

Klassifikation von Fahrerzuständen und Nebentätigkeiten über Körperposen bei automatisierter Fahrt

Dipl.-Inform. **Manuel Martin**, Fraunhofer IOSB, Karlsruhe
Dr.-Ing. Dipl.-Psych. **Frederik Diederichs**, Fraunhofer IAO, Stuttgart
M. Sc. **Kangxiong Li**, Hochschule Pforzheim
Dr.-Ing. **Michael Voit**, Fraunhofer IOSB, Karlsruhe
Dipl.-Psych. **Vivien Melcher**, Fraunhofer IAO, Stuttgart
Dipl.-Ing. **Harald Widloither**, IAT der Universität Stuttgart
Prof. Dr.-Ing. **Rainer Stiefelhagen**, Karlsruhe Institut für Technologie

Kurzfassung

Durch die fortschreitende Automatisierung von Fahrzeugen, besonders des Fahrvorgangs selbst, verändert sich die Rolle des Fahrers mehr und mehr hin zum Passagier. Damit steigt die Bedeutung von Nebenaufgaben und fahrfremden Tätigkeiten. Solange jedoch mit Rückübergaben der Fahraufgabe an den Fahrer während der Fahrt gerechnet werden muss, müssen aus Sicherheits- und Komfortgründen die Aktivitäten des Fahrers erfasst werden. Eine Möglichkeit hierfür ist die optische Erfassung und Klassifikation der Körperhaltung. In diesem Beitrag präsentieren wir ein System zur manuellen Analyse der Körperhaltung für Simulator-Studien sowie einen Ansatz zur automatischen Erfassung der Körperhaltung im Fahrzeug.

1. Einleitung

Durch die Entwicklung automatisiert fahrender Fahrzeuge verändert sich die Rolle des Fahrers weg von der klassischen Fahraufgabe hin zum Passagier, dem ähnliche Nebentätigkeiten erlaubt sind wie allen anderen Fahrzeuginsassen. Für ein automatisiert fahrendes Fahrzeug ist es wichtig, diese Nebentätigkeiten zu verstehen, zum einen aus Sicherheitsgründen, z.B. wenn der Fahrer das Steuer übernehmen soll, zum anderen, um den Insassen kontextsensitive Interaktionsmöglichkeiten bieten zu können. Nebentätigkeiten und fahrfremde Tätigkeiten bieten einen wesentlichen Mehrwert für den Fahrer und sind daher auch eine mögliche Einnahmequelle für die Anbieter. Eine Möglichkeit, Nebentätigkeiten zu klassifizieren und automatisch optisch zu erfassen, ist die Beobachtung und Klassifikation der Körperpose. Statische und dynamische Körperposen sowie natürliche Gesten können Aufschluss über das Aktivitätsniveau des Fahrers, den Fahrerzustand, seine Intentionen [1] und über die Art der Nebentätigkeit geben. Wir stellen eine Lösung für die manuelle Analyse von Nebentätigkeiten bezüglich unterscheidbarer Körperposen vor und präsentieren einen Ansatz zur automatischen optischen Auswertung auf Basis einer Tiefenkamera (MS Kinect).

2. Stand der Forschung

Der Fokus unseres Beitrags liegt auf der Erfassung der Körperpose im Fahrzeug sowie darauf aufbauend auf der Klassifizierung von Nebentätigkeiten bei automatisierter Fahrt. Wir berücksichtigen hierbei die Stufen 4-5 nach SAE [2] bzw. das »Hochautomatisierte und Vollautomatisierte Fahren« nach der Definition der BAST.

Zur Erfassung und Analyse der Körperpose gibt es unterschiedlichste Ansätze. Die genaueste Erfassung der Pose ist mit optischen Marker-basierten oder Marker-losen Multi-Kamera-Motion-Capture-Systemen möglich. Beide Methoden sind für einen Einsatz im Fahrzeug eher ungeeignet, da es durch die eingeschränkten räumlichen Verhältnisse schwierig ist, die Kameras geeignet zu platzieren. Eine andere Möglichkeit zur Erfassung ist die Anbringung von Inertialsensorik am Körper. Hierbei kommt es aber oftmals mit der Zeit zu Sensordrift und, insbesondere im Fahrzeug, zu Störungen durch Metallteile, die das Ergebnis verschlechtern. Alle diese Systeme haben zudem den Nachteil, dass Sie teuer sind, einen hohen Konfigurationsbedarf haben und nicht geeignet sind, später in Serie eingesetzt zu werden.

