

# InCarIn

## IntelligentCarInterieur

Insassen- und Innenraumkontextanalyse im Automobil zur adaptiven und kontextsensitiven Anpassung von Interaktionstechnik und Assistenzsystemen für alle

---

## Schlussbericht

### Zuwendungsempfänger:

Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung  
und

Fraunhofer-Institut für Arbeitswissenschaft und Organisation

### Förderkennzeichen:

16SV6292

### Gefördert durch:



### Betreut vom Projektträger

**VDI|VDE|IT**

Autoren: Michael Voit & Manuel Martin, Fraunhofer IOSB; Frederik Diederichs, Fraunhofer IAO

## 1 Inhalt

1	Inhalt.....	1
2	Zusammenfassung .....	3
3	Kurzdarstellung .....	5
3.1	Aufgabenstellung .....	5
3.2	Voraussetzungen .....	7
3.2.1	Fraunhofer-Institut für Arbeitswissenschaft und Organisation .....	7
3.2.2	Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung .....	8
3.2.3	Bosch.....	9
3.2.4	Johnson Controls bzw. Visteon .....	9
3.2.5	NordSys GmbH.....	9
3.2.6	Universität Stuttgart.....	10
3.2.7	Volkswagen.....	10
3.3	Planung und Ablauf des Vorhabens.....	10
3.3.1	TP 1: Nutzeranforderungen und Anwendungsszenarien .....	14
3.3.2	TP 2: Technisches Systemkonzept, Integration und Demonstratoren .....	14
3.3.3	TP 3: Sensoren und Algorithmen zur Insassen- und Innenraumkontexterkenkung .....	15
3.3.4	TP 4: Erkennung Fahrerzustand/-absicht und Anwendungen .....	16
3.4	Wissenschaftlicher-technischer Stand zu Projektbeginn .....	17
3.4.1	Verfahren und Systeme zur Posenschätzung.....	18
3.4.2	Verfahren und Systeme zur Fahrer- und Insassenmodellierung .....	19
3.5	Zusammenarbeit mit anderen Stellen.....	20

---

4	Ergebnisbericht .....	20
4.1	Erzielte Ergebnisse .....	20
1.1.1	Kamerasetup für die Gesamt Innenraumanalyse.....	20
1.1.2	Algorithmus zur Körperposenerfassung der Insassen .....	26
1.1.3	Trainieren der Modelle zur Körperposenerfassung.....	31
1.1.4	Gesamtinnenraumerfassung.....	34
1.1.5	Interaktionserkennung mit dem Fahrzeuginnenraum.....	35
1.1.6	Erkennung von Nebentätigkeiten aller Insassen .....	37
1.1.7	Entwicklung von Interaktionsprinzipien für proaktive Interaktion mit Insassen .....	38
1.1.8	Datenerhebung im T5 als Realfahrtversuch.....	40
1.1.9	Aufbau der Versuchsträger .....	41
1.1.10	Entwicklung einer Bewertungsmethode für Körperpositionen.....	43
1.1.11	Verbreitung der Ergebnisse, Halbzeit- und Abschlusspräsentation und Homepage .....	46
4.2	Fortschritt auf diesem Gebiet bei anderen Stellen.....	46
4.3	Veröffentlichung der Ergebnisse .....	47
5	Literaturangaben.....	48

## 2 Zusammenfassung

Das Vorhaben IntelligentCarInterieur („InCarIn“) hatte sich zum Ziel gesetzt, durch die Erforschung und Entwicklung neuer Lösungen zur Fahrzeuginnenraumsensorik, Innenraumkontexterfassung und Innenraumkontextinterpretation eine wichtige Grundlage für zukünftige Sicherheits-, Komfort- und Assistenzsysteme im Fahrzeug zu schaffen. Eine tiefgehende Adaption elektronischer Fahrzeugsysteme an menschliche Bedürfnisse – insbesondere an die Bedürfnisse von Menschen mit altersbedingten oder anderweitigen körperlichen Einschränkungen – ist ohne dieses grundlegende Know-How nicht realisierbar.

Die Ergebnisse des Vorhabens sollten ein besseres Verständnis über das menschliche Verhalten im Fahrzeug, die Aufmerksamkeitsverteilung sowie über Handlungsabsichten, Gewohnheiten und Wünsche liefern. Als Motivation dienten hierzu adaptive Fahrzeugsysteme, die von neuen Bedien- und Interaktionssystemen bis hin zu neuen, kontrollierten Fahrassistenz- und Fahrautomatiksystemen reichen.

Die Forschungs- und Industriepartner des InCarIn-Konsortiums brachten alle hierfür erforderlichen Sachkenntnissen zusammen: Die Universität Stuttgart sowie das Fraunhofer-Institut für Arbeitswirtschaft und Organisation (IAO) verfügten über umfangreiches Grundlagenwissen zur Bewertung von Fahreigenschaften und zur Durchführung von Nutzerstudien unter Fahrbedingungen. Die Grundlagen zum maschinellen Sehen und zur Personen-Perzeption wurden durch das Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) beigesteuert. Das IOSB verfügte zusätzlich über große Erfahrungen in der Kontext- und Situationsanalyse. Die Partner Bosch und Johnson Controls trugen mit ihren umfangreichen Erfahrungen bei den Fahrer-Fahrzeug-Interaktionssystemen zum Projekt bei. Die NordSys GmbH war auf dem Gebiet der Systemintegration und Datenfusion von Fahrerinformations- und Assistenzsystemen spezialisiert und brachte eine Vielzahl eigenentwickelter Tools zur Sensor- und Steuergeräteintegration mit ein, die bereits in zahlreichen Forschungsprojekten erfolgreich zum Aufbau und zur Integration prototypischer Fahrzeugsysteme Verwendung fanden. Und die Volkswagen AG verfügte durch eigene Serienentwicklungen und Forschungsaktivitäten über reichhaltige Erfahrung in der Fahrzeugumfeld- und Fahrerbeobachtung und sah erheblichen Bedarf am Einsatz dieser Systeme in zukünftigen Produkten.

Die Bearbeitung des Vorhabens wurde auf vier inhaltliche und ein organisatorisches Teilprojekt aufgeteilt: Im ersten Teilprojekt sollten Nutzeranforderungen analysiert werden und auf dieser Basis neue, innenraumkontextsensitive Anwendungen und ein Insassenmodell entwickelt werden. In den kontinuierlichen und iterativen Nutzerstudien wurde ein besonderer Schwerpunkt auf die Nutzerakzeptanz sowie die Einhaltung von datenschutzrechtlichen Anforderungen und Bedenken

gelegt. Das zweite Teilprojekt hatte eine gemeinsame Referenzplattform zum Ziel, die anschließend zur Validierung und Evaluation der entwickelten Lösungen durch prototypische Versuchsträger umgesetzt werden sollte.

Im dritten Teilprojekt bündelte sich die Entwicklung von Sensorsystemen und Algorithmen zur bildanalytischen Umsetzung der Insassen- und Innenraumkontexterkennung. Dazu sollten Sensorsysteme an den Fahrzeugkontext angepasst, mit vorhandenen Fahrzeugsensorsystemen sowie mit Bring-In-Devices integriert und die Daten zu einer Innenraum-Sensordatenbasis fusioniert werden. Für die Erkennung und Interpretation der Zustände und Absichten des Fahrers und weiterer Insassen wurde Teilprojekt 4 zugeschnitten. Hierzu sollte die situative Aktivität basierend auf den erfassten Sensordaten interpretiert, sowie Verfahren des maschinellen Lernens und Data-Minings eingesetzt werden, um langfristige Gewohnheiten zu ermitteln. Die starken Abhängigkeiten zwischen den inhaltlichen Teilprojekten erforderte dabei eine strukturierte, gemeinsame Arbeitsweise, die durch die Aktivitäten des begleitenden Teilprojektes 5 sichergestellt werden sollte. Ziel von TP5 war es, ein durchgängiges Projektverständnis aufzubauen und die Projektergebnisse intern wie auch nach außen verständlich, transparent und zielgruppenorientiert zu kommunizieren.

**Teilprojekt 1: Nutzeranforderungen und Anwendungsszenarien**

Frühzeitige und kontinuierliche Einbindung von Nutzern, Nutzerzentrierte Evaluation des Gesamtsystems und von Anwendungen.

**Teilprojekt 2: Technisches Systemkonzept, Integration und Demonstratoren**

Aufbau eines Versuchsfahrzeugs zur Erfassung vielfältiger Insasseneigenschaften unter realen Fahrbedingungen. Bereitstellung einer projektweit einheitlichen Referenz-Plattform zur Datenverarbeitung.

**Teilprojekt 3: Sensoren und Algorithmen zur Innenraumkontexterkennung**

Umsetzung einer sensorischen und bildanalytischen Erfassung aller Insassen im Fahrzeuginnenraum durch einen heterogenen Sensorverbund. Modellierung der 3D-Körperposen aller Insassen.

**Teilprojekt 4: Erkennung Fahrerzustand/-Absicht und Anwendungen**

Fahreridentifikation und Fahrerzustandserkennung. Klassifikation von Aktivitäten, Interaktionen, visueller Aufmerksamkeit und Ablenkung, Fahrer-Workload.

**Teilprojekt 5: Projektmanagement und Dokumentation**

Inhaltliche Abstimmung der Arbeiten, Planung und Organisation von Projekttreffen und Präsentationen, Informationsaustausch sowie partnerübergreifende Koordination. Darstellung der Projektergebnisse nach außen.

Abbildung 1: Die Teilprojekte von InCarIn

## 3 Kurzdarstellung

### 3.1 Aufgabenstellung

Ziel des Projektvorhabens InCarIn war es, mittels einer ganzheitlichen Personen- und Gestenerkennung aller Fahrzeuginsassen und einer adaptiven Innenraumkontextanalyse die individuellen Bedürfnisse aller Insassen zu erfassen und entsprechend dem jeweiligen Kontext Assistenz, Information und Komfortfunktionen anzubieten sowie die Mensch-Technik Interaktion auf natürliche und nutzerspezifische Weise auszulegen.

Mit InCarIn sollte erstmals ein substantieller Beitrag geleistet werden, mittels dem die Situation und Verfassung aller Fahrzeuginsassen wahrgenommen und interpretiert wird. Basierend auf dieser Interpretation sollte eine kontextsensitive und adaptive Interaktion zwischen Fahrzeug und Insassen erfolgen. Das Projektvorhaben sollte durch die angereicherte Personenerfassung im Innenraum die Grundlagen dafür legen, um sowohl für die Insassen, wie auch für die Gesamtbevölkerung, einen Sicherheitsgewinn zu erzielen.

Die Arbeitsziele waren daher vorrangig auf die Forschung und Entwicklung der Sensorbasis ausgerichtet, mittels der im Innenraum für alle Insassen eine Erfassungsmethodik ausgearbeitet werden musste. Das beinhaltete robuste Sensoren und Auswertelgorithmen, Methoden und Leitfäden bezüglich der Anforderungen an Sensoren, die Sensorfusion und die Definition und Umsetzung von Schnittstellen gleichermaßen. Aufgabe war es, diese, unter Berücksichtigung von Nutzerbedürfnissen und Nutzeranforderungen, zu einer kontextsensitiven und adaptiven Insassen-Fahrzeug-Interaktion zu vereinen. Essenziell waren hierzu folgende Aufgaben von Anfang an erkennbar: Konzeption und Umsetzung eines Systems zur robusten Posenerkennung von Fahrzeuginsassen, inkl. Hardware, Integration im Fahrzeug, die Anpassung für unterschiedliche Nutzergruppen (z.B. Kinder/Erwachsene), Algorithmen für die Erkennung von Position, Körperbewegungen/-haltung, sowie Detektion kritischer Zustände. Darauf aufbauend war ein Methodenbaukasten des maschinellen Lernens zur Unterstützung weiterer Anwendungsszenarien geplant.

Um den Nutzerbedürfnissen und –Qualifikationen gerecht zu werden, sollte von Anfang an eine iterative Einbeziehung von realen Nutzern erfolgen. Ergänzend dazu, und ergänzend zur technischen Validierung der Systeme, sollte die Absicherung der Anwendbarkeit unter Berücksichtigung rechtlicher Aspekte im Vordergrund stehen. Diese Aufgaben wurden durch die Universität Stuttgart gebündelt in TP1 gemeinsam mit dem Fraunhofer IAO koordiniert. Dazu wurden von Anfang an Arbeiten geplant, um die grundlegende Nutzerszenarien und -bedürfnisse zu erfassen und zu systematisieren. Typische Verhaltensweisen im Fahrzeug sowie die zu erkennenden atypischen Verhaltensweisen und Schlüsselsituationen für die geplanten Anwendungsfälle sollten dadurch ermittelt werden. Erste Datenerfassungen waren hierzu im

Fahrsimulator am Fraunhofer IAO geplant, in denen mittels Probandenstudien der Ausgangspunkt für die Auswahl der notwendigen Sensoren und Algorithmen erfolgen sollte.

Für die Entwicklung dieser Sensoren und Algorithmen zeichnete sich hingegen das Fraunhofer IOSB in TP3 verantwortlich. Kamerabasiert wurde darin die Körperpose aller Insassen erfasst und Insassenmodelle entwickelt, die in TP2, TP4 und TP5 im gemeinsamen Demonstrator in Abstimmung mit Bosch, Volkswagen, Johnson Controls (inzwischen Visteon) und der NordSys integriert und in fokussierten Anwendungen evaluiert wurden.

Neben seinem Fahrsimulationslabor brachte das Fraunhofer IAO in InCarIn seine Expertise zur Entwicklung von Referenzmethoden von Fahrer- und Insassenzuständen ein, während das IOSB sich in der Algorithmen- und Sensorbasisentwicklung verantwortlich zeigte. Die beiden Institute ergänzten den Verbund damit durch das logische Verwertungskonzept, sensornahe Integrationsarbeiten hauptverantwortlich zu begleiten und in realitätsnahen und nutzerzentrierten Studien zu validieren und evaluieren.



Abbildung 2: Ziel des Projekts war die Entwicklung alltagstauglicher und von Nutzern akzeptierte Innenraumerfassung und Assistenz im Innenraum. Dazu wurde zunächst eine Anforderungsanalyse an die zu entwickelnden Systeme erstellt. Iterative Nutzerstudien begleiteten die dreijährige Entwicklungsphase.

## 3.2 Voraussetzungen

Das Konsortium wurde mit der Absicht zusammengestellt, wissenschaftliche Arbeiten in den beiden Fraunhofer-Instituten und der Universität Stuttgart zu bündeln und zur Integration und Umsetzung die in der Automobilbranche übliche Rollenverteilung der Geschäftsmodelle widerzuspiegeln.

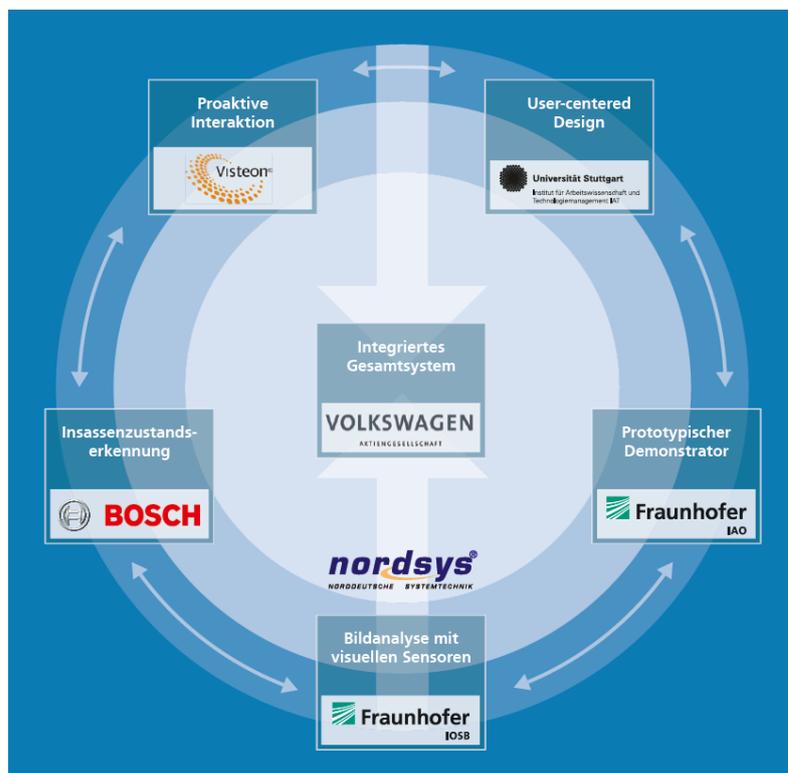


Abbildung 3: Die Verbundpartner ergänzten sich hervorragend und brachten jeweils eigene Expertise zur Zielerreichung in den Verbund ein.

### 3.2.1 Fraunhofer-Institut für Arbeitswissenschaft und Organisation

Seit über 10 Jahren befasst sich das Vehicle Interaction Team des Fraunhofer-Institut für Arbeitswissenschaft und Organisation IAO unter Leitung von Univ.-Prof. Dr.-Ing. Dr.-Ing. E.h. Dieter Spath. mit der Mensch-Fahrzeug-Interaktion, der Adaption von Fahrerassistenzsystemen und der Fahrerzustandserfassung. Grundlage für die angewandte Forschung bildet das interdisziplinär besetzte Fahrsimulationslabor. Das am Fraunhofer IAO erarbeitete Wissen wird in Kooperationen mit Industriepartnern in innovative Produkte und Systeme überführt.

Durch die enge Kooperation mit der Universität Stuttgart kann für die Entwicklung und Evaluation auf fundierte und aktuelle Erkenntnisse der Grundlagenforschung im Bereich Fahrer-Fahrzeug-Interaktion zurückgegriffen werden.

Durch die Entwicklung einer Referenzmethode sollte das Thema Fahrer- und Insassenzustand von Seiten des IAO aus wissenschaftlicher Sicht betrachtet werden und objektive Kriterien entwickelt werden, die dazu dienen sollten, das im Projekt entwickelte System hinsichtlich der validen und zuverlässigen Messung von fahrer- und Insassenzuständen zu überprüfen.

Zusätzlich stellte das IAO durch sein Fahrsimulationslabor ein realitätsnahes Umfeld zur frühen Evaluation von Sensoren und Algorithmen durch die Partner zur Verfügung. Auch erste Adaptionen und Interaktionskonzepte sollten durch die Universität Stuttgart im Labor unter realistischen aber nicht sicherheitskritischen Bedingungen getestet werden. Ergänzt durch den Alterssimulationsanzug stellte das Fraunhofer IAO somit einen realitätsnahen Laboraufbau zur Verfügung, der frühzeitig im Projekt zu nutzerzentrierten und technischen Evaluationen eingesetzt werden kann.

### **3.2.2 Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung**

Das Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB) unter der Leitung von Prof. Dr.-Ing. J. Beyerer begleitet mit seinen Kernkompetenzen der Bildgewinnung bis zur Bildauswertung die Prozesskette von der Grundlagenforschung bis zur Entwicklung. In der Abteilung ‚Interaktive Analyse und Diagnose (IAD)‘ werden hierzu innovative Interaktionsmethoden und Assistenzsysteme entwickelt, die Menschen bei ihren Aufgaben unterstützen. Die Anwendungsfelder reichen dabei von der aufklärenden Szenenanalyse bis zu Fabrikprozessen in der industriellen Produktion. Das Hauptaugenmerk liegt dabei auf der Entwicklung videobasierter Perzeptions- und Interaktionstechniken, insbesondere auf der Erfassung von Personen und der Auswertung ihrer Aktivitäten sowie der Integration und Entwicklung multimodaler Interaktionstechniken in intelligenten und proaktiven Umgebungen.

Damit sollen vorrangig Fragestellungen nach natürlichen Eingabemodalitäten, dem Berücksichtigen von Kontextbezügen und den Möglichkeiten intuitiver Benutzerschnittstellen beantwortet werden.

