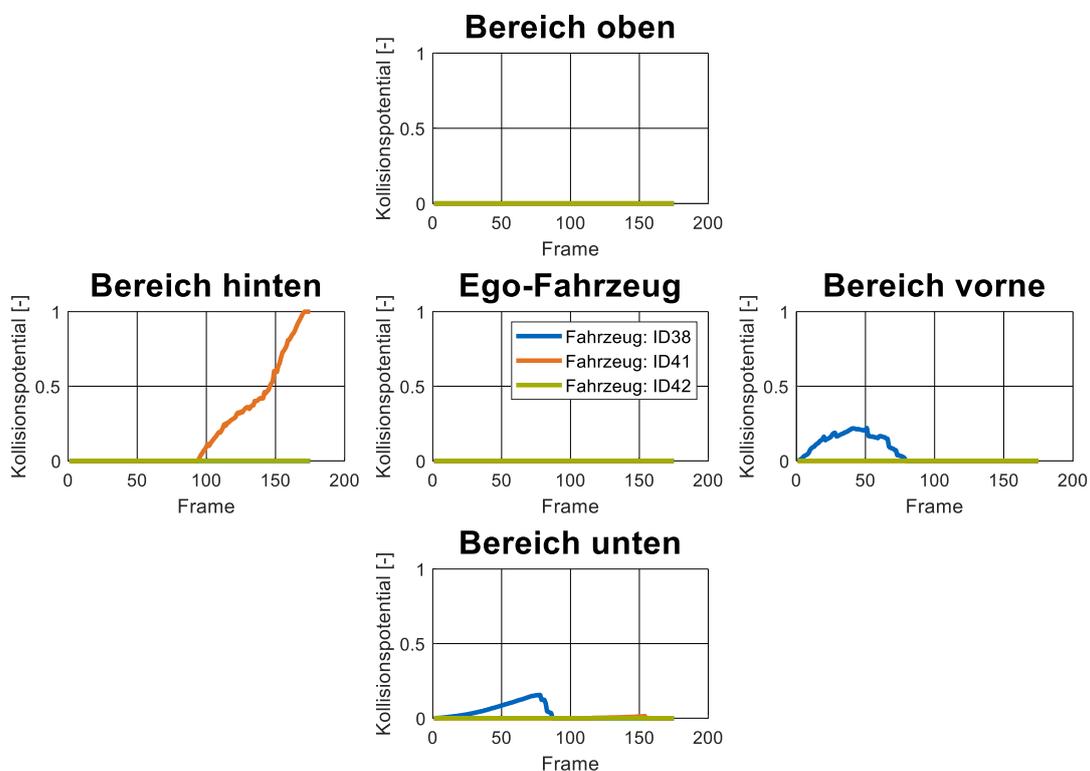


Abbildung 3.7: Repräsentative Verläufe des Kollisionspotentials in den 5 Umgebungs-bereichen des Ego-Vehikels. Challenger Erstkontaktprinzip: ID38 bei Frame 1 – Challenger Maximalwertprinzip: ID41 bei Frame 170

Im Vergleich dazu besteht im vorderen und unteren Bereich ein deutlich niedrigeres Kollisionspotential durch das Fahrzeug mit ID38, wobei diese Areale bereits zu einem früheren Zeitpunkt beeinflusst werden. Die Tatsache, dass trotz eines geringeren Maximalwerts in letztgenannten Bereichen der Erstkontakt mit einem Umgebungsfahrzeug auftritt, veranlasst das Ego stets aufgrund dieser primären Gefährdung eine kollisionsvermeidende Reaktion einzuleiten. Diese Erkenntnis begründet die primäre Verwendung des Erstkontakt-Prinzips in der vorliegenden Arbeit.

Der Durchlauf der angeführten Prädiktionsmethode in Verbindung mit der Untersuchung des Kollisionspotentials auf Basis der Überlappungsfläche sorgt für eine weitere Aussortierung irrelevanter Cluster und dadurch für eine deutliche Reduktion des Umfangs an potentiellen Testfällen. Diese müssen im Folgenden passend unter den einzelnen definierten Klassen des Testkatalogs verteilt werden. Dahingehend ist es notwendig, in den weiteren Prozessschritten ein geeignetes und möglichst exaktes Werkzeug zur korrekten Erkennung logischer Szenarien hinter den einzelnen Verkehrssituationen einzusetzen. Der nächste Abschnitt widmet sich dementsprechend der Entwicklung eines solchen Tools als Bindeglied zwischen den nicht klassifizierten Clustern und dem szenariobasierten Testkatalog.

3.2.2 Regelbasiert-semantische Klassifikation der Szenarien

Unter Kapitel 2.6 im Stand der Technik wurde gezeigt, dass in Zusammenhang mit der Klassifikation von Verkehrsszenarien vorrangig zwei Ansätze Verwendung finden, neuronale Netze und regelbasierte Verfahren. In Bezug auf die Entwicklungsdauer der beiden Prinzipien, kann a priori an keinem der beiden Methoden ein wesentlicher Vorteil festgestellt werden. Ansätze der künstlichen Intelligenz erfordern die Zusammenstellung eines umfangreichen und möglichst heterogenen Trainingsdatensatzes, wobei jede beinhaltete Verkehrssituation manuell untersucht und einer der definierten Klassen zugeordnet werden muss. Im Vergleich dazu sind zur Ermittlung und Festlegung der Regeln für die zweitgenannte Methode ebenso viele zeitliche Ressourcen einzuberechnen. Da sich demnach keines der beiden Verfahren der ersten Einschätzung nach besser für die Bearbeitung der vorliegenden Problemstellung eignet, wird hauptsächlich der regelbasierte Ansatz verfolgt und mit einem Anteil an Methoden aus dem Feld der künstlichen Intelligenz ergänzt.

Die Szenarioklassifikation wird auf alle aus der Challenger-Identifikation verbliebenen Cluster angewendet, wobei sich das entwickelte Klassifikationstool der vorliegenden Arbeit aus zwei Hauptfunktionen zusammensetzt. Diese erwarten als Input in jeder Iteration ein Paar aus 2 Fahrzeugen des momentan betrachteten Clusters, wobei jeweils das zentrale Ego-Vehikel gemeinsam mit einem der Umgebungsfahrzeuge übergeben werden.

Das Prinzip hinter dem implementierten Verfahren ist Abbildung 3.8 zu entnehmen. Die erste Hauptfunktion, das Zustandserkennungsmodul, übernimmt die Aufgabe der frame-basierten Zustandsidentifikation. Das Modul erfasst dazu in jedem diskreten Zeitschritt den vorliegenden Zustand zwischen dem Ego und dem betrachteten Umgebungsfahrzeug (SV) und erzeugt darüber Listen der Zustandsverläufe. Auf Grundlage dieser zeitdiskreten Listen leitet das zweite Modul des Klassifikators logische Manöver ab. Sofern das Verkehrsszenario aus einer Abfolge mehrerer fundamentaler Manövern zusammensetzt ist, teilt das zweite Modul die Zustandsverläufe in korrespondierende Sublisten auf und kann so die Manöversequenzen identifizieren (Manöver-Split).

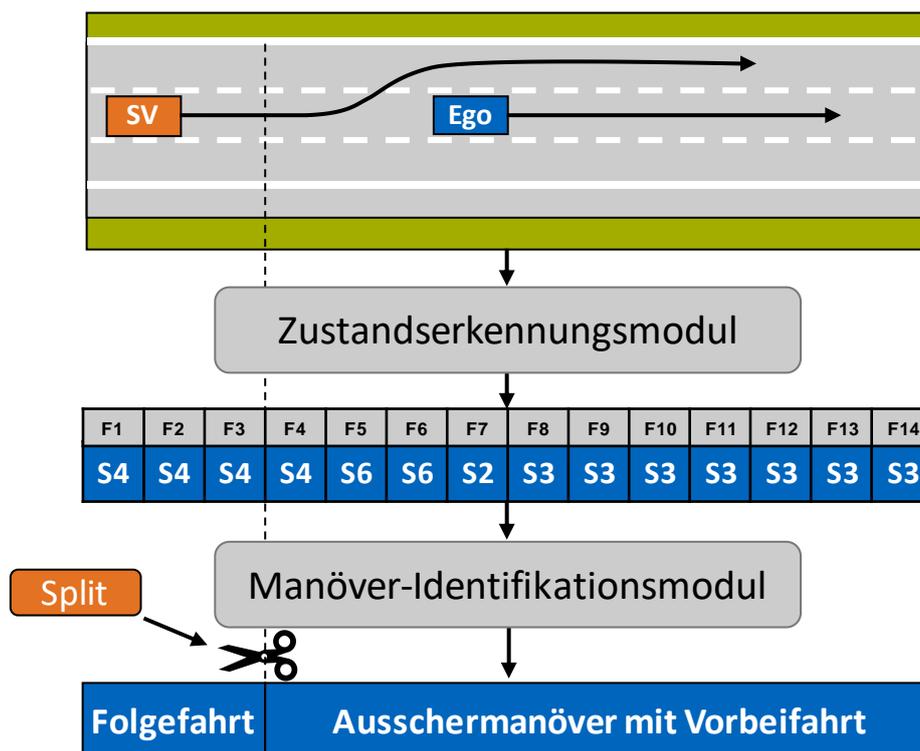


Abbildung 3.8: Funktionskette der Erkennung logischer Manöver bestehend aus einer frame-basierten Zustandsidentifikation und der Anschlussfunktion zur Ableitung logischer Manöver. Das zweite Modul teilt die Zustandsliste bei Verkettung mehrerer definierter Manöver (strichlierte Linie → Manöver-Split).

Zustandserkennungsmodul

Das von SCHNEIDER [52, S. 41-44] vorgestellte Prinzip zur regelbasierten Erkennung von Verkehrsmanövern beruht auf der Analyse von Zustandsabfolgen innerhalb einer Verkehrssituation. In Orientierung an diesem Prinzip müssen somit zur eigentlichen Identifikation von Manövern zuerst situative Zustände definiert und in der weiteren Anwendung Werkzeuge zur Erfassung derselben in den vorliegenden highD-Daten implementiert werden. Das Fundament des hier vorgestellten Moduls bilden die in Tabelle 3.1 angeführten semantischen Zustandsdefinitionen.

Die Liste setzt sich aus 8 möglichen Zustands-Grundformen zusammen. Diese charakterisieren sich mit Ausnahme der Zustände S6 und S7 in erster Linie über die relative Konstellation der Positionswerten der beteiligten Fahrzeuge. Die beiden genannten Sonderfälle beschreiben einen Fahrstreifenwechsel des Umgebungsfahrzeugs, wobei auf die konkrete Erkennungsstrategie dieses Zustands noch eingegangen wird. Anzumerken ist, dass die Grundformen aus Tabelle 3.1 bis auf Zustand S8 nur im relevanten Bereich festgelegt sind. Dieser erstreckt sich in Anlehnung an die Definition der ROI aus Kapitel 3.2.2 über den aktuell vom Ego-Fahrzeug befahrenen und einem oder zwei benachbarter Fahrstreifen.

Tabelle 3.1: Grundform der potentiellen Zustände (states) zwischen Ego und Umgebungsfahrzeug

Bezeichnung	Semantische Definition
S1	Umgebungsfahrzeug auf benachbarter Spur hinter Ego
S2	Umgebungsfahrzeug auf benachbarter Spur neben Ego
S3	Umgebungsfahrzeug auf benachbarter Spur vor Ego
S4	Umgebungsfahrzeug auf identer Spur hinter Ego
S5	Umgebungsfahrzeug auf identer Spur vor Ego
S6	Aktiver Fahrstreifenwechsel des Umgebungsfahrzeugs hinter Ego
S7	Aktiver Fahrstreifenwechsel des Umgebungsfahrzeugs vor Ego
S8	Aufenthalt des Umgebungsfahrzeugs in nicht relevantem Bereich

Das zentrale Erkennungsmodul wird in MATLAB mit Hilfe von Stateflow implementiert, wobei es sich um ein grafisches Modellierungswerkzeug handelt, das insbesondere zur Abbildung von Zustandsdiagrammen und sequenziellen Entscheidungslogiken geeignet ist [59]. Im vorliegenden Fall wird das Tool genutzt, um frame-by-frame das betrachtete Fahrzeug-Paar zu analysieren und auf Basis definierter Übergangsbedingungen für jeden Zeitpunkt den vorliegenden Zustand zu erfassen. Die in Tabelle 3.1 angeführten Zustände bilden hierfür die Basis. Die final ausgegebenen Zustände des Stateflow-Modells unterscheiden sich jedoch von den Grundvarianten dahingehend, dass den Hauptformen aus Tabelle 3.1 zusätzliche, situationsrelevante Informationen angehängt werden (Abbildung 3.9). Die genaue Beschreibung der gewählten Nomenklatur ist dem Anhang A zu entnehmen.

Dazu gehört beispielsweise die Angabe, auf welchem Fahrstreifen in Relation zum Ego-Vehikel sich das Umgebungsfahrzeug momentan befindet (SV-Fahrstreifen) und ob das Ego selbst im aktuell betrachteten Frame einen Fahrstreifenwechsel ausführt (Ego-Fahrstreifenwechsel). Von zentraler Bedeutung ist zudem die Beziehung, in der die beiden Fahrzeuge bezüglich ihrer Relativgeschwindigkeit stehen, wobei entweder eines der beiden Fahrzeuge erkennbar schneller als das andere fährt oder sich beide Vehikel mit derselben Geschwindigkeit fortbewegen.

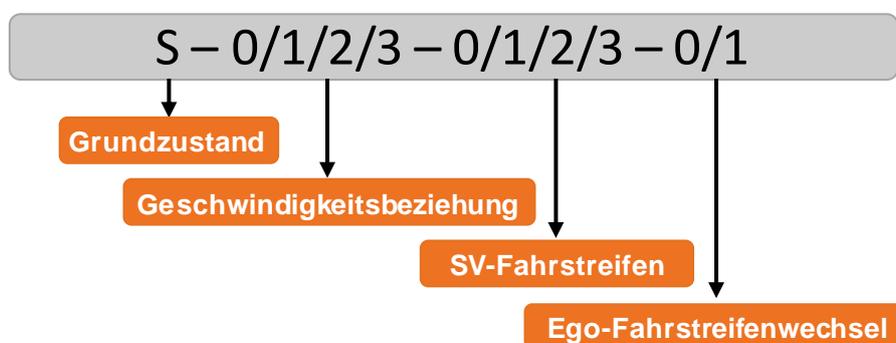


Abbildung 3.9: Struktur der Zustandsdefinition. Der Grundzustand wird in jedem Frame um die Informationen Geschwindigkeitsbeziehung, relativer Fahrstreifen des SV und Spurwechsel des Ego-Fahrzeugs erweitert.

Die Unterscheidung zwischen diesen beiden Möglichkeiten ist nicht trivial, wobei die Begründung dafür in der Verwendung von (diskretisierten) Verkehrsrealdaten liegt. In solchen Daten ist zu erwarten, dass immer ein minimaler Geschwindigkeitsunterschied vorliegt, obwohl die subjektive Betrachtung der Fahrzeugbewegung idente Geschwindigkeiten vermuten lässt.

Um diesem Problem der Unterscheidbarkeit zu begegnen, wird eine integrale Betrachtung des Relativgeschwindigkeitsverlaufs in Kombination mit einer Grenzwertüberwachung angesetzt (Abbildung 3.10). Solange die beiden Fahrzeuge näherungsweise dieselbe Geschwindigkeit besitzen, wird sich in jedem Frame die Größe des Integrals (der diskreten Summe) nur unwesentlich erhöhen oder verringern. Vielmehr schwankt in diesem Fall der betragsmäßig kleine Wert zwischen dem positiven und negativen Bereich nahe 0 (linker Abschnitt in Abbildung 3.10). Sofern sich eines der beiden Fahrzeuge jedoch für einen längeren Zeitraum minimal schneller bewegt, führt die damit verbundene Zunahme des Integrals zur Überschreitung des Grenzwerts, wodurch per Definition in eine der beiden zuvor erwähnten Geschwindigkeitsbeziehungen gewechselt wird. Ein erneutes Abfallen des Integrals unter den Grenzwert setzt den Zustand und die zu überprüfende integrale Größe unmittelbar auf die Ausgangswerte zurück (Reset).

Neben den selbsterklärenden, trivial zu identifizierenden Zuständen S1 bis S5, bedarf es zum Verständnis der Fahrstreifenwechselzustände (S6/S7) und dem letztgenannten Eintrag der Tabelle 3.1 einer gesonderten Betrachtung. Zustand S8 wird semantisch als Aufenthalt des betrachteten Umgebungsfahrzeugs außerhalb des relevanten Bereichs beschrieben. Damit wird ein Zustand bezeichnet, in welchem sich ein Umgebungsfahrzeug weder auf demselben noch einem benachbarten Fahrstreifen in Relation zum Ego-Vehikel befindet, sondern mindestens ein Fahrstreifen zwischen dem Fahrzeug-Paar vorhanden ist. Offensichtlicher Weise besteht die Möglichkeit der Existenz eines derartigen Zustands nur bei Autobahnen mit mehr als 2 Fahrstreifen. Die separate Behandlung dieses Zustands ist darauf zurückzuführen, dass Fahrzeuge außerhalb der näheren Umgebung des Ego-Vehikels definitionsgemäß keinen unmittelbaren Einfluss auf dieses ausüben und der Zustand somit keine Relevanz für die Identifikation eines logischen Manövers besitzt.

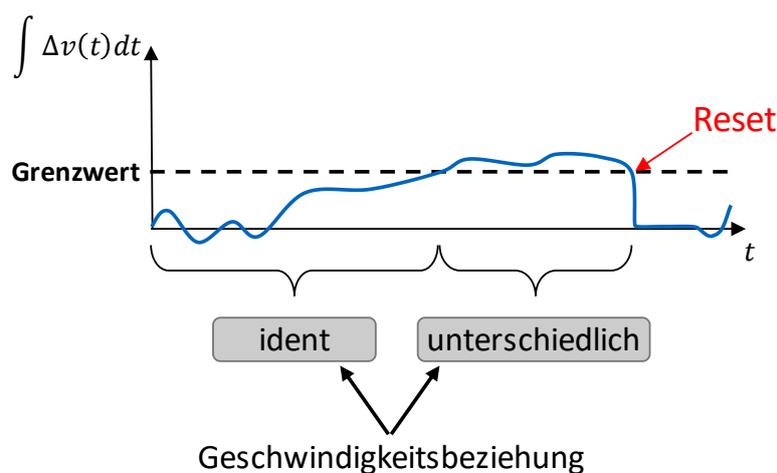


Abbildung 3.10: Prinzip der Identifikation der Geschwindigkeitsbeziehung zwischen den beiden Fahrzeugen. Solange der Integralwert unter der festgelegten Grenze bleibt, wird die Geschwindigkeit der beiden Vehikel als identisch betrachtet.

Eine besondere Herausforderung in Verbindung mit der Identifikation der Fälle S6 und S7 stellt die korrekte Erfassung des Beginns und des Endes dieser Zustände dar. Die einfache Ermittlung einer Überschreitung von Fahrstreifenbegrenzungen über die Untersuchung des Abstands Fahrzeugmittelpunkt zu Fahrstreifenmarkierung bestätigt zwar verlässlich einen Fahrstreifenwechsel, ist jedoch dahingehend beschränkt, dass die Zustände nur in einem einzelnen Frame erkannt werden können. Der gesamte Fahrstreifenwechsellvorgang erstreckt sich im Allgemeinen jedoch über einen längeren Zeitraum. Daher besteht die Notwendigkeit ein komplexeres Modell zur sicheren Erkennung des gesamten Fahrstreifenwechsels zu entwickeln und in das Zustandsidentifikationsmodul einzubinden.

Die implementierte Methode greift dazu frameweise auf einige der verfügbaren Daten des betrachteten Fahrzeugs zu. Darunter fallen vor allem die y-Position und Fahrzeuggeschwindigkeit in Querrichtung, weil sich im Zuge der Entwicklung des Tools diese Variablen als besonders geeignet zur Identifikation eines Fahrstreifenwechsels herausgestellt haben. Naheliegenderweise verringert sich zu Beginn des Manövers der Abstand des Fahrzeugmittelpunkts zur zu überschreitenden Fahrstreifenmarkierung, wohingegen derselbe gegen Ende des Fahrstreifenwechsels mit umgekehrtem Vorzeichen wieder zunimmt. Dementsprechend repräsentiert der Absolutwert dieser Größe eine wesentliche Charakteristik eines Fahrstreifenwechsels, wobei dieses Merkmal über eine einfache Grenzwertkontrolle überprüft werden kann. Dennoch zeigt sich, dass ein kleiner und stark fluktuierender Abstand zu einer Fahrstreifenmarkierung ebenso auf eine unruhige Fahrweise des Lenkers zurückzuführen sein kann. Somit gilt der absolute Abstand eines Vehikels zur nächstgelegenen Fahrstreifenmarkierung als notwendiges, aber nicht als hinreichendes Kriterium für eine sichere Erkennung eines Fahrstreifenwechsels.

Die Verknüpfung dieses Wertes mit der momentan vorherrschenden Quergeschwindigkeit liefert hingegen die benötigte Informationsgrundlage für eine korrekte und sichere Erfassung. Die vorrangige Schwierigkeit dieses Ansatzes liegt dabei in der geeigneten Wahl der Grenzwerte für die zu überprüfenden Variablen. Diesbezüglich ist dem Abschnitt Ergebnisse (Kapitel 4.1.2) eine Studie zur Untersuchung der idealen Einstellung dieser Werte zu entnehmen.

Unter Einsatz der vorgestellten Funktionen ist es nun möglich, die Abfolge der Zustände zwischen den beteiligten Fahrzeugen in jedem Frame zu extrahieren. Die derart erzeugten Zustandslisten stellen den notwendigen Input für die zweite große Funktion des Klassifikators, das Manöver-Identifikationsmodul dar, das im Folgenden detailliert beschrieben wird.

Manöver-Identifikationsmodul

Nach erfolgreicher Durchführung des vorangegangenen Bearbeitungsschritts analysiert das Manöver-Identifikationsmodul die ermittelten Listen und sucht darin definierte Muster in den Abfolgen, die den a priori festgelegten logischen Manövern zugeordnet werden können. Die dahingehende Notwendigkeit einer frühen Erfassung und Festlegung der Manöver nimmt einen großen Teil des zeitlichen Entwicklungsaufwands des beschriebenen Moduls in Anspruch.

Die semantischen Beschreibungen der Manöver werden primär auf Basis einer wissensbasierten Herangehensweise über die Konkretisierung mittels subjektiver Erfahrungen erstellt. Dieser Ansatz geht jedoch mit dem Nachteil einher, dass im ersten Definitionsprozess mit hoher Wahrscheinlichkeit nicht alle auftretenden, beziehungsweise potentiell möglichen Manöver einer

Autobahnfahrt erfasst werden. Aus diesem Grund wird auf die Notwendigkeit der iterativen Erweiterung und Verbesserung dieser Definitionen hingewiesen. Die Struktur des gesamten, hier vorgestellten Werkzeugs begünstigt durch die explizite Aussortierung nicht oder falsch klassifizierter Situationen diesen iterativen Aufbauprozess.

Die definierten logischen Manöver lassen sich aufgrund von Ähnlichkeiten in den zugrundeliegenden Zustandsverläufen in übergeordnete Manöverklassen einteilen. Im Rahmen der Arbeit werden so 12 Manöverklassen identifiziert, die in Summe 158 logische Einzelmanöver zusammenfassen (Auszug in Tabelle 3.2). Eine vollständige Liste der Manöverdefinition ist dem Anhang B zu entnehmen.

Jedes der Manöver repräsentiert eine eindeutige und unverwechselbare Abfolge der zuvor festgelegten Zustände. Diese Informationen werden in tabellarischer Form in einem File abgelegt, das im Weiteren als Basis für das Training eines Entscheidungsbaumes genutzt wird. Dies geschieht über die Verwendung des MATLAB internen Befehls `fitctree`, der auf Grundlage der übergebenen Trainingsdaten ein baumartig strukturiertes Vorhersagemodell erstellt [60]. Mit dessen Hilfe lassen sich die extrahierten Zustandsabfolgen aufgrund ihrer klaren Unterscheidbarkeit den logischen Manövern zuordnen. Ein Vorteil dieser Herangehensweise besteht vor allem darin, dass unbekannte Verläufe innerhalb einer Zustandsliste vom Entscheidungsbaum nicht zugeordnet werden können und derartige Situationen mit undefinierten Zustandsverläufen unmittelbar aus der Menge an Testfällen, die in den finale Testkatalog aufzunehmen sind, aussortiert werden. Somit bleiben nur eindeutig erfasste Zustandsketten für die Weiterverarbeitung erhalten. Die aussortierten Elemente gehen jedoch nicht verloren, sondern dienen in weiterer Folge als Basis für die iterative Verbesserung der zuvor erwähnten Manöverliste.