Für Simulator-Versuche und deren Analyse eignen sich deshalb auch Systeme, die eine effiziente manuelle Auswertung ermöglichen. Solche Systeme gibt es bereits in anderen Bereichen. Sie werden beispielsweise zur Evaluation von Ergonomie und Arbeitssicherheit eingesetzt [3]. Die Basis zur Auswertung ist dabei eine Aufzeichnung der Probanden mit einer einfachen Videokamera und eine Auswertung durch einen Beobachter über einen Analysebogen. Aktuell gibt es noch kein System, das speziell für den Einsatz im Simulator konzipiert ist und die vorhandenen Systeme sind nur eingeschränkt übertragbar.

Zum späteren Einsatz der im Simulator-Versuch erzielten Ergebnisse im Fahrzeug muss die Auswertung dann jedoch automatisch und in Echtzeit erfolgen. Aktuell beschränkt sich die Erfassung des Fahrers hauptsächlich auf die Beobachtung der Kopforientierung und Blickrichtung. In einzelnen Fällen werden auch die Hände erfasst [4]. Um darüber hinaus auch die gesamte Körperpose zu bestimmen eignen sich Verfahren, die wenige Kameras und maschinellen Lernverfahren verwenden [5]. Für viele Anwendungsfälle ist dabei eine deutlich geringere Präzision notwendig als jene, die die oben genannten Motion-Capture-Verfahren liefern.

3. Entwurf der DPAM-Methodik

Das Ziel der 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.

Es wurde dafür 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. [6] haben dieses Problem untersucht und für verschiedene Gelenke ideale Kategoriegrößen bestimmt, die wir hier als Basis verwenden (siehe Tabelle 1). 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 1: 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 [#]	Bewegungsspielraum [Grad]
Kopf Sagittalebene	5	-45 bis 45
Kopf Frontalebene	5	-55 bis 55
Kopf Transversalebene	7	-80 bis 80
Rumpf Sagittalebene	7	-55 bis 100
Rumpf Frontalebene	5	-50 bis 50
Rumpf Transversalebene	5	-50 bis 50
Schulter Sagittalebene	9	-45 bis 180
Schulter Frontalebene	9	-45 bis 180
Schulter Transversalebene	6	-60 bis 90
Ellenbogen Sagittalebene	6	0 bis 150
Handhaltung	12	-

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

1. Die Ergebnisse der Analyse lassen sich schnell auf ein anderes Programm übertragen.
2. Die Zuordnung der Ergebnisse ist erheblich einfacher.
3. Durch eine geeignete Dialogführung und Darstellung kann eine höhere Zuverlässigkeit und ein höherer Detailgrad erreicht werden.

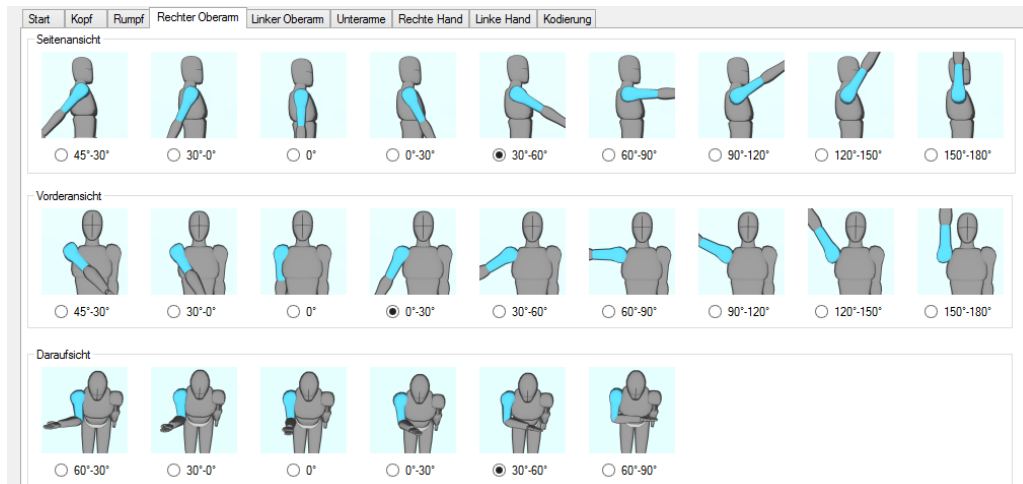


Bild 1: Ansicht des Programms zur Erstellung des DPAM-Codes.