Das Fraunhofer IOSB beschäftigt sich inzwischen seit vielen Jahren mit dem Forschungsthema intelligenter Umgebungen, in denen Menschen videobasiert, aber auch akustisch, wahrgenommen und durch intuitive Assistenzsysteme proaktiv unterstützt werden. Im Verbund steuerte das Fraunhofer IOSB seine Expertise im videobasierten Erfassen der Insassen und dem Nachvollziehen ihrer Aktivitäten sowie damit einhergehenden Ablenkungen im Fahrzeuginnenraum bei.

Die bereits bestehenden Verfahren am IOSB sollten hierzu aus ihrem bisherigen Anwendungsumfeld eines aufmerksamen Raums bzw. einer aufmerksamen Umgebung in das Fahrzeuginnere übertragen werden. In diesem Zusammenhang beteiligte sich das IOSB bei der grundlegenden Fragestellung nach geeigneter Sensorik im Fahrzeuginnenraum (ggf. im

Sensorverbund), um alle Insassen in ausreichender Auflösung und Qualität zu erfassen. Auf die im Verbund festgelegte Sensorplattform wurden schließlich die einzusetzenden Verfahren angepasst bzw. notwendigen Weiterentwicklungen vorgenommen, um von allen Personen ein Skelettmodell ihrer Körperhaltung und hiervon abgeleitetes Wissen über ihre Bewegungen, Gesteninteraktion, Aktivitäten und Ablenkungen zu deduzieren.

Die Beteiligung des IOSB war notwendig, um eine vollständige Beobachtung der Insassenaktivität und allgemein vorhandenen Innenraumsituation zu erhalten und hierauf sowohl mit Interaktions- aber auch Sicherheitskonzepten eingehen zu können.

### 3.2.3 Bosch

Bosch nahm die Rolle eines Automobilzulieferers ein, dessen Schwerpunkt in der Einbringung von Integrations-Knowhow und der Zulieferung von Technologien und Komponenten für das Gesamtsystem besteht. Die Arbeitsschwerpunkte von Bosch lagen in der Bereitstellung von Sensorsystemen, sowohl neuer Messtechnik als auch seriennaher eigener Lösungen für die Zustandserkennung von Fahrer und Passagieren. Bosch leistete darauf aufbauend Untersuchungen von Verfahren, um Fahrgewohnheiten einzulernen und damit verbundene Nutzfunktionen für Infotainment- und Assistenzsysteme zu untersuchen, die auf der Prädiktion von Fahrerhandlungen basieren.

### 3.2.4 Johnson Controls bzw. Visteon

JCI zeichnete sich für eingebettete Hardware verantwortlich, die die Integration der Funktionen Infotainment, Anbindung zu Cloud-basierten Services sowie das zu entwickelnde Fahrermodell beinhaltete. Daneben beteiligte sich Johnson Controls an der Entwicklung des Fahrermodells, testete es und integrierte es als eingebettete Komponenten in die hauseigene Hardware. In diesem Zusammenhang trug Johnson Controls eine entscheidende Rolle bei der Integration dieser Hard- und Softwarekomponenten in das Gesamtsystem. Während der Projektlaufzeit wechselte die beteiligte Abteilung von Johnson Controls zu Visteon.

### 3.2.5 NordSys GmbH

NordSys sollte bereits in der Analysephase die spätere Integrationsfähigkeit der Systemkomponenten in die Versuchsträger-Fahrzeuge erfassen und beurteilen. Darüber hinaus zeichnete sich NordSys dafür verantwortlich, das Integrationskonzept zu entwickeln, wie die in diesem Projekt entstehenden neuen Sensorknoten in bestehende Fahrzeugarchitekturen integriert werden können. Sowohl Fragen der Hardwareintegration als auch der softwareseitigen Datenfusion eine Rolle spielten dabei eine Rolle.

### 3.2.6 Universität Stuttgart

Als universitäres Forschungsinstitut brachte die Universität Stuttgart seine Kompetenz zur Erforschung der Grundlagen einer nutzergerechten Insassen- und Innenraumkontextanalyse ein. Die Universität Stuttgart setzte sich zum Ziel, Nutzeranforderungen und -bedürfnisse zu analysieren und entsprechende Gestaltungsleitlinien hinsichtlich der neuen Interaktionsmöglichkeiten abzuleiten. Während der Projektlaufzeit wurden darauf basierend Gestaltungsleitlinien neuer Insassen-Innenraum-Interaktionen definiert und evaluiert. In diesem Zusammenhang betrachtete der Partner auch die kulturellen und ethischen Fragestellungen aus Sicht der Nutzer.

### 3.2.7 Volkswagen

Volkswagen übernahm als Partner aus der Automobilindustrie die Projektleitung sowie die Leitung der Teilprojekte zur Integration und Projektmanagement. Durch die Übernahme der Gesamtkoordination und der technischen Koordination des Aufbaus einer Versuchsträgerplattform sollte ein früher Bezug zu den realen Anforderungen in Fahrzeugsystemen sichergestellt werden. Ziel Volkswagens war es, Nutzeranforderungen und technische Herausforderungen insbesondere im Hinblick auf zukünftige Fahrzeugkonzepte zu betrachten und hierfür Lösungen zu entwickeln, während die Automobiltauglichkeit der entwickelten Sensorkonzepte sichergestellt und auch in realen Fahrttests geprüft werden sollte.

## 3.3 Planung und Ablauf des Vorhabens

Das Projekt InCarIn gliederte sich in 5 Teilprojekte (kurz TP), welche in Arbeitspaketen (kurz AP) organisiert waren. Das Projektmanagement und die Koordination des Gesamtvorhabens wurden dabei durch VW übernommen, die wissenschaftliche Koordination durch die Universität Stuttgart. Die Aufgaben hierzu wurden in TP5 gebündelt.

Koordiniert durch die Universität Stuttgart wurden in TP1 zunächst grundlegende Nutzerszenarien und -bedürfnisse erfasst und systematisiert. Daraus wurden typische Verhaltensweisen im Fahrzeug sowie die zu erkennenden atypischen Verhaltensweisen und Schlüsselsituationen für die geplanten Anwendungsfälle ermittelt.

Die Sensor- und Algorithmenentwicklung wurden durch Bosch und das Fraunhofer IOSB in TP3 und 4 entwickelt und eingebracht, während Volkswagen das technische Integrationskonzept in TP2 beisteuerte. Darin spielte insbesondere auch das durch die NordSys entwickelte Toolkonzept

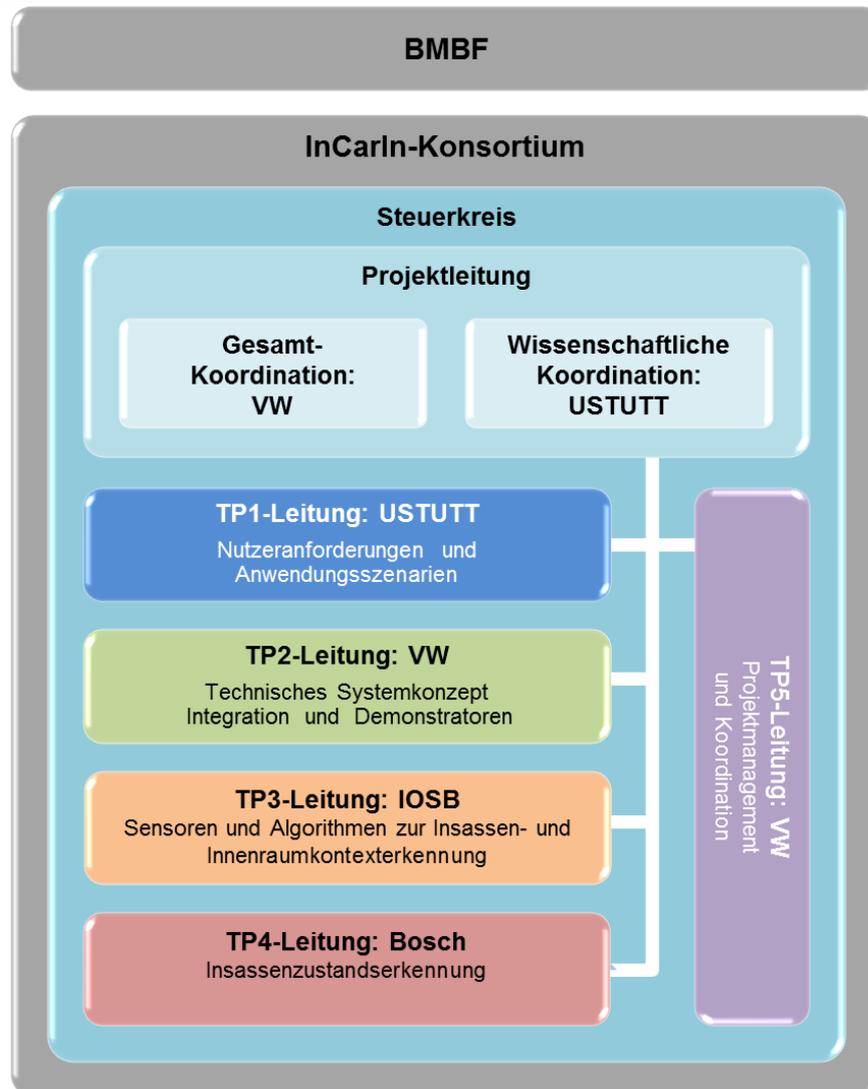


Abbildung 4: Struktur und Verantwortliche der Teilprojekte.

und die Dokumentation und Umsetzung der gemeinsam genutzten Softwareschnittstelle eine zentrale Rolle für die Datenfusion und die angestrebte Wiederverwendbarkeit aller Arbeiten. Johnson Controls, Bosch und das Fraunhofer IOSB entwickelten auf dieser Datenbasis in TP4 das situationsbezogene Insassenmodell, das dann im gemeinsamen Demonstrator mit den fokussierten Anwendungen in Probandenstudien zu evaluieren war.

Zwischen den Teilprojekten und Arbeitspaketen existierten mehrere Abhängigkeiten und Wechselwirkungen. TP1 bildete die Grundlage für die Arbeiten in TP2 und TP3. Basierend auf den hier entwickelten Gestaltungsrichtlinien sollten Sensoren ausgewählt werden, Algorithmen entwickelt und ein Gesamtsystem konzipiert werden. Basierend auf den definierten Eigenschaften der Sensoren und Algorithmen sollten dann wiederum Interaktionskonzepte entwickelt werden. Diese komplettierten schließlich das Gesamtsystem, das im Fahrzeug umgesetzt wird. Begleitet werden sollten diese Arbeiten durch eine iterative Evaluation zur Überprüfung der Einhaltung von Nutzeranforderungen und –Bedürfnissen, der Vereinbarkeit mit kulturellen und ethischen Faktoren

## Schlussbericht

---

sowie der Analyse der Nutzersicht auf das Thema Datenschutz. TP2 stellte die zentrale Referenzplattform dar und wies somit deutliche Abhängigkeiten zu allen anderen Teilprojekten auf. Die Konzeption der Referenzplattform erfolgte auf den Vorgaben aus Teilprojekt 1 (insbesondere den in AP 1.1 festzulegenden Anforderungen und Szenarien) und Teilprojekt 3 (in dem die Sensorbasis spezifiziert wurde). Über die Kamerasensorik hinaus stellte TP3 dabei insbesondere den Kern der Innenraumerfassung und erfassbarer Insassenzustände dar, womit eine deutliche Wechselwirkung zu TP4 gegeben war, in welchem darauf aufbauend situative Handlungen und das Einlernen individueller Gewohnheiten geplant war. Gleichwohl sollte die Sensorauswahl in TP3 in Abhängigkeit von den in TP1 extrapolierten Anforderungen an die Innenraumerfassung durchgeführt werden.

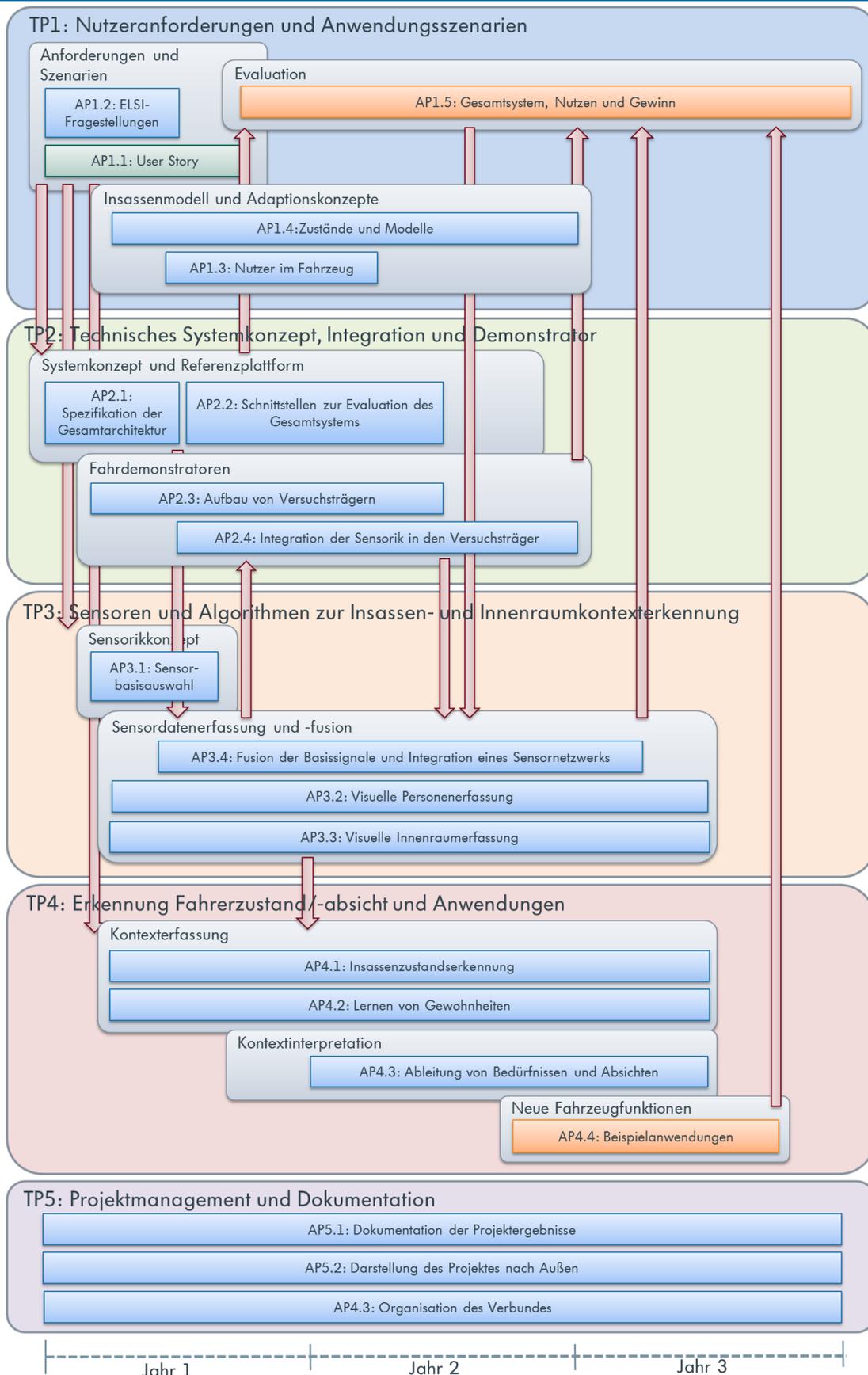


Abbildung 5: Abhängigkeiten der Teilprojekte und Arbeitspakete über die drei Projektjahre.

### 3.3.1 TP1: Nutzeranforderungen und Anwendungsszenarien

Dieses Teilprojekt bildete durch seine Analyse der Nutzeranforderungen und -bedürfnisse die Grundlage zur Entwicklung einer innenraumkontextsensitiven und insassenadaptiven Interaktion im Fahrzeug. Im Fokus standen von Beginn an Zielgruppen mit spezifischen Eigenschaften, wie die Generation 65+, Fahranfängern oder Personen mit speziellen Einschränkungen durch Krankheiten. Bei der Erhebung und Evaluation sollte ein besonderes Augenmerk auf ethische Aspekte und kulturelle Unterschiede gelegt werden. Gerade das Thema Gesten und Interaktion mittels Gesten erforderte die Betrachtung unterschiedlicher kultureller Herkunft. Die Einbindung der Nutzer erfolgte dabei durchgängig von Beginn des Projektes bis zu einer finalen Evaluation des Gesamtsystems durch iterative Befragungen, Fokusgruppen, Verhaltensdatenanalyse, Evaluationen von Konzepten und ähnlichen Methoden. Die Auswahl der Anwendungsszenarien sowie die Festlegung der zu berücksichtigenden Personas stand dabei zunächst im Vordergrund. Darauf und auf den zu erwartenden und erfassbaren Körperposen und Gesteninteraktionen aufbauend, wurden Gestaltungsrichtlinien und nutzerbezogenen Adaptionmöglichkeiten von Fahrerinformationssystemen im Dialog zwischen den verantwortlichen Partnern ausgearbeitet.

### 3.3.2 TP2: Technisches Systemkonzept, Integration und Demonstratoren

Teilprojekt 2 war dazu konzipiert worden, die Forschungsarbeiten zur Insassenzustandserkennung und Situationserkennung in einem Gesamtsystem zusammenzuführen. Um die in TP1 und TP4 entwickelten Insassen- und Kontextmodelle in einer realitätsnahen Umgebung evaluieren zu können, sollte das System als prototypischer Versuchsträger in ein echtes Fahrzeug umgesetzt werden. Der Versuchsträger sollte dazu dienen, die Integrierbarkeit der Teilsysteme zur Sensordatenerfassung validieren und die Sensorik in realitätsnahen Fahrsituationen analysieren zu können. Eine vollständig integrierte Sensorik sollte damit die Grundlage bieten, um weitergehende Studien zur Entwicklung des Kontextmodells und zur Validierung des Insassenmodells auszuarbeiten.

Für den Entwurf des technischen Systemkonzepts wurde zunächst in mehreren Arbeitstreffen die Systemtopologie in gemeinsamer Abstimmung mit Bosch, NORDSYS, JCI und Volkswagen abgestimmt. Zur Reduktion von Redundanzen und dem Einsparen von Rechenleistung stand dabei insbesondere die Mehrfachnutzung derselben Sensorik und die Auswahl eines gemeinsamen Entwicklungsframeworks im Mittelpunkt. Mit ADF wurde ein plattformübergreifendes, modulares Framework ausgewählt, in dem schließlich alle technischen Komponenten zu integrieren waren. Die NORDSYS hat hierzu alle betroffenen Verbundpartner geschult und in gemeinsamer Abstimmung wurden Schnittstellen der jeweiligen Komponenten definiert und in den Folgejahren implementiert. Parallel wurde mit den Aufbauarbeiten des

Versuchsträgers bei Volkswagen begonnen und erste Sensorik und die zugehörigen Erfassungskomponenten integriert.

Mit Ende des Jahres 2014 konnten alle Implementierungsarbeiten der ADTF-Schnittstellen abgeschlossen und im März 2015 zu ersten Evaluationen der ausgewählten Sensorpositionen im VW-Versuchsträger übergegangen werden. Die Bestätigung der ausgewählten Sensorpositionen führte schließlich dazu, dass im Mai 2015 die Kameras im Versuchsträger fest verbaut und erste Laufzeittests der Sensorik und des Netzwerks im Versuchsträger durchgeführt werden konnten. Ergänzend zur verbauten Sensorik wurden die entsprechenden Auswerte-PCs eingebaut und angeschlossen.