Die Güte des Verfahrens zur Zuordnung der Verkehrssituationen mit Hilfe des Entscheidungsbaums hängt somit in erster Linie von der Qualität der Trainingsdaten ab. Außerdem bilden die Zustandsverläufe nicht zwangsmäßig von vornherein eines der definierten Basismanöver ab, sondern können auch Aneinanderreihungen derselben repräsentieren (Abbildung 3.11). In Folge dessen werden dem eigentlichen Klassifikationsschritt Maßnahmen vorgelagert, die derartige Verkettungen der Manöver erfassen und geeignet aufteilen.

Als Beispiel für einen solchen Fall kann der zwischenzeitliche Aufenthalt eines Umgebungsfahrzeugs außerhalb des relevanten Bereichs angeführt werden. Eine derartige Situation ist in Abbildung 3.11 zu sehen. Das Umgebungsfahrzeug geht darin zwischenzeitliche in eine Phase des Zustands S8 über (rot markierter Bereich). Dadurch ergibt sich der angegebene Zustandsverlauf, wobei diese explizite Abfolge a priori mit keinem der definierten Manöver vereinbar ist. Vielmehr muss die gesamte Liste am Eintrag S8 in zwei unabhängige Teillisten aufgetrennt werden (Split), sodass diese anschließend in gewohnter Weise über die Vorhersage des Entscheidungsbaumes den logischen Manövern zugeordnet werden können.

Tabelle 3.2: Auszug aus der Liste an logischen Manövern

Manöver	Beschreibung
FBA	Umgebungsfahrzeug befindet sich in annähernder Folgefahrt hinter dem Ego
SPCCIF	Umgebungsfahrzeug passiert das Ego und schert vor diesem ein
SOF	Vollständiges Überholmanöver des Umgebungsfahrzeugs

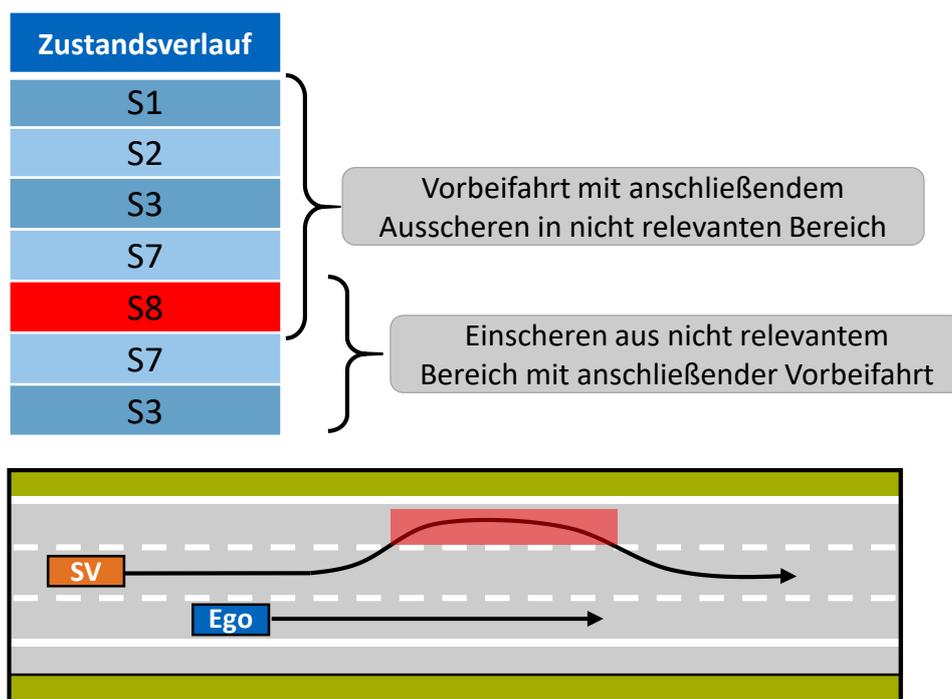


Abbildung 3.11: Verketzung von zwei logischen Manövern. Die Zustandsliste wird aufgrund des zwischenzeitlichen Auftretens von Zustand S8 in zwei separate Teillisten aufgesplittet, deren Manöver über das Identifikationsmodul erkannt werden können.

Neben dem beschriebenen Schnittkriterium aufgrund des Vorkommens des Zustands S8, induziert ein weiterer spezieller Verlauf der Zustände ein Aufsplitten der Liste. Dabei ist das entscheidende Kriterium die Veränderung der Geschwindigkeitsbeziehung zwischen den Fahrzeugen. Als Beispiel kann das Manöver FBA (Follow-Drive Behind Approach) aus Tabelle 3.2 herangezogen werden. Die annähernde Folgefahrt durch das SV charakterisiert sich mitunter durch eine deutlich höhere Geschwindigkeit des Umgebungsfahrzeugs verglichen mit jener des Ego-Vehikels. Bremsst das SV im Laufe der Fahrt über den Autobahnabschnitt jedoch ab und passt dadurch die Geschwindigkeit an die Bewegung des Ego-Vehikels an, so beendetet dieser Vorgang und die dadurch induzierte Veränderung der Geschwindigkeitsbeziehung das Manöver FBA. In der korrespondierenden Zustandsliste einer solchen Situation ist in diesem Moment derselbe Grundzustand S4 (SV hinter Ego auf identer Spur) in zwei aufeinanderfolgenden Frames angeführt, jedoch liegt im Eintrag zur Geschwindigkeitsbeziehung aus Abbildung 3.9 ein veränderter Wert vor, was den Schnitt der Zustandsliste notwendig macht.

Die Integration dieser Funktionsanteile in den Klassifikator komplementiert das vorletzte Glied der Werkzeugkette und liefert somit die fertige Sammlung an geclusterten und klassifizierten Situationen für den zu erstellenden Testkatalog. Die endgültige Zusammenführung der Informationen aus den vorangegangenen Bearbeitungsschritten übernimmt das Fusionsmodul, auf das im folgenden Abschnitt zur Vervollständigung der Methodik hinter der vorliegenden Arbeit eingegangen wird.

3.2.3 Aufbau des Szenarienkatalogs über Fusion von Challenger-Prädiktion und Manöveridentifikation

Rückblickend auf die Darstellung der Prozesskette in Abbildung 3.1 wird aufbauend auf dem vorangegangenen Prozessteil ein letzter Verarbeitungsschritt benötigt, um zum angestrebten Ziel des szenariobasierten Testkatalogs zu gelangen. Dabei handelt es sich um eine Funktion, die per se keinen direkten Einfluss auf die extrahierten und klassifizierten Verkehrssituationen hat, sondern vielmehr die gewonnenen Informationen der vorangegangenen Verarbeitungskette zusammenführt. Die einzelnen unabhängigen Manöver werden darüber den wenigen definierten Klassen des Szenariokatalogs zugeordnet.

Das methodische Fundament für diesen Fusionsprozess bildet die Definition der Testszenarien und die damit verknüpfte Verteilungslogik des PEGASUS-Projekts [16, S. 62-64], wie in Abbildung 2.15 und Tabelle 2.8 dargestellt. Daran angelehnt wird für die vorliegende Arbeit ein Konkretisierungsschritt implementiert, der die zentrale Verteilungsaufgabe der Verkehrssituationen zu den Szenarien übernimmt. Diese Verteilung lässt sich anhand eines baumartig-strukturierten Entscheidungsverfahrens visualisieren (Abbildung 3.12) [16, S. 63].

Die Ausgangslage dieses finalen Prozesses bildet die Menge an relevanten Verkehrssituationen, die über die vorangegangenen Verarbeitungsschritte aus den rohen highD Daten extrahiert werden konnten. Jede dieser Situationen beinhaltet ein Fahrzeug, das im Laufe der Prozesskette als Challenger, also als kollisionskritisches Umgebungsfahrzeug identifiziert wurde. Nachdem nur die Einflüsse dieser Vehikel die Relevanz einer Situation zur Aufnahme in den Testkatalog begründet, beschränkt sich der vorgestellte Fusionsprozess daher auf die alleinige Betrachtung der Challenger-Fahrzeuge zur Identifikation des zugrundeliegenden Szenarios. Dadurch werden die Verkehrssituationen unabhängig von den Einflüssen der restlichen Umgebungsfahrzeuge in die 9 Szenarioklassen (Tabelle 2.8) zugeordnet.

Es können jedoch, abseits der unmittelbaren Kollisionsgefahr durch die Challenger, weitere Umgebungsfahrzeuge einen indirekten Einfluss auf die Kritikalität einer Verkehrssituation und damit auf die Beschreibung des Szenarios besitzen. Im PEGASUS-Projekt wird in diesem Zusammenhang von Aktions-Restriktionen gesprochen, wobei darüber Fahrzeuge beschrieben werden, die den Handlungsraum des Ego-Vehikels für kollisionsvermeidende Manöver einschränken [17]. So verhindert beispielsweise ein Umgebungsfahrzeug neben dem Ego-Vehikel ein Ausweichmanöver auf diesen benachbarten Fahrstreifen. Aus genannten Gründen tragen die erwähnten Aktions-Restriktionen wesentlich zur Kritikalitätseinstufung einer Verkehrssituation bei. Dieser Aspekt wird in der vorliegenden Arbeit aus zeitlichen Gründen jedoch nicht in die Verteilungslogik des vorgestellten Fusionsprozesses miteinbezogen, weshalb die Möglichkeit zur zusätzlichen Integration der Aktions-Restriktionen hier als Erweiterungspotential des Moduls in Ausblick gestellt wird.

Der folgenden Darstellung in Abbildung 3.12 ist zu entnehmen, dass die zentralen Kriterien für die Zuordnung der Challenger-Situationen zu den Szenarioklassen einerseits der von der potenziellen Kollision betroffene Fahrzeugbereich und andererseits der Frame der kollisionskritischen Prädiktion sind. Die Bedeutung dieser Kriterien, beziehungsweise die Festlegung der zugehörigen Werte ist im Folgenden erörtert.

Kollisionskritischer Fahrzeugbereich

Dieses Kriterium ist aus dem vorangegangenen Prozessschritt der Challenger-Prädiktion bekannt und entspricht einem der 5 Areale um das Ego-Vehikel (Abbildung 3.6), das als kollisionskritischer Fahrzeugbereich betrachtet wird. Zur weiteren Zuordnung der Situationen in die Szenarioklassen ist es ausreichend, die 5 Bereiche auf die Areale vorne, hinten und seitlich zusammenzufassen (Abbildung 3.12).

Frame der kollisionskritischen Prädiktion

Ebenso lässt sich die zweite notwendige Größe aus den Ergebnissen der Challenger-Identifikation ableiten. Konkret beschreibt diese Größe jenen diskreten Zeitpunkt (Frame), in dem im Zuge des Prädiktionsvorgangs aus Kapitel 3.2.1 eine potentielle Kollision des Ego-Vehikels mit einem Umgebungsfahrzeugs vorausgesagt wird. Dieser konkrete Zeitpunkt wird im Folgenden als Frame der kollisionskritischen Prädiktion verstanden. Es ist zu beachten, dass dieser Frame vom gewählten Prinzip zur Challenger-Identifikation (Erstkontakt/Maximalwert, Abschnitt 3.2.1) abhängt. Dieser Aspekt lässt sich anhand Abbildung 3.7 nachvollziehen, in der nach dem Erstkontaktprinzip bereits im ersten Frame der Situation eine Kollision mit Fahrzeug 38 vorhergesagt wird, während nach dem Maximalwertprinzip erst bei Frame 170 das entscheidende Challenger-Potential von Fahrzeug 41 prädiziert wird.

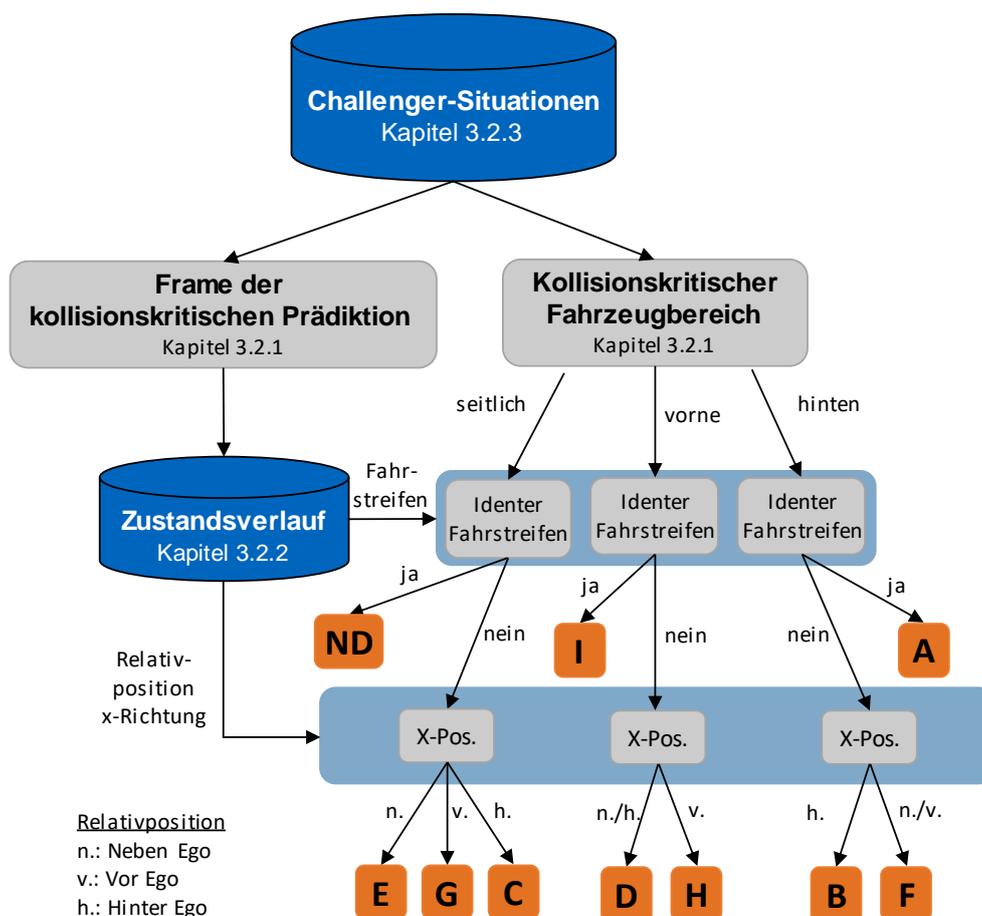


Abbildung 3.12: Entscheidungslogik des Fusionsschritts in Anlehnung an [16, S. 63]. Die notwendigen Informationen stammen aus dem Challenger-Prädiktionsmodul (Kapitel 3.2.1) und dem Manöver-Klassifikationsschritt (Kapitel 3.2.2).

Abbildung 3.12 ist zu entnehmen, dass das der beschriebene Frame der kollisionskritischen Prädiktion herangezogen wird, um aus dem frame-basierten Zustandsverlauf zwischen dem betrachteten Challenger und dem korrespondierenden Ego-Fahrzeug, die Relativposition in Längsrichtung und den Vergleich der momentan befahrenen Fahrstreifen in diesem kritischen Zeitpunkt zu extrahieren. Der benötigte Zustandsverlauf ist dabei ein Nebenprodukt der vorangegangenen Manöverklassifikation (Kapitel 3.2.2).

Unter Einbezug aller vorgestellter Kriterien und Größen kann der präsentierte Zuordnungsprozess der Challenger-Situationen zu den 9 Szenarioklassen realisiert werden, wobei anzumerken ist, dass die Kombination aus einem seitlichen Kollisionspotential bei identer Fahrspur nicht definiert ist. Derartige Situationen werden analog zu nicht klassifizierbaren Manövern (Kapitel 3.2.2) aussortiert, um bei Bedarf gesondert analysiert werden zu können.

3.3 Zusammenfassung der Methodik

Das vorangegangene Kapitel liefert einen Einblick in die implementierten Ansätze, um aus unstrukturierten Rohdaten mit geringem Informationsgehalt einen Katalog an definierten und absicherungsrelevanten Verkehrsszenarien zu bilden. Der zugrundeliegende Verarbeitungsprozess lässt sich dabei geeignet in die Hauptmodule Vorverarbeitung, Clustering, Challenger-Identifikation, Klassifikation und Fusion untergliedern. Diese Strukturierung ermöglicht es, die Daten zum einen von einem einzelnen Modul auf eine spezielle Informationen zu untersuchen, oder zum anderen die gesamte Prozesskette durchlaufen zu lassen, um einen validen Testkatalog zu erstellen. Vorteilhaft ist zudem der universelle Aufbau des Werkzeugs, sodass neue Rohdaten, bei identer Struktur verglichen mit dem highD Datensatz, automatisiert verarbeitet, analysiert und bestehenden Szenariokatalogen angehängt werden können.

4 Ergebnisse

Der folgende Abschnitt befasst sich mit den Ergebnissen, die unter Anwendung des entwickelten Werkzeugs aus dem highD Datensatz gewonnen werden. Dahingehend wird im ersten Unterkapitel eine umfangreiche Parameterstudie angeführt, die die notwendige Entscheidungsgrundlage für die geeignetste Wahl der essentiellen Parameter des Tools bildet. Anschließend daran wird anhand einer Stichprobenanalyse der Ergebnisse aus dem Klassifikationsmodul dessen Güte bewertet, woraufhin im letzten Punkt auf das zentrale Ziel der Arbeit, der Erzeugung eines szenariobasierten Testkatalogs auf Basis der zur Verfügung stehenden Datenbank eingegangen wird.

4.1 Parameterstudie

Dem vorangegangenen Kapitel 3 zur Methodik ist zu entnehmen, dass in den einzelnen Prozessschritten des Tools an diversen Stellen variable Parameter eingesetzt werden. Diese haben nicht nur auf den jeweiligen Verarbeitungsvorgang einen maßgeblichen Einfluss, sondern wirken sich vielmehr auf den finalen Output des Werkzeugs aus. Dementsprechend ist es sinnvoll und auch notwendig, vor der eigentlichen Benützung des Tools, eine Untersuchung der einzelnen Parameter und derer Einflüsse auf die Zusammenstellung des finalen Testkatalogs durchzuführen. Auf diese Weise soll für zukünftige Anwendungen der geeignetste Parametersatz ermittelt werden und anhand der folgenden Studien fundiert nachvollziehbar sein. Entsprechend des sequentiellen Ablaufs des Werkzeugs sind in Tabelle 4.1 die einflussreichsten Parameter angeführt.

Tabelle 4.1: Parameter des Tools und Argumentationsgrundlage für deren Festlegung

Modul	Parameter	Wert	Einheit	Argumentation
Clustering	k_{SD}	2	[-]	Doppelter Sicherheitsabstand zur Erfassung von 2 relevanten Fahrzeugen vor dem Ego (Methode)
Challenger-Prädiktion	t_{pred}	undef.	[s]	Untersuchung Prädiktionszeit
	b_{coll}	0.75	[m]	Abstand zwischen Fahrzeugaußenkante und Fahrstreifenmarkierung bei idealer Fahrt in Spurmitte
	$l_{coll, fact}$	undef.	[-]	Untersuchung Prädiktionslänge
Klassifikation	$v_{y, lim, LC}$	undef.	[m/s]	Untersuchung Fahrstreifenwechsel-Quergeschwindigkeit

Die Werte für k_{SD} und b_{coll} können a priori mittels einer schlüssigen Argumentation festgelegt und begründet werden, weshalb nur der restliche Teil der angeführten Parameter im Zuge einer Studie genauer betrachtet werden muss. Die daraus resultierenden Erkenntnisse werden zu Evaluierungszwecken mit Literaturwerten verglichen.

Um die Parameterstudien auf einer möglichst heterogenen und repräsentativen Datengrundlage aufzubauen, die Untersuchungen der Parameter jedoch in einem vertretbaren Zeitraum umsetzen zu können, beruhen die folgenden Überprüfungen nur auf einem Auszug der vollständigen highD Datenbank. Es werden die ersten 10 Aufnahmen des Datensatzes herangezogen, die sowohl 2-, als auch 3-spurige Autobahnabschnitte beinhalten. Zudem bildet diese Auswahl unterschiedliche Verkehrsgeschehnisse ab. Darunter fallen Stausituationen und Phasen mit zügig fließendem Verkehr.

Wird in diesem Zusammenhang die relative Häufigkeitsverteilung der Fahrzeug-Durchschnittsgeschwindigkeit in Längsrichtung der ersten 10 Datensätze mit jener der gesamten highD Datenbank verglichen (Abbildung 4.1), so zeigt sich eine ausgeprägte Ähnlichkeit der Verläufe. Trotz eines geringen Unterschieds in den beiden Mittelwerten der Durchschnittsgeschwindigkeiten von 2.286m/s, kann der verwendete Teil der Rohdaten als valide Repräsentation der gesamten highD Datenbank betrachtet werden. Dementsprechend besitzen die folgenden Argumentationen bei der Auswahl einzelner Parameter auf dem gesamten Datensatz Gültigkeit.

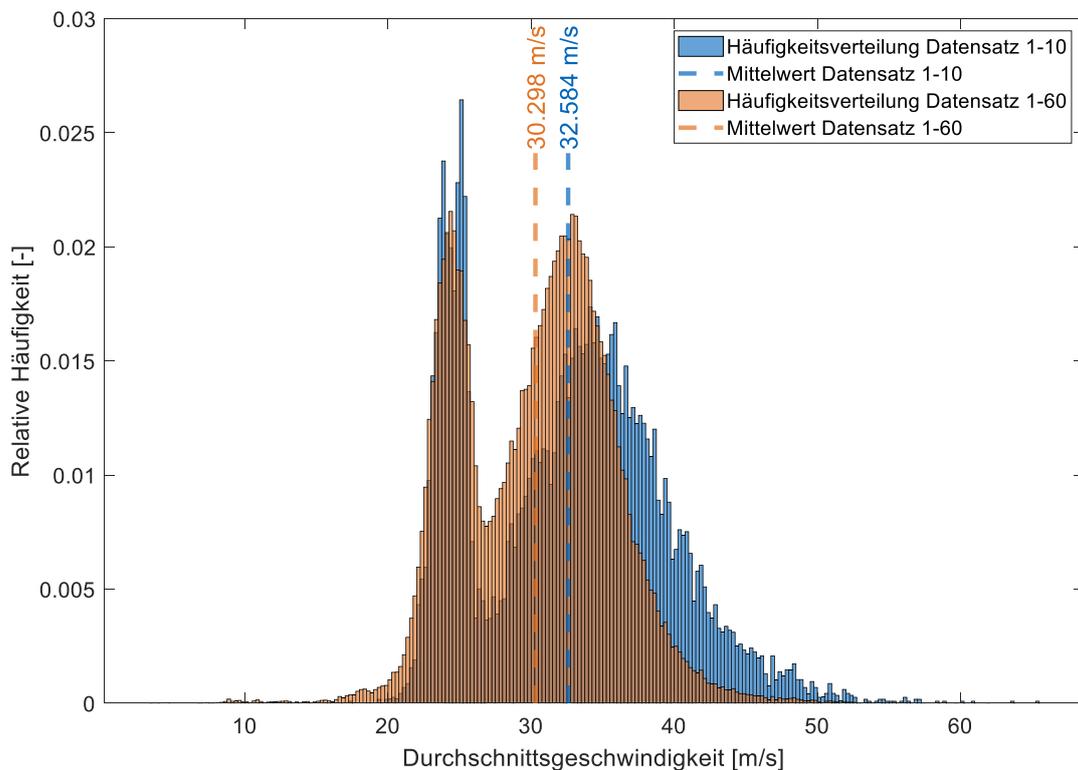


Abbildung 4.1: Relative Häufigkeitsverteilung der Durchschnittsgeschwindigkeit inklusive der Mittelwerte für die ersten 10 Datensätze und die vollständige HighD-Datenbank (60 Datensätze).

4.1.1 Parameterstudie zur Challenger-Prädiktion

Tabelle 4.1 ist zu entnehmen, dass allen voran die Prädiktionszeit (t_{pred}) und die Größe des relevanten Kollisionsbereichs um das Ego-Fahrzeug (definiert über $l_{\text{coll, fact}}$) die zentralen Parameter des Challenger-Erkennungsmoduls darstellen. Bezogen auf den übergeordneten Gesamtprozess des Werkzeugs beeinflussen diese Parameter vorrangig die Beurteilung der Kritikalität der betrachteten Verkehrssituationen und damit das Verhältnis aus relevanten zu irrelevanten Situationen im finalen Testkatalog.

Die Breite des Kollisionsbereichs b_{coll} wird auf Basis der logischen Argumentation aus Kapitel 3.2.1 konstant bei 0.75 m gehalten und hat somit einen unveränderlichen statischen Einfluss auf die Ergebnissen der Challenger-Prädiktion. In Folge dessen besteht nur eine Notwendigkeit die Werte von t_{pred} und $l_{\text{coll, fact}}$ einer detaillierteren Untersuchung zu unterziehen.