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 Bild 1). Durch die Piktogramme wird die Auswertung stark vereinfacht, da dies einen direkten Vergleich zwischen dem Piktogramm und dem zu annotierenden Videobild ermöglicht.

4. Skeletterfassung im Fahrzeug

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. Dieser zweite Schritt wird in diesem Beitrag aber nicht betrachtet.

Das hier vorgestellte Verfahren zur Skeletterfassung verwendet eine Tiefenkamera (Microsoft Kinect v2) und funktioniert marker-los. Es baut auf dem Verfahren von J. Shotton et al. [5] auf. Es werden keine Auswertelgorithmen der Hersteller-SDKs verwendet. Grundsätzlich basiert das Verfahren auf dem Labeling verschiedener Körperregionen im Tiefenbild mit Hilfe von maschinellen Lernverfahren in Form von Entscheidungsbäumen.

Das Gesamtsystem kann in mehrere Schritte unterteilt werden, die im Folgenden kurz beschrieben werden:

Bei der *Vorverarbeitung* der Sensordaten wird das Tiefenbild der Kamera zuerst geglättet, um das Rauschen zu unterdrücken. Anschließend wird das Tiefenbild mit Hilfe der Spezifikation der Kamera in eine Punktwolke umgewandelt.

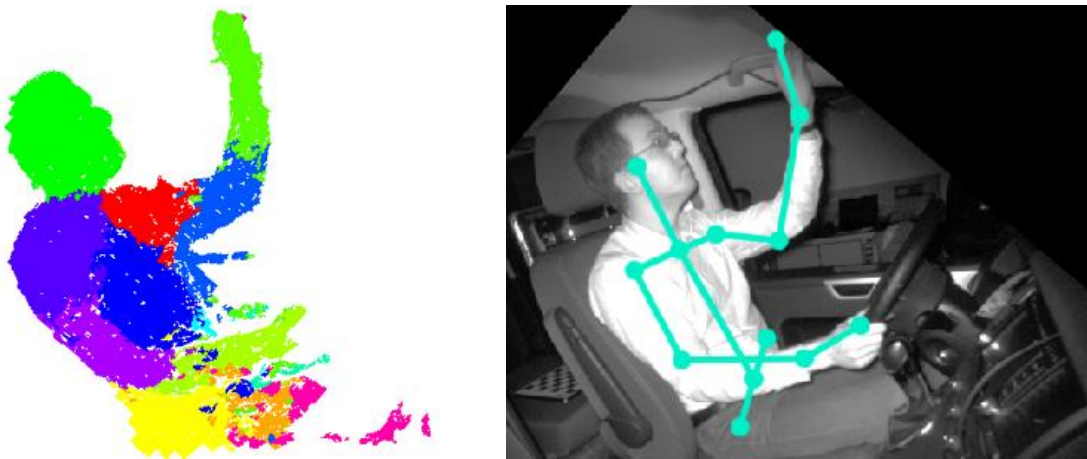


Bild 2: Links: Ergebnis der Segmentierung und der Körperteilklassifikation;
rechts: Ergebnis der Skelettschätzung.

Die *Segmentierung* des Tiefenbilds beruht auf einem dreidimensionalen Hintergrundmodell des leeren Fahrzeugs in Form eines Voxel-Grids. Ein Voxel stellt einen kleinen Würfel im dreidimensionalen Raum dar; es entspricht einem Pixel im zwei-dimensionalen Raum. Jeder Voxel des Hintergrundmodells codiert, ob das Volumen Teil des Fahrzeugs, also des Hintergrunds, ist oder nicht. Zum Einlernen des Modells wird im aktuellen System mit der Kamera das leere Fahrzeug aufgenommen und die aufgezeichneten Sensordaten zur Generierung des Hintergrundmodells verwendet. Alternativ könnte auch ein CAD-Modell des Fahrzeugs verwendet werden. Nach dem Einlernen des Modells kann für jeden Punkt der Punktwolke effizient der Voxel bestimmt werden, in den der Punkt fällt. Ist dieser Voxel nicht Teil des Hintergrundes, so wird der Punkt in die Segmentierung mit aufgenommen.

Das *Labeling* der Körperregionen arbeitet auf den durch den vorherigen Schritt segmentierten Tiefendaten, die den Fahrer beschreiben. Es wird hierfür durch mehrere Entscheidungsbäume für jeden Punkt der Segmentierung bestimmt, zu welcher Körperregion er gehört. Die Entscheidungsbäume werden dafür zuerst auf einer Vielzahl segmentierter und gelabelter Kamerabilder aus derselben Kameraperspektive mit unterschiedlichen Posen eingelernt. Die Entscheidungsbäume liefern ein Histogramm, das die Wahrscheinlichkeit der Zugehörigkeit eines Punktes zu einer Region beschreibt. Der Eintrag mit dem größten Wert definiert dann, welcher Region der Punkt am wahrscheinlichsten zuzuordnen ist (siehe Bild 2).