### 3.3.3 TP3: Sensoren und Algorithmen zur Insassen- und Innenraumkontexterkennung

Die Sensor- und Algorithmen Entwicklung konzentrierte sich zunächst auf eine Sensorbedarfserhebung im Innenraum und verschiedene Sensorpositionen wurden in gemeinsamer Absprache entworfen und untersucht. In diesem Zusammenhang wurden verschiedene Sensoren für den Einsatz im Fahrzeuginnenraum betrachtet und getestet und in einer vorläufigen Auswertung der Fokus auf den Einsatz von Stereo- und 3D-Kameras gelegt. In den Tests wurde insbesondere deutlich, dass aktive Beleuchtung der Sensorik notwendig werden würde, um der variantenreichen Beleuchtungsschwankungen im Innenraum Rechnung tragen zu können. Die finale Auswahl fiel daher auf die Nutzung aktiver 3D-Kameras, die ferner ermöglichte mit nur vier verbauten Sensoren den gesamten Innenraum erfassen und wahrnehmen zu können. Nachdem die Spezifikation der Sensorbasis fristgerecht abgeschlossen werden konnte, begannen Datenaufnahmen an den definierten Positionen im Innenraum und erste Portierungsarbeiten der am IOSB vorhandenen Arbeiten zur Körperposenerfassung, um anhand der aufgezeichneten Daten die notwendigen Weiterführungs- und Anpassungsarbeiten identifizieren zu können. Erste Experimente mit Verfahren des maschinellen Lernens, insbesondere Random Decision Forests, zur Verbesserung des Systems zeigten sich an dieser Stelle vielversprechend.

Im zweiten Projektjahr konnte frühzeitig eine erste lauffähige Integration der Körperposenerfassung im Fahrzeuginnenraum sowohl für die Vordersitz- als auch die Rückbankinsassen live getestet werden. Das lauffähige System wurde dazu eingesetzt, im Versuchsträger einen neuen Datensatz aufzuzeichnen, um das nun vorliegende Verfahren auf den Versuchsträger anzupassen und zu robustifizieren. Neben den Annotationsarbeiten des aufgezeichneten Datensatzes wurde mit den Implementierungsarbeiten zur Visualisierung der erfassten Daten im Innenraum begonnen.

Aufgrund des erfolgreichen Machbarkeitsnachweises mit der bestehenden Sensorbasis konnte darauf verzichtet werden, weitere redundante Kameraansichten in den Innenraum integrieren zu

müssen. Aufgrund der geforderten Praxistauglichkeit und einzusparenden Rechenleistung war dies ein sehr positives Ergebnis.

Die Arbeiten im dritten Jahr konzentrierten sich auf die Stabilisierung der Erfassungsalgorithmen und dem Aufbau eines eigenen Frameworks zum Eintrainieren der dahinterstehenden Verfahren maschinellen Lernens. Dabei wurde insbesondere die Problemstellung berücksichtigt, ausreichend Trainingsdaten für die Erfassungskomponenten akquirieren zu können, um eine hinreichend genaue Erfassungsqualität gewährleisten zu können. Die Wahl fiel dabei auf die Nutzung synthetisch generierter Trainingsdaten, da diese im Vergleich zu realen Datensammlungen den Vorteil einer automatisierbaren Generierung und automatischen Annotation bieten. Insbesondere durch die automatische Annotation konnte ein deutlicher Zuwachs der erreichten Posenschätzungsgenauigkeit verzeichnet werden.

Die Integration in das unter den Verbundpartnern festgelegte ADTF-Framework wurde abgeschlossen und Komponenten fertigstellend in den Versuchsträger integriert werden. Dabei wurde insbesondere ein gemeinsames Fahrzeugkoordinatensystem festgelegt, in dem alle Insassen und Innenraumobjekte, die durch das CAD-Modell des Fahrzeugs bekannt vorlagen, modelliert wurden. Mit Abschluss der Arbeiten und der fertigen Integration konnten in diesem Fahrzeugkoordinatensystem alle Insassen sowie erkannte und nicht dem Fahrzeug zugehörigen Objekte im Innenraum beschrieben werden.

### **3.3.4 TP4: Erkennung Fahrerzustand/-absicht und Anwendungen**

Die Erkennung der Fahrerzustände und –absichten baut unmittelbar auf der Insassenerfassung in TP3 auf. Vorausblickend wurden hierzu im ersten Projektjahr erste Konzepte entworfen und Experimente im Versuchsträger geplant, um die zu berücksichtigenden Innenraumsituationen durch Nutzerdaten abbilden zu können. Eine Abstimmung mit den in TP1 beschriebenen Personas stand dabei im Vordergrund.

Um Aktivitäten der Nutzer unmittelbar aus ihrer Körperhaltung interpretieren zu können, wurden im Innenraum zonenbasierte Interaktionsmodelle definiert (Sonnenblende, Rückspiegel, Lenkrad, usw.). Bei Annäherung der Gliedmaßen einer erfassten Person konnten damit direkte Manipulationen und zugehörige Handlungen nachvollzogen werden. Die Interaktionsmodelle wurden dabei durch geometrische Primitive beschrieben, die in das später gemeinschaftliche Fahrzeugkoordinatensystem approximativ um die jeweiligen Interaktionsbereiche gelegt wurden.

Aufbauend auf den ersten Ergebnissen der integrierten Erfassungskomponenten in TP3 wurden vorläufige Tests verschiedener Körperposenklassifikationen durchgeführt und ein Drehbuch als Grundlage für die geplante Datensammlung verschiedener Innenraumaktivitäten erstellt.

Aufgrund der andauernden Integrationsarbeiten der Erfassungskomponenten und der Erstellung des Trainingsframeworks in TP3 kam es in TP4 bis zuletzt zu deutlichen Verzögerungen in der

Datensammlung. Fortschritte in AP 4.1 und AP 4.2 konnten daher nur langsam erreicht werden, weswegen im dritten Projektjahr eine Laufzeitverlängerung beantragt wurde. Zwischenzeitlich wurden das Zonenmodell zur Erkennung von Fahrzeuginteraktionen verbessert und damit ermöglicht, auch sich annähernde Gliedmaßen erkennen zu können. Zum Beispiel konnte hierdurch nun ermittelt werden, wie groß der Körperteilabstand einer erfassten Person zu den Airbags war. Das Zonenmodell wurde in den Versuchsträger integriert und alle Schnittstellen in den Entwicklungsframeworks implementiert. Parallel wurden die Szenarien der Beispielanwendungen gemeinsam mit den verantwortlichen Projektpartnern konzipiert und diese schließlich in Datensammlungen zum Einlernen der Verhaltenserkennung ausgearbeitet. Neben Untersuchungen verschiedener Verfahren maschinellen Lernens wurden schließlich künstliche Neuronale Netze zum Verwerten der erfassten Körperposendaten aus TP3 und zur Klassifikation verschiedener Nutzeraktivitäten eingelernt. Die Klassifikation erfolgt dabei zunächst framebasiert für jedes Kameraeinzelbild. Über ein Zeitfenster werden die Klassifikationsergebnisse aber schließlich gemittelt und geglättet.

### 3.4 Wissenschaftlicher-technischer Stand zu Projektbeginn

Durch die Entwicklung einer Insassen- und Innenraumkontextanalyse entstehen neue Herausforderungen und Möglichkeiten zur Gestaltung von intuitiven MMI-Konzepten. Zwar kann auf ergonomische Erfahrungswerte und Grundsätze in der MMI-Entwicklung zurückgegriffen werden (z.B. European Statement of Principles 2006), doch erforderte gerade das Ziel einer insassenbedarfsoptimierten Informationsdarbietung infolge ihrer hohen Variabilität – etwa in Bezug auf Informationsgehalt, Kritikalität, Situation – die Entwicklung neuer Lösungsansätze.

Diese beziehen sich zum einen auf die Gestaltung der Fahrer-Fahrzeug-Interaktion auf Basis umfassender bedarfsorientierter Analysen des gesamten Innenraums. Dabei kam in InCarIn auch die Frage auf, wie viele Informationen ein Fahrer in einer konkreten Situation überhaupt verarbeiten kann, insbesondere, wenn diese von unterschiedlichen Informationsquellen (z.B. Car2X, Fahrerassistenzsysteme, Navigation) simultan oder in dichter Folge angeboten werden. Bisherige MMI-Konzepte wurden überwiegend pro System erarbeitet, d.h. vernetzende, systemübergreifende Ansätze finden bislang nur unzureichend Beachtung.

Durch eine Innenraumerfassung erschienen zu Projektbeginn Innovationen beispielsweise im Kontext von Anzeigeort, Anzeigemodalität, Anzeigezeitpunkt und Anzeigemedium wie auch Möglichkeiten zur Reduktion der Anzeigenvielfalt durch systemübergreifende Konzeptionalisierungen möglich.

Die Arbeiten von InCarIn sollten nicht bei Null ansetzen. Zuvor existierten bereits MMI-Konzepte, die sich entweder schon in Serie befanden oder unter Einbezug neuartiger Fahrzeugkomponenten bzw. Rückmeldekanäle noch Forschungsstatus hatten. Des Weiteren

wurden in nationalen sowie internationalen Forschungsprojekten zu Fahrerassistenzsystemen bereits einzelne, spezielle Aspekte zur MMI-Gestaltung untersucht. Die darin gewonnenen Erkenntnisse bildeten den Ausgangspunkt für die Gestaltung der InCarIn Fahrer-Fahrzeug-Interaktion.

Deutlich war jedoch, dass bei der damaligen MMI-Gestaltung im Fahrzeug lediglich der Fahrer und der Fahrkontext im Mittelpunkt standen. Weitere Insassen und deren Bedürfnisse wurden kaum oder gar nicht beachtet. Neben der Fahrer-Fahrzeug-Interaktion spielt bei der Insassen- und Innenraumkontextanalyse jedoch auch die Insassen-Fahrzeug-Interaktion eine wichtige Rolle. Erste Ansätze, wie beispielsweise der Mercedes Benz SPLITVIEW [1], zeigten damals die Möglichkeiten einer insassenbedarfsoptimierten Informationsdarbietung.

### 3.4.1 Verfahren und Systeme zur Posenschätzung

Bisherige Ansätze vor Projektbeginn untersuchten nur einzelne Aspekte der Personen, d.h. meist lokal eingegrenzte Bereiche einer Person. Im Vordergrund standen Gesicht und Kopf, Hände, Füße und umfassende Körperposen meist auf den Oberkörper beschränkt. Eine zum Projektbeginn vorliegende Übersicht ist in [6] zu finden.

Die Erfassung des Kopfes und des Gesichts war einer der bereits am intensivsten erforschten Bereiche bei der videobasierten Erfassung im Fahrzeuginnenraum. Im Fokus standen vor Projektbeginn besonders die Müdigkeitserkennung [7] [8] und Erkennung der Aufmerksamkeit des Fahrers [9]. Beide Aspekte lassen sich sowohl basierend auf der Kopfdrehung bzw. Kopforientierung [7] als auch basierend auf der Blickrichtung bzw. auf augenspezifischen Merkmalen [9] erkennen. Weitere Aspekte umfassten die Identifizierung zur Aktivierung personalisierter Einstellungen und der Emotionserkennung, um mehr über die Gemütsverfassung des Fahrers und damit seinen möglichen Reaktionen zu erfahren.

Die Erfassung der Hände diente unterschiedlichen Zwecken. Zum einen ließen sich Informationen über die Abgelenktheit und das Reaktionsvermögen erfassen. Sind beide Hände am Lenkrad, dann ist die Aufmerksamkeit wahrscheinlich höher als wenn mit einer Hand eine Interaktion mit dem Handy stattfindet. Zusätzlich konnte bereits über Gestenerkennung eine direkte Interaktion mit dem Boardsystem (Navigation, Radio) realisiert werden, wodurch ein Mehrwert entstand und individuell konfigurierbare Oberflächen möglich wurden. Da Gesten Bestandteil der zwischenmenschlichen Kommunikation sind, konnte durch die Erkennung kommunikativer Gesten zusätzlich Informationen über die Interaktion im Fahrzeug gewonnen werden [10] [11].

Die Erfassung der Füße zielte bisher auf die Erkennung des Bremsverhaltens ab [12]. Die Motivation dahinter war, dass ein System, das erkennt, wenn sich der Fuß in eine bremsbereite Position bewegt, ggf. bremsunterstützende Systeme entsprechend anpassen kann. Zudem konnte über die Zeit das Bremsverhalten gelernt werden.

Die Körperpose, als Beschreibung der Haltung des Fahrers, wurde bisher dazu genutzt, um Sicherheitssysteme anzupassen (z.B. Airbagauslösung [4]) oder um mehr über die Intention des Fahrers [5] bzw. den aktuellen Zustand zu erfassen (z.B. Müdigkeit oder Aufmerksamkeit). Die Körperposenerkennung bezog sich bei aktuellen Systemen zum Projektbeginn meist nur auf den Oberkörper ([2] [3] [5]) und ist in der Regel auf den Fahrer beschränkt. Zur Bestimmung der Körperpose konnte jedoch prinzipiell auf bekannte Verfahren zurückgegriffen werden, wobei jedoch unbedingt die besonderen Umgebungsbedingungen im Auto berücksichtigt werden mussten. Eine allgemeine Übersicht über verschiedene Ansätze war bereits in [13] zu finden.

### 3.4.2 Verfahren und Systeme zur Fahrer- und Insassenmodellierung

Im Rahmen der Fahrermodellierung konnten zu Projektbeginn drei Hauptströmungen unterschieden werden. Diese Hauptströmungen waren Fahrermodellierung im engeren Sinne, d.h. Modellierung der Fahrzeugführung (vor allem Längs- und Querführung), Fahrerzustandsmodellierung bzw. -erkennung und Intentionserkennung. Die Ansätze konzentrierten sich dabei auf den Fahrer und berücksichtigten meist keine weiteren Fahrzeuginsassen, so dass ein allgemeines Insassenmodell bis dato nicht vorlag.

#### **Fahrermodellierung i.e.S.**

Ziel dieser Modellierungsrichtung ist die Modellierung des menschlichen Fahrverhaltens, wobei der Schwerpunkt traditionell bei der Längs- und Querführung liegt. Regelungstechnische [14] und optimal theoretische Ansätze [15] [16] postulieren hierbei eine Solltrajektorie und fassen den Fahrer als Regler bzw. Optimierer auf, der für eine möglichst geringe Abweichung von der Solltrajektorie sorgt. Weitere Ansätze nutzen Verfahren des maschinellen Lernens [17] [18] um aus empirisch ermittelten Fahrdaten ein Verhaltensmodell zu entwickeln. Darüber hinaus gab es auch Bestrebungen, mittels neurophysiologischer [19] oder kognitiver Modellierung [20] das menschliche Fahrverhalten nachzubilden. Der Vorteil dieser Methoden liegt in der expliziten Berücksichtigung menschlicher Beschränkungen.

#### **Fahrerzustandserkennung**

Im Rahmen der Fahrerzustandsmodellierung wird versucht, aus messbaren Parametern wie z.B. Lidschlussfrequenz oder Lenkwinkeländerung auf hypothetische Zustände des Fahrers, beispielsweise Müdigkeit oder Ablenkung, zu schließen [21] [22] [23]. Zu Projektbeginn war der aktuelle Stand der Technik hierbei Verfahren des maschinellen Lernens einzusetzen, die eine Klassifikation des Fahrerzustands in diskrete Klassen (müde vs. wach, abgelenkt vs. aufmerksam) erlaubten.

## **Verfahren und Systeme zur Intentionserkennung**

Intentionserkennung im Fahrzeug erfolgte bis dato vorwiegend im Kontext Manövererkennung. Ziel war dabei vorherzusagen, welches Fahrmanöver (z.B. Abbiegen, Überholen, Spurwechsel) der Fahrer beabsichtigte, um diese Intention für Assistenzsysteme nutzbar zu machen. Die Modelle kombinierten vor allem Daten aus der indirekten Fahrerbeobachtung, bei der keine spezifisch auf den Fahrer gerichteten Sensoren zum Einsatz kamen. So konnte beispielsweise über fahrzeugseitige Parameter wie Lenkwinkel, Spurposition oder Betätigung des Blinkers auf eine Abbiegeintention geschlossen werden.

## **Verfahren zum inkrementellen Lernen von Gewohnheiten**

Zu Projektbeginn existierten bereits Algorithmen, welche fähig waren, um das Lernverfahren sukzessive mit neuen Daten zu befüllen, siehe beispielsweise [24]. Auch hat sich bereits gezeigt, dass die Bayes'schen Verfahren dafür besonders geeignet sind, siehe [25]. Um die Parameter der Modelle inkrementell zu lernen, bietet sich der Expectation Maximization (EM)-Algorithmus an, welcher in verschiedenen Varianten umgesetzt wurde [26]. Auch wurden Bayes'sche Verfahren angewendet, um das Verhalten eines Fahrers online zu lernen [27]. Bis dato wurden diese Verfahren jedoch nur auf Situationen außerhalb des Fahrzeugs, also z.B. die aktuelle Verkehrssituation angewendet, nicht jedoch auf das Verhalten des Fahrers selbst. InCarIn hatte u.a. zum Ziel, dies beizutragen.

## **3.5 Zusammenarbeit mit anderen Stellen**

Das Projekt InCarIn hat viele Kontakte in Forschung und Industrie eröffnet. Dies spiegelte sich unter anderem in den Teilnehmerlisten der Halbzeitpräsentation und Abschlusspräsentation wieder. Auf Konferenzen wurde das Thema einer breiten Fachöffentlichkeit mehrfach präsentiert. Fraunhofer nutzte die Zwischen- und Endergebnisse auch um neue Förderprojekte mit neuen Partnern und Partnern aus dem Projekt anzustoßen. Indirekt floss die Erfahrung aus dem Projekt zudem in Kooperationen mit Industriepartnern ein.

## **4 Ergebnisbericht**

### **4.1 Erzielte Ergebnisse**

#### **1.1.1 Kamerasetup für die Gesamt Innenraumanalyse**

Um die Algorithmen für die Gesamtinnenraumerfassung entwickeln zu können musste zuerst festgelegt werden, welche Kameras an welchen Positionen verbaut werden können. Hier fand im Projekt eine ausführliche Evaluation verschiedener Sensorsysteme statt. Hauptziel dieses Kamerasystems war die Erfassung der Körperpose aller Insassen in 3D. Entsprechend muss an einem Punkt in der Prozesskette aus den Kamerainformationen 3D Daten gewonnen werden. Je

nach verwendeter Technologie entsteht hierbei größerer oder kleinerer Rechenaufwand. Die Evaluation der Sensoren umfasste diesen Rechenaufwand, die Eignung für wechselnde Beleuchtungsbedingungen wie auch Tag und Nachtfahrten, Verbaupositionen und gegenseitige Störung der Systeme. Im Folgenden werden die verschiedenen Sensortypen vorgestellt, die evaluierten Verbaupositionen und das finale Kamera Setup.

### **1.1.1.1 Evaluation unterschiedlicher Kamerasysteme**

#### **RGB-Farbkameras**

Ein Vorteil von Farbkameras ist die große Auswahl verfügbarer Modelle. Sie sind kostengünstig zu erwerben und können im Verbund ebenfalls dazu eingesetzt werden, Umgebungen aus unterschiedlichen Perspektiven dreidimensional zu rekonstruieren. Der große Nachteil dieser Technik ist aber, dass zwangsweise Licht im sichtbaren Spektrum vorhanden sein muss, um Farbinformation zu erhalten. Dies ist gerade bei Nachtfahrten im Fahrzeug nicht gegeben da hier eine zusätzliche Lichtquelle im Innenraum nötig wäre, die den Fahrer stört. Farbkameras sind daher vom Einsatz ausgeschlossen.