Untersuchung Prädiktionszeit

Der Parameter t_{pred} wird im Tool in Form eines Vektors verwendet, der die im Zuge des Prädiktionsvorgangs zu untersuchenden Zeitschritte beinhaltet. Die vektorielle Ausführung dieses Parameters erlaubt die Vorgabe einer beliebigen Anzahl an Prädiktionszeitpunkten, wobei die Angabe der Werte in Sekunden und explizit in aufsteigender Reihenfolge zu erfolgen hat. Tabelle 4.2 listet die in der vorliegenden Parameterstudie eingesetzten Vektoren auf. In jeder Iteration der Studie erfolgt eine Verlängerung des Prädiktionshorizonts um 1 Sekunde. Darüber soll ermittelt werden, wie sich der Anteil an Challenger-Fahrzeugen bei zunehmender Prädiktionsdauer verändert.

Es ist zu erwarten, dass bei steigender Prädiktionszeit die Anzahl an Challenger-Fahrzeugen zunimmt, wodurch sich in gleichem Maße die Menge an relevanten Fällen für den Testkatalog vergrößert. Hohe Prädiktionszeiten gehen in Verbindung mit dem verwendeten, beschleunigungs-basierten Vorhersagemodell im Allgemeinen jedoch mit einer signifikanten Zunahme der Prädiktionsungenauigkeit bezogen auf die tatsächlich auftretenden Zustände einher. Dies begründet sich vor allem darin, dass für eine exaktere Vorhersage von Fahrzeugzuständen über längere Zeiträume wesentlich mehr Informationen, wie beispielsweise der Straßenverlauf, in einem Prädiktionsmodell berücksichtigt werden müssen.

Dahingehend soll die vorliegende Parameterstudie insbesondere zeigen, ob die Annahme kurzer Prädiktionszeiten zu einem ausreichenden Anteil an Challenger-Situationen für den Testkatalog führt. Ist dem nicht so, muss von vornherein eine Prädiktion über längere Zeiträume stattfinden, wobei zur Gewährleistung einer angemessenen Genauigkeit der Vorhersagen in diesem Fall auf ein komplexeres Prädiktionsmodell zurückgegriffen werden muss.

Die Resultate in Tabelle 4.2 bestätigen die Erwartung der Zunahme an Challenger-Situationen bei steigenden Prädiktionszeiten, wobei neben der Variation von t_{pred} die beiden anderen essentiellen Parameter der Challenger-Prädiktion mit $b_{\text{coll}} = 0.75 \text{ m}$ und $l_{\text{coll, fact}} = \frac{5}{10}$ konstant gehalten werden. Bezogen auf die gesamte Menge an untersuchten Fahrzeugen (11370) bewegt sich der prozentuale Anteil an Challenger-Vehikel in einem Bereich zwischen ca. 30 und 45 %.

Tabelle 4.2: Parametervariation t_{pred}

Variation	t_{pred} [s]	Anzahl Challenger-Fahrzeuge (% von 11370 Fahrzeugen)
1	[1]	3314 (29.19%)
2	[1, 2]	3404 (29.94%)
3	[1, 2, 3]	3736 (32.86%)
4	[1, 2, 3, 4]	4348 (38.24%)
5	[1, 2, 3, 4, 5]	5038 (44.31%)

Insbesondere die Variationen 4 und 5 gehen mit einer starken Zunahme an Challenger-Vehikel verglichen mit den kurzzeitigen Prädiktion bis 3 Sekunden einher. Letztere zeigen untereinander keine signifikante Erhöhung des Prozentwerts.

Im Allgemeinen ist die Verwendung eines möglichst kurzen Prädiktionszeitraums anzustreben, weil einerseits die Rechenzeit während der Challenger-Identifikation dadurch minimiert werden kann und andererseits die Abweichung der Vorhersage des beschleunigungsbasierten Prädiktionsmodells von der Realität für kurze Vorhersagezeiträume vertretbar gering ist. Zur dementsprechende Festlegung des Parameters mit einem niedrigen Wert (Variation 1 oder 2 in Tabelle 4.2), muss jedoch ein zusätzliches Kriterium erfüllt werden. Dieses fordert, dass die Einstellung des Parameters t_{pred} die Erkennung von Challenger-Situationen aus allen 9 Szenarioklassen (Abbildung 3.12) ermöglicht.

Offensichtlicher Weise muss zum Nachweis dieser Bedingung garantiert werden, dass aus allen Klassen kritische Szenarien im betrachteten Datensatz vorkommen. Die Auswertung der Studie zeigt dahingehend, dass für hohe Prädiktionszeiten wie in Variation 5 aus Tabelle 4.2 dieses Kriterium erfüllt wird. Dementsprechend kann davon ausgegangen werden, dass ein vollständiger Szenarien-katalog erstellbar ist und daher die Verletzung des genannten Kriteriums negativ in die Beurteilung der jeweiligen Parameterwahl einfließen muss.

Dieser Sachverhalt zeigt sich anhand der Ergebnisse für zwei beispielhaft gewählte Einstellungen des untersuchten Parameters in Abbildung 4.2. Darin sind die Verteilungen der relevanten Fälle unter den Szenarioklassen für die beiden Vektoren $t_{\text{pred}} = [1]$ und $t_{\text{pred}} = [1, 2]$ zu sehen. Die Verteilung in der linken Grafik verdeutlicht, dass in den Klassen D und F bei Wahl einer kurzen Prädiktionszeit keine repräsentativen Situationen identifiziert werden können. Im Vergleich dazu veranschaulicht der rechte Teil aus Abbildung 4.2, dass die zweite Einstellung des Parameters zu einer Zuordnung von 6 Fällen in Szenarioklasse D führt.

Obwohl die längere Prädiktionszeit ebenso keinen vollständigen Szenarien-katalog erzeugt, erweist sich die Einstellung der Prädiktion auf einen Vorhersagehorizont von 2 Sekunden dennoch als ausreichend. Dies begründet sich darin, dass nach Abbildung 2.15 die Fälle D und F im Wesentlichen dasselbe Szenario gespiegelt um die Querachse des Ego-Vehikels durch dessen Fahrzeugmittelpunkt repräsentieren. Diese rotationssymmetrische Eigenschaft lässt die Annahme zu, dass bei einer Identifikation von Fällen der einen Klasse ebenso Situationen der anderen Klasse potentiell erkannt werden können. Dieses Argument untermauert demnach die Wahl von $t_{\text{pred}} = [1, 2]$ als optimale Parametrierung der Prädiktionszeit.

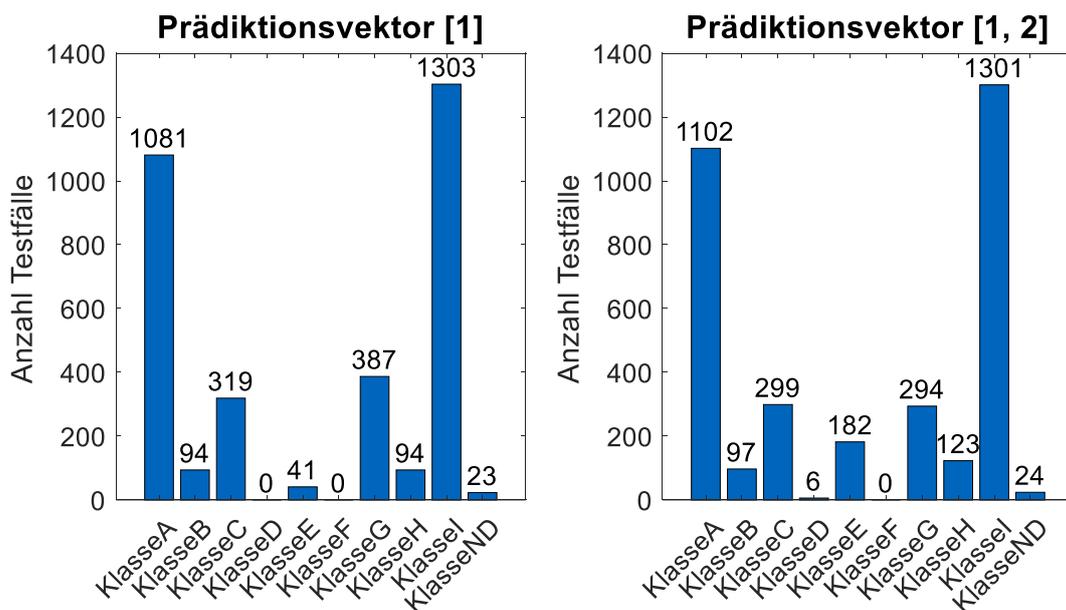


Abbildung 4.2: Beispielhafte Verteilung der relevanten Testfälle unter den Szenarioklassen A-I und der Klasse ND (Nicht Definiert) bei Verwendung unterschiedlicher Einstellungen des Parameters t_{pred} in Sekunden.

Neben dieser Eigenschaft findet sich der gewählte Parameterwert in der Literatur in Verbindung mit einem interessanten Aspekt der menschlichen Informationsverarbeitung beim Autofahren wieder. In [61, S. 126] wird die Fahrzeugführung als Gegenwartserlebnis für den Menschen beschrieben. Ein solches wird als eine wahrnehmbare Zeitspanne von 2 Sekunden beschrieben, wobei dieser Literaturwert somit exakt der in der Parameterstudie ermittelten optimalen Einstellung der Prädiktionszeit entspricht. Dieser Zeitraum beschreibt im Allgemeinen die Dauer, in der ein Mensch während des Autofahrens Entscheidungen fällt, Handlungen plant und Reaktionen umsetzt.

Wird diese Erkenntnis auf das beschleunigungsbasierte Vorhersagemodell des Tools umgelegt, so kann der Einfluss der Prädiktionszeit auf die Erkennung einer Challenger-Situation folgendermaßen interpretiert werden. Wird zu einem beliebigen Moment über die Prädiktion des Ego-Fahrzeugs ein Umgebungsvehikel als kollisionskritisch eingestuft, so muss nach beschriebenem Prinzip des Gegenwartserlebnisses der menschliche Fahrer spätestens zu diesem Zeitpunkt mit der Entscheidungsfindung und Umsetzung geeigneter Gegenmaßnahmen zur Kollisionsvermeidung beginnen. Dies entspricht der theoretischen Annahme, dass ein Lenker erst ab diesem Zeitpunkt ein Bewusstsein zur Situationsinterpretation erlangt und daher lediglich 2 Sekunden zur vollständigen Informationsverarbeitung und adäquaten Handlungsumsetzung zur Verfügung hat. Diese fiktive Situation tritt beispielsweise dann ein, wenn ein AutobahnpiLOT die Aufgabe der Fahrzeugführung ohne Übernahmezeit vollständig an einen unaufmerksamen Lenker übergibt. Schafft es der Lenker unter genannten Randbedingungen geeignet zu reagieren und dadurch eine Kollision zu vermeiden, beziehungsweise das Kollisionsrisiko zu senken, so gilt die Situation als sicher bewältigbar.

Wird in derselben Situation und unter denselben Bedingungen ein autonomes Fahrzeug im Zuge eines Absicherungsprozesses getestet, so kann dessen Reaktion und Bewältigungsstrategie unmittelbar der des Menschen gegenübergestellt werden. Somit eignen sich Challenger-Situationen auf Grundlage des 2-sekündigen Prädiktionshorizonts besonders, um die Sicherheit des autonomen Systems verglichen mit dem menschlichen Durchschnittsfahrer bewerten zu können.

Untersuchung Prädiktionslänge

Neben der Vorgabe des Zeithorizonts bei der Positionsvorhersage, hat die Wahl der Länge l_{coll} des untersuchten Kollisionsbereichs vor und hinter dem Ego-Fahrzeug einen maßgeblichen Einfluss darauf, welche und wie viele Umgebungsfahrzeuge als potenzielle Unfallpartner eingestuft werden (Abbildung 3.6). In Abschnitt 3.2.1 wird bereits darauf hingewiesen, dass diese Länge geschwindigkeitsabhängig gestaltet werden muss. Dazu wird ähnlich zur Berechnung der ROI aus Kapitel 3.2.2 der gesetzlich vorgeschriebene Sicherheitsabstand (SA) zur Definition der Kollisionslänge l_{coll} herangezogen. Im Vergleich zur Region-of-Interest wird der Wert für l_{coll} über die Multiplikation des Sicherheitsabstands mit dem parametrierbaren Faktor $l_{\text{coll, fact}}$ festgelegt, wobei dieser Wert die Ausdehnung des Kollisionsbereichs in x-Richtung auf einen definierten Bruchteil des Sicherheitsabstands beschränkt (Formel (4.1)).

$$l_{\text{coll}} = l_{\text{coll, fact}} SA \quad (4.1)$$

Somit wird die vorliegende Parameterstudie zur Kollisionslänge im Wesentlichen über die Variation des Faktors $l_{\text{coll, fact}}$ beschrieben. Die untersuchten Werte des Parameters sind Tabelle 4.3 zu entnehmen. Die eingetragenen Variationen beruhen auf den gesetzlich geregelten Bußgeldstufen für das Vergehen der Unterschreitung des vorgeschriebenen Sicherheitsabstands auf deutschen Autobahnen [62]. Analog zur vorherigen Studie werden die beiden anderen Parameter als Konstanten behandelt, $b_{\text{coll}} = 0.75 \text{ m}$ und $t_{\text{pred}} = [1, 2]$.

Im Zuge der Parameterstudie kann der veränderliche Faktor als Kritikalitätsmaß für die Aufnahme eines Umgebungsfahrzeugs in die Menge der Challenger-Vehikel interpretiert werden. Dies begründet sich darin, dass kleine Werte für $l_{\text{coll, fact}}$ zu geringeren Längen des Kollisionsbereichs führen, weshalb im Zuge der Prädiktion nur besonders nahe und somit kollisionskritische Umgebungsfahrzeuge erfasst werden. Analog zu den Untersuchungen der Prädiktionszeit t_{pred} , soll die Wahl des geeignetsten Faktors ebenfalls unter Beachtung der resultierenden Verteilung der Fälle innerhalb der Szenarioklassen erfolgen.

Tabelle 4.3: Parametervariation von $l_{\text{coll, fact}}$ basierend auf dem Bußgeldkatalog für das Vergehen der Unterschreitung des gesetzlich vorgeschriebenen Sicherheitsabstand [62].

Variation	$l_{\text{coll, fact}}$ [-]	Anzahl Challenger-Fahrzeuge (% von 11370 Fahrzeugen)
1	0.5	3404 (29.94%)
2	0.4	2457 (21.61%)
3	0.3	1390 (12.23%)
4	0.2	580 (5.10%)
5	0.1	178 (1.57%)

Die Analyse der in Tabelle 4.3 angegebenen Ergebnisse der Parameterstudie von $l_{\text{coll, fact}}$ bestätigt in erster Linie die Erwartung, dass sich die Anzahl an Challenger-Vehikel deutlich verringert, wenn die Länge des überprüften Kollisionsbereichs in der Umgebung des prädizierten Ego-Fahrzeugs verkürzt wird. Zudem kann festgehalten werden, dass die Länge des Kollisionsbereichs im Vergleich zur Prädiktionszeit einen wesentlich stärkeren Einfluss auf die Reduktion der Challenger-Anzahl hat. Diese Eigenschaft ist wünschenswert, weil sich dadurch der Umfang des finalen Testkatalogs verringert und gleichzeitig die Relevanz der verbleibenden Testfälle für die Absicherung autonomer Fahrfunktionen höher eingestuft wird. Dennoch kann der minimale Faktor von 0.1 aus Tabelle 4.3 nicht ohne Weiteres als optimale Einstellung des Parameters angenommen werden.

Der Grund dafür ist, dass unter denselben argumentativen Aspekten wie in Verbindung mit der Variation der Prädiktionszeiten, die resultierende Szenarierverteilung als Entscheidungskriterium bei der Wahl von $l_{\text{coll, fact}}$ beachtet werden muss. Nur so lässt es sich vermeiden, dass die Anzahl an ermittelten Challenger-Situationen zwar auf ein Minimum reduziert wird, der resultierende Testkatalog jedoch im Sinne nicht vorhandener Fälle in einzelnen Szenarioklassen unvollständig ist. Dahingehend ist die Betrachtung der Variation 1 in Tabelle 4.3 von besonderem Interesse, weil diese eine identische Parameterkonfiguration wie jene der Variante 2 in Tabelle 4.2 darstellt ($b_{\text{coll}} = 0.75 \text{ m}$, $l_{\text{coll, fact}} = \frac{5}{10}$, $t_{\text{pred}} = [1, 2]$). Die Resultate der vorangegangenen Studie zeigen, dass eben diese Parameterzusammenstellung die Extraktion eines vollständigen Testkatalogs auf Grundlage der ersten 10 highD Datensätze ermöglicht. Folglich ist das Kriterium der Vollständigkeit für die erste Variation von $l_{\text{coll, fact}}$ erfüllt. Demensprechend lässt sich die aktuelle Studie als Parameteroptimierung im Sinne der Suche nach der maximalen Verkleinerung des Werts für $l_{\text{coll, fact}}$ (ausgehend von 0.5) interpretieren, bei der beschriebenes Kriterium gerade noch erfüllt wird.

Abbildung 4.3 vergleicht hierfür exemplarisch die Verteilungen für die Faktoren 0.1 und 0.3. Die Säulendiagramme spiegeln bedingt die erwartete Distribution der Fälle zwischen den Szenarioklassen wider.

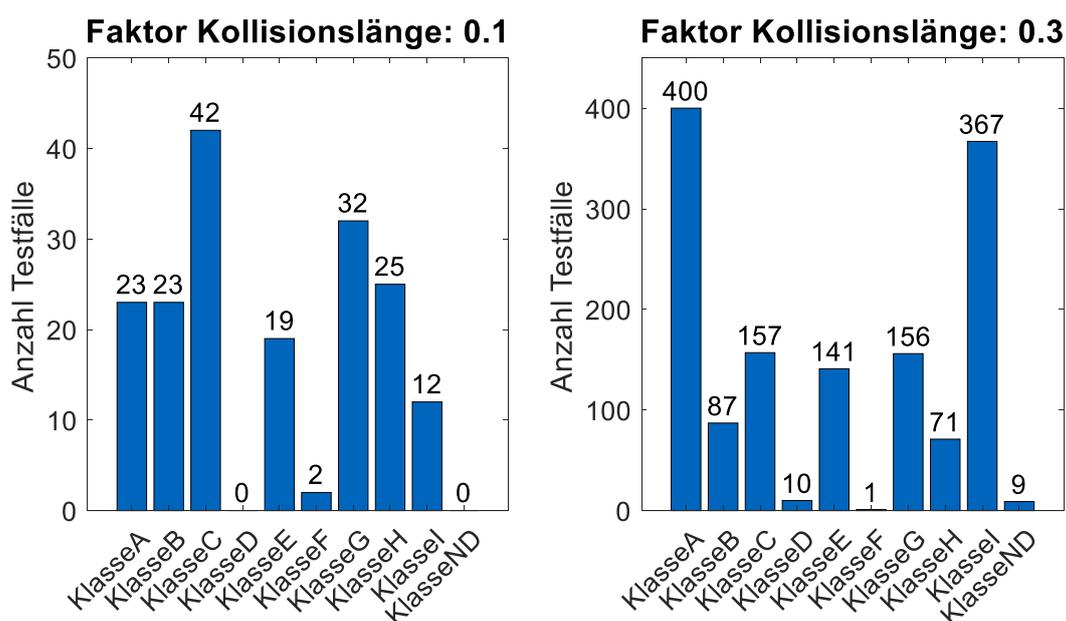


Abbildung 4.3: Verteilung der relevanten Testfälle unter den Szenarioklassen A-I und ND (nicht definiert) für zwei unterschiedliche Einstellungen des Parameters $l_{\text{coll, fact}}$.

Global betrachtet nimmt die Gesamtanzahl an Challenger-Vehikel mit steigender Länge des Kollisionsbereichs zu. So führt die Verdreifachung des Parameterwerts wie in Abbildung 4.3 dargestellt zur Erkennung von ca. der 8-fachen Menge an kollisionsgefährlichen Fahrzeugen. Entgegen den Erwartungen kann dieser Zusammenhang jedoch nicht direkt auf die einzelnen Szenarioklassen des Testkatalogs übertragen werden, was anhand von zwei bemerkenswerten Erkenntnissen aus der Analyse der Resultate in Abbildung 4.3 nachvollziehbar ist. Für diese Erkenntnisse ist die Klasse ND (nicht definiert) nicht von Belang, weshalb die undefinierten Fälle in den folgenden Ausführungen vernachlässigt werden.

Die erste Besonderheit in den parameterbedingten Entwicklungen der Fallzahlen stellt der variierende Zuwachs in den einzelnen Klassen dar. Mit Ausnahme der Klassen D und F kann eine positive Korrelation zwischen Bereichslänge und Challenger-Anzahl festgestellt werden. Die Analyse des prozentualen Zuwachses der Testfälle zeigt jedoch, dass genannte Korrelation in den einzelnen Klassen unterschiedlich stark ausgeprägt ist. Insbesondere hebt sich die Entwicklung der Fallzahlen in den beiden Klassen A und I deutlich von den restlichen ab. So zeigt die Gegenüberstellung der Testfallzahlen in Abbildung 4.3 für die beiden genannten Klassen eine parameterbedingte Erhöhung um eine Faktor 17.4 (Klasse A), beziehungsweise 30.6 (Klasse I) bezogen auf die jeweiligen Resultate bei $l_{\text{coll, fact}} = \frac{1}{10}$. Im Vergleich dazu liegen die entsprechenden Faktoren für die restlichen Klassen in einem Bereich zwischen 3.7 und 7.4, wodurch die ausgeprägte Dominanz der Klassen A und I verdeutlicht wird.

Dieser Sachverhalt lässt sich über die Betrachtung der Szenariodefinitionen der Klassen A und I erklären. Wie in Abbildung 2.15 dargestellt, handelt es sich bei den Szenarien A und I um Auffahrunfälle durch das Ego-Fahrzeug, beziehungsweise das Challenger-Vehikel. Es ist naheliegend, dass eine Verlängerung des Kollisionsbereichs über die Variation von $l_{\text{coll, fact}}$ bei ansonsten identen Parametern der Prädiktion zu einer vermehrten Identifikation von kritischen Folgefahrtsituationen führt. Im Gegensatz dazu wirkt sich eine Verlängerung des Kollisionsareals deutlich geringer auf die Erkennung der restlichen Szenarien B-H (Abbildung 2.15) aus. Der Hintergrund dafür liegt in deren Ausgangspositionen, die ausnahmslos auf dem benachbarten Fahrstreifen des Ego-Fahrzeugs liegen. In Folge dessen hat ein verlängerter Kollisionsbereich keinen direkt erkennbaren Einfluss auf die Identifikation dieser Szenarien.

Als zweite Besonderheit in den Verteilungen aus Abbildung 4.3 ist auffällig, dass den Veränderungen der Fallzahlen in den Klassen D und F kein erkennbares Muster zugrunde liegt. Es lässt sich einzig feststellen, dass wenige Situationen den genannten Klassen zugeordnet werden und größere Werte des Faktors $l_{\text{coll, fact}}$ mit einer tendenziell rückläufigen Entwicklung der Fallzahlen dieser Klassen einhergehen.

Anhand eines repräsentativen Beispiels kann nachvollzogen werden, wie diese Eigenschaften der Verteilungen innerhalb der Szenarioklassen D und F zustande kommen. In Abbildung 4.4 ist exemplarisch die Erfassung eines Umgebungsfahrzeugs (orange) als Challenger auf Basis der Prädiktion des Ego-Vehikels (blau) dargestellt, wobei 3 verschiedene Kollisionsbereichslängen (grün-rote Bereiche um Ego-Prädiktion) gegenübergestellt werden.

Die Grafik in Abbildung 4.4 oben links verdeutlicht, dass eine zu kurze Länge des Kollisionsbereichs die Umgebungsfahrzeuge trotz einer subjektiv beurteilt kritischen Nähe zum Ego nicht als Challenger wahrnimmt. Im Gegensatz dazu werden dieselben Fahrzeuge bei einem langen Kollisionsbereich vermehrt als kritische Verkehrsteilnehmer erfasst (Abbildung 4.4 oben rechts).

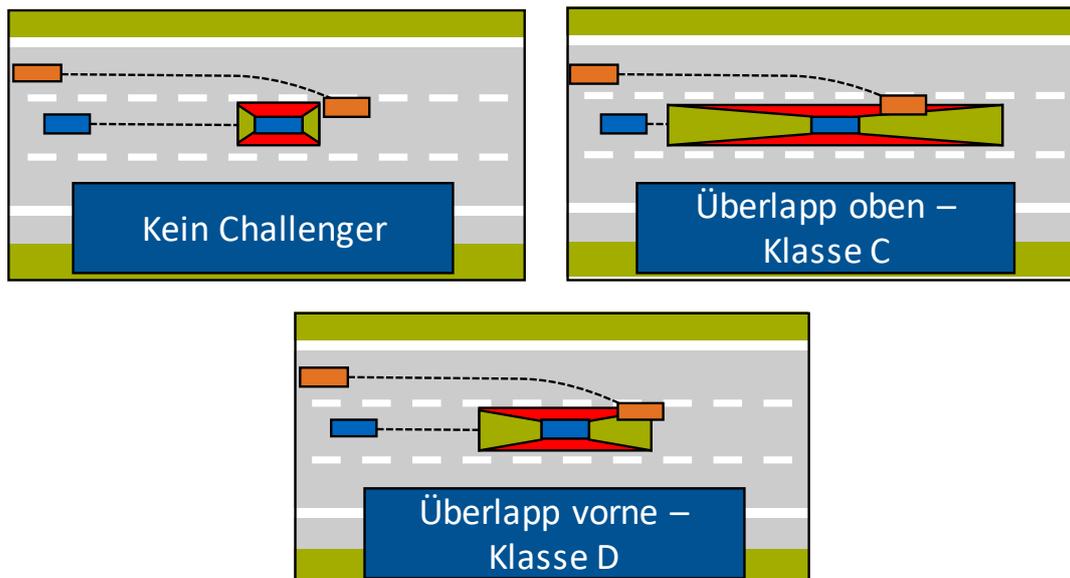


Abbildung 4.4: Auswirkung der Länge des Kollisionsbereichs (rot-grüne Areale) auf die Zuordnung der Challenger-Situationen zu den Klassen D und F.