Das *Clustern* fusioniert die zuvor pixelweise gelabelten Punkte zu größeren Regionen. Hierfür wird eine Variante der Connected Component Analysis verwendet. Dabei wird zur Bestimmung der Cluster sowohl die euklidische Distanz als auch die geschätzte Körperregion genutzt. Nur Punkte, die der gleichen Region zugeordnet sind und in räumlicher Nähe liegen, können in einem Cluster zusammengefügt werden. Von jedem resultierenden Cluster wird nachfolgend

der Schwerpunkt berechnet. Die Schwerpunkte bilden zusammen mit der Clustergröße und den entsprechenden Körperteillabeln die Hypothesen für die einzelnen Gelenke.

Aufgabe der *Skelettschätzung* ist es, aus den Hypothesen für die einzelnen Gelenke die richtigen auszuwählen. Bei perfekten Körperteillabeln und perfektem Clustering würde es für jedes Gelenk nur eine Hypothese geben, die einfach ausgewählt werden könnte. In der Praxis erhält man aufgrund von Fehlern bei den vorangegangenen Schritten eine größere Anzahl Gelenkhypothesen für jedes Gelenk. Zusätzlich kommt es im Fahrzeug durch die beengten Verhältnisse und meist nur suboptimal anbringbaren Kameras zu großen Verdeckungen, so dass Gelenke zeitweise nicht sichtbar, aber durch Rauschen eventuell trotzdem zugehörige Hypothesen vorhanden sind. Um Verdeckungen festzustellen und die richtigen Gelenke zu finden, wird ein iteratives Verfahren eingesetzt. Im ersten Schritt wird dabei für jedes Gelenk die Hypothese, die zum größten Cluster gehört, als korrekt angenommen. Im zweiten Schritt wird dann entlang der kinematischen Kette überprüft, ob es entlang der Verbindung zwischen zwei Gelenken Punkte in der Punktwolke gibt, die in der Nähe der Verbindung liegen. Ist dies nicht der Fall, wird für das jeweilige Gelenk die nächstbeste Hypothese gewählt und Schritt 2 wiederholt. Auf diese Weise wird sichergestellt, dass das Skelett innerhalb der Punktwolke liegt.

5. Evaluation

Die DPAM-Methode wurde in einer Fahrsimulator-Studie evaluiert. Es nahmen 23 Probanden teil (10 weiblich, 13 männlich; Alter 15 bis 58 Jahre). Da bei der Betrachtung der Körperpose die Anthropometrie des Fahrzeugführers wichtig ist, wurde zudem sowohl die Verteilung der Größenverhältnisse der Bevölkerung abgebildet als auch Linkshänder ausgewählt. Der Versuch wurde in einem VW Passat durchgeführt und der Fahrer wurde mit Kameras an jeder A-Säule gefilmt. Ziel der Studie war es zu untersuchen, inwieweit verschiedene Nebentätigkeiten bei automatisierter Fahrt über die Pose des Oberkörpers identifizierbar sind. Es wurden 14 häufige Nebentätigkeiten ausgewählt: SMS schreiben, Tablet benutzen, mit Beifahrer sprechen, ausruhen, Buch lesen, Makeup anbringen, mit Person auf 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 genutzt, keine Simulator-Software, um eine automatisierte Fahrt realistischer Abzubilden.

Der Versuch bestand aus zwei Teilen, die nacheinander durchgeführt wurden. Jeder Teil dauerte ca. 20 Minuten. Beim ersten Versuch ging es um eine manuelle Fahrt, der Fahrer

musste während der gesamten Fahrt mindestens eine Hand am Lenkrad belassen und dabei die Nebentätigkeiten durchführen.

Der zweite Teil simulierte eine hochautomatisierte Fahrt der Stufe 4 nach SAE [1]. Währenddessen mussten die Probanden auf die Anweisungen des Versuchsleiters reagieren. Sobald der Proband eine Anweisung wie zum Beispiel »Bitte trinken Sie den Tee« hörte, musste er diese ausführen.

Bei der Versuchsdurchführung wurden insgesamt 770 Minuten Videomaterial bei der manuellen Fahrt und 623 Minuten bei hochautomatisierter Fahrt gesammelt.