#### **Infrarotkameras**

In einem dunklen Innenraum bieten sich Infrarotkameras zur Insassenbeobachtung an. Mit einer aktiven Infrarotbeleuchtung, lässt sich der zu beobachtende Bereich um die jeweilige Person ausleuchten, ohne sie selbst zu stören oder im sichtbaren Lichtspektrum für Störlicht zu sorgen. Die Kamera kann damit im nahen Infrarotspektrum eine ausreichend ausgeleuchtete Aufnahme liefern, auf der eine Insassenbeobachtung auch unter schwachen Lichtverhältnissen möglich wird. Gegenlicht kann indes noch zu einer Überbelichtung führen, womit bei einer adaptiv-aktiven Infrarotbeleuchtung zu einem gewissen Grad entgegengewirkt werden kann. Prinzipiell bieten sich Infrarotkameras daher zur Innenraumbeobachtung an und wurden in der Evaluation unter anderem in Form von Stereokameras berücksichtigt.

#### **Stereokameras**

Stereokamerasysteme bestehen aus zwei dicht nebeneinander angebrachten RGB-, Grauwert-, oder Infrarotkameras. Der Unterschied in der Perspektive auf dieselbe beobachtete Szene ermöglicht eine dreidimensionale Rekonstruktion derselben, indem der Pixelversatz in den jeweils zeitsynchronen Bildpaaren geometrisch zur Berechnung der Distanz des entsprechenden Objektpunktes ausgenutzt wird.

Stereokamerasysteme können mit marktüblichen Kameras aufgebaut werden, was ihnen eine große Flexibilität in ihrem geringen Formfaktor und den verfügbaren und einsetzbaren Optiken gibt. Damit heben sich Stereosysteme insbesondere gegenüber anderen RGB-D-Sensoren ab, da sie durch Weitwinkelobjektive einen deutlich größeren Erfassungsbereich bei gleichzeitig geringerer Verbaudistanz zulassen – ein Aspekt, der im eingeschränkten Fahrzeuginnenraum nicht zu vernachlässigen ist. Ebenfalls entfällt durch die prinzipiell nicht vorhandene

Aktivbeleuchtung die Gefahr einer Interferenz zu anderen verbauten Systemen (bspw. im nahen Infrarot). Dieser Aspekt ist jedoch insbesondere im PKW-Innenraum durch die geschilderte Notwendigkeit einer ergänzenden Beleuchtung in dunklen Situationen zu vernachlässigen. Die flexible Zusammensetzung macht indes eine Berechnung der Tiefeninformation in Software nötig. Um ein dichtes Disparitätsbild in hoher Qualität zu errechnen ist zum Zeitpunkt der Evaluation mit hohem Rechenaufwand eine maximale Framerate von ca. 10-15 Frames pro Sekunde möglich. Dadurch wird die Echtzeitfähigkeit stark eingeschränkt sowie der Verbau entsprechender Rechenleistung im Fahrzeug erforderlich gemacht und somit der Vorteil des kleinen Formfaktors des Sensorkopfes relativiert. Darüber hinaus weisen Stereosysteme eine hohe Abhängigkeit von ausreichend texturierten Oberflächen auf, um einen Pixelversatz in den beiden Kamerabildern nachvollziehen zu können – bei unifarbenen Oberflächen scheitern die gängigen Stereogeometrieverfahren und liefern unzuverlässige Tiefendaten. Dieser Nachteil wird besonders bei entsprechender Kleidung, Hautoberflächen oder einem unifarbenen Interieur im Fahrzeuginnenraum deutlich. Es wurde eine selbst gebaute Stereokamera mit extremen Weitwinkelobjektiven an den unterschiedlichen Verbauplätzen und mit unterschiedlichen Algorithmen zur Disparitätsberechnung in die Evaluation mit einbezogen.

### **Kameraverbund**

Ergänzend zu RGB-D-Kameras und Stereokameras besteht die Möglichkeit monokulare Sensoren im Verbund im Fahrzeuginnenraum zu verbauen. Ähnlich zur Stereogeometrie werden verschiedene Ansichten aus unterschiedlichen Perspektiven genutzt, um die beobachtete Umgebung dreidimensional zu rekonstruieren. Durch den größeren Verbauplatz unterscheiden sich die Perspektiven einzelner Kameraansichten allerdings stark. Eine direkte Korrespondenzsuche zur Bestimmung der Tiefe wird dadurch erschwert. Ein Standardverfahren zur 3D-Rekonstruktion in Kameraverbänden verwendet deshalb einen Triangulationsansatz auf einer Vorder-/Hintergrund Segmentierung jedes Kamerabildes im Verbund. Eine solche Segmentierung jedes Bildes ist aufwändig und bei großen Beleuchtungsvarianzen im Auto nur schwer robust zu erreichen. Zusätzlich muss für eine artefaktfreie Rekonstruktion jeder Punkt des Autoinnenraums zu jedem Zeitpunkt in mindestens zwei Kameraansichten sichtbar sein. Im Autoinnenraum wird es aber leicht zu Verdeckungen kommen, wodurch diese Voraussetzung auch mit einer großen Kamerazahl nur schwer zu erreichen ist. Zusätzlich steigt durch eine hohe Kameraanzahl der Rechenaufwand für die Rekonstruktion enorm an. Auf Grund dieser Probleme ist diese Lösung unattraktiv für den Einsatz im Fahrzeuginnenraum und wurde nicht weiter betrachtet.

### **RGB-D-Kameras**

RGB-D-Kameras liefern im Gegensatz zu einer regulären Farbkamera ein Tiefenbild, indem der Pixelwert pro Bildpunkt nicht dem Farbwert, sondern dem Abstand des entsprechenden Objektpunkts von der Kamera entspricht. Diese Tiefeninformation müsste beim Einsatz von sonst

üblichen Farb- oder Infrarotkameras durch Stereogeometrie oder einen Betrieb im Kameraverbund erst rechenintensiv rekonstruiert werden. Der Vorteil von RGB-D-Kameras liegt durch den reduzierten Aufwand für eine dreidimensionale Rekonstruktion damit deutlich auf der Hand. Es gibt allerdings nur ein beschränktes Angebot dieser Kameras die zwar oft preislich günstig sind, aber Einschränkungen gerade für den Fahrzeuginnenraum bieten. Das Objektiv dieser Kameras ist zumeist fest verbaut und kann nicht verändert werden. Für den Fahrzeuginnenraum und die Erfassung der ganzen Person sind weitwinkelige Kameras notwendig. Dies schränkt das Angebot an verwendbaren Tiefenkameras weiter ein. Zusätzlich haben alle diese Kameras sowohl eine minimal als auch eine Maximaldistanz. Nur in diesem Bereich werden Tiefendaten erzeugt. Gerade die Minimaldistanz stellt im Fahrzeuginnenraum oft eine Herausforderung für die Sensorplatzierung dar.

Es gibt verschiedene Techniken mit denen die Tiefendaten gewonnen werden. Die zwei bekanntesten und preislich attraktivsten für den Projektraum sind *Structured Light* und *Time of Flight*. Beiden Techniken ist gemein, dass aktiv Licht ausgestrahlt wird um die Szene dreidimensional zu erfassen. Hierdurch ist ein Betrieb auch nachts problemlos möglich, allerdings ergeben sich dadurch Schwierigkeiten bei Sonneneinstrahlung, da das vom Sensor erzeugte Licht überstrahlt werden kann. Die Robustheit gegenüber Sonnenlicht unterscheidet sich stark von Kamera zu Kamera und wurde im Projekt untersucht. Im Folgenden werden die beiden untersuchten Erfassungsprinzipien erklärt:

*Structured Light-Kameras* basiert auf dem Ausstrahlen eines Musters (bspw. eines Gitters) in die zu beobachtende und dreidimensional zu rekonstruierende Szene. Die Darstellung und Reflektion des Musters wird schließlich dazu benutzt, um Tiefeninformation zu berechnen und die Oberflächen in der bestrahlten Umgebung nachzuvollziehen. In der Regel wird das Muster im nahen Infrarotbereich ausgestrahlt um den Nutzer nicht zu stören. Durch die Projektion des Musters kommt es beim Betrieb mehrerer Sensoren im gleichen Bereich allerdings oft zu Interferenzen die zu erhöhtem Rauschen oder in Teilbereichen des Bildes zum vollständigen Ausfall von Information führt.

*Time of Flight-Kameras* (TOF), oder auch Photomischdetektor-Kameras (PMD) genannt, geben üblicherweise Lichtimpulse (bspw. in Infrarot) an die Umgebung ab und messen pro Bildpunkt die Laufzeit des Lichts, bis es von einem Objekt reflektiert wieder auf dem Sensor auftrifft. Die Lichtimpulse sind moduliert was den Einfluss störender Hintergrundbeleuchtung in Grenzen unterdrückt. Um mehrere Sensoren zeitgleich im Verbund einzusetzen, haben verschiedene Hersteller unterschiedliche Verfahren entwickelt, die Lichtimpulse zu differenzieren. Wird keine aktive Synchronisierung verwendet kann es zu Störungen zwischen Sensoren im gleichen Arbeitsbereich kommen. Weil die Erhebung der Tiefeninformation in Hardware geschieht, sind TOF-Kameras mit Frameraten von bis zu 160 Frames pro Sekunde ausgestattet.

### 1.1.1.2 Evaluation der Kameratechniken

Um eine erste Einschätzung der die prinzipielle Eignung der unterschiedlichen Sensoren zu bekommen wurden außerhalb des Fahrzeugs mit den unterschiedlichen Sensoren im Innenraum und draußen bei Sonnenlicht Bilder aufgenommen und deren Qualität subjektiv bewertet. Tabelle 1 gibt eine Übersicht über die Ergebnisse.

**Tabelle 1: Die im Projekt getesteten Tiefensensoren mit ihren Eigenschaften**

	Kameratyp	Öffnungswinkel	Erfassungsdistanz	Auflösung	Sonnenlichteinfluss
Stereo Kamera	stereo	variabel	variabel	variabel	gering
Asus Xtion	structured	58x45	80-350cm	320x240	stark
Kinect One	tof	70x60	50-450cm	512x424	gering
Intel Realsense F200	structured	73x59	2-120cm	640x480	stark
PMD Camboard Nano	tof	90x68	-	160x120	Sehr stark
Softkinetic DS325	tof	74x58	15-100cm	320x240	stark
Softkinetic DS311	tof	57,3x42	15-100cm	160x120	gering

Anhand dieser Ergebnisse wurde beschlossen alle weiteren Tests nur mit der Kinect One und der Stereokamera durchzuführen. Die Kinect One zeigt im Vergleich der RGB-D Sensoren in fast allen Punkten das beste Verhalten. Sie ist recht robust gegenüber Sonnenlicht, zeigt wenig Tiefenrauschen, hat die beste Auflösung und hat einen relativ weiten Öffnungswinkel. Die Stereokamera wird weiter betrachtet, da Sie die größte Flexibilität in jeder Hinsicht bietet und auch im späteren kommerziellen Einsatz voraussichtlich die billigste und am flexibelsten zu verbauende Lösung wäre.

### 1.1.1.3 Evaluation der Sensorpositionen

Mit den zwei im letzten Schritt selektierten Sensoren (Kinect One und Stereo Kamera) wurden weitere Tests im Fahrzeug durchgeführt, um mögliche Verbauorte zu bestimmen und die Anzahl notwendiger Kameras festzulegen. Abbildung 4.1.1 stellt die mit der Volkswagen AG festgelegten, möglichen Verbauorte für Kameras zur Innenraumerfassung dar.

Die Ergebnisse sind in Tabelle 2 aufgeführt. In unseren Experimenten hat sich gezeigt, dass der Blickwinkel aus dem Infotainment Bereich und dem Instrumentencluster auf Grund von Verdeckungen unabhängig von der Sensorik ungeeignet ist. Eine möglichst frontale Sicht von

## Schlussbericht

Oben ist am besten für die Körperposenschätzung geeignet, da es hier zu den geringsten Verdeckungen kommt. Für Fahrer und Beifahrer eignen sich daher am besten die Sonnenblenden und der Rückspiegel. Diese Positionen sind andererseits auch sehr nah bei den Insassen was die Erfassung gerade mit RGB-D Sensoren auf Grund des nötigen Öffnungswinkels und Mindestabstands erschwert. Die A-Säulen führen schon zu mehr Verdeckungen. Wenn allerdings von der Beifahrer A-Säule der Fahrer und von der Fahrer A-Säule der Beifahrer beobachtet wird ergibt sich ein großer Abstand zur beobachteten Person, was sehr günstig für den Verbau von RGB-D Sensoren ist. Die B-Säulen und der Dachhimmel eignen sich nur für die Rückbank, da der Torso der Insassen im Frontbereich aus dieser Perspektive nicht erfassbar ist. Der Dachhimmel bietet hier den Vorteil, dass die ganze Rückbank mit einer Kamera abgedeckt werden kann während man aus der Sicht der B-Säulen nur einen der Insassen erfassen kann.

**Tabelle 2: Eignung der Sensoren für die verschiedenen Positionen im Fahrzeug**

	Stereo	Kinect One
A-Säulen	Fahrer, Beifahrer	Fahrer, Beifahrer
Infotainment Bereich	-	-
Instrumentencluster	-	-
Dachhimmel	Rückbank	Rückbank
Sonnenblende	Fahrer, Beifahrer	-
Rückspiegel	Fahrer, Beifahrer	-
B-Säule	Rückbank links/rechts	-



**Abbildung 4.1.1** Mögliche Verbau-Orte von Kamerasensorik im Fahrzeuginnenraum

### 1.1.1.4 Finales Kamerasetup

Als finales Kamerasetup für den gesamten Innenraum sind sowohl Stereokameras als auch Kinects geeignet. In beiden Fällen sind 3 Sensoren notwendig. Im Kinect Setup würden die Sensoren an den A-Säulen verbaut, wobei die Kinect auf der Beifahrerseite den Fahrer beobachtet und umgekehrt. Der dritte Sensor wird im Dachhimmel für die Rückbank verbaut. Im Stereo Setup können die Sensoren für Fahrer und Beifahrer am Innenspiegel oder in den Sonnenblenden verbaut werden. Dies liefert einen günstigeren Blickwinkel. Die Kamera für die Rückbank würde genauso im Dachhimmel verbaut.

Für das Stereosetup wäre zusätzlich noch Infrarotbeleuchtungen notwendig, die in der Evaluation nicht mit berücksichtigt wurden. Außerdem wäre zusätzliche Rechenleistung alleine für die Berechnung der Disparität Bilder notwendig. Aus diesen Gründen wurde entschieden im Projekt den Aufbau mit drei Kinects zu verwenden, da hier weder weitere Beleuchtung noch erhöhter Rechenaufwand entsteht und der Gesamtaufwand dadurch sinkt.

### **1.1.2 Algorithmus zur Körperposenerfassung der Insassen**

Ein wesentlicher Teil der Arbeiten des Fraunhofer IOSB war die Entwicklung eines Verfahrens zur Körperposenerfassung von Insassen im Fahrzeug. Die Körperpose sollte im Projekt als abstrahiertes Merkmal der Insassen für die weitere Verarbeitung zum Beispiel zur Erkennung von Nebentätigkeiten genutzt werden.

Das entwickelte Verfahren nutzt Tiefensensoren um ein 3D Skelett der sichtbaren Körperteile der Insassen zu bestimmen. Ein Tiefensensor liefert für jeden Pixel den Abstand zur Szene an Stelle der Farbinformation. Weiß man zusätzlich noch die Brennweite des Objektivs und die Sensorgröße, so kann aus dem Tiefenbild die 3D Szene rekonstruiert werden.

Das im Projekt entwickelte Verfahren läuft in 4 Schritten ab. Zuerst werden die Rohdaten vorverarbeitet, dann werden die Insassen im Tiefenbild segmentiert, um unabhängig vom Fahrzeuginnenraum zu werden. Jeder Punkt dieser Segmentierung wird dann mit Hilfe von maschinellen Lernverfahren einem Körperteil zugeordnet und im letzten Schritt wird über Heuristiken aus diesen geschätzten Körperteilen die Körperpose abgeleitet.

#### **1.1.2.1 Vorverarbeitung der Daten**

Tiefenbilder können abhängig vom eingesetzten Sensor und der erfassten Szene starkes Rauschen aufweisen. Dieses stört die folgende Verarbeitung stark. Die Daten werden deshalb mit unterschiedlichen Filtern vorverarbeitet. Zuerst werden alle Punkte des Tiefenbildes ausgefiltert bei denen der entsprechende Pixel im Luminanzbild unterhalb eines klein gewählten Schwellwerts liegt. Diese Bereiche entsprechen Punkten der Szene die schlecht belichtet sind. Dies kann mehrere Gründe haben. Entweder das Material hat schlechte Reflektionseigenschaften, weil es zum Beispiel matt schwarz ist, oder der Punkt ist weit weg, wie es zum Beispiel bei der Sicht aus den Fenstern des Fahrzeugs der Fall ist. Gerade bei schlecht reflektierenden Materialien weisen viele Tiefensensoren ein stark erhöhtes Rauschverhalten auf durch das Ausfiltern dieser Bereiche werden Störungen effektiv reduziert. Um das Rauschen im übrigen Bild weiter zu reduzieren wird danach sowohl ein Medianfilter als auch ein Gaussfilter eingesetzt.

#### **1.1.2.2 Segmentierung**

Nach der Vorverarbeitung trennt der Algorithmus die Insassen vom Innenraum. Hierfür wird ein 3D Hintergrundmodell eingesetzt. Dieses Hintergrundmodell wird durch ein Voxelgrid repräsentiert. Ein Voxelgrid unterteilt den Raum in viele kleine Würfel mit fester Kantenlänge. Im

Falle des Hintergrundmodells ist in jedem Voxel gespeichert, ob der Raumbereich Freiraum ist oder zu einer Oberfläche der Innenausstattung gehört. Um dieses Modell zu erzeugen wird mit den Sensoren das leere Fahrzeug aufgenommen. In mehreren aufeinander folgenden Bildern wird dann in einem Akkumulator Voxelgrid  $A_i$  ein Zähler erhöht, falls ein Punkt der Punktwolke in den Voxel fällt. Daraus wird mit folgender Formel dann das Hintergrundmodell erzeugt:

$$M_i = \begin{cases} \text{Innenraum}, & \frac{A_i}{n} > \theta \\ \text{Freiraum}, & \text{sonst} \end{cases}$$

Wobei  $M$  das Hintergrundmodell ist,  $i$  ist der Index eines Voxels,  $n$  ist die Zahl der akkumulierten Bilder und  $\theta$  ist ein Schwellwert zwischen 0 und 1. Durch diesen Schwellwert wird der Einfluss des übrigen Rauschens auf das Modell festgelegt. Je höher das Rauschen desto niedriger muss der Schwellwert gewählt werden um nicht Teile der Innenausstattung aus dem Modell zu verlieren. Wird der Schwellwert aber zu klein gewählt, dann werden Teile des Freiraums als belegt identifiziert.

Das gelernte Hintergrundmodell kann danach effizient genutzt werden um die Insassen zu segmentieren. Hierfür wird für jeden Punkt der Eingabe Punktwolke bestimmt in welchen Voxel der Punkt fällt. Ist dieser Voxel als Freiraum gekennzeichnet, so gehört der Punkt zu einem Insassen.

### 1.1.2.3 Klassifikation von Körperteilen

Die Körperteilklassifikation setzt die Segmentierung voraus. Jeder Punkt der Segmentierung wird einzeln anhand seiner Umgebung klassifiziert und einem Körperteil zugeordnet. Durch dieses Vorgehen ist das Verfahren trivial parallelisierbar und kann sowohl auf Mehrkernprozessoren als auch auf Grafikkarten effizient implementiert werden. Zur Klassifikation werden Random Decision Forests verwendet. Der Random Decision Forest besteht aus mehreren binären Random Decision Trees. Die inneren Knoten des Baums beinhalten einfache Klassifikatoren, die den zu klassifizierenden Testpunkt entweder in den Linken oder Rechten Unterbaum leiten. In jedem Blattknoten des Baums ist eine Wahrscheinlichkeitsverteilung gespeichert, die beschreibt, wie wahrscheinlich der Punkt zu welchem Körperteil gehört. Während des Trainings eines Random Decision Trees werden sowohl die Parameter der einfachen Klassifikatoren als auch die Wahrscheinlichkeitsverteilungen der Blattknoten gelernt. Ziel des Lernverfahrens ist es die Trainingsbeispiele möglichst gut zu trennen und somit möglichst reine Verteilungen in den Blattknoten zu generieren.