Die trapezförmige Ausführung der separaten Kollisionsareale führt jedoch dazu, dass diesen Challenger-Fahrzeugen zumeist ein Gefahrenpotential bezüglich seitlichen Kollisionen zugeschrieben wird (Klasse C). Die Grafik oben rechts in Abbildung 4.4 veranschaulicht diesen Sachverhalt, weil nach dem Erstkontaktprinzip das Challenger-Fahrzeug initial mit dem rot dargestellten Kollisionsareal überlappt.

Über die passende Einstellung des Wertes für $l_{\text{coll, fact}}$ in einem Bereich zwischen den beiden angeführten Extrema kann die Identifikation der dargestellten Verkehrssituationen als Szenarien der Klassen D oder F realisiert werden (Abbildung 4.4 unten). In diesem Fall überlappt das kollisionskritische Umgebungsfahrzeug bei Erstkontakt hauptsächlich mit dem grünen Bereich vor dem prädierten Ego-Vehikel, weshalb dieses Areal im Weiteren als zentral kollisionskritischer Bereich betrachtet wird.

Diese Erörterungen, in Kombination mit der Rahmenbedingung die Anzahl an Testfällen möglichst gering zu halten, führen auf die optimale Wahl des Parameters $l_{\text{coll, fact}} = 0.2$. Dadurch wird gewährleistet, dass Situationen ähnlich jener aus der unteren Grafik in Abbildung 4.4 definitiv den Szenarioklassen D oder F zugeordnet werden können und nicht fälschlicher Weise als Abdränger-, beziehungsweise Seitenaufprall-Szenarien (Klasse C) interpretiert werden.

Studie zum Einfluss des Challenger-Identifikationsprinzips

Unabhängig von den vorherigen Parameterstudien soll in diesem Abschnitt analysiert werden, welche Bedeutung dem Prinzip hinter dem Auswahlprozess des primären Challenger-Fahrzeugs in einer Verkehrssituation zugeschrieben wird. Wie in Kapitel 3.2.1 erörtert, wird in der vorliegenden Arbeit hauptsächlich das Erstkontaktprinzip verfolgt, also jenes Umgebungsfahrzeug als primärer Challenger für ein Ego-Vehikel betrachtet, bei dem zum frühesten Zeitpunkt ein Kollisionspotential ermittelbar ist. Alternativ dazu kann jenes Vehikel als relevanter

Challenger eingestuft werden, dessen Kollisionspotential im Zuge des Verkehrsszenarios verglichen mit allen anderen Umgebungsfahrzeugen einen Maximalwert annimmt.

Somit kommt es zu einer abweichenden Interpretation von Verkehrssituationen durch die beiden grundverschiedenen Ansätze, wodurch gegebenenfalls für dasselbe Ego-Vehikel unterschiedliche Fahrzeuge als relevante Challenger identifiziert werden. Dies kann in weiterer Folge zu einer variablen Klassifikation der Szenarien und dadurch zu Veränderungen in der Verteilung der Fälle innerhalb des Testkatalogs führen.

Um den Umfang dieser Auswirkungen auf den Output des gesamten Tools abschätzen zu können, wird in zwei Durchläufen die gesamte Verarbeitungskette unter Variation des genannten Challenger-Identifikationsprinzips auf die repräsentativ gewählten Datensätze 1-10 angewendet. Die restlichen Parameter werden auf Basis der Resultate des vorangegangenen Abschnitts eingestellt.

In Abbildung 4.5 sind die Ergebnisse dieser Untersuchungsreihe zu sehen. Die Anzahl an extrahierten Challenger-Situationen in den jeweiligen Katalogen des Erstkontaktverfahrens und des Maximalwertansatzes ist mit 577 Szenarien erwartungsgemäß identisch. Das begründet sich darin, dass das frame-basierte Ermitteln der Kollisionspotentiale (Abbildung 3.6) und damit die entscheidende Challenger-Erkennung für beide Methoden unverändert abläuft, was in der identen Anzahl an Fällen resultiert. Der wesentliche Unterschied zwischen den beiden Prinzipien manifestiert sich erst in der Interpretation der erfassten Kollisionspotentialverläufe (Abbildung 3.7).

Die Häufigkeitsverteilungen in Abbildung 4.5 verdeutlichen, dass die Gesamtanzahl an Testfällen in den beiden Katalogen zwar identisch ist, die interne Verteilung der Fälle unter den Szenarioklassen jedoch eine ausgeprägte Ungleichheit zwischen dem Erstkontakt- und dem Maximalwertprinzip aufzeigen.

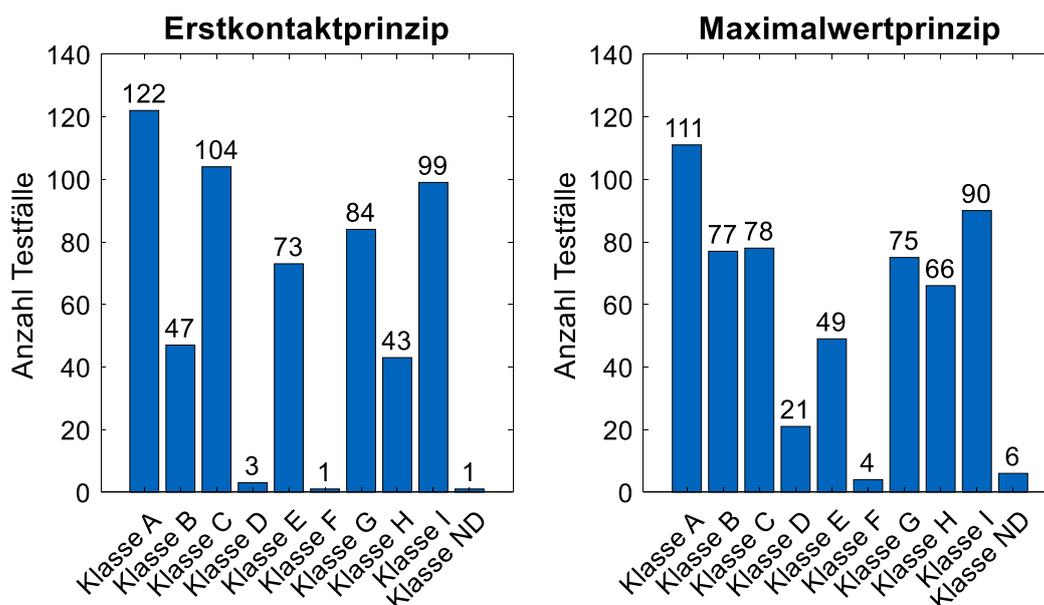


Abbildung 4.5: Vergleich der Szenarioverteilung für die beiden Challenger-Identifikationsprinzipien. Es zeigen sich deutliche Unterschiede in den Verteilungen der beiden Verfahren.

Eine erste Auffälligkeit beschreibt die höhere Anzahl an nicht definierten Szenarien (Klasse ND) bei Verwendung des Maximalwertverfahrens (Abbildung 4.5, rechts). Die Differenz zwischen den beiden Methoden in dieser Klasse ist jedoch gering und mit Anteilen von 0.17 % (Erstkontakt) und 1.04 % (Maximalwert) bezogen auf die Gesamtzahl an Testfällen kann diese Szenarioklasse in den folgenden Erörterungen vernachlässigt werden.

Eine bedeutendere Erkenntnis zeigt sich dahingehend, dass bei Verwendung des Maximalwertprinzips wesentlich mehr Situationen in den Klassen D und F identifiziert werden. Verglichen mit der Verteilung des Erstkontaktprinzips steigt außerdem die Anzahl der Fälle in den Klassen B und H, wohingegen die Bereiche C, E und G an Situationen verlieren. Zusammengefasst verdeutlicht das, dass sich die Gesamtanzahl an Challenger-Fahrzeugen in den Klassen A-I nur unwesentlich ändert und dennoch große Unterschiede in den Verteilungen innerhalb der Szenarioklassen auftreten. Diese Tatsache legt die Vermutung nahe, dass die Variation des Identifikationsprinzips primär zu einer Umverteilung der erkannten Challenger-Fälle innerhalb der Szenarioklassen führt.

Um diese These zu untersuchen, wird ein Vergleich zwischen den beiden Häufigkeitsverteilungen aus Abbildung 4.5 gezogen. Ziel ist es dabei, idente Challenger-Vehikel in den Verteilungen zu identifizieren und deren Zuordnung innerhalb der Szenarioklassen abzugleichen. Beispielhaft zeigt Abbildung 4.6 die Umverteilung der Fälle aus der Szenarienkategorie B nach dem Erstkontaktprinzip auf den Testkatalog des Maximalwertverfahrens. Anhand dieses repräsentativen Beispiels lässt sich feststellen, dass der Anteil an Challenger-Situationen der Klasse B primär zwischen den Klassen B und C aufgeteilt wird, wenn anstelle des Erstkontaktprinzips das maximale Kollisionspotential als entscheidende Größe betrachtet wird.

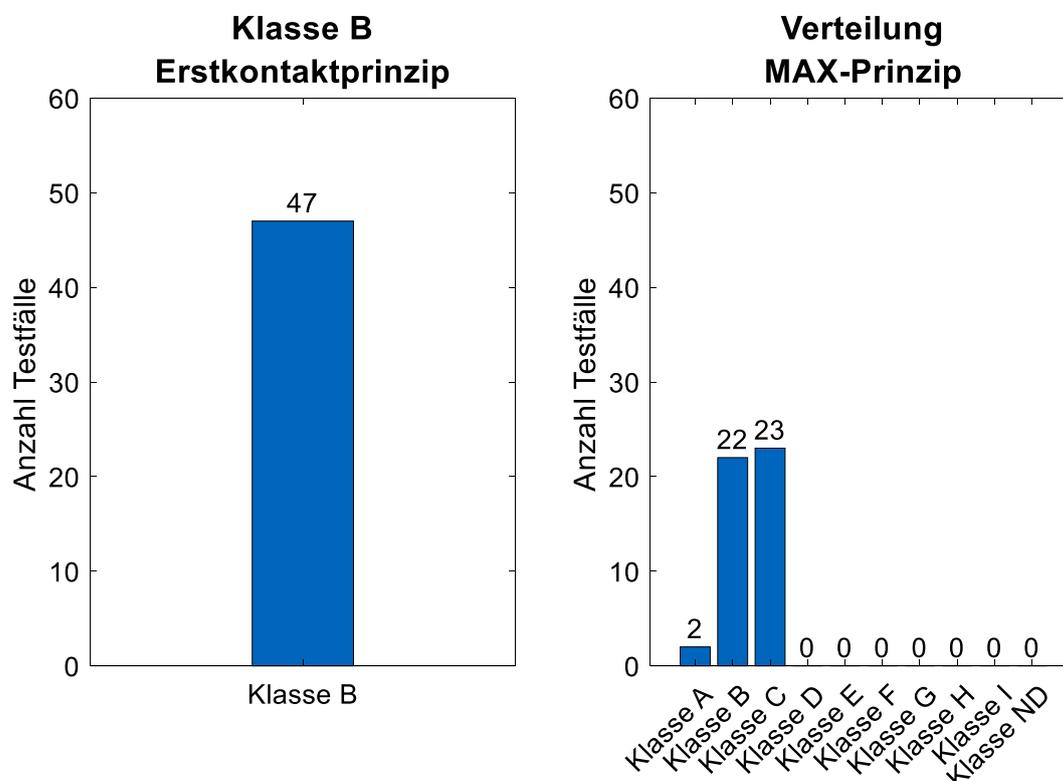


Abbildung 4.6: Verteilung der Challenger-Fälle von Szenarioklasse B. Die 47 Testfälle aus Klasse B nach dem Erstkontaktprinzip verteilen sich entsprechend der rechten Grafik innerhalb des Katalogs nach dem Maximalwertprinzip.

Diese Verteilung kann anhand eines Vergleichs der Challenger-Trajektorien dieser beiden Klassen (Abbildung 2.15) nachvollzogen werden. Die entsprechenden Trajektorien B und C besitzen einen einheitlichen Ausgangspunkt und unterscheiden sich hauptsächlich über variable Kollisionsbereiche am Ego-Vehikel. Diesem Aspekt zugrundeliegend identifiziert das Maximalwertprinzip innerhalb der identischen Challenger-Fälle offenbar vermehrt ein erhöhtes seitliches Kollisionspotential während der Verkehrssituation, auch wenn sich nach dem Erstkontaktprinzip zu einem früheren Zeitpunkt bereits eine Tendenz zu Trajektorie B aus Abbildung 2.15 ausgebildet hat.

Unter einer analogen Betrachtung der Umverteilung von Challenger-Situationen für alle 9 Szenarioklassen, lassen sich die folgenden Zusammenhänge ableiten. Die Angaben beziehen sich jeweils auf die Verteilung der Fälle einer Klasse des Erstkontaktprinzips auf den Katalog des Maximalwertverfahrens analog zum Beispiel in Abbildung 4.6.

- Fälle aus Klasse A verbleiben primär in Klasse A
- Fälle aus Klasse B verteilen sich primär auf die Klassen B und C
- Fälle aus Klasse C verteilen sich primär auf die Klassen B und C
- Klasse D tritt primär bei Maximalwertprinzip auf
- Fälle aus Klasse E verteilen sich auf restliche Szenarioklassen außer A und I
- Klasse F tritt primär bei Maximalwertprinzip auf
- Fälle aus Klasse G verteilen sich primär auf die Klassen G und H
- Fälle aus Klasse H verteilen sich primär auf die Klassen G und H
- Fälle aus Klasse I verbleiben primär in Klasse I

Fazit zur Variation des Challenger-Identifikationsprinzips

Grundsätzlich haben beide untersuchten Auswahlprinzipien der Challenger-Fahrzeuge eine Daseinsberechtigung, weil alle zwei Verfahren zur Extraktion eines validen Testkatalogs aus den Rohdaten führen. Somit obliegt es dem Anwender des vorgestellten Werkzeugs das gewünschte Prinzip zu wählen. Entscheidend ist dahingehend, wozu die extrahierten Testfälle im Zuge eines Absicherungsprozesses autonomer Fahrfunktionen dienen sollen. Besteht das primäre Ziel in der Überprüfung der Reaktionen des Systems in überaus kritischen Situationen, so ist das Maximalwertprinzip dem Erstkontaktverfahren vorzuziehen. Soll hingegen untersucht werden, wie ein autonomes Vehikel im Allgemeinen auf gewöhnliche Einflüsse durch umgebende Fahrzeuge reagiert, stellt das Erstkontaktprinzip die bessere Wahl dar.

4.1.2 Parameterstudie zur Klassifikation

In Kapitel 3.2.2 wird das Modul zur Identifikation der Verkehrsmanöver hinter den geclusterten Situationen vorgestellt. Es wird im speziellen darauf hingewiesen, dass die Genauigkeit des Klassifikators zum größten Teil von der Qualität und der Vollständigkeit der manuellen Manöverdefinitionen abhängt. Zudem hat der Parameter $v_{y,\text{lim},LC}$ im weiteren Sinn einen bedeutenden Einfluss auf die Klassifikation, weil er die zentrale Größe zur Erfassung von Fahrstreifenwechselzuständen repräsentiert und daher maßgeblich zur korrekten Erfassung der Verkehrsmanöver beiträgt. Auf beide Einflussfaktoren wird in den folgenden Unterkapiteln eingegangen.

Untersuchung Fahrstreifenwechsel-Quergeschwindigkeit

Im Zuge der Entwicklung des Zustandserkennungsmoduls hat sich gezeigt, dass zur Identifikation einer Fahrstreifenwechselsituation die alleinige Betrachtung der y-Position eines Fahrzeugs nicht ausreicht, um genannte Manöver sicher und vollständig erfassen zu können. Wesentlich genauere Aussagen über den tatsächlichen Beginn und das Ende eines Fahrstreifenwechsels lassen sich bei zusätzlicher Betrachtung der Dynamikwerte des Vehikels in Querrichtung treffen.

Ein Fahrstreifenwechsel geht im Vergleich zu einer stabilen Geradeausfahrt erwartungsgemäß mit höheren Werten der Querbeschleunigung einher. Jedoch unterscheidet sich diese Größe betragsmäßig zu undeutlich von den Werten, die während Korrekturmaßnahmen der Lenker zum Halten des aktuell befahrenen Fahrstreifens auftreten. Somit ist die Querbeschleunigung als Maß zur Erkennung der Fahrstreifenwechsel ungeeignet.

Diese Eigenschaft betrifft die Geschwindigkeit in y-Richtung jedoch nicht, weil im Vergleich zu erwähnten Korrekturmaßnahmen die Quergeschwindigkeit im Zuge eines Fahrstreifenwechsels deutlich höhere Werte annimmt und folglich als charakteristische Größe eines solchen Manövers betrachtet werden kann. Dies begründet sich primär darin, dass die Querbeschleunigung im Zuge eines Fahrstreifenwechsels deutlich länger in dieselbe Richtung wirkt als bei Korrekturmaßnahmen, wodurch sich eine zunehmend höhere Quergeschwindigkeit aufbaut.

Es muss explizit darauf hingewiesen werden, dass die betrachteten Autobahnabschnitte durchwegs gerade verlaufen, weshalb Dynamikwerte in Querrichtung in keinem Fall auf einen gekrümmten Straßenverlauf zurückzuführen sein können. Daher ist die Verwendung der Quergeschwindigkeit zur Identifikation eines Fahrstreifenwechsels in diesem Fall zulässig.

Der Fokus dieses Teils der Parameterstudien richtet sich auf die Ermittlung eines geeigneten Schwellwerts $v_{y,\text{lim},LC}$ bezüglich der y-Geschwindigkeit zur Unterscheidung einer Geradeausfahrt von einem Fahrstreifenwechsel. Dazu sind im Wesentlichen zwei Fragen zu beantworten, die in separaten Untersuchungen der highD Daten behandelt werden:

- Hängt die Quergeschwindigkeit während eines Fahrstreifenwechsels von der Längsgeschwindigkeit des Fahrzeugs ab und wenn ja, in welcher Beziehung stehen die Größen zueinander?

- Mit welchem Wert muss demnach die Grenze zur Zustandserkennung des Fahrstreifenwechsels angesetzt werden?

Zur Beantwortung der ersten Frage müssen aus den Rohdaten jene Fahrzeuge extrahiert werden, die mit Sicherheit einen Fahrstreifenwechsel ausführen. Dazu werden die Daten der einzelnen Vehikel anhand eines numerischen Eintrags separiert, der angibt, wie häufig das zugehörige Fahrzeug im Laufe der Fahrt über den Autobahnabschnitt den Fahrstreifen wechselt. Die Teilmenge der über diesen Wert ermittelten Fahrzeuge mit einfachem Fahrstreifenwechsel wird anschließend dahingehend analysiert, ob sich ein korrelativer Zusammenhang zwischen den Längs- und Quergeschwindigkeiten finden lässt. Dazu werden die Geschwindigkeitswerte jenes Frames entnommen, in dem sich der Mittelpunkt des betrachteten Fahrzeugs während des Fahrstreifenwechsels am nächsten an der überfahrenen Fahrstreifenmarkierung befindet. Es wird die Annahme getroffen, dass an dieser Position die maximale Quergeschwindigkeit während des Manövers auftritt.

Abbildung 4.7 zeigt das Resultat der durchgeführten Regressionsanalyse, wobei entlang der Abszisse des Diagramms die Längsgeschwindigkeit der Fahrzeuge abgelesen werden kann. Auf der Ordinate sind die zugehörigen Quergeschwindigkeiten eingetragen.

Wird die Gleichung der resultierenden Regressionsgeraden betrachtet (Formel (4.2)), so lässt sich anhand der geringen Steigung von ca. $4.5 \cdot 10^{-4}$ kein signifikanter Zusammenhang zwischen den gerichteten Geschwindigkeiten während eines Fahrstreifenwechsels feststellen. Dementsprechend ist im Weiteren die Verwendung eines konstanten Grenzwerts zur Manövererkennung gerechtfertigt.

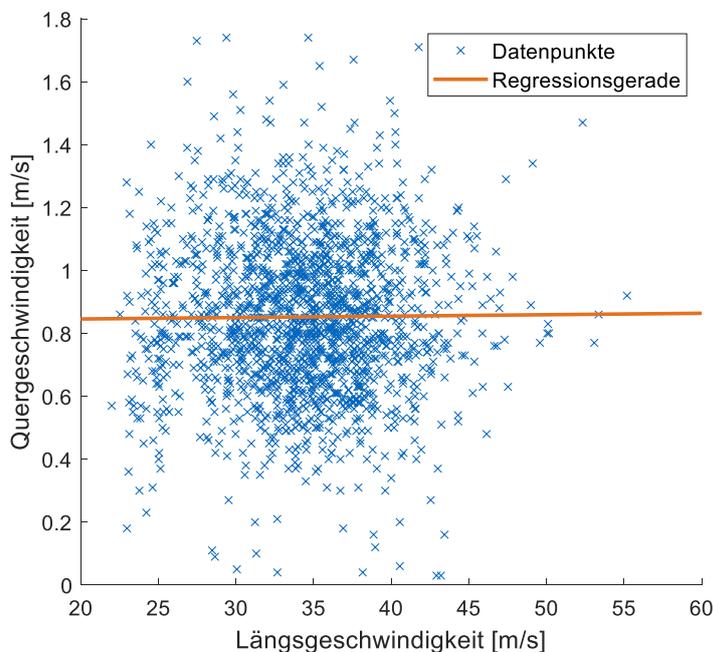


Abbildung 4.7: Regressionsanalyse der Spurwechselsituationen in den Datensätzen 1-10

$$y = 0.00045179x + 0.83673 \quad (4.2)$$

Aufbauend auf der Erkenntnis über die Unabhängigkeit zwischen Quer- und Längsgeschwindigkeit während eines Fahrstreifenwechsels, ist ein geeigneter konstanter Grenzwert für die Identifikation dieses Manövers zu finden. Die Schwellwertbetrachtung soll die Einleitung eines Fahrstreifenwechsels möglichst sicher und dennoch so früh wie möglich erkennen können. Um dahingehend eine geeignete Parametrierung von $v_{y,lim,LC}$ zu erreichen, wird im Folgenden untersucht, wie sich die Variation dieses Wertes (Tabelle 4.4) auf die Güte der Fahrstreifenwechselidentifikation auswirkt. Dazu werden aus den Fahrzeugdaten der ersten 10 highD Datensätze zufällig je 200 Fahrten mit und ohne Fahrstreifenwechsel ausgewählt. Mit Hilfe des implementierten Zustandserkennungsmoduls werden die als unbekannt angenommenen Daten unter Verwendung des variablen Parameters den Klassen Fahrstreifenwechsel und kein Fahrstreifenwechsel zugewiesen. Dabei ist die Untersuchung derart konzipiert, dass eine Situation bereits nach der erstmaligen Erkennung eines Fahrstreifenwechselzustands (S6/S7 Tabelle 3.1) als ein solches Manöver klassifiziert wird.

Die Verteilung zwischen den zwei Klassen muss entsprechend der getroffenen Auswahl 1:1 sein, also je 200 Fahrzeuge pro Gruppe. Tabelle 4.4 listet neben den Parameterwerten der Variationen, die jeweilige Anzahl der Zuordnungen zu den einzelnen Klassen durch das Zustandserkennungsmodul auf. Daraus können folgende Schlüsse gezogen werden.

Die Verteilung der Datenpunkte in Abbildung 4.7 streut stark, jedoch liegt die Mehrheit der Quergeschwindigkeitswerte in einem Bereich zwischen 0.5 und 1.2 m/s. Folglich führt ein zu hoch angesetzter Wert für $v_{y,lim,LC}$ zu einer drastischen Reduktion der Anzahl gefundener Fahrstreifenwechselsituationen, weil nur mehr besonders schnelle Wechsel erfasst werden. Ist der Wert hingegen zu niedrig, werden gegebenenfalls Situationen aufgrund einer unruhigen Geradeausfahrt fälschlicherweise den Fahrstreifenwechselmanövern angerechnet.

Die Wahl von 0.5 m/s für den Parameter erweist sich als geeignetster Wert, bei dem der Fehler bezüglich der Erkennung von Fahrstreifenwechselmanövern auf ein vertretbares Maß reduziert wird. Es ist anzumerken, dass die Werte aus Tabelle 4.4 für die Variation 3 eine ideale und fehlerfreie Identifikation von Fahrstreifenwechsel suggerieren, obwohl die Erkennungsrate tatsächlich nicht exakt 100 % beträgt.