Für die Klassifikation der Körperteile wird ein schwacher Klassifikator in den inneren Knoten verwendet, der auf dem Vergleich von Tiefenwerten in der Umgebung des Testpunktes basiert. Der Klassifikator ist wie folgt definiert:

$$S_i = \{x | x \in S \text{ und } d(x + \frac{o_1}{d(x)}) - d(x + \frac{o_2}{d(x)}) < t\}$$

$$S_r = S/S_l$$

$S$  ist die Menge an Testpunkten die vom jeweiligen Knoten auf den linken  $S_l$  und rechten  $S_r$  Teilbaum verteilt werden. Jeder Testpunkt besteht aus seiner Position  $x$  im Tiefenbild und der entsprechenden Grundwahrheit in der Trainingsphase.  $d(x)$  ist der Tiefenwert an Position  $x$ ,  $o_1$  und  $o_2$  sind gelernte Offsets in Bildkoordinaten und  $t$  ist ein Schwellwert der auch in der Trainingsphase gelernt wird.  $d(x)$  wird auf einen großen Strafwert gesetzt, falls der Punkt außerhalb der Segmentierung liegt. Die Wahrscheinlichkeitsverteilung in den Blattknoten wird in Form von Histogrammen aus der Grundwahrheit der Trainingspunkte generiert, die einen Blattknoten erreichen. Um das Ergebnis des gesamten Random Decision Forests für einen Testpunkt zu erhalten wird der Durchschnitt der Ergebnisse jedes Decision Trees gebildet.

Die Körperteilklassifikation liefert als Ergebnis für jeden Punkt der Segmentierung eine Wahrscheinlichkeitsverteilung über alle Körperteile, die beschreibt wie wahrscheinlich ein Punkt zu einem Körperteil gehört.

#### 1.1.2.4 Bestimmung der Körperpose

Um aus der Körperteilklassifikation die Körperpose abzuleiten müssen die einzeln klassifizierten Punkte fusioniert werden. Die einfachste Möglichkeit hierfür ist die Bestimmung des Mittels aus allen Punkten der Segmentierung gewichtet mit den Wahrscheinlichkeiten des jeweiligen Körperteils. Dies führt aber leider nicht zu einem robusten Ergebnis da es bei der Klassifikation immer zu Fehldetektionen kommt, die das Ergebnis bei diesem Vorgehen stark verschlechtern.

Stattdessen wird ein Clustering der einzeln klassifizierten Punkte genutzt. Hierfür wird zuerst für jeden Punkt der Segmentierung das wahrscheinlichste Körperteil bestimmt und der Rest der Wahrscheinlichkeitsverteilung verworfen. Danach wird Connected Component Clustering angewendet um Punkte der Segmentierung zu gruppieren, die sowohl im Tiefenbild benachbart sind, dem gleichen Körperteil zugeordnet sind und einen kleinen Euklidischen Abstand haben:

$$C_i = \{p, q | l(p) = l(q) \text{ und } ||p - q|| < t\}$$

$C_i$  ist der Cluster mit Index  $i$ ,  $p$  und  $q$  sind benachbarte Punkte,  $l(p)$  ist das Label von  $p$  und  $t$  ist der maximale Abstand den zwei Punkte im gleichen Cluster haben können. Für jedes dieser Cluster wird nachfolgend sowohl der Zentroid  $J_i$  als auch ein Gewicht  $W_i$  bestimmt:

$$W_i = \sum_{p \in C_i} w(p)$$

$$J_i = \frac{1}{W_i} \sum_{p \in C_i} w(p)p$$

$w(p)$  ist die Wahrscheinlichkeit, dass Punkt  $p$  zum entsprechenden Körperteil gehört.

Jedes Paar aus Zentroid und Gewicht stellt eine Hypothese eines Körperteils dar. Häufig gibt es aufgrund von Fehldetektionen und Rauschen mehrere Hypothesen jedes Körperteils. Das Endergebnis wird deshalb dadurch erzeugt, dass für jedes Körperteil der Zentroid mit dem

höchsten Gewicht genutzt wird. Liegt auch das größte Gewicht unterhalb eines Schwellwertes so wird dieses Körperteil im aktuellen Bild als nicht sichtbar eingestuft.

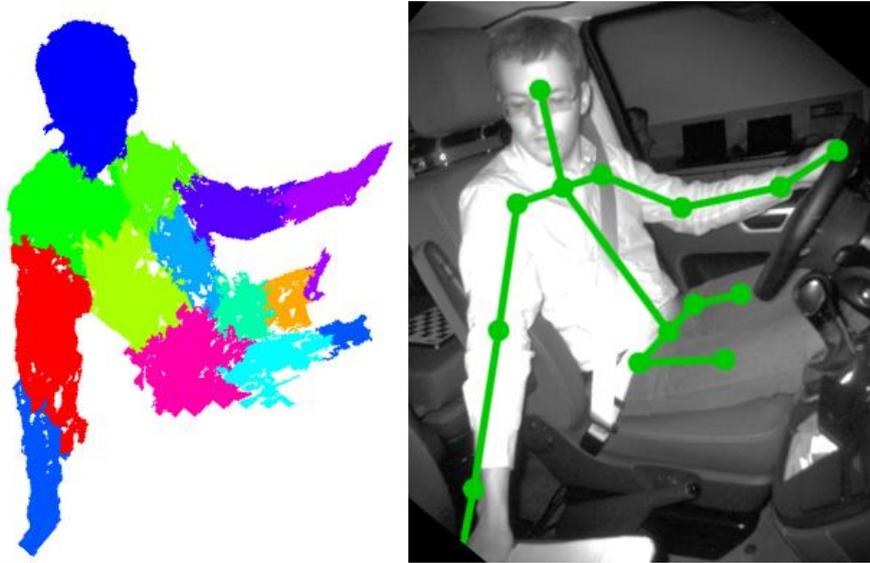


Abbildung 2: Links: Ergebnis der Körperteilklassifikation. Rechts: Die Abgeleitete Körperpose

#### 1.1.2.5 Evaluation

Im Rahmen des Projektes wurden keine Daten aus dem Fahrzeug mit der Position der Körperteile annotiert. Es liegt deshalb keine Grundwahrheit vor gegen die quantitativ evaluiert werden könnte. Eine manuelle Annotation der Videodaten wäre sehr zeitaufwändig gewesen. Es wurden verschiedene Möglichkeiten untersucht um mit einem anderen System automatisch eine genaue Annotation zu erzeugen. Hierbei wurde ein markerbasiertes Trackingsystem von Vicon in Betracht gezogen. Es ergeben sich mit diesem System aber mehrere Probleme. Es müssen reflektierende Marker an den Probanden angebracht werden, die sowohl das Tiefenbild stören, als auch im Infrarotbild sichtbar sind. Des Weiteren müssen auch zusätzliche große Kameras verbaut werden, was im Fahrzeug nicht möglich war. Als zweite Möglichkeit wurde ein System mit Inertialsensoren (Myomotion von Noraxon) untersucht. Nachteil dieser Systeme ist, dass sie durch Sensordrift nach einiger Zeit fehlerhafte Daten liefern. Diese Systeme sind prinzipbedingt auch nicht in der Lage die Position der Körperteile im Raum zu bestimmen. Es ist nur möglich die relative Bewegung zwischen den einzelnen Gelenken festzustellen. Dadurch wäre die Annotation mit diesem System aber auch nur eingeschränkt brauchbar und es wurde deshalb auch darauf verzichtet.

Die Güte der Körperposenschätzung wurde deshalb im Projektverlauf qualitativ beurteilt durch Visualisierung der Ergebnisse. Abbildung 3 zeigt einige Beispielergebnisse. Bei guter Sichtbarkeit wird das Skelett meistens korrekt erkannt. Es kommt allerdings zu Schwierigkeiten, wenn Verdeckungen auftreten. In diesem Fall versucht der Algorithmus über Schwellwerte zu erkennen, welche Körperteile verdeckt sind. Dies kann aber auch fehlschlagen, wie im letzten Beispiel Bildpaar zu sehen.

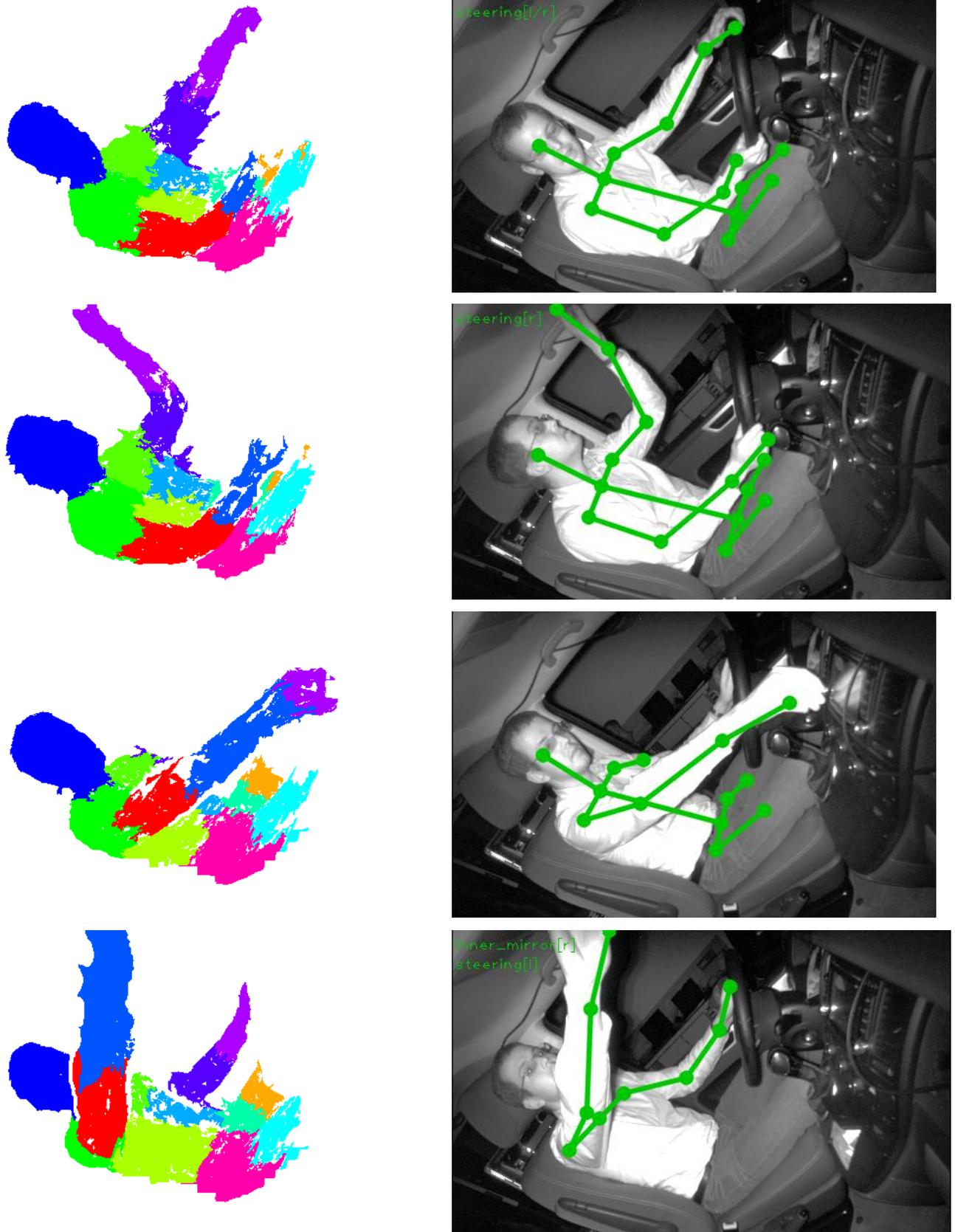


Abbildung 3: Beispielergebnisse der Körperposenschätzung. Links: Körperteilklassifikation. Rechts: Abgeleitete Körperpose

### 1.1.3 Trainieren der Modelle zur Körperposenerfassung

Der Algorithmus zur Posenschätzung verwendet Random Decision Forests. Es müssen für die Funktionalität des Algorithmus Modelle für jede Kameraposition eingelernt werden. Für das Lernen sind verschiedene Funktionen notwendig. Im Folgenden wird zuerst der Trainingsalgorithmus beschrieben, dann wird auf verschiedene Methoden eingegangen mit denen im Laufe des Projektes Trainingsdaten gesammelt wurden und anschließend wird die Software beschrieben mit der diese Trainingsdatenmenge erfolgreich verarbeitet werden konnte.

#### 1.1.3.1 Trainingsverfahren für Decision Forests

Das Training von Random Decision Forests für die Klassifikation setzt annotierte Trainingsdaten voraus. Diese Trainingsdaten bestehen für die Körperposenschätzung aus segmentierten Tiefenbildern und der Annotation des Körperteils für jeden Pixel der Segmentierung. Der Algorithmus bestimmt das Label für jeden Pixel einzeln. Entsprechend ist ein Trainingsbeispiel auch nur ein Pixel einer Segmentierung mit seinem jeweiligen Label. Ein Tiefenbild liefert also bereits viele Trainingsbeispiele.

Das Trainingsverfahren selbst baut jeden Baum des Decision Forests unabhängig auf. Für jeden binären Baum wird bei der Wurzel mit allen Trainingsdaten gestartet. Es werden dann eine Menge an zufällig generierten Klassifikatoren (siehe 1.1.2.3) erzeugt. Hierbei werden die Offsets und der Schwellwert innerhalb festgelegter Grenzen zufällig gewählt. Jeder Klassifikator wird dann mit den annotierten Trainingsbeispielen auf seine Güte untersucht indem alle Trainingsdaten mit jedem Klassifikator ausgewertet werden. Dadurch werden die Daten durch jeden Detektor in zwei Teile für die jeweiligen Teilbäume aufgeteilt. Für jeden Detektor wird dann auf dieser Aufteilung eine Gütefunktion berechnet, um den besten Detektor zu finden. Hierfür wird der Information Gain über die Verteilung der Trainingsbeispiele bestimmt:

$$I = H(S) - \sum_{c \in \{l,r\}} \frac{|S_c|}{|S|} H(S_c)$$

$S$  ist die Menge an Testdaten für den Knoten,  $H(S)$  ist die Shannon Entropie von  $S$ ,  $S_c$  stellt die Aufteilung der Trainingsbeispiele durch den Detektor für den linken- und rechten Teilbaum dar.

Der Detektor mit dem größten Information Gain wird dann für diesen Knoten genutzt und das Training für beide Unterbäume mit den aufgeteilten Trainingsdaten fortgesetzt.

#### 1.1.3.2 Trainingsdaten

Das Erzeugen von Trainingsdaten ist ein sehr aufwändiger Prozess. Je größer die Varianz im zugrundeliegenden Problem desto größer muss auch die Menge der Trainingsdaten sein. Die Schätzung von Körperteillabels hat eine sehr große Komplexität, da der menschliche Körper sehr viele Körperposen einnehmen kann. Zusätzlich zur Körperpose müssen auch andere Faktoren wie die Größe von Personen und die Körperproportionen mit modelliert werden. Gerade die Körperproportionen werden durch Kleidung noch weiter verändert. Um alle diese Variablen in den

Trainingsdaten abzubilden sind viele Probanden notwendig, die dann auch jeweils viele verschiedene Körperposen einnehmen sollten. Diese Daten müssen nicht nur aufgezeichnet, sondern auch annotiert werden. Zeichnet man reale Kameras auf und verwendet keine weiteren Hilfsmittel, so muss die Annotation manuell erfolgen, was für Körperteilklassifikation auf mehreren tausend Trainingsbildern kaum realisierbar ist. Im Projekt wurden zwei verschiedene Systeme zur Erzeugung von Trainingsdaten entwickelt. Zu Anfang des Projekts wurden reale Kameradaten aufgezeichnet und diese mit Hilfe eines anderen automatisierten Systems annotiert. Im Laufe des Projektes hat sich aber herausgestellt, dass dieses Vorgehen immer noch zu schlecht skaliert und es damit nicht möglich ist systematisch eine gute Trainingsdatenbasis zu sammeln. Deshalb wurde zusätzlich auch noch ein System zur Generierung vollständig synthetischer Trainingsdaten umgesetzt und somit eine deutlich reproduzierbarere und erweiterbare Basis für Trainingsdaten geschaffen. Im Folgenden werden beide Systeme beschrieben.

#### **Trainingsdaten aus realen Kamerabildern:**

Zur Erzeugung von Trainingsdaten auf Basis der im Projekt verwendeten Tiefenkameras wurde ausgenutzt, dass das Verfahren eine Segmentierung voraussetzt. Hierdurch wird eine Unabhängigkeit von der Umgebung erreicht, die es ermöglicht Trainingsdaten auch außerhalb eines Fahrzeugs aufzuzeichnen. Außerhalb des Fahrzeugs können auch andere Systeme zur Körperposeschätzung genutzt werden, die im Fahrzeug aufgrund der Platzverhältnisse nicht funktionieren. Der Aufbau für das Sammeln der Trainingsdaten bestand aus einem fest stehenden Stuhl im Labor und zwei Kameras. Die Tiefenkamera aus dem Fahrzeug wurde so positioniert, dass sie den gleichen Blickwinkel wie im Fahrzeug aufzeichnet. Die zweite Kamera war nur für die Annotation notwendig. Hierfür wurde eine weitere Microsoft Kinect mit dem Microsoft SDK zur Körperposeschätzung verwendet. Diese Kamera kann frei platziert werden so dass die Körperposeschätzung die besten Ergebnisse liefert. Beide Kameras wurden zueinander kalibriert. Bei der Aufzeichnung der Trainingsdaten wurden dann beide Kameras und das vom Microsoft SDK erzeugte Skelett aufgezeichnet. Dieses Skelett wurde dann mit Hilfe der Kalibrierung in die Zielkamera transformiert. Somit ist das Tiefenbild dann mit dem Skelett annotiert. Zusätzlich wird aber auch die Annotation der Körperteile benötigt. Hierfür wurde ein Einfaches auf Zylindern basiertes Modell des Menschen genutzt um die Körperteile zu annotieren.

Mit diesem System wurde zu Anfang des Projekts von mehreren Personen Daten aufgezeichnet und annotiert. Die Personen wurden instruiert verschiedene fahrrelevante Bewegungen nachzuahmen. Des Weiteren führten alle Probanden noch möglichst zufällig Bewegungen aus um auch sonst mögliche Körperhaltungen abzudecken. Dadurch wurden etwa 10.000 Trainingsdatenbilder aufgezeichnet. Die auf diesen Daten erzeugten Modelle lieferten bereits vielversprechende Ergebnisse, die sich im Projektverlauf aber als nicht genügend herausgestellt haben. Das Verfahren hat mehrere Nachteile: Wird die Position des Sensors verändert, muss der gesamte Prozess wiederholt werden, die Bewegungen der einzelnen Probanden sind schlecht

kontrollierbar und nicht reproduzierbar und die Körperteile können mit diesem Verfahren nur ungenau annotiert werden.

### **Trainingsdaten aus synthetischen Modellen:**

Um genauere Modell erzeugen zu können und den Aufwand für die Erzeugung der Trainingsdaten weiter zu senken wurde ein System implementiert um die Trainingsdaten vollständig synthetisch zu generieren. Es sind mehrere Komponenten notwendig um dieses Ziel zu erreichen. Es muss eine Datenbasis von möglichen Körperbewegungen im Fahrzeug aufgezeichnet werden, es müssen 3D-Modelle von Personen mit unterschiedlichen Proportionen und Größen generiert werden und es muss ein System zur Animation der Menschmodelle und zum Rendern der Tiefenbilder und der Grundwahrheit entwickelt werden. Für die Einzelschritte konnte hier auf verfügbaren Programmen aufgebaut werden. Die Datenbasis von Motion Capture-Daten wurde mit Hilfe einer Kinect und der Software Brekel Pro Body [28] aufgezeichnet. Diese Daten wurden im Anschluss analysiert um ähnliche Posen zu entfernen und um fehlende Posen noch hinzuzufügen. Hierdurch konnte eine Datenbasis von etwa 80.000 unterschiedlichen Einzelposen generiert werden die viele der im Sitzen möglichen Körperposen abdeckt. Zur Generierung der Menschmodelle wurde das Programm MakeHuman [29] eingesetzt. Es ermöglicht Menschmodelle mit unterschiedlichen Parametern zu erzeugen. Hiermit wurden 6 weibliche und 6 männliche Modelle mit unterschiedlichen Proportionen exportiert. Auf diesen Menschmodellen wurde mit Hilfe einer Textur jedes Körperteil farblich annotiert (siehe Abbildung 9). Die Daten wurden abschließend mit dem freien Rendering Programm Blender [30] erzeugt. Normalerweise werden mit Blender Farbbilder erzeugt. Um stattdessen Tiefenbilder zu erzeugen wurde ein Shader integriert, der aus dem z-Puffer der Rendering Engine das Tiefenbild generiert. Für jede Kameraposition wurde eine entsprechende Szene erstellt bei der die Menschmodelle im gleichen Blickwinkel zur synthetischen Kamera positioniert wurden, wie die echte Kamera auch im Fahrzeug positioniert ist. Über die Scripting Engine von Blender war es möglich quasi beliebig viele unterschiedliche Trainingsbilder zu erzeugen. Für jedes Trainingsbild wurde dafür zufällig eines der Menschmodelle ausgewählt. Dieses wurde für größere Varianz noch zufällig in der Größe skaliert und leicht rotiert und verschoben. Anschließend wurde das Menschmodell in eine zufällig gewählte Pose aus der Datenbasis gebracht und dann das Tiefenbild und die Annotation generiert. Mit diesem System wurde für das abschließende Training ein annotierter Datensatz mit 100.000 Trainingsbildern sowohl für die Fahrerkamera an der A-Säule als auch für die Rückbankkamera im Dachhimmel erzeugt. Das Ergebnis der Körperposenschätzung konnte durch dieses Verfahren deutlich verbessert werden.

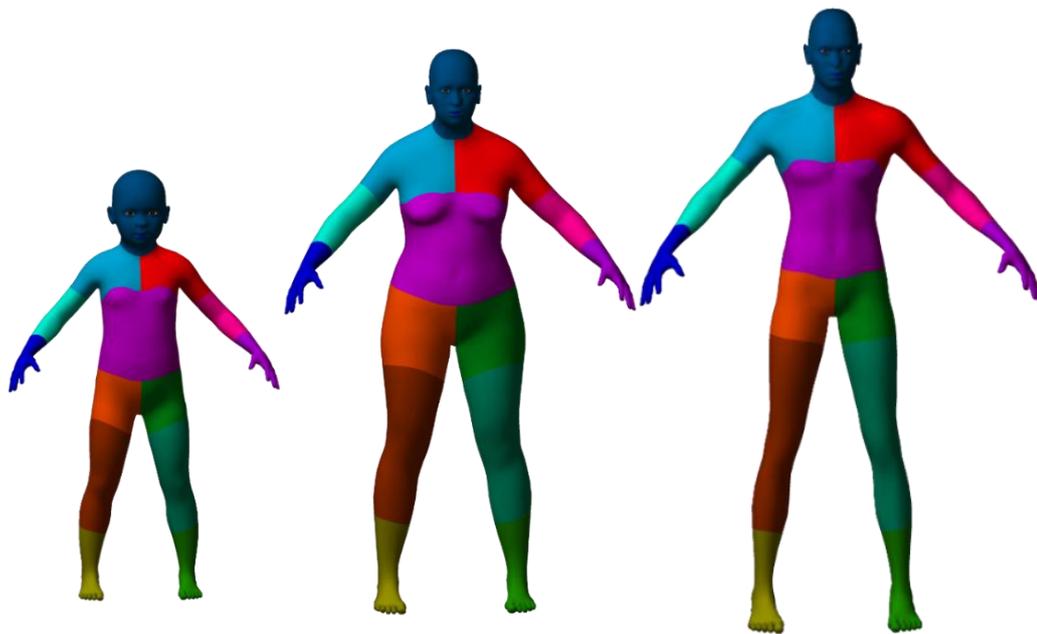


Abbildung 9: Verschiedene synthetische Menschmodelle mit Textur zur Annotation

### 1.1.3.3 Verteiltes Trainingssystem

Viele öffentlich verfügbare Bibliotheken für das Training und die Evaluation von Random Decision Forests haben Einschränkungen. Oftmals kann der Detektor in den einzelnen Knoten nicht selbst bestimmt werden und zusätzlich skalieren die wenigsten Bibliotheken über Rechengrenzen hinweg. Beides war für dieses Projekt notwendig. Deshalb wurde ein eigenes System für das Training und die Evaluation von Decision Forests implementiert. Dieses System kann hochparallel und effizient auf vielen Rechnern gleichzeitig arbeiten. Ein Zentraler Server verwaltet den Zustand des Trainings und verteilt Aufgabenpakete an alle aktiven Knoten. Für die Verteilung der Arbeit und für Ausfallsicherheit wurde der Message Broker RabbitMQ eingesetzt.

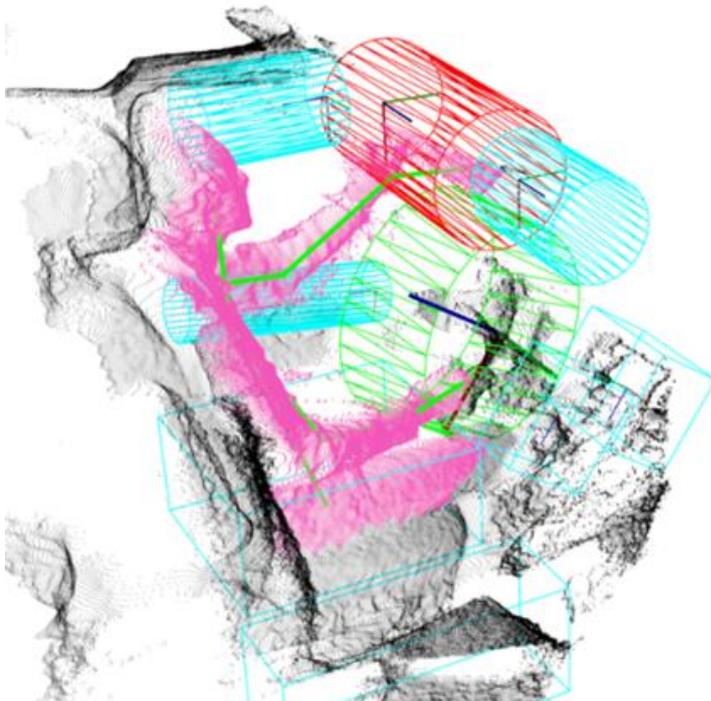
### 1.1.4 Gesamtinnenraumerfassung

Für die Erfassung des gesamten Innenraums wurden drei Kameras eingesetzt. Die geschätzten 3D-Körperposen liegen zuerst im Koordinatensystem des Sensors vor. Dies ist für eine Auswertung der Gesamtinnenraumsituation nachteilig, da die Ergebnisse nicht in einem gemeinsamen Modell des Fahrzeuginnenraums verortet sind. Um dies zu erreichen wurden alle Kameras kalibriert und ein gemeinsamer Ursprung an der Mitte der Vorderachse Kameraunabhängig festgelegt. Alle erfassten Körperposen werden in dieses Koordinatensystem transformiert. Durch dieses Vorgehen ergeben sich mehrere Vorteile. Wird der Sensor an eine andere Position gesetzt ändern sich die Daten nicht wenn das System neu angemessen wurde. Zusätzlich kann jetzt auch eindeutig zugeordnet werden auf welchem Sitz eine Person sitzt indem

die Position des Steißbeins mit der bekannten Position der Sitze im inneren des Fahrzeugs abgeglichen wird. Diese optische Bestimmung der Sitzbelegung bietet schon einen ersten Mehrwert für folgende Systeme. Es kann dadurch auch eine erste Verifikation erfolgen. Optische Systeme sind oft anfällig für Fehldetektionen. Wird erkannt, dass ein Skelett nicht auf einem Sitz sitzt, sondern zum Beispiel im Raum schwebt oder eher an der Oberfläche der Decke liegt, so handelt es sich mit aller Wahrscheinlichkeit um eine Fehldetektion, die durch diese Verifikation unterdrückt werden kann.

### 1.1.5 Interaktionserkennung mit dem Fahrzeuginnenraum

Im Projekt sollte unter anderem erfasst werden, was die Insassen im Fahrzeug tun. Hierbei ist vor allem für den Fahrer auch die Interaktion mit dem Fahrzeug selbst von Interesse. Hierzu gehören Fragestellungen wie zum Beispiel: Hat er die Hände am Lenkrad? Stellt er den Innenspiegel ein? Sucht er etwas auf dem Beifahrersitz?



**Abbildung 5: Visualisierung der Interaktionserkennung mit dem Fahrzeuginnenraum**

Auf Basis der 3D-Körperpose können diese Fragestellungen einfach beantwortet werden. Hierzu wird die im letzten Abschnitt eingeführte kameraunabhängige Lokalisierung der Insassen im Innenraum erweitert. Es werden 3D-Primitive wie Zylinder und Quader um interessante Teile des Fahrzeuginnenraums gelegt. Das Lenkrad wird beispielsweise als Zylinder repräsentiert. Im Folgenden wird dann für jeden Insassen und jedes Körperteil der Abstand zur Oberfläche dieser Volumen bestimmt. Der Abstand ist positiv, wenn sich das Körperteil außerhalb des Volumens befindet und negativ wenn es sich innerhalb befindet. Der Abstand kann genutzt werden um auch

die Annäherung an ein Objekt zu erkennen, um dadurch beispielsweise früher als jetzt möglich bereits zu reagieren.

Um eine Interaktion zum Beispiel mit dem Lenkrad zu erfassen wird auf den berechneten Abstand der Hände ein Schwellwert angewendet. Ist der Abstand kleiner als der Schwellwert findet eine Interaktion statt. Ist er größer findet keine Interaktion statt.

Dieses System wurde auf Daten aus dem Versuchsträger evaluiert. Hierfür wurde die Interaktion von 5 Probanden mit unterschiedlichen Regionen im Fahrzeuginnenraum manuell annotiert und diese Grundwahrheit mit dem Ergebnis des Algorithmus verglichen. Abbildung 6 und Abbildung 7 zeigen die Ergebnisse für die Rechte und die linke Hand. Im Allgemeinen wird die linke Hand zur Interaktion mit weniger Regionen verwendet, weil viele Fahrzeugbereiche nicht bequem erreichbar sind. Der Anteil nicht erfasster Interaktionen ist für die linke Hand auch leicht erhöht, da es aufgrund der Kameraposition häufiger zu Verdeckungen des linken Arms hauptsächlich durch den rechten Arm kommt. Für beide Hände gilt, dass es kaum zu Verwechslungen der einzelnen Interaktionsbereiche kommt. Es werden nur Bereiche verwechselt, die dicht beieinanderliegen. Dies ist vor allem zwischen dem Schaltheben und dem Infotainmentbereich der Fall, da beim VW T5 der Schalthebel im Armaturenbrett direkt unterhalb des Infotainmentbereichs angebracht war. In einigen Fällen wurde sogar mit dem Infotainmentbereich interagiert während die Hand auf dem Schalthebel lag. Eine eindeutige Unterscheidung ist dann nicht mehr möglich. In vielen Fällen in denen eine Interaktion nicht richtig erkannt wurde, wurden die Hände entweder vom Algorithmus nicht erkannt oder waren gerade im Falle der linken Hand verdeckt.

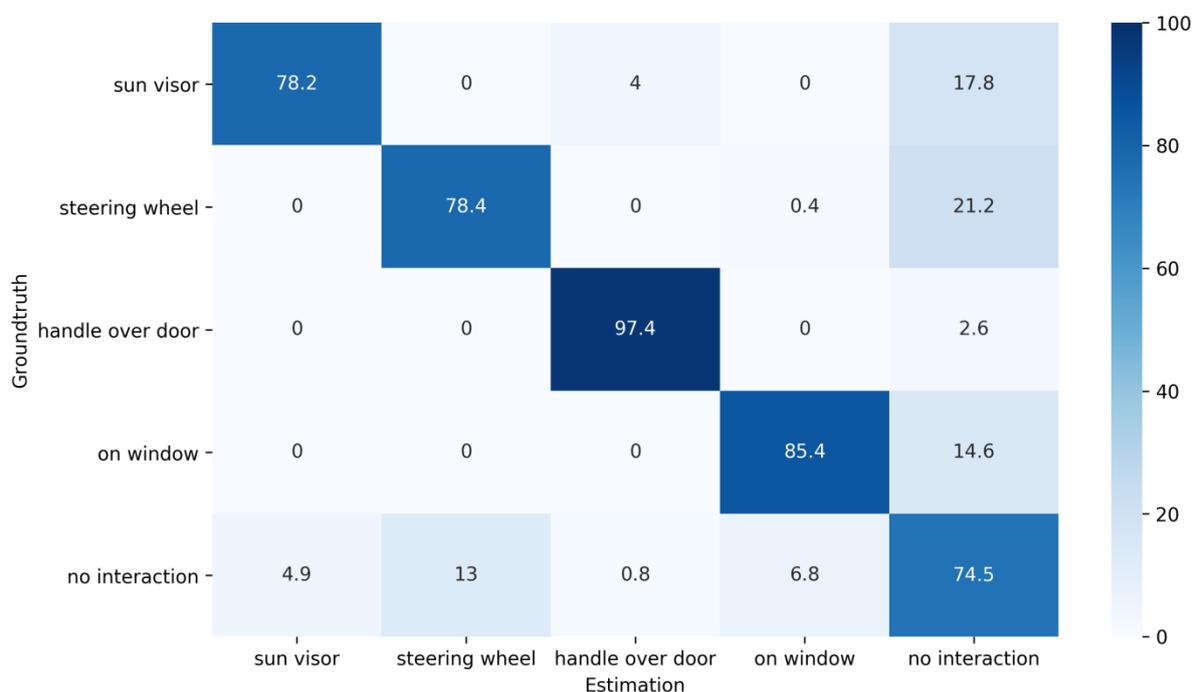


Abbildung 6: Ergebnisse der Interaktionserkennung der linken Hand

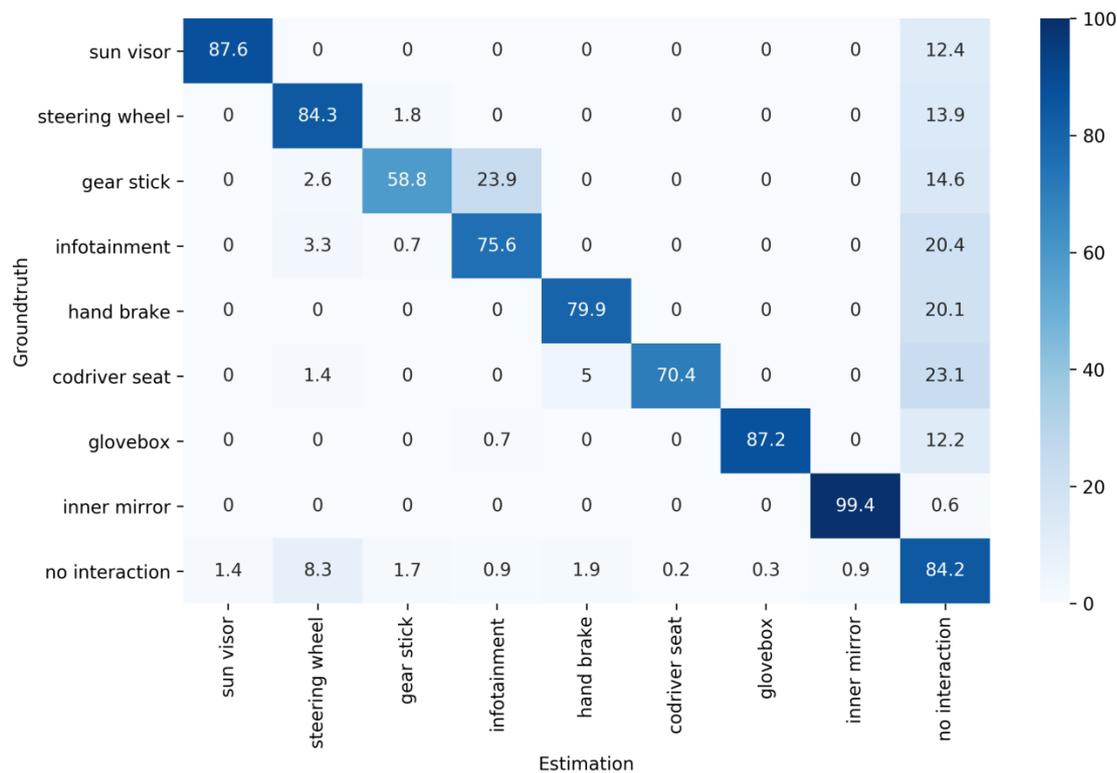


Abbildung 7: Ergebnisse der Interaktionserkennung der rechten Hand

### 1.1.6 Erkennung von Nebentätigkeiten aller Insassen

Um die Situation im Fahrzeuginnenraum klassifizieren zu können, ist es von enormer Bedeutung die Aktivitäten der einzelnen Insassen bestimmen zu können. Als Aktivitäten sind hier beispielsweise »Essen«, »Trinken« oder »Smartphone bedienen« definiert. Die Liste der Aktivitäten kann beliebig erweitert werden, es müssen jedoch ausreichend Trainingsdaten zur Verfügung stehen.

Die Aktivitätserkennung findet auf Basis der Einzelbilder der Posenschätzungen statt. Der schematische Ablauf der Verarbeitungskette ist in Abbildung 8 zu sehen. Als Eingang wird jeweils die Pose betrachtet. Die Daten aller Körperteile werden vorverarbeitet. Zu diesem Schritt gehört die Ausrichtung der Pose, sowie eine Hauptkomponentenzerlegung, um die Anzahl der betrachtenden Variablen zu verringern. Die eigentliche Klassifikation wird durch künstliche neuronale Netze in Form von Multilayer Perceptrons durchgeführt. Für jede zu erkennende Aktivität wurde hierzu ein neuronales Netz eingelernt. Als Klassifikationsstrategie kommt die »One vs. Rest« Strategie zum Einsatz. Jedes Neuronale Netz klassifiziert also eine Aktivität gegen alle anderen. Als Ergebnis der Klassifikation wird die Aktivität herausgegeben, die mit der größten Konfidenz angezeigt wird. Über die Bildfolge werden die Ergebnisse gemittelt, um so gegenüber Ausreißern resistent zu werden. Das System wurde auf den Daten trainiert, die in Griesheim mit dem Versuchsträger erfasst wurden. Abbildung 9 zeigt einige Beispielergebnisse des Systems auf diesen Daten.