Begründet wird dieser Trugschluss damit, dass das Erkennungsmodul bei den gewählten Daten zufälliger Weise exakt gleich viele Fälle der jeweils falschen Klasse zuweist. Deshalb gleicht sich die Bilanz in Tabelle 4.4 mit jeweils 200 identifizierten Fällen aus, die tatsächlich korrekten Identifikationen belaufen sich aber nur auf 96.5 %.

Tabelle 4.4: Parametervariation $v_{y,lim,LC}$

Variation	$v_{y,lim,LC}$ [m/s]	Anzahl Fahrstreifenwechsel	Anzahl kein Fahrstreifenwechsel
1	0.1	323	77
2	0.3	210	190
3	0.5	200	200
4	0.8	117	283
5	1.0	56	344

Eine Darstellung dieses Sachverhalts ist in Form einer Confusion-Matrix in Abbildung 4.8 zu sehen. Die Matrix beinhaltet die absoluten Zahlen der korrekten und falschen Klassifikationen, wobei in horizontaler Richtung die erwarteten Klassen und entlang der Vertikalen die Vorhersagen angeordnet werden. Im vorliegenden Fall werden jeweils 193 Situationen korrekt als Fahrstreifenwechsel (grünes Feld unten rechts), beziehungsweise kein Fahrstreifenwechsel (grünes Feld oben links) identifiziert, während jeweils 7 Fälle den falschen Klassen zugeordnet werden. Bezogen auf die gesamte Anzahl von 400 untersuchten Situationen ergibt sich der bereits erwähnte Prozentsatz von 96.5%, der somit als Maß für die Güte des Fahrstreifenwechseleerkennungsmoduls herangezogen werden kann.

Confusion Matrix

Vorhersage	kein FSW	FSW	
	kein FSW	FSW	Zielklasse
kein FSW	193 48.3%	7 1.8%	96.5% 3.5%
FSW	7 1.8%	193 48.3%	96.5% 3.5%
	96.5% 3.5%	96.5% 3.5%	96.5% 3.5%

Abbildung 4.8: Confusion-Matrix der Güte des Erkennungsmoduls von Fahrstreifenwechsel (FSW)

4.1.3 Resultat Parameterstudien

Die vorangegangenen Untersuchungen verdeutlichen die entscheidenden Einflüsse der betrachteten Parameter auf die Ergebnisse des Moduls zur Identifikation von Challenger-Fahrzeugen und den Resultaten des Klassifikationsschritts. Die Studien haben gezeigt inwiefern die Einstellungen dieser Werte die Zusammenstellung der relevanten Szenarien innerhalb des resultierenden Testkatalogs bestimmen.

Für die maßgeblichen Parameter der beiden Funktionsmodule hat sich dahingehend im Zuge der durchgeführten Analysen folgende Wertkombination als geeignete Basis für weitere Untersuchungen herausgestellt (Tabelle 4.5).

Unter Verwendung dieser Parameterkonfiguration können Testfälle aus allen Szenarioklassen gefunden werden, wodurch sich stets ein valider und vor allem vollständiger Testkatalog bei ausreichendem Umfang der untersuchten Rohdaten bilden lässt.

Tabelle 4.5: Optimale Parametereinstellung als Resultat der durchgeführten Parameterstudien

Modul	Parameter	Wert	Einheit
Clustering	k_{SD}	2	[-]
Challenger-Prädiktion	t_{pred}	[1, 2]	[s]
	b_{coll}	0.75	[m]
	$l_{coll, fact}$	0.2	[-]
Klassifikation	$v_{y, lim, LC}$	0.5	[m/s]

4.2 Stichprobenanalyse zur Überprüfung der Güte des Klassifikators

In folgendem Abschnitt wird die Güte des entworfenen Moduls zur Klassifikation der Verkehrsszenarien anhand einer stichprobenartig ausgeführten Analyse ermittelt. Dazu wird im ersten Schritt berechnet, welchen Umfang die Stichprobe für ein global repräsentatives Ergebnis besitzen muss. Daraufhin wird aus dem Katalog an gewonnenen Testfällen die notwendige Anzahl an Verkehrssituationen extrahiert, deren Abläufe visualisiert und die zuvor ermittelten Manöver der einzelnen Fahrzeuge anhand einer subjektiven Kontrolle auf ihre Korrektheit überprüft. Die Stichproben werden dabei aus der Gesamtheit aller klassifizierbaren Testfälle entnommen, also aus den einzelnen Situationen innerhalb der 9 Szenarioklassen des Testkatalogs. Als Basis für die Untersuchung dienen erneut die Ergebnissätze 1-10, wobei als Parameterkonfiguration bei der Erzeugung der Resultate die optimalen Werte aus Tabelle 4.5 verwendet werden.

Aus den 10 Rohdatensätzen werden im Zuge des Verarbeitungsprozesses von allen geclusterten Situationen unter der angegebenen Parameterkonfiguration 5038 Verkehrsszenarien in den finalen Testkatalog aufgenommen. Die einzelnen Szenarien beinhalten dabei das Ego-Vehikel und ein oder mehrere Umgebungsfahrzeuge, wobei sich in Summe eine Anzahl von 23815 Fahrzeugen ergibt. Aus dem durchgeführten Klassifikationsschritt ist von jedem dieser Vehikel bekannt, welches Manöver das Fahrzeug innerhalb seines Verkehrsszenarios bezogen auf das zugehörige Ego-Vehikel ausführt. Somit liegt eine Menge von 23815 klassifizierten Manövern vor (das Ego wird auch klassifiziert), aus denen die Stichprobe zur Evaluierung der Güte dieser Klassifikationen gezogen wird.

Nachdem die Resultate der Stichprobenuntersuchung eine Aussage über die funktionale Qualität des Klassifikators bezogen auf die gesamte Populationsgröße von 23815 Vehikel ermöglichen soll, muss aus dieser Anzahl an Fahrzeugen eine statistisch repräsentative Menge im Sinne des Stichprobenumfangs entnommen und untersucht werden.

Dieser berechnet sich anhand folgender Formel (4.3) [63].

$$\text{Stichprobenumfang} = \frac{\frac{z^2 p(1-p)}{e^2}}{1 + \frac{z^2 p(1-p)}{e^2 N}} \quad (4.3)$$

Darin beschreibt N die Populationsgröße, also im vorliegenden Fall 23815 Fahrzeuge. Die Variable z basiert auf dem Konfidenzniveau, das wiederum angibt, wie gut die Gesamtpopulation durch das Resultat der Stichprobe repräsentiert wird. Im vorliegenden Fall wird $z = 1.96$ angesetzt, was einem Konfidenzniveau von 95% entspricht. Die dritte veränderliche Größe e beschreibt die Fehlermarge, die ein Maß der Präzision der über die Stichprobe gewonnenen Information darstellt. Die Variable p ermöglicht die Angabe einer erwarteten Standardabweichung der Ergebnisse, wobei in vielen Fällen $p = 0.5$ angegeben wird, weil zumeist vor der eigentlichen Durchführung der Studie keine genauere Kenntnis über diese Größe vorhanden ist [63].

In Zusammenhang mit der vorliegenden Stichprobenanalyse zur Erfassung der Güte des entwickelten Klassifikators werden folgende Parameter für Formel (4.3) gewählt:

$$z = 1.96$$

$$e = 0.05$$

$$p = 0.5$$

$$N = 23815$$

Die Güte des Klassifikators kann somit, bezogen auf die Gesamtpopulation, mit einer Wahrscheinlichkeit von 95% (z-Wert) in einem Bereich von $\pm 5\%$ (Fehlermarge e) der Klassifikationsgüte der Stichprobe erwartet werden.

Die Berechnung des Stichprobenumfangs bei den gewählten Parametern ergibt eine Mindestanzahl an zu untersuchenden Fahrzeugmanövern von rund 378. Aufgrund der Struktur der Cluster hinter den Verkehrsszenarien wird die zufällige Auswahl der Stichprobenelemente nicht auf Einzelfahrzeugebene durchgeführt, sondern der Einfachheit halber per Zufallsprinzip 100 der 5038 Verkehrsszenarien ausgewählt. Die variierende Anzahl an Fahrzeugen innerhalb dieser 100 Cluster führt auf einen endgültigen Stichprobenumfang von 477 Einzelfahrzeugen.

Im Zuge der subjektiven Analyse werden die Trajektorien und Geschwindigkeitsverläufe dieser Fahrzeuge untersucht und über einen Vergleich der Werte mit dem zugehörigen Ego-Vehikel das korrespondierende Manöver identifiziert. Der Abgleich der subjektiven Manöverzuordnung mit den Ausgaben des entwickelten Klassifikationsmoduls ermöglicht die Berechnung der prozentualen Erkennungsgenauigkeit des Systems. Abbildung 4.9 präsentiert das Ergebnis der Stichprobenanalyse. Der Grafik ist zu entnehmen, dass rund 95% (94.76%) der betrachteten Fahrzeugen durch den Klassifikator den korrekten Manövern zugeordnet werden, während der Anteil an inkorrekten Klassifizierungen bei knapp über 5% (5.24%) liegt.

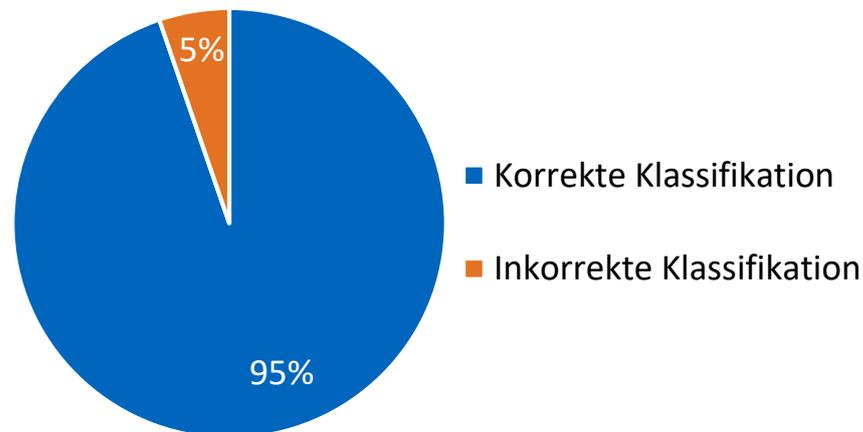


Abbildung 4.9: Güte des Klassifikators bezogen auf den Stichprobenumfang von 477 Fahrzeugen. Konkret wird in 94.76% der Untersuchungen die Verkehrssituation dem korrekten Manöver zugeordnet, während in 5.24% der Fälle eine falsche Klassifikation erfolgt.

Unter den gewählten Rahmenbedingungen der Stichprobenanalyse bedeutet das, dass die Genauigkeit des Klassifikators bei Untersuchungen der Gesamtpopulation (Datensätze 1-10) mit einer Wahrscheinlichkeit von 95% in einem Bereich zwischen 89.76% und 99.76% zu finden sein wird.

Zu Beginn von Kapitel 4.1 wird anhand eines Vergleichs der Häufigkeitsverteilungen der Durchschnittsgeschwindigkeiten zwischen der gesamten highD Datenbank und den gewählten Datensätzen 1-10 argumentiert, dass der verwendete Teil der Rohdaten eine allgemeingültige Repräsentationsmenge der gesamten Datenbank darstellt. Auf derselben argumentativen Basis kann die Erkenntnis zur Güte des Klassifikators unmittelbar auf die Gesamtheit des highD Datensatzes übertragen werden, sofern die zuvor angeführte Stichprobe über 477 Fahrzeuge einen ausreichenden Umfang besitzt, um für die gesamte Datenbank ebenfalls repräsentativ zu sein. Die globale Fahrzeuganzahl aller 60 highD Datensätze liegt bei 110500 Vehikel. Unter der Annahme, dass jedes dieser Fahrzeuge klassifiziert werden kann, berechnet sich der notwendige Stichprobenumfang bei dieser Populationsgröße über Gleichung (4.3) zu rund 384 zu untersuchende Vehikel. Offensichtlicher Weise wird dieser Stichprobenumfang bereits in der vorgestellten Analyse von 477 Fahrzeugen in ausreichendem Maß abgedeckt.

Obwohl die Stichprobe demnach eine Aussage über die globale Güte des Klassifikators zulässt, muss zur Bestätigung dieser Erkenntnis ein zusätzlicher Aspekt beachtet werden. Es wurden die Fahrzeuge der Stichprobe nach dem Zufallsprinzip ausgewählt, weshalb die Analyse Gültigkeit besitzt, jedoch gewährleistet das nicht das Vorkommen aller potentieller Fahrmanöver innerhalb der betrachteten Menge. Somit kann die ermittelte Güte nur dann als global repräsentativ betrachtet werden, wenn die Stichprobe möglichst das gesamte Spektrum an Fahrmanövern abdeckt. Wie in Kapitel 3.2.2 angeführt setzt sich dieses aus 12 übergeordneten Klassen zusammen.

Zur Validierung der globalen Aussagekraft des Gütewerts wird deshalb eine Häufigkeitsverteilung der Stichprobenelemente erstellt (Abbildung 4.10).

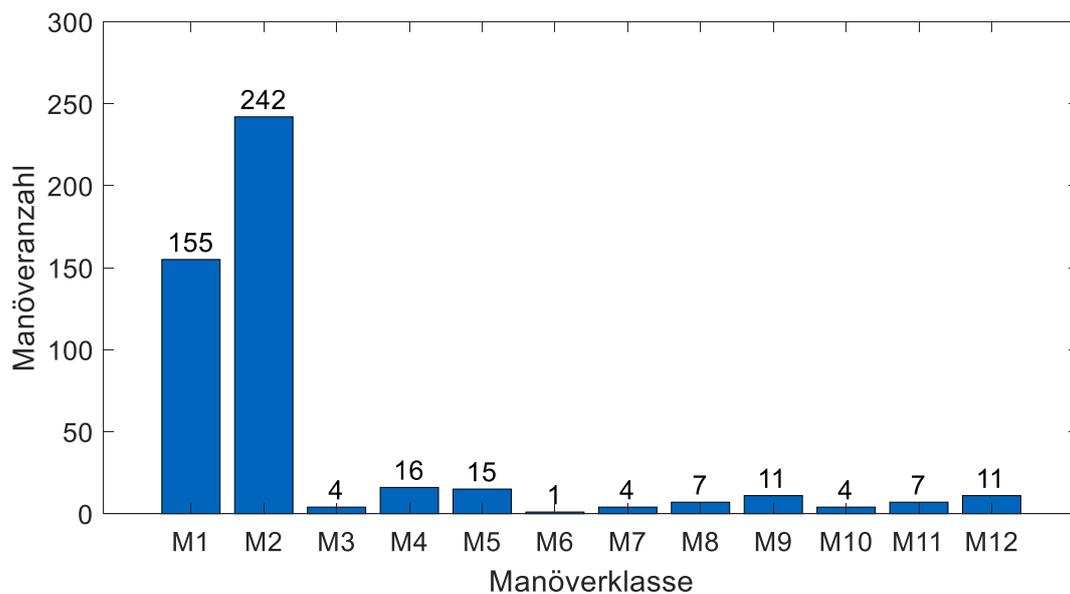


Abbildung 4.10: Verteilung der Stichprobe innerhalb der 12 Hauptmanöverklassen. Die Beschreibungen der Klassen M1 bis M12 sind dem Anhang zu entnehmen.

Abbildung 4.10 zeigt die resultierende Verteilung unter den Hauptmanöverklassen (eine Liste der Manöverklassen ist dem Anhang zu entnehmen). Erwartungsgemäß kommen die Manöver M1, Folgefahrten, und M2, Vorbeifahrten, mehrmals in der Stichprobe vor. Die Anzahl der Fälle innerhalb der anderen 10 Manöverklassen ist im Gegensatz dazu vergleichsweise niedrig. Beispielsweise tritt ein vollständiges Überholmanöver M6 nur einmal innerhalb der 477 gewählten Situationen auf.

Grundsätzlich folgt aus einer derart einseitigen Verteilung einer Stichprobe eine Verringerung der Repräsentativität derselben. Nach HULLINGER [64, S. 4-5] ist eine erhobene Stichprobe in einer solchen Form dennoch repräsentativ für die Gesamtpopulation, wenn die ungleiche Häufigkeitsverteilung als eine kennzeichnende Eigenschaft des Datensatzes angesehen werden kann.

In Zusammenhang mit der Untersuchung der Fahrmanöver trifft diese Regel nach HULLINGER zu, weil eine Manöververteilung analog jener der Stichprobe (Abbildung 4.10) für einen Datensatz aus Aufzeichnungen des Autobahnverkehrs erwartet wird. Insbesondere der dominierende Anteil an Vorbei- und Folgefahrten (M1 und M2) gegenüber den restlichen Manöverklassen ist ein deutliches Erkennungsmerkmal einer Sammlung aus solchen Daten. Unter dieser Argumentation kann die Verteilung aus Abbildung 4.10 als charakteristisches Kennzeichen des highD Datensatzes angesehen werden, weshalb die Erkenntnisse aus der Stichprobenanalyse ihre Gültigkeit beibehalten.

4.3 Gesamtauswertung highD-Datensatz

Der folgende Abschnitt befasst sich abschließend für den Ergebnisteil der vorliegenden Arbeit mit der Auswertung der gesamten highD Datenbank. Wie in Kapitel 2.3 beschrieben, setzt sich der highD Datensatz aus 60 Einzelaufnahmen mehrerer Autobahnabschnitte zusammen, wobei 2 der Dateien (25_tracks.csv und 26_tracks.csv) in den folgenden Auswertungen außer Acht gelassen werden. Dies begründet sich darin, dass diese beiden Aufnahmen primär dreispurige Stausituationen abbilden, wobei dieser Verkehrszustand in Verbindung mit dem zu extrahierenden Testkatalogs nur bedingt relevant ist. Zudem sind die beiden genannten Dateien im Vergleich zu den restlichen 58 Aufnahmen der Datenbank deutlich größer, was durch eine wesentlich höhere Anzahl an Fahrzeugen innerhalb der beiden Aufnahmen induziert wird. Dieser Unterschied führt zu einer unverhältnismäßig längeren Berechnungsdauer der beiden genannten Dateien, weshalb sie im Zuge des hier vorgestellten Gesamtdurchlaufs der highD Datenbank ignoriert werden.

In den Studien der vorangegangenen Unterkapitel konnte gezeigt werden, dass die Wahl einiger zentraler Parameter des Tools eine wesentliche Rolle bei der Identifikation und Klassifikation von Challenger-Situationen spielt. Die Einstellung dieser Werte wirkt sich primär auf die Anzahl der extrahierten, relevanten Szenarien und die finale Verteilung derselben innerhalb des Testkatalogs aus. Eine optimale Wahl der Parameter hat sich im Zuge der vorangegangenen Studien ergeben, jedoch konnte aus zeitlichen Gründen im Rahmen dieser Arbeit nur ein einzelner Parametersatz auf der gesamten highD Datenbank getestet werden.

Die Zusammenstellung dieser Parameterauswahl ist in folgender Tabelle 4.6 aufgelistet. Es ist anzumerken, dass die Werte für t_{pred} und $l_{\text{coll, fact}}$ nicht der identifizierten optimalen Einstellung aus Tabelle 4.5 entsprechen. Vielmehr werden bewusst die größten Werte aus der Parameterstudie 4.1.1 zur Untersuchung des gesamten highD-Datensatzes herangezogen. Diese suboptimale Wahl wird damit argumentiert, dass exemplarisch der parameterbedingt größtmögliche Testkatalog aus der vollständigen Datenbank generiert werden soll.

Tabelle 4.6: Parameterkonfiguration der Gesamtauswertung

Modul	Parameter	Wert	Einheit
Clustering	k_{SD}	2	[-]
Challenger-Prädiktion	t_{pred}	[1,2,3,4,5]	[s]
	b_{coll}	0.75	[m]
	$l_{\text{coll, fact}}$	0.5	[-]
Klassifikation	$v_{y, \text{lim, LC}}$	0.5	[m/s]

Wie angegeben bildet eine Auswahl von 58 Rohdatensätzen der highD Datenbank die Ausgangslage dieser Gesamtauswertung. Im Detail umfassen diese Rohdaten 104955 einzelne Fahrzeuge. Auf Basis dieser Ausgangssituation wird die vollständige Prozesskette entsprechend Abbildung 3.1 durchgearbeitet.

Der erste wesentliche Extraktionsschritt der Situationen für den Testkatalog besteht in der Bildung der Fahrzeuggruppen mittels ROI-basierter Clusteranalyse. Entsprechend der Methodik hinter diesem Prozessschritt werden initial 104955 Vehikel als separate Clusterelemente (Ego-Fahrzeuge) angesehen. Im Zuge der Analyse wird jedem dieser Elemente eine Menge an Umgebungsfahrzeugen zugewiesen, sofern sich diese zeitweise innerhalb eines definierten Bereichs um das Ego aufhalten. Einigen dieser Ego-Vehikel nähern sich jedoch zu keinem Zeitpunkt der Aufnahme etwaige Umgebungsfahrzeuge, weshalb diese Situationen im Sinne von Freifahrten aus der Menge an potentiell relevanten Clustern aussortiert werden. Diese Fälle treten jedoch äußerst selten innerhalb der gewählten highD Daten auf, weshalb der Clusteringschritt lediglich zu einer Reduktion auf 104434 mögliche Situationen führt. Die verbliebenen Cluster beinhalten bis zu 23 beteiligte Umgebungsfahrzeuge, wobei dahingehend eine Verteilung entsprechend der Grafik in Abbildung 4.11 vorliegt. Die extrahierten Verkehrssituationen setzen sich im Mittel aus einem Ego-Vehikel und 5.87 Fahrzeugen (gesamt 6.87 Fahrzeuge) zusammen.

Dieser Satz an geclusterten Verkehrssituationen wird im nächsten Verarbeitungsschritt auf die Existenz von Challenger-Vehikel untersucht. Im Zuge der vorliegenden Gesamtauswertung wird das zuvor erwähnte Erstkontaktprinzip zur Identifikation der Challenger-Fahrzeuge eingesetzt. Diese Untersuchung teilt die Cluster in zwei Teilmengen auf, den irrelevanten Situationen (38848) und jenen mit vorkommenden Challenger-Vehikeln (65586). Über diesen Prozessschritt können bei eingestelltem Parametersatz im Zuge der Challenger-Prädiktion somit weitere 37% der vernachlässigbaren Situationen aussortiert werden. Hier ist anzumerken, dass dieser prozentuale Anteil bei Einsatz der optimalen Parameterkonfiguration wie sie in den Parameterstudien identifiziert wurde, bedeutend höher ausfällt. Angesichts des Ergebnisses der Variation 4 aus Tabelle 4.3 (optimaler Parameter) kann dahingehend mit einer Reduktion des relevanten Datenumfangs durch die Challenger-Prädiktion auf einen einstelligen Prozentwert bezogen auf den gesamten initialen Datenumfang gerechnet werden.

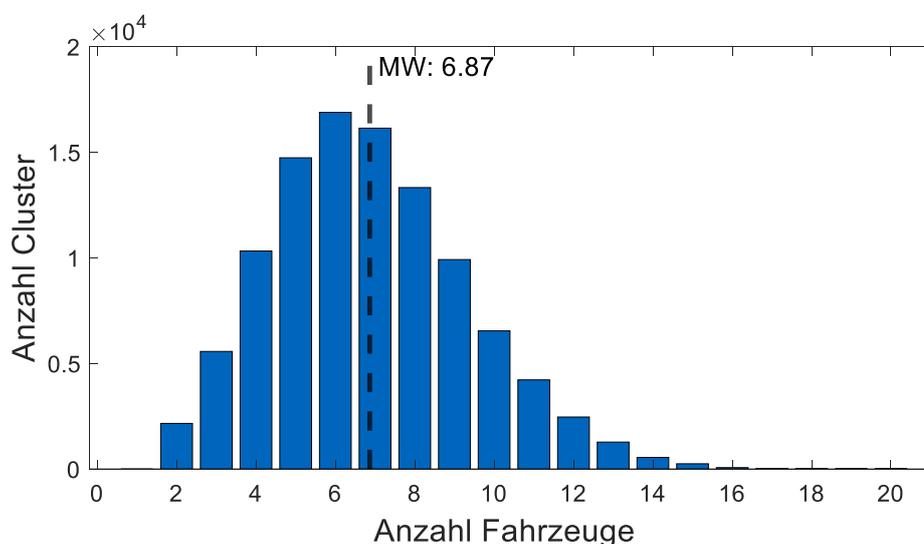


Abbildung 4.11: Häufigkeitsverteilung der Cluster bezogen auf die Anzahl an beteiligten Fahrzeugen. Im Mittel sind in einem Cluster 6.87 Fahrzeuge involviert.