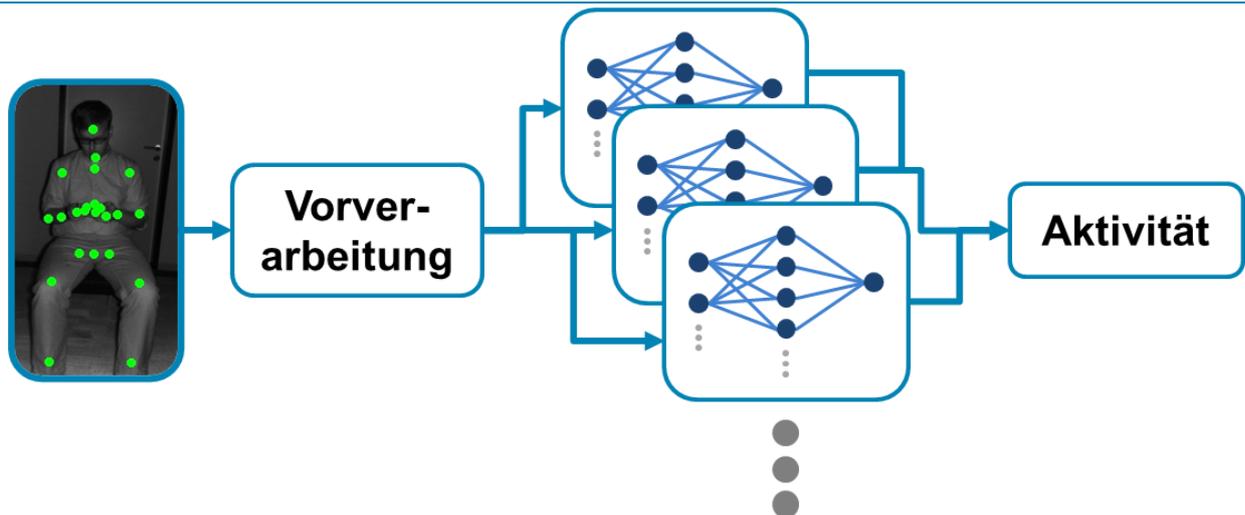


Abbildung 8 Genereller Ablauf der Aktivitätserkennung



Abbildung 9: Beispielergebnisse der Aktivitätserkennung

### 1.1.7 Entwicklung von Interaktionsprinzipien für proaktive Interaktion mit Insassen

Für die proaktive Interaktion aufgrund der InCarIn Zustandsbewertung wurden Interaktionskonzepte mit Experten entwickelt, die proaktive und situations-/posenangemessene Interaktionen seitens des Fahrzeugs ermöglichen sollen. Die iterative Entwicklung wurde mit Bosch durchgeführt und zuerst als Klick-Prototyp umgesetzt, später im Forschungs-HMI. Dafür wurde ein neues Layout für InCarIn typische Situationen für das Bosch-HMI entworfen und in einer iterativen Nutzereinbindung evaluiert. Hierbei wurden insbesondere unterschiedliche Automatisierungsstufen für die proaktive Interaktion untersucht.

Zusammen mit dem IAT der Uni Stuttgart wurde eine Probandenstudie im Fahrsimulator des IAO zur Evaluierung des proaktiven Konzepts und der Umsetzung im Bosch-HMI realisiert.

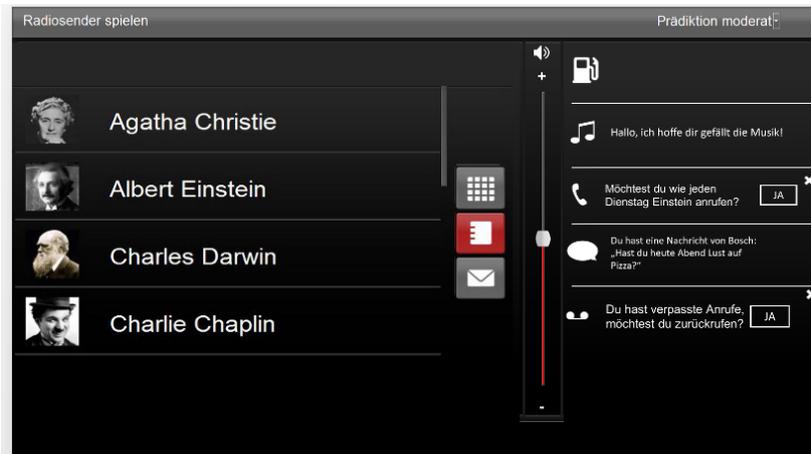
Das Interaktionskonzept basiert dabei auf einer abgestuften Automatisierungsskala, die in Abbildung 10 dargestellt ist.

Grad	Bezeichnung	Nutzer	System
1	Manuell	Der Nutzer muss selbstständig auf Aktionen zugreifen	Das System bietet keine Assistenz.
2	Unterstützend	Entscheidung beim Nutzer ob er mit dem System interagieren will	System bietet eine wahrnehmbare Interaktion an.
3	Teilautomatisiert	Nutzer hat die Möglichkeit durch eine Entscheidung eine Interaktion auszuführen.	System bietet Interaktion an und zeigt bei möglicher Entscheidung des Nutzers die Interaktion an.
4	Hoch Automatisiert	Nutzer wird nur informiert	System entscheidet über Interaktion.
5	Vollautomatisiert	Keine Entscheidungsmöglichkeit	System arbeitet vollständig autonom und aktiviert eigenständig Funktionen

Abbildung 10: Automatisierungsskala des proaktiven Interaktionskonzepts

Im ersten Automatisierungsgrad muss der Anwender selbst, aktiv auf eine Information zugreifen. Diese Stufe kann der Nutzer zu jedem Zeitpunkt selbst anwenden, wenn er eine Information in der Statusleiste auswählt. Eine Umsetzungsmöglichkeit ist eine fest stehende Symbolreihe, welche nur durch aktives Handeln weitere Informationen übermittelt. Die zweite Automatisierungsstufe gibt dem Nutzer direkt die Information des Hinweises aus, fordert aber keine weitere Interaktion mit dem Insassen. Es werden keine Interaktionsbuttons vorgesehen. In der dritten Automatisierungsstufe zeigt das System mögliche Interaktionsoptionen an, die der Nutzer aktivieren kann. Bei einer positiven Entscheidung durch den Nutzer führt das System die jeweilige Funktion aus. Die vierte Stufe ist eine informative Stufe, die den Nutzer über eine, vom System getroffene Entscheidung informiert. Dabei hat der Nutzer keinen Einfluss auf die Entscheidung des Systems. Im Gegensatz zur fünften Stufe erhält er aber eine Information, die bei der letzten Stufe entfällt. Hier arbeitet das System völlig eigenständig.

Die Umsetzung der proaktiven Interaktion erfolgte im Bosch Forschungs-HMI. Diesem wurde eine weitere Spalte im rechten Bildschirmrand hinzugefügt, in der proaktive Vorschläge unterschiedlicher Automatisierungsstufen dargestellt wurden. (vgl. Abbildung 11)



**Abbildung 11: Bosch Forschungs-HMI mit proaktiven Vorschlägen auf der rechten Seite**

In dieser Leiste ist hier das Nachrichtensymbol sichtbar, welches mit einer 1 versehen, eine neue Nachricht anzeigt. Das Musiksymbol veranschaulicht eine Automatisierung vierten Grades, da der Nutzer auf die Musikauswahl des Systems hingewiesen wird, aber keinen direkten Einfluss nehmen kann. Die Musikauswahl erfolgt dabei autonom auf Basis der gesammelten Daten, die Auskunft über häufig gehörte Musik geben.

Der Hinweis mit dem Telefonsymbol zeigt einen Vorschlag für Anrufe. Es handelt sich um die dritte Automatisierungsstufe, die dem Nutzer eine Interaktionsauswahl bietet, die dieser nutzen kann um vereinfacht Funktionen auszuführen. Dabei ist es durch die JA oder NEIN Auswahl möglich eine schnelle Entscheidung zu treffen. Mit der gleichen Funktion ist das Anrufbeantworter Symbol hinterlegt, durch das der Fahrer darauf hingewiesen wird, dass neue Anrufe vorliegen. Diese können durch die Schnellauswahl gezielt geöffnet werden.

Das zweite Nachrichtensymbol zeigt die eingegangene Nachricht an. Dabei handelt es sich um den zweiten Automatisierungsgrad, der Informationen für den Anwender darstellt, aber keine Nutzerinteraktion erfordert. Dadurch hat der Nutzer auch keine Auswahlmöglichkeit für eine schnelle Interaktion gegeben, sondern muss selbst aktiv auf die Information zugreifen um weiter interagieren zu können.

### 1.1.8 Datenerhebung im T5 als Realfahrtversuch

Als technische Unterstützung durch für VW und IAT half das IAO während einer Realfahrtuntersuchung auf dem Fluggelände in Griesheim/Darmstadt zur Erhebung von Verhaltensdaten. Das Fraunhofer IAO übernahm die technisch Betreuung des Versuchsträgers T5 während der 2-wöchigen Versuche und unterstützte dabei die Versuchsleitung vom IAT der Uni Stuttgart und VW. Die Aufbereitung der erhobenen Daten erfolgte ebenfalls durch das Fraunhofer IAO, so dass diese den Partnern zur Verfügung gestellt werden konnten.



### 1.1.9 Aufbau der Versuchsträger

Aufbau eines VW Passat als InCarIn Demonstrator für die Halbzeitpräsentation (HZZP).

Zur Halbzeitpräsentation wurde ein VW Passat mit Innenraumsensorik auf Basis der IOSB Sensoren ausgestattet. Zusätzlich wurde ein HMI System von Visteon integriert. Mit dem Versuchsträger wurde die DPAM Methode in einer Studie zu Gesten und Posen beim manuellen und automatisierten Fahren aller Insassen entwickelt. Auf der Basis des VW Passat wurde im Projektverlauf auch ein Porsche Macan aufgebaut und in eine Fahrsimulation am Fraunhofer IAO integriert. Das proaktive HMI vom IAT wurde in diesen Versuchsträgern iterativ nutzerzentriert entwickelt und evaluiert. Im Porsche Macan führte das IAT zusammen mit dem IAO zudem eine Anwendungsstudie mit Probanden zum Thema Schlafen im Level 4 durch. Am IAO wurden dafür die Fahrscenarien für eine Nachtfahrt programmiert. Nach dem Aufwachen durchfuhren die Probanden im manuellen Betrieb anspruchsvolle Streckenabschnitte. Ihre Fahrleistung in den Manövern wurde aufgezeichnet und mit einer Vergleichsgruppe verglichen. Weitere Angaben zu den Studien sind auch dem Abschlussbericht des IAT der Universität Stuttgart zu entnehmen.



Abbildung 12: VW Passat Versuchsträger.

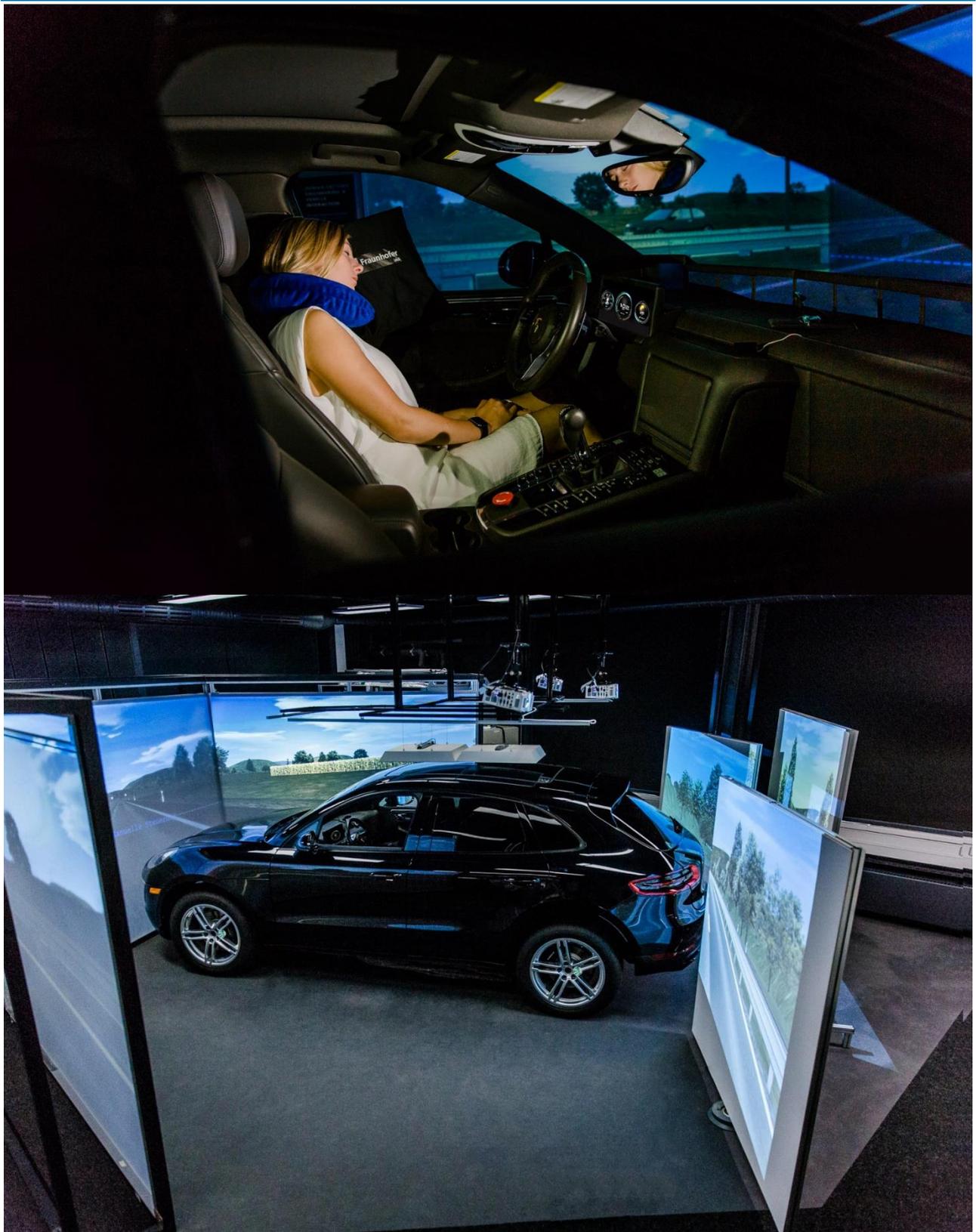


Abbildung 13: Porsche Macan Versuchsträger.

### 1.1.10 Entwicklung einer Bewertungsmethode für Körperpositionen

Entwicklung einer Referenzmethode am IAO und Durchführung einer Fahrsimulatorstudie mit 20 Probanden mit dem IAT zur Validierung der Referenzmethode. Der Versuch liefert gleichzeitig

Referenzdaten für die Posenerkennung des IOSB Kinect Sensors. Das Ziel der in InCarIn entwickelten DPAM (Driver Posture Analysis Method)-Methode ist, ein Verfahren zu entwickeln, welches ermöglicht, Videoaufnahmen der Körperpose der Insassen ohne komplexe Technik, wie zum Beispiel Marker-basierte Trackingverfahren, effizient zu analysieren und ein Ergebnis zu erzeugen, das mathematisch vergleichbar ist. Dafür wurde ein 16-stelliger Code entwickelt der pro analysiertem Bild die Pose des Kopfes, der oberen Extremitäten und des Torsos beschreibt. Jeder Ziffer des Codes ist dabei ein Freiheitsgrad eines Gelenkes zugeordnet. Die Anzahl der Freiheitsgrade und deren Bewegungsspielraum unterscheiden sich je nach Gelenk aufgrund von anatomischen Gegebenheiten. Der Bewegungsspielraum der einzelnen Gelenke wird zur Kodierung in Kategorien unterschiedlicher Größe eingeteilt. Ein Problem hierbei ist die Wahl der Größe der Kategorien für jedes Gelenk. Werden zu kleine Kategorien gewählt, entstehen später mehr Fehler bei der Annotation, bei zu großen Kategorien ist die Auswertung dagegen nicht aussagekräftig, weil zu ungenau. P.M. van Wyk et al. (2009) haben dieses Problem untersucht und für verschiedene Gelenke ideale Kategoriegrößen bestimmt, die wir hier als Basis verwenden (siehe Tabelle 3). Die Anzahl der Kategorien ergibt sich dabei aus der Kategoriegröße und dem anatomisch begrenzten Bewegungsspielraum des Gelenks. Die Kategorisierung der Hände stellt eine Ausnahme dar, da hier die Kategorien keine Winkelbereiche, sondern verschiedene Grifftypen beschreiben.

**Tabelle 3: Kodierung der Gelenke im DPAM-Code mit ihren Kategorien und Winkelbereichen. Schulter, Ellenbogen und Handhaltung sind jeweils für Rechts und Links vorhanden.**

Kategorie Beschreibung	Kategorien [#]	Winkelbereich [Grad]
Kopf Sagittalebene	5	-45 – 45
Kopf Frontalebene	5	-55 – 55
Kopf Transversalebene	7	-80 – 80
Rumpf Sagittalebene	7	-55 - 100
Rumpf Frontalebene	5	-50 – 50
Rumpf Transversalebene	5	-50 – 50
Schulter Sagittalebene	9	-45 – 180
Schulter Frontalebene	9	-45 – 180
Schulter Transversalebene	6	-60 – 90
Ellenbogen Sagittalebene	6	0 – 150
Handhaltung	12	-

Anstatt eines Analyseprotokolls in Form von Papier wie bei RULA wird die DPAM Referenzmethodik als Programm entwickelt. Das Programm hat folgende Vorteile gegenüber dem Papierprotokoll:

- Die Ergebnisse der Analyse lassen sich schnell auf ein anderes Programm übertragen.

## Schlussbericht

- Die Zuordnung der Ergebnisse ist viel einfacher.
- Es ist kein Papier nötig.
- Durch eine geeignete Dialogführung und Darstellung kann eine höhere Zuverlässigkeit und ein höherer Detailgrad erreicht werden.

Das Auswerteprogramm stellt immer alle Kategorien aller Freiheitsgrade eines Gelenks als Piktogramme auf einmal dar. Der Nutzer kann frei zwischen den einzelnen Gelenken wechseln (siehe Abbildung 14).

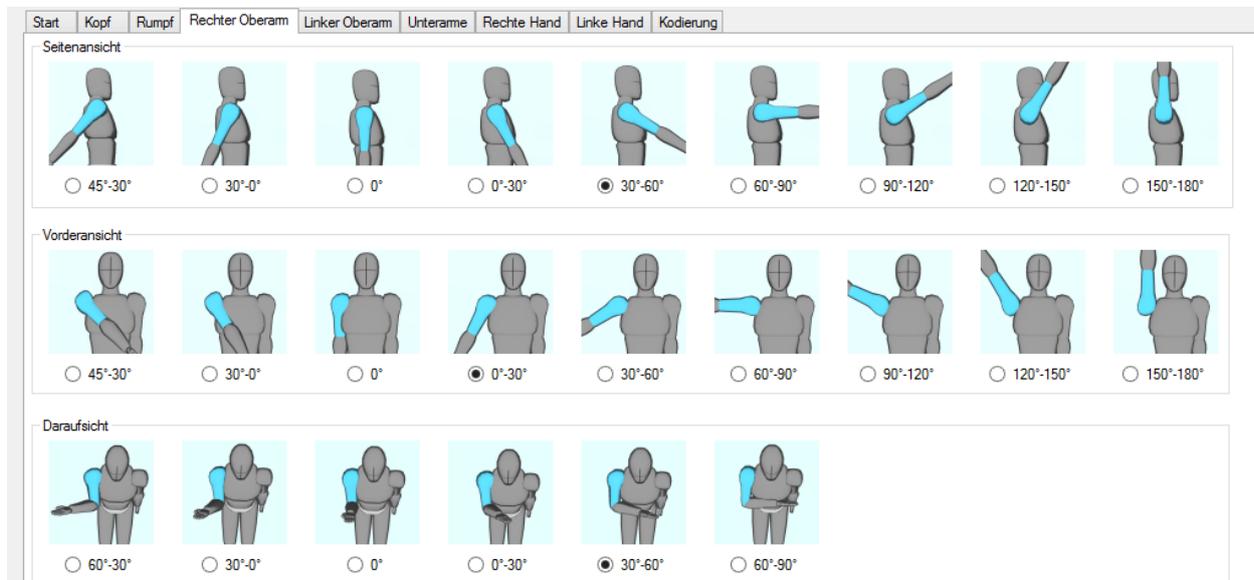


Abbildung 14: Ansicht des Programms zur Erstellung des DPAM-Codes

Durch die Piktogramme wird die Auswertung stark vereinfacht, da dies einen direkten Vergleich zwischen dem Piktogramm und dem zu annotierenden Videobild ermöglicht. Im Simulatorversuch wurde auch festgestellt, dass die Qualität der Ergebnisse steigt, wenn die Piktogramme aus der gleichen Perspektive dargestellt werden, wie auch die Videobilder aufgezeichnet wurden, da das den Vergleich weiter erleichtert.