Alle als relevant identifizierten Verkehrssituationen bilden im weiteren Verlauf der Prozesskette den notwendigen Input für das nächste Modul, den Klassifikator (Abbildung 3.1). Von zentraler Bedeutung in diesem Schritt ist nicht nur die Güte des Klassifikationsvorgangs bezogen auf bekannte Verkehrsmanöver, die bereits im Zuge einer Stichprobenanalyse validiert wurde (Kapitel 4.2), sondern auch die Betrachtung des globalen Abdeckungsgrades des Moduls. Darunter wird das Verhältnis aus bekannten zu unbekanntem Manövern verstanden, die in den realen Verkehrsaufzeichnungen potenziell auftreten können.

Jene Verkehrssituationen die a priori nicht in der Liste identifizierter Manöver vorkommen, werden von vornherein an einen definierten Ablageort verschoben und daher vom Klassifikator nicht weiterbearbeitet. Diese undefinierten Fälle müssen in einem iterativen Erweiterungsprozess der erwähnten Manöverliste hinzugefügt werden, wodurch die Anzahl an unbekanntem Situationen zunehmend reduziert wird. Bezogen auf die Ergebnisse der vorliegenden Gesamtauswertung ergibt sich ein Anteil von 24.7% an nicht definierten und somit unklassifizierbaren Verkehrssituationen. Es ist in diesem Kontext zu beachten, dass dieser Prozentsatz tendenziell indirekt proportional von der Anzahl an extrahierten Challenger-Fahrzeugen abhängt. Folglich steigt der verhältnismäßige Anteil an nicht zugeordneten Situationen, wenn sich die Anzahl an Challenger-Fällen aufgrund einer optimierten Parameterkonstellation verringert.

Zu sehen ist dieser Sachverhalt in den Resultaten der Parameterstudie zur Prädiktionlänge in Abschnitt 4.1.1. Wird darin die prozentuale Entwicklung der Anzahl an Challenger-Fahrzeugen mit jener der Anzahl an inkorrekten Klassifikationen verglichen, so ergibt sich die in Abbildung 4.12 dargestellte Beziehung. Es zeigt sich, dass der prozentuale Anteil an Challenger-Vehikeln bezogen auf die gesamte Menge an Fahrzeugen (11370) bei Verlängerung des Kollisionsbereichs über die Erhöhung des Faktors $l_{\text{coll, fact}}$ (Abszisse) von rund 3% auf ca. 46% anwächst. Im Gegensatz dazu fällt der prozentuale Anteil an nicht klassifizierbaren Manövern aus der Menge an Challenger-Vehikeln tendenziell ab, was eben beschriebenen Sachverhalt widerspiegelt.

Aus diesem Grund verdeutlicht sich die Notwendigkeit die Manöverliste kontinuierlich zu erweitern, um diese Schwäche des Klassifikationssystems zu verringern und dadurch den Anteil an undefinierten Fälle (Abbildung 4.12 ,orange) auf einen möglichst geringen Prozentwert zu senken.

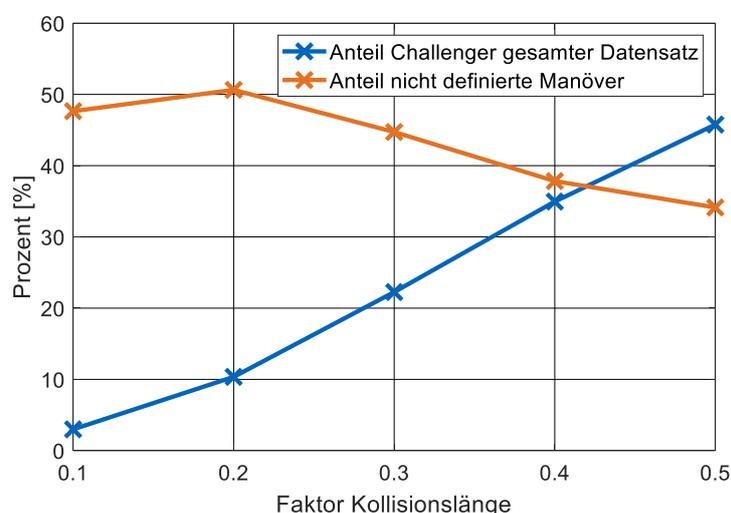


Abbildung 4.12: Prozentuale Entwicklung der Challenger-Anzahl (bezogen auf Gesamtanzahl Fahrzeuge) und der Anzahl nicht definierter Manöver (bezogen auf die Challenger-Anzahl)

Der letzte Verarbeitungsschritt der Prozesskette fusioniert die vorangegangenen Ergebnisse und erstellt darüber den gewünschten Szenarienkatalog. Analog des zuvor erwähnten Sachverhalts der Aussortierung von nicht klassifizierbaren Verkehrsmanövern im Zuge der Manöverbestimmung (Abbildung 4.12), scheidet in diesem letzten Prozessschritt ein weiterer Anteil der verbliebenen Situationen aus. In diesem Zusammenhang ist die Auslese jedoch nicht auf einen mangelnden Abdeckungsgrad der Manöverdefinitionen zurückzuführen, sondern stellt jenen Anteil der Challenger-Fahrzeuge dar, die im Zuge des Fusionschritts nach Abbildung 3.12 explizit der Szenarioklasse ND (Tabelle 2.8, Nicht definiert) zugeordnet werden.

Auf Grundlage der 58 Rohdatensätze ergibt sich über den Fusionsprozess ein Katalog aus 49562 Verkehrssituationen innerhalb der 9 definierten Szenarioklassen A bis I und der Klasse ND. Das finale Ergebnis am Ende des gesamten Prozesses kann anhand der in Abbildung 4.13 präsentierten Häufigkeitsverteilung nachvollzogen werden. Die nicht definierten Szenarien (Klasse ND) tragen mit einem Anteil von 0.42% bezogen auf den Gesamtkatalog nur gering zu diesem Endergebnis bei.

Es zeichnet sich ähnlich der Ergebnisse der vorangegangenen Parameterstudien eine stark ausgeprägte Ungleichmäßigkeit der Fallzahlen innerhalb der 9 Klassen ab. Besonders auffällig ist der Anteil an Fällen in den Klassen A und I, der mit rund 63% nahezu 2/3 der Gesamtanzahl relevanter Szenarien ausmacht. Dieser Wert ist jedoch nicht überraschend, wenn die Manöververteilung der Stichprobe aus Abschnitt 4.2 (Abbildung 4.10) zu Vergleichszwecken herangezogen wird. Bereits in diesem Zusammenhang hat sich ein vermehrtes Auftreten von Folgefahrtmanövern M1 herausgestellt. Die Szenarien der Klassen A und I (Auffahrunfälle) sind zu einem großen Teil mit genanntem Manöver M1 in Verbindung zu bringen, was sich folglich in den hohen Testfallzahlen widerspiegelt.

Im Vergleich dazu ist der Umfang der Klassen D und F verhältnismäßig niedrig. Bezogen auf die gesamte Testkataloggröße fallen lediglich ca. 1.09% auf diese beiden Szenarioklassen. Dieser geringe Wert ist mitunter dadurch zu begründen, dass es sich bei genannten Klassen um die komplexesten Szenariovarianten des Katalogs handelt. Dieser Umstand kann durch die Betrachtung der notwendigen Trajektorien dieser Klassen (Abbildung 2.15) nachvollzogen werden.

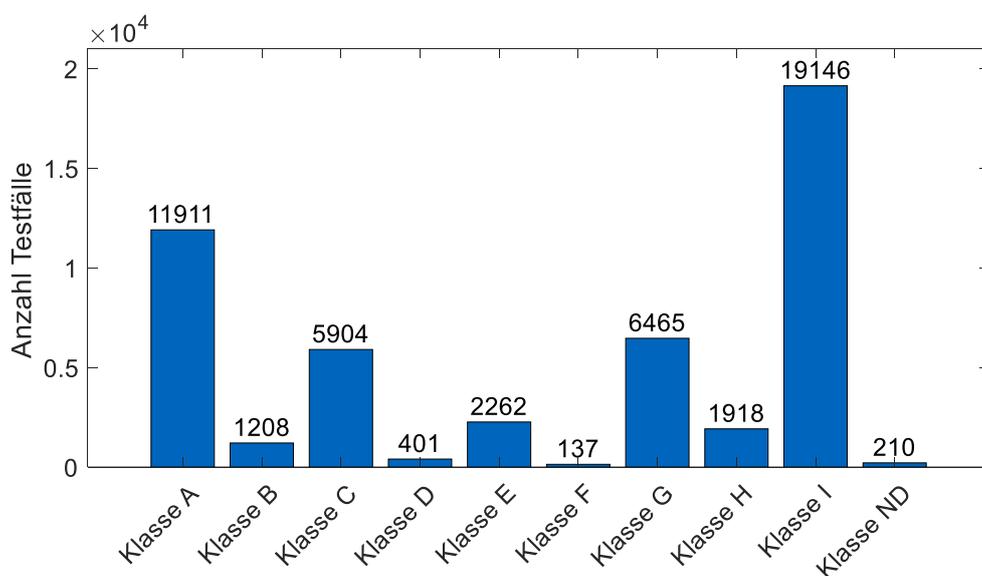


Abbildung 4.13: Finale Verteilung innerhalb des extrahierten Szenarienkatalogs auf Basis der highD-Datenbank. Von 49562 Testfällen in diesem Katalog entfällt ein vernachlässigbar kleiner Anteil von ca. 0.42% auf die Klasse undefinierter Szenarien (ND).

Beispielsweise kreuzt die Trajektorie D jene der Szenarien E, F und G und besitzt zudem aufgrund derselben Startposition eine gewisse Ähnlichkeit mit den deutlich einfacher zu erreichenden Fällen B und C. Somit ist unter Verwendung des Erstkontaktprinzips die Wahrscheinlichkeit einer der trivialeren Klassen (B, C, E) zugeordnet zu werden deutlich höher, als tatsächlich eine vollständige Trajektorie D auszubilden.

Überblick zur Gesamtauswertung

Der abschließende Überblick über die Resultate die im Zuge der Verarbeitung der highD-Datenbank gewonnen werden konnten, komplementiert das Kapitel 4 zu den Ergebnissen der vorliegenden Arbeit. Mit Hilfe von Abbildung 4.14 und der Tabelle 4.7 soll die Interpretation der angeführten Ergebnisse und die kontextuelle Zuordnung der einzelnen Resultate in die gesamte Prozesskette des Tools vereinfacht werden.

Über die Elimination der Freifahrten im Zuge des ROI-basierten Clusteringschritts (-0.49 %) und dem Aussortieren irrelevanter Szenarien im Challenger-Prädiktionsmodul (-37 %) wird eine angestrebte Reduktion des Datenumfangs bezogen auf die Ausgangsmenge erreicht.

Nicht definierte Fahrmanöver innerhalb der verbleibenden Menge von 65586 Challenger-Situationen führen in weiterer Folge im Prozessabschnitt der Manöverklassifikation zu einem unerwünschten Verlust von ca. 24.7 % der relevanten Testfälle.

Aus dem Rest werden weitere 0.42 % Testfälle aussortiert, die keiner der 9 Szenarioklassen aus Abbildung 2.15 zugeordnet werden können.

Somit setzt sich der endgültige Szenarienkatalog aus 49352 Verkehrssituationen verteilt auf 9 Klassen zusammen. Das entspricht in Verbindung mit der gewählten Parameterkonfiguration einem Anteil von 47 % der ursprünglichen Anzahl an Fahrzeugen, die potenziell die Basis für einen relevanten Testfall des szenariobasierten Testkatalogs repräsentieren.

Tabelle 4.7: Auswirkung der einzelnen Prozessschritte des Tools auf den Umfang der relevanten Testfälle

Modul	Ausgangsmenge	Endmenge	Reduktion
Preprocessing	104955	104955	0%
Clustering	104955	104434	-0.49%
Challenger-Prädiktion	104434	65589	-37%
Manöverklassifikation	65586	49562	-24.7%
Fusion	49562	49352	-0.42%

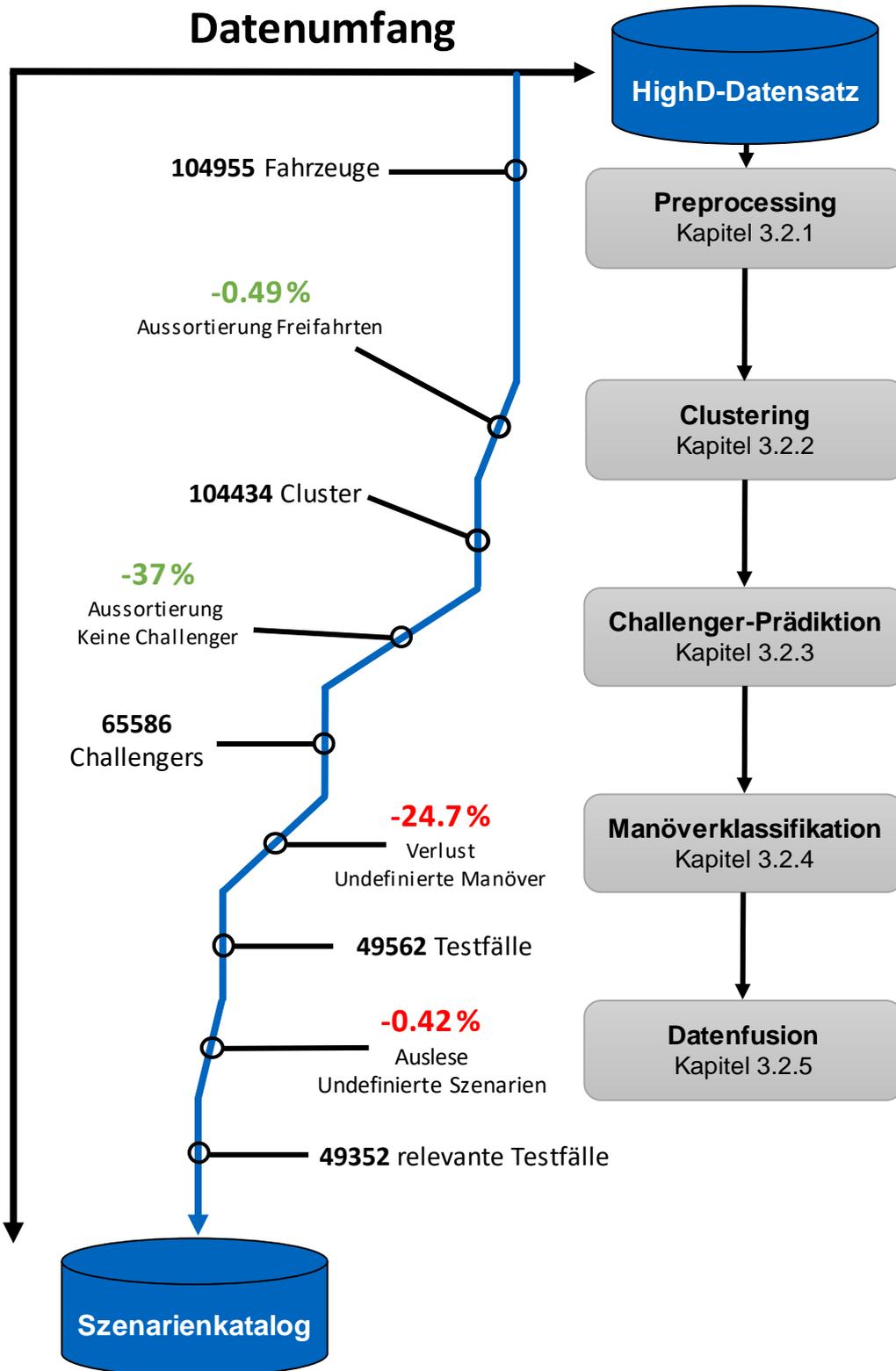


Abbildung 4.14: Überblick über die Gesamtauswertung des vollständigen highD-Datensatzes entsprechend der Reihenfolge der Prozesskette

5 Diskussion

Das entwickelte Werkzeug zur Extraktion und Klassifikation von Verkehrsszenarien aus realen Rohdaten stellt die Schnittstelle zwischen der Datenaufnahme und der Absicherung autonomer Fahrfunktionen dar. Dies ist ein essenzieller Aspekt eines Absicherungsprozesses, weil sich nur durch eine sorgfältige und sinnvolle Aufbereitung von Testfällen ein schneller und einfacher Einsatz derselben im Zuge simulierter oder realer Fahrzeugüberprüfungen gewährleisten lässt.

Insbesondere die Absicherung höchstautomatisierter Fahrfunktionen bis hin zum vollautonomen Fahrzeug erfordern klar strukturierte Testkataloge, weil die Verwendung eines ausgereiften Katalogs eine ressourcenschonende und dennoch aussagekräftige Untersuchung genannter Systeme begünstigt. Ein derartiger Testkatalog charakterisiert sich vor allem durch einen hohen Abdeckungsgrad bezogen auf die Gesamtheit der potenziell eintretenden Verkehrsszenarien.

Die Ergebnisse der Arbeit verdeutlichen, dass diese Anforderungen an einen szenariobasierten Testkatalog durch das entwickelte Werkzeug grundsätzlich erfüllt werden. Entsprechend der gewählten Einteilung der Verkehrssituationen in 9 Szenarioklassen zeigt sich, dass vom Tool jeder dieser Klassen eine adäquate Menge an Testfällen zugeordnet wird. Somit deckt der dadurch erzeugte Testkatalog das gesamte Spektrum an definierten Verkehrsszenarien ab, was als besonders positiv bezüglich der Funktionalität der entwickelten Werkzeugkette dieser Arbeit betrachtet werden kann.

Die Ergebnisse haben jedoch auch gezeigt, dass insbesondere beim Klassifikationsmodul aus Kapitel 3.2.2 ein deutliches Verbesserungspotential gegeben ist. Das begründet sich im aktuell hohen Anteil an nicht klassifizierbaren Fällen bezogen auf die Menge an absicherungsrelevanten Szenarien. Diese Schwachstelle des Tools ist hauptsächlich auf den gewählten Ansatz des Klassifikators zurückzuführen.

Die angewandte regelbasierte Strategie erfordert einen hohen Aufwand zur breiten Erfassung potenziell auftretenden Verkehrssituationen im Autobahnverkehr. Aus zeitlichen Gründen und der primär wissensbasierten Herangehensweise an die Manövererfassung, ist der Abdeckungsgrad der vom Klassifikator identifizierbaren Situationen aktuell nur bedingt zufriedenstellend. Es zeigt sich zwar eine Hohe Güte bezüglich der korrekten Manöveridentifikation der a priori bekannten Situationen, jedoch Bedarf es in Verbindung mit der Weiterentwicklung des vorliegenden Werkzeugs einer zunehmenden Schließung der Lücken nicht identifizierbarer Verkehrsmanöver. Dahingehend ist vom rein wissensbasierten Ansatz abzusehen und stattdessen ein iterativer Prozess anzustreben, in dem unbekannt deklarierte Situationen analysiert, deren Zustandsverläufe identifiziert und die Erweiterung der Manöverliste um diese Fälle realisiert wird.

Die Resultate der Gesamtauswertung des vollständigen highD-Datensatzes bestätigen die grundsätzliche Funktion der Prozesskette, wobei sich die dokumentierten Ergebnisse auf den angeführten suboptimalen Parametersatz beziehen, weshalb ein weiterer Durchlauf mit bestmöglich angepassten Parametern sinnvoll erscheint. Unter Vorgabe der optimalen Parameterkonfiguration, die über die angeführten Studien ermittelt wurde, kann ein noch

aussagekräftigerer Szenarienkatalog abgeleitet werden. Dies begründet sich in den deutlich strengeren Relevanzkriterien zur Aufnahme einer Verkehrssituation in den Testkatalog. In Folge dessen wird der Umfang eines Szenarienkatalogs auf Basis des optimalen Parametersatzes wünschenswerter Weise reduziert. Im Umkehrschluss können die wenigen verbliebenen Testfälle jedoch als besonders herausfordernde Repräsentanten der einzelnen Szenarioklassen angesehen werden.

6 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wird ein Werkzeug zur automatisierten Extraktion und Klassifikation von Verkehrsszenarien aus einer Datenbank realer Verkehrsaufzeichnungen entwickelt. Das übergeordnete Ziel ist die Erstellung eines Katalogs an Verkehrsszenarien, die eine besondere Relevanz bezüglich der Absicherung autonomer Fahrzeuge aufweisen. Die gesamte Verarbeitungskette wird als modularer Baukasten entworfen, um die Hauptfunktionen des Tools separat entwickeln und unabhängig voneinander einsetzen zu können.

Zu Beginn von Kapitel 2 werden gängige Methoden und Ansätze zur Absicherung autonomer Fahrfunktionen präsentiert, wobei insbesondere die damit verbundene Notwendigkeit zur Erfassung relevanter Verkehrssituationen hervorgehoben wird. Dahingehend werden unterschiedliche Datenerhebungsformen von realen Straßenverkehrssituationen angeführt, wobei der Fokus auf der Methode hinter dem in der vorliegenden Arbeit verwendeten highD-Datensatz liegt. Entsprechend der Aufgabe des ersten zentralen Moduls der entwickelten Werkzeugkette werden anschließend gängige Clusteranalysemethoden eingeführt, die primär zur Identifikation verborgener Muster und Gruppierungen innerhalb großer Datenmengen herangezogen werden können. Zum besseren Verständnis dieser Verfahren wird anhand eines repräsentativen Beispiels die Funktionsweise der Methoden erörtert. Im darauffolgenden Abschnitt werden Prädiktionsverfahren zur Vorhersage von Fahrzeugzuständen vorgestellt, wobei insbesondere auf Methoden zur Erkennung von zukünftigen kollisionsgefährlichen Situationen im Autobahnverkehr eingegangen wird. Eine weitere zentrale Aufgabe des realisierten Tools betrifft die Klassifikation von Verkehrsszenarien. Dahingehend wird der momentane Stand der Technik angeführt, wobei vor allem regelbasierte Verfahren und Methoden aus dem Feld des maschinellen Lernens als die beiden primär vorkommenden Ansätze evaluiert werden. Das Kapitel 2 wird mit einer kritischen Hinterfragung der vorgestellten Anwendungsverfahren zur Erzeugung und Bereitstellung relevanter Testfälle für die Absicherung autonomer Fahrzeuge abgeschlossen, worauf aufbauend der zentrale Forschungsschwerpunkt für die vorliegende Arbeit abgeleitet wird.

Im Anschluss an den Stand der Technik wird in Kapitel 3 detailliert auf die Methode hinter dem realisierten Werkzeug eingegangen. Dazu wird zuerst ein Überblick über die Modulkette gegeben und die übergeordneten Aufgaben der einzelnen Subfunktionen dargestellt. Anschließend werden die einzelnen funktionalen Elemente tiefergehend erörtert. Zu Beginn dieser Beschreibungen wird auf die Aufbereitung der highD-Daten eingegangen, die als wesentliches Ziel die Vereinheitlichung der gesamten Datenbank zur einfacheren Weiterverarbeitung hat. Auf Grundlage dieser harmonisierten Daten wird anschließend die Vorgehensweise zur Clusterfindung innerhalb der Verkehrsaufnahmen präsentiert. Hierzu wird auf das Prinzip der Region-of-Interest verwiesen, die eine zentrale Rolle bei der Gruppierung der Fahrzeuge zu kontextuell zusammengehörigen Clustern spielt. Aufbauend auf den daraus resultierenden Verkehrssituationen wird im nächsten Abschnitt der Methode das Prinzip der Challenger-Identifikation auf Basis eines beschleunigungsbasierten Prädiktionsmodells vorgestellt. Im Anschluss daran wird auf den

regelbasierten Klassifikator eingegangen, der über ein Zustandsidentifikationsmodul Manöver hinter den relevanten Verkehrssituationen erkennen kann. Dahingehend wird der Einsatz eines regelbasierten Entscheidungsbaums zur kontrollierten Zuordnung von Zustandslisten zu definierten Manövern angeführt. Abgeschlossen wird Kapitel 3 mit den Erörterungen zum letzten Verarbeitungsprozessschritt der Werkzeugkette, dem Fusionsmodul. Dieses erstellt auf Grundlage der Ergebnisse der vorangegangenen Funktionsabschnitte den finalen szenariobasierten Testkatalog.

In Kapitel 4 werden die Ergebnisse der vorliegenden Arbeit vorgestellt. Ein wesentlicher Anteil dieses Abschnitts befasst sich tiefergehend mit der Analyse der Einflüsse zentraler Parameter des Tools. Ziel dieser Untersuchungen ist es, die bestmögliche Konstellation an Parameterwerten zu identifizieren, die die Erfassung eines qualitativ hochwertigen Szenariokatalogs begünstigen. Dazu werden vor allem Parameter des Prädiktionsmodells zur Identifikation der kollisionskritischen Verkehrsteilnehmer variiert und die damit einhergehenden Veränderungen im resultierenden Testkatalog untersucht. Im Anschluss daran wird eine Stichprobenanalyse vorgestellt, auf deren Basis eine Aussage über die Güte des realisierten Klassifikators für Verkehrsszenarien getätigt werden kann. Hierzu wird der notwendige Stichprobenumfang berechnet und über einen subjektiven Vergleich die Verteilung zwischen korrekt und inkorrekt erfassten Manöverbeschreibungen ermittelt. Der Ergebnisabschnitt wird mit der Präsentation der Resultate eines vollständigen Durchlaufs von 58 der 60 zur Verfügung stehenden Datensätze komplementiert. Darin kann die Erfüllung der zentralen Aufgabenstellung der vorliegenden Arbeit bestätigt werden, indem im Besonderen auf die deutliche Reduktion des Datenumfangs und die sinnvolle Verteilung der Szenarien innerhalb des gewonnenen Testkatalogs eingegangen wird.

Im letzten Kapitel dieser Arbeit wird das entwickelte Verfahren kritisch begutachtet und die exakte Einordnung des Tools in einen vollständigen Absicherungsprozesses für autonome Fahrzeuge angeführt. Zudem werden in dieser Diskussion Schwachstellen des realisierten Werkzeugs betrachtet und inwieweit diese in zukünftigen Arbeiten verbessert und kompensiert werden können, um die Genauigkeit des Tools und die Qualität des resultierenden Szenariokatalogs zunehmend zu erhöhen.

Ausblick

Entsprechend der Ausführungen in Kapitel 4 besteht eine zentrale Aufgabe in Verbindung mit der Weiterentwicklung des vorliegenden Werkzeugs in der sukzessiven Erfassung weiterer Verkehrsmanöver, um die Güte des Klassifikators schrittweise zu erhöhen. Hierzu kann ein iterativer Prozess angesetzt werden, der aus den nicht klassifizierbaren Clustern Gruppen ähnlicher Verkehrssituationen extrahiert, deren Zustandsabfolgen identifiziert und so nach und nach den Manöverkatalog erweitert. Diesem Prinzip zufolge wird sich der Anteil an unbekanntem Manövern stetig verringern, wodurch zusätzliche Testfälle in die Szenarioklassen des Testkatalogs eingebunden werden können.

Ein weiterer Punkt bezüglich der Erweiterungen des Werkzeugs behandelt die Integration eines komplexeren Vorhersagemodells im Zuge der Challenger-Prädiktion. Ziel hierbei soll der Einbezug realer Fahrverhaltensmuster und Einflüssen durch vorherrschende Umgebungsbedingungen sein, um realitätsnähere Vorhersagen des Ego-Fahrzeugs zu ermöglichen. Auf diese Weise können kollisionskritische Fahrzeuge wesentlicher genauer bestimmt werden.

Zusätzlich zur Verbesserung des Vorhersagemodells kann sich eine Folgearbeit mit der Weiterentwicklung des restlichen Challenger-Identifikationsmoduls beschäftigen. Vor allem wird in der aktuellen Version angenommen, dass Challenger-Vehikel unabhängig von ihrer Position relativ zum Ego-Fahrzeug immer denselben Einfluss auf das Ego ausüben. Wird jedoch eine allgemeine Verkehrssituation eines sich nähernden Umgebungsfahrzeug hinter dem Ego betrachtet, so liegt deutlich mehr Verantwortung geeignete Maßnahmen zur Unfallvermeidung umzusetzen beim Folgefahrzeug. Im Vergleich dazu muss bei einer umgekehrten Konstellation das Ego-Vehikel vergleichsweise stärker eingreifen, um eine Kollision abzuwenden. Folglich kann eine mögliche Erweiterung des Challenger-Moduls die Integration eines Maßes für die situativen Verantwortung einzelner Verkehrsteilnehmer in Zusammenhang mit der Aufnahme von Umgebungsfahrzeugen in die Klasse der Challenger-Vehikel sein.

Abbildungsverzeichnis

Abbildung 1.1:	Reduktion unbekannter und damit risikobehafteter Szenarien durch Integration von SOTIF-Aktivitäten [4, S. 16].	2
Abbildung 2.1:	Vereinfachte Darstellung des V-Modells nach [15].	8
Abbildung 2.2:	Schematischer Ablauf der im Zuge des PEGASUS-Projekts entwickelten Methode [17].	10
Abbildung 2.3:	Prinzip der Datenerfassung mittels Drohne [18].	11
Abbildung 2.4:	Visualisierung der Aufzeichnungen. Mittels Bildverarbeitungsalgorithmen werden die Verkehrsteilnehmer extrahiert [18, S. 2123].	12
Abbildung 2.5:	Beispiel für den Ablauf des <i>k-means</i> Clustering nach [27, S. 34].	15
Abbildung 2.6:	Ergebnisse k-means mit $k = 3$ (oben) und $k = 4$ (unten)	18
Abbildung 2.7:	Dendrogramme des HAC auf Basis der Fahrzeugmassen. Die subjektive Wahl des Cut-off Werts zu 150 liefert das intuitiv erwartete Ergebnis von 3 Clustern.	19
Abbildung 2.8:	Prozesskette Clusterverfahren in Anlehnung an [33, S. 41].	20
Abbildung 2.9:	Auszug der identifizierten Unfallklassen für T-förmige Kreuzungen nach [40, S. 148].	21
Abbildung 2.10:	Polygonüberschneidung zweier Objekte und Grad der Kollision nach [42, S. 50].	23
Abbildung 2.11:	Erhöhung der Kollisionsgefahr durch zunehmende Überlappung der Hüllvolumen nach [42, S. 51], beispielhaft ausgeführt als Kugelvolumen mit Radien R_1 und R_2 .	23
Abbildung 2.12:	Risk-Map einer Kreuzungssituation aus [47, S. 706].	24
Abbildung 2.13:	Beispiel eines verborgenen Markov Modells in Anlehnung an [51, S. 83]. Zwischen den Zuständen V und B kann mit den Wahrscheinlichkeiten a/b gewechselt werden.	26
Abbildung 2.14:	Zustandsabfolge des Überholmanövers nach [52, S. 42]. Das betrachtete Manöver ist anhand des definierten Zustandsverlaufs (Tabelle 2.7) eindeutig identifizierbar.	27
Abbildung 2.15:	Trajektorien der Challenger-Fahrzeuge zur Klassifikation der Verkehrsszenarien (A-I) basierend auf [16, S. 63]. Aus Gründen der Symmetrie ist nur die linke Seite bezogen auf das Ego-Fahrzeug dargestellt.	29

Abbildung 2.16:	Prinzip des maschinellen Lernens nach [54, S. 5]. Der Prozess des Ausgabenvergleichs und der Parameteranpassung wird iterativ bis zu einem gewünschten Grad der Modellgenauigkeit ausgeführt.	30
Abbildung 3.1:	Überblick über die gesamte Prozesskette des Tools.	33
Abbildung 3.2:	Darstellung der einzelnen Verarbeitungsschritte des Preprocessing	35
Abbildung 3.3:	Anpassung der Fahrstreifennummerierung	38
Abbildung 3.4:	Darstellung der Region-of-Interest (ROI) des aktuellen Ego-Fahrzeugs (rot) auf Basis der geschwindigkeitsabhängigen Größen $lROI, front$ und $lROI, back$. Umgebungsfahrzeuge (weiß) außerhalb des rot markierten Bereichs werden nicht in das Cluster aufgenommen.	40
Abbildung 3.5:	Dendrogramm der Minimalabstände zum Ego-Fahrzeug eines exemplarisch gewählten Frames. Die Clusterbildung beruht auf den Cut-off Werte, die über die Länge der ROI ($lROI, front$, $lROI, back$.) festgelegt sind.....	40
Abbildung 3.6:	Darstellung des Prädiktionsprinzips und der Kollisionsbereiche (1-5) inklusive der beschreibenden Parameter.	43
Abbildung 3.7:	Repräsentative Verläufe des Kollisionspotentials in den 5 Umgebungsbereichen des Ego-Vehikels. Challenger Erstkontaktprinzip: ID 38 bei Frame 1 – Challenger Maximalwertprinzip: ID 41 bei Frame 170	44
Abbildung 3.8:	Funktionskette der Erkennung logischer Manöver bestehend aus einer frame-basierten Zustandsidentifikation und der Anschlussfunktion zur Ableitung logischer Manöver. Das zweite Modul teilt die Zustandsliste bei Verkettung mehrerer definierter Manöver (strichlierte Linie → Manöver-Split).....	46
Abbildung 3.9:	Struktur der Zustandsdefinition. Der Grundzustand wird in jedem Frame um die Informationen Geschwindigkeitsbeziehung, relativer Fahrstreifen des SV und Spurwechsel des Ego-Fahrzeugs erweitert.....	47
Abbildung 3.10:	Prinzip der Identifikation der Geschwindigkeitsbeziehung zwischen den beiden Fahrzeugen. Solange der Integralwert unter der festgelegten Grenze bleibt, wird die Geschwindigkeit der beiden Vehikel als identisch betrachtet.	48
Abbildung 3.11:	Verkettung von zwei logischen Manövern. Die Zustandsliste wird aufgrund des zwischenzeitlichen Auftretens von Zustand S8 in zwei separate Teillisten aufgesplittet, deren Manöver über das Identifikationsmodul erkannt werden können.....	51
Abbildung 3.12:	Entscheidungslogik des Fusionsschritts in Anlehnung an [16, S. 63]. Die notwendigen Informationen stammen aus dem Challenger-Prädiktionsmodul (Kapitel 3.2.1) und dem Manöver-Klassifikationsschritt (Kapitel 3.2.2).	53
Abbildung 4.1:	Relative Häufigkeitsverteilung der Durchschnittsgeschwindigkeit inklusive der Mittelwerte für die ersten 10 Datensätze und die vollständige HighD-Datenbank (60 Datensätze).....	56
Abbildung 4.2:	Beispielhafte Verteilung der relevanten Testfälle unter den Szenarioklassen A-I und der Klasse ND (Nicht Definiert) bei Verwendung unterschiedlicher Einstellungen des Parameters $tpred$ in Sekunden.	59

Abbildung 4.3:	Verteilung der relevanten Testfälle unter den Szenarioklassen A-I und ND (nicht definiert) für zwei unterschiedliche Einstellungen des Parameters $l_{coll, fact}$	61
Abbildung 4.4:	Auswirkung der Länge des Kollisionsbereichs (rot-grüne Areale) auf die Zuordnung der Challenger-Situationen zu den Klassen D und F.....	63
Abbildung 4.5:	Vergleich der Szenarioverteilung für die beiden Challenger-Identifikationsprinzipien. Es zeigen sich deutliche Unterschiede in den Verteilungen der beiden Verfahren.....	64
Abbildung 4.6:	Verteilung der Challenger-Fälle von Szenarioklasse B. Die 47 Testfälle aus Klasse B nach dem Erstkontaktprinzip verteilen sich entsprechend der rechten Grafik innerhalb des Katalogs nach dem Maximalwertprinzip.....	65
Abbildung 4.7:	Regressionsanalyse der Spurwechselsituationen in den Datensätzen 1-10	68
Abbildung 4.8:	Confusion-Matrix der Güte des Erkennungsmoduls von Fahrstreifenwechsel (FSW).....	70
Abbildung 4.9:	Güte des Klassifikators bezogen auf den Stichprobenumfang von 477 Fahrzeugen. Konkret wird in 94.76 % der Untersuchungen die Verkehrssituation dem korrekten Manöver zugeordnet, während in 5.24 % der Fälle eine falsche Klassifikation erfolgt.....	73
Abbildung 4.10:	Verteilung der Stichprobe innerhalb der 12 Hauptmanöverklassen. Die Beschreibungen der Klassen M1 bis M12 sind dem Anhang zu entnehmen.....	74
Abbildung 4.11:	Häufigkeitsverteilung der Cluster bezogen auf die Anzahl an beteiligten Fahrzeugen. Im Mittel sind in einem Cluster 6.87 Fahrzeuge involviert.....	76
Abbildung 4.12:	Prozentuale Entwicklung der Challenger-Anzahl (bezogen auf Gesamtanzahl Fahrzeuge) und der Anzahl nicht definierter Manöver (bezogen auf die Challenger-Anzahl).....	77
Abbildung 4.13:	Finale Verteilung innerhalb des extrahierten Szenarienkatalogs auf Basis der highD-Datenbank. Von 49562 Testfällen in diesem Katalog entfällt ein vernachlässigbar kleiner Anteil von ca. 0.42 % auf die Klasse undefinierter Szenarien (ND).....	78
Abbildung 4.14:	Überblick über die Gesamtauswertung des vollständigen highD-Datensatzes entsprechend der Reihenfolge der Prozesskette.....	80
Abbildung A.1:	Struktur der Zustandsdefinition.....	xiv
Abbildung B.2:	Hauptmanöverklassen 1 - 12.....	xvii

Tabellenverzeichnis

Tabelle 2.1:	Grade der Automatisierung nach BASt [1]	6
Tabelle 2.2:	Ausschnitt aus dem highD Datensatz in Form der ersten 5 Frames des Fahrzeugs mit ID 1 [23].	13
Tabelle 2.3:	Klassen der Clusteranalysemethoden nach Wu [24, S. 3-5]	14
Tabelle 2.4:	Datenpunkte Clusterbeispiel.....	17
Tabelle 2.5:	Ergebnis k-means bei Wahl $k = 3$	18
Tabelle 2.6:	Vergleich der Ergebnisse beider Clusterverfahren.....	20
Tabelle 2.7:	Zustandsabfolge des Überholmanövers nach [52, S. 42].....	27
Tabelle 2.8:	Logische Szenarien nach [16, S. 64].....	28
Tabelle 3.1:	Grundform der potentiellen Zustände (states) zwischen Ego und Umgebungsfahrzeug	47
Tabelle 3.2:	Auszug aus der Liste an logischen Manövern.....	50
Tabelle 4.1:	Parameter des Tools und Argumentationsgrundlage für deren Festlegung. 55	
Tabelle 4.2:	Parametervariation t_{pred}	58
Tabelle 4.3:	Parametervariation von l_{coll} , $fact$ basierend auf dem Bußgeldkatalog für das Vergehen der Unterschreitung des gesetzlich vorgeschriebenen Sicherheitsabstand [62].	60
Tabelle 4.4:	Parametervariation v_y , lim , LC	69
Tabelle 4.5:	Optimale Parametereinstellung als Resultat der durchgeführten Parameterstudien	71
Tabelle 4.6:	Parameterkonfiguration der Gesamtauswertung.....	75
Tabelle 4.7:	Auswirkung der einzelnen Prozessschritte des Tools auf den Umfang der relevanten Testfälle	79
Tabelle A.1	Grundzustand der States.....	xiv
Tabelle A.2	Geschwindigkeitsbeziehung.....	xv
Tabelle A.3	SV-Fahrstreifen	xv
Tabelle A.4	Ego-Fahrstreifenwechsel.....	xv
Tabelle B.1	Liste Hauptmanöverklassen (beispielhaft Fahrzeuge Nr. 1 und 2).....	xvi
Tabelle B.2	Liste Subklassen - Einzelmanöver	xviii

Literaturverzeichnis

- [1] Bundesanstalt für Straßenwesen, „Rechtsfolgen zunehmender Fahrzeugautomatisierung“, *Forschung kompakt*, 11/12, 2012. [Online]. Verfügbar unter: https://www.bast.de/BAST_2017/DE/Publikationen/Foko/Downloads/2017-2010/2012-11.pdf?__blob=publicationFile&v=1
- [2] M. Hillenbrand, „Funktionale Sicherheit nach ISO 26262 in der Konzeptphase der Entwicklung von Elektrik/Elektronik Architekturen von Fahrzeugen“, Institut für Technik der Informationsverarbeitung, Karlsruher Institut für Technologie (KIT), D-76131 Karlsruhe. [Online]. Verfügbar unter: <https://publikationen.bibliothek.kit.edu/1000025616/2325028>
- [3] L. Schnieder und R. S. Hosse, *Leitfaden Safety of the Intended Functionality*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019.
- [4] T. Doms, B. Rauch, B. Schrammel, C. Schwald, E. Spahovic, C. Schwarzl, „Highly Automated Driving: The new challenges for Functional Safety and Cyber Security“, TÜV AUSTRIA and VIRTUAL VEHICLE, 2018. [Online]. Verfügbar unter: https://www.tuv.at/fileadmin/user_upload/docs/group/innovation/tuv-austria-white-paper-iv-highly-automated-driving_web.pdf. Zugriff am: 16. April 2020.
- [5] J. Schäuffele und T. Zurawka, Hg., *Automotive Software Engineering: Grundlagen, Prozesse, Methoden und Werkzeuge effizient einsetzen*, 6. Aufl. Wiesbaden: Springer Fachmedien Wiesbaden GmbH, 2016.
- [6] Bundesanstalt für Straßenwesen, „Marktdurchdringung von Fahrzeugsicherheitssystemen“, *Daten & Fakten kompakt 03/2017*, Nr. 03, 2017. [Online]. Verfügbar unter: https://www.bast.de/BAST_2017/DE/Publikationen/DaFa/Downloads/2017-03.pdf?__blob=publicationFile&v=3
- [7] H. Wagner und S. Kabel, *Mobilität 4.0 – neue Geschäftsmodelle für Produkt- und Dienstleistungsinnovationen*. Wiesbaden: Springer Fachmedien Wiesbaden, 2018.
- [8] K. Reif, *Automobilelektronik*. Wiesbaden: Springer Fachmedien Wiesbaden, 2014.
- [9] SAE - Society of Automotive Engineers, *J3016B: Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles - SAE International*. [Online]. Verfügbar unter: https://www.sae.org/standards/content/j3016_201806/. Zugriff am: 6. April 2020.
- [10] H. Winner, S. Hakuli, F. Lotz und C. Singer, *Handbuch Fahrerassistenzsysteme*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
- [11] M. Butenuth, R. Kallweit und P. Prescher, „Vehicle-in-the-Loop: Reale Fahrzeugtests mit virtuellen Szenen kombiniert“, *ATZ - Automobiltechnische Zeitschrift*, Jg. 119, Nr. 9, S. 58–63, 2017, doi: 10.1007/s35148-017-0084-y.

- [12] M. Maurer, J. C. Gerdes, B. Lenz und H. Winner, *Autonomes Fahren*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015.
- [13] Bundesministerium des Innern, für Bau und Heimat, *V-Modell-XT*. [Online]. Verfügbar unter: https://www.cio.bund.de/Web/DE/Architekturen-und-Standards/V-Modell-XT/vmodell_xt_node.html. Zugriff am: 20. April 2020.
- [14] D. Firesmith, *Using V Models for Testing*. [Online]. Verfügbar unter: https://insights.sei.cmu.edu/sei_blog/2013/11/using-v-models-for-testing.html. Zugriff am: 15. April 2020.
- [15] Deutsche Gesetzliche Unfallversicherung, *Das V-Modell™ (Entwicklungsstandard)*. [Online]. Verfügbar unter: http://regelwerke.vbg.de/vbg_dguvi/di215-450/di215-450_106_.html. Zugriff am: 20. April 2020.
- [16] Verbundprojekt, „PEGASUS - Schlussbericht für das Gesamtprojekt“, 2020. [Online]. Verfügbar unter: https://www.pegasusprojekt.de/files/tmpl/pdf/PEGASUS_Abschlussbericht_Gesamtprojekt.PDF
- [17] Deutsches Zentrum für Luft- und Raumfahrt e.V., *Pegasus-Projekt*. [Online]. Verfügbar unter: <https://www.pegasusprojekt.de/en/pegasus-method>. Zugriff am: 4. Januar 2020.
- [18] R. Krajewski, J. Bock, L. Kloecker und L. Eckstein, „The highD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems“, *IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*, S. 2118–2125, 2018.
- [19] *GIDAS (German In-Depth Accident Study)*. [Online]. Verfügbar unter: <https://www.gidas.org/willkommen/>. Zugriff am: 8. April 2020.
- [20] *Testfeld Niedersachsen – Mobilitätsforschung mit Realdaten von der A39: DLR Testfeld Niedersachsen Mobilitätsforschung mit Realdaten*. [Online]. Verfügbar unter: https://www.dlr.de/content/de/artikel/news/2019/02/20190503_testfeld-niedersachsen-mobilitaetsforschung-realdaten.html. Zugriff am: 8. April 2020.
- [21] M. Treiber und A. Kesting, *Verkehrsdynamik und -simulation*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [22] J. Bock *et al.*, „The inD Dataset: A Drone Dataset of Naturalistic Road User Trajectories at German Intersections“, 18. Nov. 2019. [Online]. Verfügbar unter: <http://arxiv.org/pdf/1911.07602v1>. Zugriff am: 3. April 2020.
- [23] R. Krajewski, J. Bock, L. Kloecker und L. Eckstein, *HighD Dataset Format*. [Online]. Verfügbar unter: <https://www.highd-dataset.com/format>. Zugriff am: 2. Januar 2020.
- [24] J. Wu, *Advances in K-means Clustering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.
- [25] P. M. T.M. Kodinariya, „Review on Determining of Cluster in K-means Clustering“, *International Journal of Advance Research in Computer Science and Management Studies*, Jg. 2013, S. 90–95. [Online]. Verfügbar unter: <http://www.ijarcsms.com/docs/paper/volume1/issue6/V1I6-0015.pdf>
- [26] R. Wyrzykowski, E. Deelman, J. Dongarra und K. Karczewski, *Parallel Processing and Applied Mathematics*. Cham: Springer International Publishing, 2020.

- [27] A. Kalmár, G. Öllös, R. Vida, „Analysis of an Event Forecasting Method for Wireless Sensor Networks“ in *Acta Universitatis Sapientiae Electrical and Mechanical Engineering*, S. 26–38.
- [28] B. Everitt, S. Landau, M. Leese, D. Stahl, *Cluster Analysis*. United Kingdom: WILEY SERIES IN PROBABILITY AND STATISTICS, 2011.
- [29] J. Bacher, A. Pöge und K. Wenzig, *Clusteranalyse: Anwendungsorientierte Einführung in Klassifikationsverfahren*, 3. Aufl. München: Oldenbourg, 2010. [Online]. Verfügbar unter: <http://www.oldenbourg-link.com/isbn/9783486710236>
- [30] A. S. A.M. Dai, „The Grouped Author-Topic Model for Unsupervised Entity Resolution“ in *Artificial Neural Networks and Machine Learning – ICANN 2011*, S. 241–249, doi: 10.1007/978-3-642-21735-7.
- [31] J. S. M. Ester, *Knowledge Discovery in Databases: Techniken und Anwendungen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000.
- [32] M. Hudec, „Einführung in die CLUSTERANALYSE“. Skriptum, Institut für Statistik und Decision Support Systems, Universität Wien, Wien, 2003.
- [33] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, „From Data Mining to Knowledge Discovery in Databases“, *AI Magazine*, Nr. 17, S. 37–54, 1996.
- [34] H. Watanabe, T. Malý, J. Wallner, T. Dirndorfer, M. Mai, G. Prokop, „Methodology of Scenario Clustering for Predictive Safety Functions“.
- [35] G. von Buxhoeveden, „Indikatorbasierte multimodale Risikobetrachtung im Verkehr“. Dissertation, Fakultät für Maschinenbau, Technische Universität Carolo-Wilhelmina zu Braunschweig, Braunschweig, 2019.
- [36] E. Necula, „Analyzing Traffic Patterns on Street Segments Based on GPS Data Using R“, *Transportation Research Procedia*, Jg. 10, S. 276–285, 2015, doi: 10.1016/j.trpro.2015.09.077.
- [37] J. Kianfar und P. Edara, „A data mining approach to creating fundamental traffic flow diagram“, *Procedia - Social and Behavioral Sciences*, Jg. 104, S. 430–439, 2013, doi: 10.1016/j.sbspro.2013.11.136.
- [38] J. Lenard, R. Danton, M. Avery, A. Weekes, D. Zuby, M. Kühn, „Typical pedestrian accident scenarios for the testing of autonomous emergency braking systems“, *22nd Enhanced Safety of Vehicles (ESV) Conference*, 2011.
- [39] U. Sander und N. Lubbe, „The potential of clustering methods to define intersection test scenarios: Assessing real-life performance of AEB“ (eng), *Accident; analysis and prevention*, Jg. 113, S. 1–11, 2018, doi: 10.1016/j.aap.2018.01.010.
- [40] P. Nitsche, P. Thomas, R. Stuetz und R. Welsh, „Pre-crash scenarios at road junctions: A clustering method for car crash data“ (eng), *Accident; analysis and prevention*, Jg. 107, S. 137–151, 2017, doi: 10.1016/j.aap.2017.07.011.
- [41] R. Gnerlich, „Entwicklung eines Konzepts zur digitalen Untersuchung von Bauzeitverzögerungen auf Grundlage einer BIM - basierten Bauablaufsimulation“. Dissertation, Institut für Bauwirtschaft, Universität Kassel, Kassel, 2019.