Die Basis der automatischen, videobasierten Auswertung der mit der DPAM-Methode gewonnenen Ergebnisse ist eine Skeletterfassung der Fahrzeuginsassen. Aufbauend auf der Skeletterfassung kann dann mit weiteren Verfahren eine Kategorisierung, zum Beispiel von Nebentätigkeiten, erfolgen. Hierfür kann entweder vollautomatisch ein DPAM-Code generiert werden und dieser mit vorher gewonnenen Codes verglichen werden, oder es können andere Verfahren wie zum Beispiel Hidden-Markov-Modelle oder Dynamic Time Warping verwendet werden.

Die DPAM-Methode wurde in einer Fahrsimulatorstudie evaluiert. Es nahmen 23 Probanden teil, davon 13 Männer und 10 Frauen (Alter 15 bis 58 Jahre). Da bei der Betrachtung der Körperpose die Anthropometrie des Fahrzeugführers wichtig ist wurde sowohl die Verteilung der Größenverhältnisse der Bevölkerung abgebildet, als auch jeweils 2 Linkshänder ausgewählt. Der Versuch findet in einem VW Passat statt und der Fahrer wird mit Kameras an jeder A-Säule

gefilmt. Ziel der Studie war es zu untersuchen, in wie weit verschiedene Nebentätigkeiten bei automatisierter Fahrt über die Pose des Oberkörpers identifizierbar sind. Es wurden 16 häufige Nebentätigkeiten ausgewählt (SMS schreiben, Tablet benutzen, mit Beifahrer sprechen, sich ausruhen, Buch lesen, Makeup verwenden, mit Rückbank sprechen, telefonieren, ein Kaltgetränk trinken, ein Heißgetränk trinken, Essen mit einer/zwei Händen, Head Unit bedienen, Smartphone per Bluetooth mit dem Fahrzeug verbinden). Es wurde ein Realvideo einer Autobahnfahrt und keine Simulator Software genutzt. Der Versuch besteht aus zwei Abschnitten, die nacheinander durchgeführt werden. Jeder Abschnitt dauert ca. 20 Minuten. Beim ersten Versuch geht es um eine Driver Only-Fahrt, der Fahrer muss während der ganzen Fahrt das Auto steuern. Trotzdem müssen die Nebentätigkeiten durchgeführt werden. Darüber hinaus soll der Proband mit dem Lenkrad während der Fahrt gleichzeitig dem Straßenverlauf folgen. Der zweite Teil simulierte eine hochautomatisierte Fahrt der Stufe 4. Die Zeitdauer des Versuches beträgt ebenfalls ca. 20 Minuten. Während des Versuches müssen die Probanden auf die Anweisungen des Versuchsleiters reagieren. Sobald der Proband eine Anweisung, wie zum Beispiel „Bitte trinken Sie den Tee“ hört, muss er dieses Ausführen. Bei der Versuchsdurchführung wurden insgesamt 770 Minuten Videomaterial bei der Driver Only-Fahrt und 623 Minuten bei hochautomatisierter Fahrt gesammelt.

Nach dem Versuch wurden insgesamt 1.130 Bilder aus den Videos selektiert und mit der DPAM-Methodik annotiert. Einige der dynamischeren Nebentätigkeiten wurden dafür zur Kodierung in mehrere Phasen unterteilt. Typische Posen sollen dabei zwei Charakteristiken haben. Sie sollen repräsentativ und eindeutig sein. Durch die Auswertung der Kodierungen konnten die Nebentätigkeiten „Mit dem Handy anrufen“, „ein Kaltgetränk trinken“, ein Heißgetränk trinken“, „Tablet bedienen“ und „Head Unit bedienen“ gefunden werden.

Die präsentierte DPAM-Methode ist eine kosteneffiziente Möglichkeit, mit Hilfe von Videodaten manuell die Pose des Fahrers auszuwerten. In einem Simulator-Versuch wurde aufgezeigt, wie diese Methodik eingesetzt werden kann, um Nebentätigkeiten bei automatisierter Fahrt auf ihre Unterscheidbarkeit hin zu untersuchen.

#### **1.1.11 Verbreitung der Ergebnisse, Halbzeit- und Abschlusspräsentation und Homepage**

Die InCarIn-Homepage wurde durch das Fraunhofer IAO erstellt. Nach der HZP wurden die dort gezeigten Poster und Präsentationen im öffentlichen Bereich auf der Homepage eingestellt. Die vollständige Homepage ist verfügbar: [www.incarin.de](http://www.incarin.de).

## **4.2 Fortschritt auf diesem Gebiet bei anderen Stellen**

Zu Anfang des Projekts war die Erfassung des Fahrzeug-Innenraums noch wenig im Fokus. Viele der Arbeiten haben sich nur auf die Beobachtung des Gesichts des Fahrers konzentriert (siehe

3.4.1). Dieser Trend hat sich im Laufe des Projektes fortgesetzt. Die Erfassung fokussiert auf den Fahrer und im speziellen auf den Kopfbereich. Der gesamte Innenraum mit allen Insassen wird sonst immer noch wenig betrachtet.

Das Ziel dieser Analysen des Fahrers hat sich aber verändert. War früher das Ziel Assistenzsysteme für die manuelle Fahrt zu entwickeln, so liegt der Fokus heutzutage stark auf der Entwicklung von Systemen für automatisierte Fahrzeuge. Hier sind auch Nebentätigkeiten stärker in den Fokus gerückt die vor allem für Übergabeszenarien große Relevanz haben. Im ersten Schritt kann man hier den Blick des Fahrers beobachten um die Länge der Abwendung von der Straße zu bestimmen und so auf die Ablenkung zu schließen. Dies erlaubt aber keine detaillierte Detektion der Tätigkeit. Die Erfassung der Hände gewinnt deshalb an Bedeutung. Einzelne Systeme setzen hier Farbkameras ein um die Interaktion mit verschiedenen Innenraumbereichen zu erkennen. Die in InCarIn entwickelten Verfahren gehen über diese Systeme hinaus, da Sie zum einen Hardware verwendet, die sowohl Tagsüber als auch Nachts funktioniert und die auch größere Flexibilität in der Erkennung erlaubt. Die Firma Gestigon hat inzwischen mit einer automotive-tauglichen Körperposenschätzung ein System das vergleichbar zu den Systemen aus InCarIn ist. Allerdings geht vor allem das Verfahren zur Erkennung von Nebentätigkeiten über das hinaus, was Gestigon bereitstellt. Gestigon wurde inzwischen auch von Valeo aufgekauft, was das steigende Interesse an solchen Verfahren nochmals unterstreicht.

Die Verfügbarkeit von Automobil-geeigneten Tiefensensoren hat sich in der Projektlaufzeit auch verbessert. Es gibt beispielsweise im BMW 7 inzwischen eine Tiefenkamera zur Gestensteuerung im Dachhimmel. Der erfasste Bereich ist allerdings beschränkt auf das Gebiet über der Mittelkonsole. Die Sensoren werden im Vergleich zur im Projekt verwendeten Kinect auch immer kleiner was den Einbau im Fahrzeug vereinfacht und auch eine bessere Positionierung ermöglicht.

### 4.3 Veröffentlichung der Ergebnisse

- Veröffentlichungen:
  - o VDI Tagung Fahrerassistenz [33]
  - o Automobiltechnische Zeitschrift (ATZ) [32]
  - o International Conference on Intelligent Transportaion Systems (ITSC) [31]
- Vorstellung des Projekts im Rahmen von Vorträgen auf Konferenzen:
  - o VI-Summit 2014
  - o IQPC Automotive Cockpit HMI 2014
  - o IQPC Automotive Cockpit HMI 2015
  - o IQPC Automotive Cockpit HMI 2017
  - o UKPI Media: International Symposium of Automotive Testing 2016

- UKPI Media: International Symposium of Automotive Interiors 2017
- Halbzeitpräsentation am Fraunhofer IAO in Stuttgart
- Pressemitteilung zur Halbzeitpräsentation
- Abschlusspräsentation bei VW in Wolfsburg
- Erstellung einer InCarIn-Homepage [www.incarin.de](http://www.incarin.de)
- Erstellung von 2 gedruckten Projektflyern die auch auf der Homepage zur Verfügung stehen
- Erstellung von Projektpostern die auch auf der Homepage zur Verfügung stehen

## 5 Literaturangaben

- [1] Internetquelle: [m.mercedes-benz.de/de\\_DE/splitview/detail.html](http://m.mercedes-benz.de/de_DE/splitview/detail.html) (Abgerufen am 09.08.2013)
- [2] Datta, A.; Sheikh, Y.; Kanade, T. (2008): Linear motion estimation for systems of articulated planes. In: Proc. Computer Vision and Pattern Recognition 2008, S. 1-8
- [3] Demirdjian, D.; Varri, C. (2009): Driver pose estimation with 3D Time-of-Flight sensor. In Proc. Computational Intelligence in Vehicles and Vehicular Systems 2009, S. 16-22
- [4] Krumm, J.; Kirk, G. (1998): Video occupant detection for airbag deployment. In Proc. Applications of Computer Vision 1998, S. 30-35
- [5] Ito, T.; Kanade, T. (2008): Predicting driver operations inside vehicles. In Proc. Automatic Face & Gesture Recognition 2008, S. 1-6
- [6] Tran, C.; Trivedi, M.M. (2011): Vision for Driver Assistance: Looking at People in a Vehicle. In Visual Analysis of Humans, S. 597-614
- [7] Ji, Q.; Yang, X. (2002): Real-Time Eye, Gaze, and Face Pose Tracking for Monitoring Driver Vigilance. In: Real-Time Imaging, 8(5):357-377
- [8] Zhu, Z.; Ji, Q. (2004): Real time and non-intrusive driver fatigue monitoring. In: Proc. Intelligent Transportation Systems, 657-662
- [9] Fletcher, L.; Loy, G.; Barnes, N.; Zelinsky, A. (2005): Correlating driver gaze with the road scene for driver assistance systems. In: Robotics and Autonomous Systems, 52(1):71-84
- [10] Bach, K.M.; Jaeger, M.G.; Skov, M.B.; Thomassen, N.G. (2008): You can touch, but you can't look: interacting with in-vehicle systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems 2008. Pages 1139—1148.
- [11] Pickering, C.A.; Burnham, K.J.; Richardson, M.J. (2007): A Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction. In

- Proc. Institution of Engineering and Technology Conference on Automotive Electronics 2007. Pages 1-15.
- [12] McCall, J.C.; Trivedi, M.M. (2007): Driver Behavior and Situation Aware Brake Assistance for Intelligent Vehicles. In Proc. of the IEEE, 95(2):374-387
- [13] Holte, M.B.; Tran, C.; Trivedi, M.M.; Moeslund, T.B.(2012): Human Pose Estimation and Activity Recognition From Multi-View Videos: Comparative Explorations of Recent Developments. In IEEE Journal of Selected Topics in Signal Processing (5): 538-552
- [14] McRuer, D.T.; Allen, R.W.; Weir, D.H.; Klein, R.H.(1977): New results in driver steering control models. Human Factors, 19(4):381–397
- [15] Prokop, G. (2001). Modeling human vehicle driving by model predictive online optimization. Vehicle System Dynamics, 35(1):19–53.
- [16] König, L.; Neubeck, J.; Wiedemann, J. (2006): Nichtlineare Methoden zur Trajektorienplanung und automatischen Spurführung im querdynamischen Grenzbereich. In AUTOREG 2006, 2006.
- [17] Guo,K.; Pan, F.; Cheng, Y.; Ding, H.(2002): Driver model based on the preview optimal artificial neural network. In AVEC Proceedings.
- [18] Jürgensohn, T. (1997): Modellierung von Mikro-Steuerstrategien mit Hilfe künstlicher neuronaler Netze. Automobiltechnische Zeitschrift, 99(6):348–354.
- [19] Cole, D.J.(2008): Knowledge of neuromuscular dynamics can help us understand steering feel. In steering.tech.
- [20] Tsimhoni, O.; Liu, Y. (2001): Modeling Steering using the Queueing Network – Model Human Processor (QN-MHP). IN PROCEEDINGS OF THE 45TH ANNUAL MEETING OF THE HUMAN FACTORS AND ERGONOMICS SOCIETY.
- [21] Hartley, L.; Horberry, T.; Mabbott, N.(2000): Review of Fatigue Detection and Prediction Technologies. Report for the National Road Transport Commission.
- [22] Krajewski, J. ; Golz, M. ; Sommer, D. (2009): Detecting Sleepy Drivers by Pattern Recognition based Analysis of Steering Wheel Behaviour. In K. Reschke, U. Kranich & W. Schubert (Hrsg.), Tagungsband Sommeruniversität Verkehrspsychologie (S. 41). Leipzig: Merkur Druck,
- [23] Arun, S., Sundaraj, K. & Murugappan, M. (2012). Driver Inattention Detection Methods: A Review. IEEE Conference on Sustainable Utilization and Development in Engineering and Technology, Kuala Lumpur, Malaysia, 6-9 October 2012, 1-6
- [24] Khreich, W.; Granger, E.; Miri, A. ; Sabourin,R. (2012): A survey of techniques for incremental learning of HMM parameters, Information Sciences, Vol. 197, pp. 105-130
- [25] Murphy,K. (2012): Machine learning: a probabilistic perspective, MIT Press

- 
- [26] Liang,P.; Klein, D. (2009): Online EM for unsupervised models, In Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL '09), pp. 611-619
- [27] Maye, J.; Triebel, R.; Spinello, L.; Siegart, R. (2011): Bayesian on-line learning of driving behaviors, 2011 IEEE International Conference on Robotics and Automation (ICRA), pp.4341-4346, 2011
- [28] Internetquelle: <http://brekel.com/brekel-kinect-pro-body/> (Abgerufen am 13.10.17)
- [29] Internetquelle: <http://www.makehuman.org/> (Abgerufen am 13.10.17)
- [30] Internetquelle: <https://www.blender.org/> (Abgerufen am 13.10.17)
- [31] Martin M., Stuehmer S., Voit M., Stiefelhagen R. (2017): Real Time Driver Body Pose Estimation for Novel Assistance Systems, In Proc. International Conference on Intelligent Transportaion Systems (ITSC): 1738-1744
- [32] Martin M., Voit M.(2017): Insassen-Körperposenerfassung für Innenraum-Assistenzsysteme der Zukunft, In Automobiltechnische Zeitschrift (ATZ): 70ff
- [33] Martin M., Diederichs F., Li K., Voit M., Melcher V., Widloither H., Stiefelhagen R. (2016: Klassifikation von Fahrerzuständen und Nebentätigkeiten über Körperposen bei automatisierter Fahrt, In VDI Tagung Fahrerassistenz

## Berichtsblatt

1. ISBN oder ISSN	2. Berichtsart (Schlussbericht oder Veröffentlichung) Abschlussbericht
3. Titel Forschungsprojekt InCarIn: Intelligent Car Interieur	
4. Autor(en) [Name(n), Vorname(n)] Manuel Martin Michael Voit Frederik Diederichs	5. Abschlussdatum des Vorhabens 31. März 2017
	6. Veröffentlichungsdatum November 2017
	7. Form der Publikation Bericht
8. Durchführende Institution(en) (Name, Adresse) <ul style="list-style-type: none"> <li>• Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB Fraunhoferstraße 1, 76131 Karlsruhe</li> <li>• Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO Nobelstraße 12, 70569 Stuttgart</li> </ul>	9. Ber. Nr. Durchführende Institution
	10. Förderkennzeichen 16SV6292
	11. Seitenzahl 51
12. Fördernde Institution (Name, Adresse)  Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. Literaturangaben 33
	14. Tabellen 3
	15. Abbildungen 14
16. Zusätzliche Angaben	
17. Vorgelegt bei (Titel, Ort, Datum)	
18. Kurzfassung Dies ist der Abschlussbericht zu den Arbeiten des Fraunhofer IOSB und des Fraunhofer IAO im BMBF-geförderten Forschungsprojekt InCarIn – Intelligent Car Interieur. Die Hauptaufgabe des Fraunhofer IOSB im Projekt war die Entwicklung von Algorithmen zur Erfassung des gesamten Fahrzeuginnenraums. Zu Beginn des Projektes wurden hierfür verschiedene Sensorsysteme evaluiert um ein Konzept für eine Gesamtinnenraumerfassung zu erstellen. Aufbauend darauf wurden dann Systeme zur Körperposenschätzung und Aktivitätserkennung für alle Insassen umgesetzt, im Versuchsträger integriert, getestet und in Demonstrationen vorgeführt. Das Fraunhofer IAO hat im Projektrahmen mehrere Simulator Fahrzeuge für Versuche im Projektrahmen aufgebaut. Das Fraunhofer IAO hat in diesen Versuchsträgern sowohl Konzepte für die Klassifikation von Körperposen als auch für die Proaktive Interaktion mit Insassen entwickelt. Des Weiteren wurde auch die Abschlussevaluation im Versuchsträger vom VW durch das Fraunhofer IAO betreut.	
19. Schlagwörter	
20. Verlag	21. Preis

## Document Control Sheet

1. ISBN or ISSN	2. type of document (e.g. report, publication) Final Report
3. title Forschungsprojekt InCarIn: Intelligent Car Interieur	
4. author(s) (family name, first name(s)) Manuel Martin Michael Voit Frederik Diederichs	5. end of project 31. März 2017
	6. publication date November 2017
	7. form of publication Bericht
8. performing organization(s) (name, address) <ul style="list-style-type: none"> <li>• Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung IOSB Fraunhoferstraße 1, 76131 Karlsruhe</li> <li>• Fraunhofer-Institut für Arbeitswirtschaft und Organisation IAO Nobelstraße 12, 70569 Stuttgart</li> </ul>	9. originator's report no.
	10. reference no. 16SV6292
	11. no. of pages 51
12. sponsoring agency (name, address)  Bundesministerium für Bildung und Forschung (BMBF) 53170 Bonn	13. no. of references 33
	14. no. of tables 3
	15. no. of figures 14
16. supplementary notes	
17. presented at (title, place, date)	
18. abstract <p>This is the final report about the work carried out by the Fraunhofer IOSB and the Fraunhofer IAO for the BMBF funded project InCarIn – Intelligent Car Interieur.</p> <p>The main task of the Fraunhofer IOSB in the project was the development of algorithms for the monitoring of the entire interior of the car. To achieve this goal the first step in the project was an evaluation of different sensor systems to develop a sensor concept for the monitoring of the interior. On top of the resulting sensor setup the Fraunhofer IOSB developed systems for body pose estimation and activity recognition for all passengers. These systems were then integrated in the projects test vehicle for testing and demonstration purposes.</p> <p>The Fraunhofer IAO built multiple simulator vehicles in the course of the project. They were used by the Fraunhofer IAO to develop and test concepts for the classification of the posture of passengers as well as for proactive interaction with the passengers of the car. In addition the Fraunhofer IAO also supervised the final evaluation in the projects test vehicle.</p>	
19. keywords	
20. publisher	21. price