- [42] C. Woenckhaus, „Rechnergestütztes System zur automatisierten 3D-Layoutoptimierung“. Forschungsbericht, Institut für Werkzeugmaschinen und Betriebswissenschaften, Technische Universität München, München, 1994.
- [43] A. Lambert, D. Gruyer, G. S. Pierre und A. N. Ndjeng, „Collision Probability Assessment for Speed Control“ in *2008 11th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Beijing, China, 12.10.2008 - 15.10.2008, S. 1043–1048, doi: 10.1109/ITSC.2008.4732692.
- [44] T. Batz, K. Watson und J. Beyerer, „Recognition of dangerous situations within a cooperative group of vehicles“ in *IEEE Intelligent Vehicles Symposium*, Xi'an, China, 2009, S. 907–912, doi: 10.1109/IVS.2009.5164400.
- [45] C. Frese, „Planung kooperativer Fahrmanöver für kognitive Automobile“. Dissertation, Fakultät für Informatik, Karlsruher Institut für Technologie, Karlsruhe, 2011.
- [46] C. Blaschke, J. Schmitt und B. Färber, „Überholmanöver-Prädiktion über CAN-Bus-Daten“, *ATZ - Automobiltechnische Zeitschrift*, Jg. 110, Nr. 11, S. 1022–1028, 2008, doi: 10.1007/BF03222033.
- [47] F. Damerow und J. Eggert, „Predictive risk maps“ in *2014 IEEE 17th International Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China, 08.10.2014 - 11.10.2014, S. 703–710, doi: 10.1109/ITSC.2014.6957772.
- [48] M. Althoff, O. Stursberg und M. Buss, „Model-Based Probabilistic Collision Detection in Autonomous Driving“, *IEEE Trans. Intell. Transport. Syst.*, Jg. 10, Nr. 2, S. 299–310, 2009, doi: 10.1109/TITS.2009.2018966.
- [49] *Katalog*. [Online]. Verfügbar unter: <https://www.duden.de/rechtschreibung/Katalog>. Zugriff am: 14. April 2020.
- [50] N. Deo, A. Rangesh und M. M. Trivedi, „How would surround vehicles move? A Unified Framework for Maneuver Classification and Motion Prediction“, *IEEE Trans. Intell. Veh.*, Jg. 3, Nr. 2, S. 129–140, 2018, doi: 10.1109/TIV.2018.2804159.
- [51] T. Toledo und R. Katz, „State Dependence in Lane-Changing Models“, *Transportation Research Record*, Jg. 2124, Nr. 1, S. 81–88, 2009, doi: 10.3141/2124-08.
- [52] J. Schneider, „Modellierung und Erkennung von Fahrsituationen und Fahrmanövern für sicherheitsrelevante Fahrerassistenzsysteme“. Dissertation, Fakultät für Elektrotechnik und Informationstechnik, Technischen Universität Chemnitz, Chemnitz, 2009. [Online]. Verfügbar unter: <https://monarch.qucosa.de/api/qucosa%3A19384/attachment/ATT-0/>
- [53] A. Erdogan *et al.*, „Real- World Maneuver Extraction for Autonomous Vehicle Validation: A Comparative Study“ in *2019 IEEE Intelligent Vehicles Symposium (IV)*, Paris, France, 09.06.2019 - 12.06.2019, S. 267–272, doi: 10.1109/IVS.2019.8814254.
- [54] P. Kim, *MATLAB Deep Learning: With Machine Learning, Neural Networks and Artificial Intelligence*. Berkeley, CA: Apress, 2017.
- [55] H. Beglerovic, T. Schloemicher, S. Metzner und M. Horn, „Deep Learning Applied to Scenario Classification for Lane-Keep-Assist Systems“, *Applied Sciences*, 2018, doi: 10.3390/app8122590.

- [56] *Bußgeldkatalog: Abstand & Abstandsvergehen*. [Online]. Verfügbar unter: <https://www.bussgeldkatalog.org/abstand/?gclid=CO2ckuGVycMCFanMtAodpwkAcw>. Zugriff am: 18. April 2020.
- [57] T. Richter, *Planung von Autobahnen und Landstraßen*. Wiesbaden: Springer Fachmedien Wiesbaden, 2016.
- [58] H. F. H. Reinsberg, *Leitfaden Fahrbahnmarkierung*. Bonn: Deutscher Verkehrssicherheitsrat e.V., 2014. [Online]. Verfügbar unter: <https://www.dvr.de/download/publikationen-schriftenreihe-17.pdf>
- [59] *Stateflow: Modellierung und Simulation von Entscheidungslogiken mithilfe von Zustandsautomaten und Flussdiagrammen*. [Online]. Verfügbar unter: <https://de.mathworks.com/products/stateflow.html>. Zugriff am: 22. April 2020.
- [60] *Fit binary decision tree for multiclass classification - MATLAB fitctree*. [Online]. Verfügbar unter: <https://de.mathworks.com/help/stats/fitctree.html>. Zugriff am: 22. April 2020.
- [61] H. Bubb, K. Bengler, R. E. Grünen und M. Vollrath, *Automobilergonomie*. Wiesbaden: Springer Fachmedien Wiesbaden, 2015.
- [62] *Sicherheitsabstand auf der Autobahn - Abstand 2020*. [Online]. Verfügbar unter: <https://www.bussgeldkatalog.de/sicherheitsabstand-autobahn/>. Zugriff am: 27. April 2020.
- [63] *Stichprobenrechner*. [Online]. Verfügbar unter: <https://www.surveymonkey.de/mp/sample-size-calculator/>. Zugriff am: 16. April 2020.
- [64] B. Hulliger, „Einführung in die statistischen Methoden von Stichprobenerhebungen“. Skript, Hochschule für Wirtschaft FHNW, Olten, 2016.

Anhang

Anhang A	Nomenklatur der Zustandsbeschreibung.....xiv
Anhang B	Liste Manöverdefinitionen.....xvi

Anhang A Nomenklatur der Zustands- beschreibung

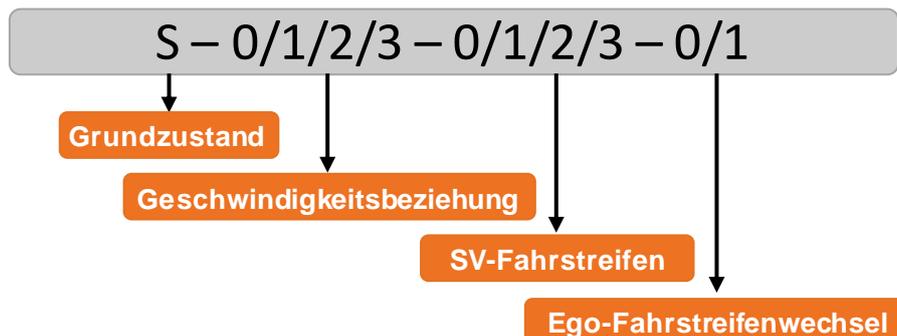


Abbildung A.1: Struktur der Zustandsdefinition

Tabelle A.1 Grundzustand der States

Bezeichnung	Semantische Definition
S1	Umgebungsfahrzeug auf benachbarter Spur hinter Ego
S2	Umgebungsfahrzeug auf benachbarter Spur neben Ego
S3	Umgebungsfahrzeug auf benachbarter Spur vor Ego
S4	Umgebungsfahrzeug auf identer Spur hinter Ego
S5	Umgebungsfahrzeug auf identer Spur vor Ego
S6	Aktiver Fahrstreifenwechsel des Umgebungsfahrzeugs hinter Ego
S7	Aktiver Fahrstreifenwechsel des Umgebungsfahrzeugs vor Ego
S8	Aufenthalt des Umgebungsfahrzeugs in nicht relevantem Bereich

Tabelle A.2 Geschwindigkeitsbeziehung

Bezeichnung	Semantische Definition
0	Relativgeschwindigkeit nicht relevant für Zustandsbeschreibung
1	Umgebungsfahrzeug langsamer als Ego
2	Identische Geschwindigkeit
3	Umgebungsfahrzeug schneller als Ego

Tabelle A.3 SV-Fahrstreifen

Bezeichnung	Semantische Definition
0	Identischer Fahrstreifen
1	Umgebungsfahrzeug auf Fahrstreifen in Fahrtrichtung rechts von Ego
2	Umgebungsfahrzeug auf Fahrstreifen in Fahrtrichtung links von Ego
3	Umgebungsfahrzeug auf Fahrstreifen außerhalb des Relevanzbereichs

Tabelle A.4 Ego-Fahrstreifenwechsel

Bezeichnung	Semantische Definition
0	Ego führt keinen Fahrstreifenwechsel aus
1	Ego führt Fahrstreifenwechsel aus

Anhang B Liste Manöverdefinitionen

Tabelle B.1 Liste Hauptmanöverklassen (beispielhaft Fahrzeuge Nr. 1 und 2)

Manöver	Beschreibung
M1	Folgefahrt
M2	Vorbeifahrt
M3	Parallelfahrt
M4	Schnelle Vorbeifahrt 1 mit Einscherer 1 vor 2
M5	Folgefahrt 1 hinter 2 mit Ausscherer 1 und Vorbeifahrt
M6	Vollständiges Überholen
M7	Schnelle Vorbeifahrt 1 und langsamer Einscherer 2 vor 1
M8	Folgefahrt 1 und Ausscherer 2 vor 1
M9	Schnelle Vorbeifahrt 1 mit Einscherer 1 hinter 2 und Folgefahrt
M10	Vorbeifahrt mit Fahrstreifenwechsel auf übernächsten Fahrstreifen (out of range)
M11	Fahrt auf übernächstem Fahrstreifen mit Fahrstreifenwechsel auf benachbarte Spur
M12	Gegensätzlicher/Paralleler Fahrstreifenwechsel

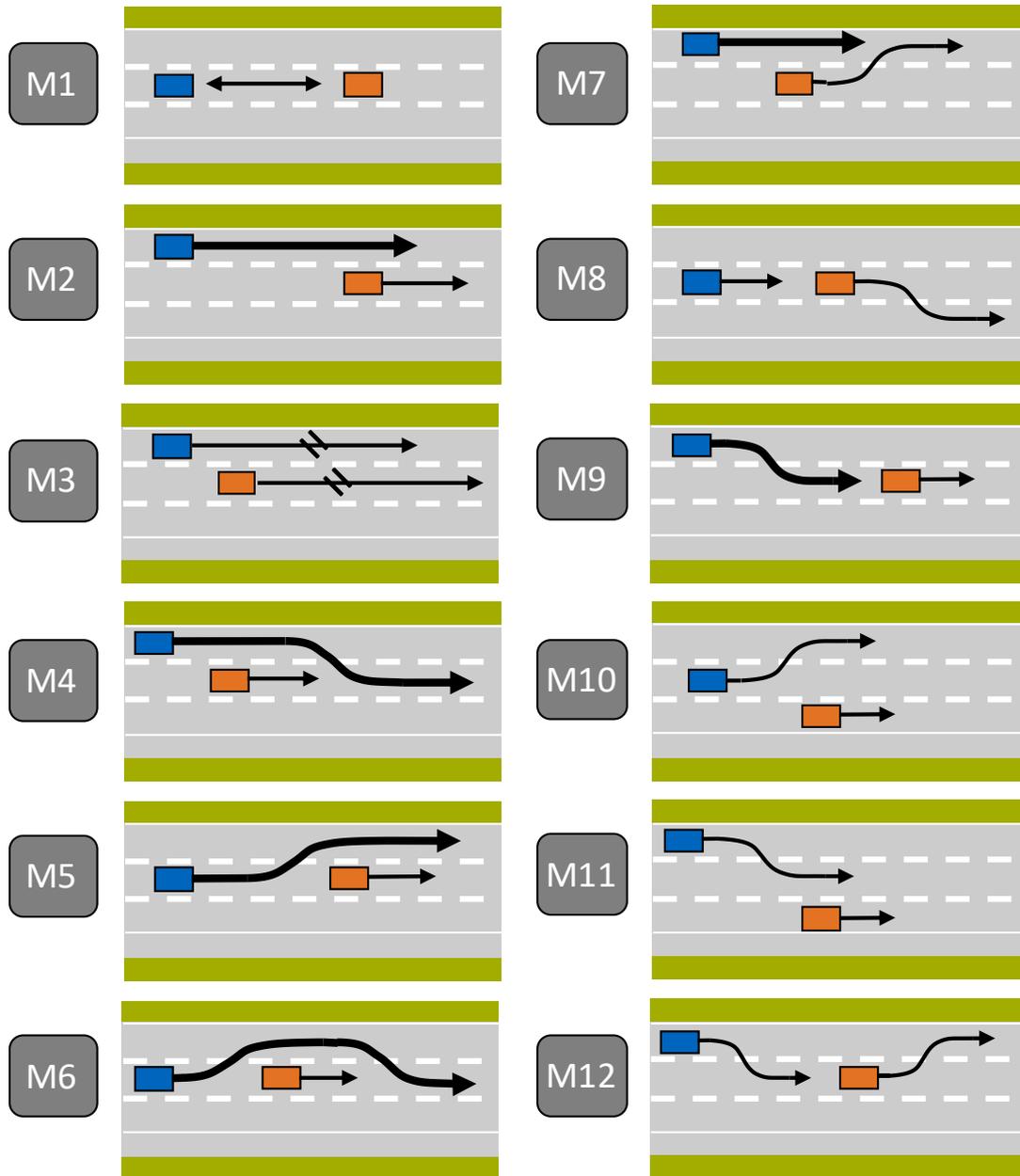


Abbildung B.2: Hauptmanöverklassen 1 - 12

Tabelle B.2 Liste Subklassen - Einzelmanöver

Manöver	Beschreibung
FFA	FollowDrive Front Approach
FFI	FollowDrive Front Identical
FFB	FollowDrive Front Backdrop
FBA	FollowDrive Back Approach
FBI	FollowDrive Back Identical
FBB	FollowDrive Back Backdrop
FFD	FollowDrive Front Diverse
FBD	FollowDrive Back Diverse
SPB	SV Passing Behind
SPM	SV Passing Middle
SPF	SV Passing Front
SPBM	SV Passing Behind Middle
SPMF	SV Passing Middle Front
SPC	SV Passing Complete
EPB	Ego Passing Behind
EPM	Ego Passing Middle
EPF	Ego Passing Front
EPBM	Ego Passing Behind Middle
EPMF	Ego Passing Middle Front
EPC	Ego Passing Completely
PSB	ParallelDrive SV Behind
PSM	ParallelDrive SV Middle
PSF	ParallelDrive SV Front
SPCCIFFA	SV Passing Complete Cut-In FollowDrive Front Approach
SPCCIFFI	SV Passing Complete Cut-In FollowDrive Front Identical
SPCCIFFB	SV Passing Complete Cut-In FollowDrive Front Backdrop
SPMFCIFFA	SV Passing Middle/Front Cut-In FollowDrive Front Approach
SPMFCIFFI	SV Passing Middle/Front Cut-In FollowDrive Front Identical
SPMFCIFFB	SV Passing Middle/Front Cut-In FollowDrive Front Backdrop
SPFCIFFA	SV Passing Front Cut-In FollowDrive Front Approach
SPFCIFFI	SV Passing Front Cut-In FollowDrive Front Identical
SPFCIFFB	SV Passing Front Cut-In FollowDrive Front Backdrop
SCIFFA	SV Cut-In FollowDrive Front Approach
SCIFFI	SV Cut-In FollowDrive Front Identical
SCIFFB	SV Cut-In FollowDrive Front Backdrop
SPCCI	SV Passing Complete Cut-In
SPMFCI	SV Passing Middle/Front Cut-In
SPFCI	SV Passing Front Cut-In
EPCCIFBA	Ego Passing Complete Cut-In FollowDrive Back Approach
EPCCIFBI	Ego Passing Complete Cut-In FollowDrive Back Identical
EPCCIFBB	Ego Passing Complete Cut-In FollowDrive Back Backdrop
EPMFCIFBA	Ego Passing Middle/Front Cut-In FollowDrive Back Approach
EPMFCIFBI	Ego Passing Middle/Front Cut-In FollowDrive Back Identical
EPMFCIFBB	Ego Passing Middle/Front Cut-In FollowDrive Back Backdrop

Manöver	Beschreibung
EPFCIFBA	Ego Passing Front Cut-In FollowDrive Back Approach
EPFCIFBI	Ego Passing Front Cut-In FollowDrive Back Identical
EPFCIFBB	Ego Passing Front Cut-In FollowDrive Back Backdrop
ECIFBA	Ego Cut-In FollowDrive Back Approach
ECIFBI	Ego Cut-In FollowDrive Back Identical
ECIFBB	Ego Cut-In FollowDrive Back Backdrop
EPCCI	Ego Passing Complete Cut-In
EPMFCI	Ego Passing Middle/Front Cut-In
EPFCI	Ego Passing Front Cut-In
SFBCOPC	SV FollowDrive Behind Cut-Out Passing Complete
SFBCOPBM	SV FollowDrive Behind Cut-Out Passing Behind/Middle
SFBCOPB	SV FollowDrive Behind Cut-Out Passing Behind
SFBCOPSB	SV FollowDrive Behind Cut-Out ParallelDrive SV Behind
SFBCOEPF	SV FollowDrive Behind Cut-Out Ego Passing Front
SFBCO	SV FollowDrive Behind Cut-Out
SCOPC	SV Cut-Out Passing Complete
SCOPBM	SV Cut-Out Passing Behind/Middle
SCOPB	SV Cut-Out Passing Behind
SCOPSB	SV Cut-Out ParallelDrive SV Behind
SCOEPF	SV Cut-Out Ego Passing Front
EFBCOPC	Ego FollowDrive Behind Cut-Out Passing Complete
EFBCOPBM	Ego FollowDrive Behind Cut-Out Passing Behind/Middle
EFBCOPB	Ego FollowDrive Behind Cut-Out Passing Behind
EFBCOPSF	Ego FollowDrive Behind Cut-Out ParallelDrive SV Front
EFBCOSPF	Ego FollowDrive Behind Cut-Out SV Passing Front
EFBCO	Ego FollowDrive Behind Cut-Out
ECOPC	Ego Cut-Out Ego Passing Complete
ECOPBM	Ego Cut-Out Passing Behind/Middle
ECOPB	Ego Cut-Out Passing Behind
ECOPSF	Ego Cut-Out ParallelDrive SV Front
ECOSPF	Ego Cut-Out SV Passing Front
SLCB	SV Lane Change Behind
SLCF	SV Lane Change Front
ELCB	Ego Lane Change Behind
ELCF	Ego Lane Change Front
SFBOFFA	SV FollowDrive Behind Overtake FollowDrive Front Approach
SFBOFFI	SV FollowDrive Behind Overtake FollowDrive Front Identical
SFBOFFB	SV FollowDrive Behind Overtake FollowDrive Front Backdrop
SOFFA	SV Overtake FollowDrive Front Approach
SOFFI	SV Overtake FollowDrive Front Identical
SOFFB	SV Overtake FollowDrive Front Backdrop
SFBO	SV FollowDrive Behind Overtake

Manöver	Beschreibung
EFBOFBA	Ego FollowDrive Behind Overtake (SV) FollowDrive Back Approach
EFBOFBI	Ego FollowDrive Behind Overtake (SV) FollowDrive Back Identical
EFBOFBB	Ego FollowDrive Behind Overtake (SV) FollowDrive Back Backdrop
EOFBA	Ego Overtake (SV) FollowDrive Back Approach
EOFBI	Ego Overtake (SV) FollowDrive Back Identical
EOFBB	Ego Overtake (SV) FollowDrive Back Backdrop
EFBO	Ego FollowDrive Behind Overtake
EPBSCIFFA	Ego Passing Behind and SV Cut-In with FollowDrive Front Approach
EPBSCIFFI	Ego Passing Behind and SV Cut-In with FollowDrive Front Identical
EPBSCIFFB	Ego Passing Behind and SV Cut-In with FollowDrive Front Backdrop
EPBMSCIFFA	Ego Passing Behind/Middle and SV Cut-In with FollowDrive Front Approach
EPBMSCIFFI	Ego Passing Behind/Middle and SV Cut-In with FollowDrive Front Identical
EPBMSCIFFB	Ego Passing Middle and SV Cut-In with FollowDrive Front Backdrop
EPMSCIFFA	Ego Passing Middle and SV Cut-In with FollowDrive Front Approach
EPMSCIFFI	Ego Passing Middle and SV Cut-In with FollowDrive Front Identical
EPMSCIFFB	Ego Passing Middle and SV Cut-In with FollowDrive Front Backdrop
EPBSCI	Ego Passing Behind and SV Cut-In
EPBMSCI	Ego Passing Behind/Middle and SV Cut-In
EPMSCI	Ego Passing Middle and SV Cut-In
SPBECIFBA	SV Passing Behind and Ego Cut-In with FollowDrive Back Approach
SPBECIFBI	SV Passing Behind and Ego Cut-In with FollowDrive Back Identical
SPBECIFBB	SV Passing Behind and Ego Cut-In with FollowDrive Back Backdrop
SPBMECIFBA	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Approach
SPBMECIFBI	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Identical
SPBMECIFBB	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Backdrop

Manöver	Beschreibung
EPBSCI	Ego Passing Behind and SV Cut-In
EPBMSCI	Ego Passing Behind/Middle and SV Cut-In
EPMSCI	Ego Passing Middle and SV Cut-In
SPBECIFBA	SV Passing Behind and Ego Cut-In with FollowDrive Back Approach
SPBECIFBI	SV Passing Behind and Ego Cut-In with FollowDrive Back Identical
SPBECIFBB	SV Passing Behind and Ego Cut-In with FollowDrive Back Backdrop
SPBMECIFBA	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Approach
SPBMECIFBI	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Identical
SPBMECIFBB	SV Passing Behind/Middle and Ego Cut-In with FollowDrive Back Backdrop
SPMECIFBA	SV Passing Middle and Ego Cut-In with FollowDrive Back Approach
SPMECIFBI	SV Passing Middle and Ego Cut-In with FollowDrive Back Identical
SPMECIFBB	SV Passing Middle and Ego Cut-In with FollowDrive Back Backdrop
SPBECI	SV Passing Behind and Ego Cut-In
SPBMECI	SV Passing Behind/Middle and Ego Cut-In
SPMECI	SV Passing Middle and Ego Cut-In
SFBECOSPC	SV FollowDrive Behind with Ego Cut-Out and SV Passing Complete
SFBECOSPBM	SV FollowDrive Behind with Ego Cut-Out and SV Passing Behind/Middle
SFBECOSPB	SV FollowDrive Behind with Ego Cut-Out and SV Passing Behind
SFBECOPSB	SV FollowDrive Behind with Ego Cut-Out and ParallelDrive SV Behind
SFBECOPF	SV FollowDrive Behind with Ego Cut-Out and (Ego) Passing Front
SFBECO	SV FollowDrive Behind with Ego Cut-Out
ECOSPC	Ego Cut-Out and SV Passing Complete
ECOSPBM	Ego Cut-Out and SV Passing Behind/Middle
ECOSPB	Ego Cut-Out and SV Passing Behind
ECOPSB	Ego Cut-Out and ParallelDrive SV Behind
ECOPF	Ego Cut-Out and (Ego) Passing Front
EFBSCOEPC	Ego FollowDrive Behind with SV Cut-Out and Ego Passing Complete

Manöver	Beschreibung
EFBSCOEPBM	Ego FollowDrive Behind with SV Cut-Out and Ego Passing Behind/Middle
EFBSCOEPB	Ego FollowDrive Behind with SV Cut-Out and Ego Passing Behind
EFBSCOPSF	Ego FollowDrive Behind with SV Cut-Out and ParallelDrive SV Front
EFBSCOPF	Ego FollowDrive Behind with SV Cut-Out and (SV) Passing Front
EFBSCO	Ego FollowDrive Behind with SV Cut-Out
SCOEPC	SV Cut-Out and Ego Passing Complete
SCOEPBM	SV Cut-Out and Ego Passing Behind/Middle
SCOEPB	SV Cut-Out and Ego Passing Behind
SCOPSF	SV Cut-Out and ParallelDrive SV Front
SCOPF	SV Cut-Out and (SV) Passing Front
EPBCIFFA	Ego Passing Behind Cut-In FollowDrive Front Approach
EPBCIFFI	Ego Passing Behind Cut-In FollowDrive Front Identical
EPBCIFFB	Ego Passing Behind Cut-In FollowDrive Front Backdrop
PSFECIFFA	ParallelDrive SV Front Ego Cut-In FollowDrive Front Approach
PSFECIFFI	ParallelDrive SV Front Ego Cut-In FollowDrive Front Identical
PSFECIFFB	ParallelDrive SV Front Ego Cut-In FollowDrive Front Backdrop
SPFECIFFA	SV Passing Front Ego Cut-In FollowDrive Front Approach
SPFECIFFI	SV Passing Front Ego Cut-In FollowDrive Front Identical
SPFECIFFB	SV Passing Front Ego Cut-In FollowDrive Front Backdrop
SPBCIFBA	SV Passing Behind Cut-In FollowDrive Back Approach
SPBCIFBI	SV Passing Behind Cut-In FollowDrive Back Identical
SPBCIFBB	SV Passing Behind Cut-In FollowDrive Back Backdrop
PSBCIFBA	ParallelDrive SV Behind Cut-In FollowDrive Back Approach
PSBCIFBI	ParallelDrive SV Behind Cut-In FollowDrive Back Identical
PSBCIFBB	ParallelDrive SV Behind Cut-In FollowDrive Back Backdrop
EPFSCIFBA	Ego Passing Front SV Cut-In FollowDrive Back Approach
EPFSCIFBI	Ego Passing Front SV Cut-In FollowDrive Back Identical
EPFSCIFBB	Ego Passing Front SV Cut-In FollowDrive Back Backdrop