Nach dem Versuch wurden insgesamt 1130 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 sollten dabei zwei Charakteristiken haben: Sie sollen repräsentativ und eindeutig sein. Durch die Auswertung der Kodierungen konnten den Nebentätigkeiten »Mit dem Handy anrufen«, »ein Kaltgetränk trinken«, »ein Heißgetränk trinken«, »Tablet bedienen« und »Head Unit bedienen« eindeutige Posen zugeordnet werden.

Zur Evaluation des Systems zur Körperposenschätzung des Fahrers auch bei Realfahrten wurde ein VW T5 mit einer Kinect v2 an der Beifahrer A-Säule ausgestattet. Die etwas unübliche Position für eine Beobachtung des Fahrers ist dem Öffnungswinkel und dem benötigten Mindestabstand geschuldet. Diese Kamera wurde nach einem Test vieler Consumer-Tiefenkameras ausgewählt, weil Sie die größte Robustheit gegenüber Sonnenlicht zeigte. Es ist somit auch möglich, das System auf offener Straße bei moderater Sonneneinstrahlung erfolgreich zu nutzen. Die Entscheidungsbäume wurden mit etwa 10.000 annotierten Bildern unterschiedlicher Posen trainiert. Die Bilder wurden im Labor außerhalb des Fahrzeugs aber mit ähnlicher Perspektive aufgezeichnet. Dies ist problemlos möglich, da nach der Segmentierung das Fahrzeug selbst keine Rolle mehr spielt. Mit diesem Setup wurden sowohl während der Fahrt als auch im Stand mehrere Stunden Daten aufgenommen. Leider gibt es für diese Daten aktuell noch keine Groundtruth. Eine quantitative Auswertung ist deshalb nicht möglich. In der Zukunft planen wir, Daten manuell zu annotieren. Bis jetzt wurden die Daten qualitativ durch Überlagerung des Skeletts mit den Kamerabildern evaluiert. Es hat sich gezeigt, dass das System die Pose bei manueller Fahrt und auch bei verschiedenen Nebentätigkeiten gut erkennt. Bild 3 zeigt einige Beispielergebnisse des Systems.

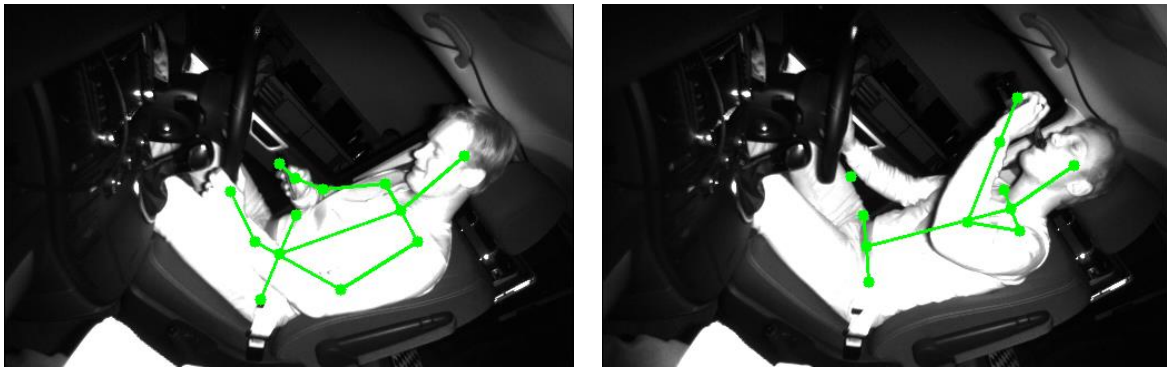


Bild 3: Beispielbilder der Ergebnisse des Systems zur Skeletterkennung für die Aktivitäten »SMS schreiben« und »ein Kaltgetränk trinken«

5. Zusammenfassung

Ziel des Beitrags ist aufzuzeigen, wie Nebentätigen bei automatisierter Fahrt manuell analysiert und im nächsten Schritt automatisch erfasst werden können. Hierfür wurden zwei Systeme vorgestellt. Die präsentierte DPAM-Methode ist eine kosteneffiziente Möglichkeit, mit Hilfe von Videodaten manuell die Pose des Fahrers auszuwerten. In einem Simulator-Versuch wird aufgezeigt, wie diese Methodik eingesetzt werden kann, um Nebentätigkeiten bei automatisierter Fahrt auf ihre Unterscheidbarkeit hin zu untersuchen. Das zweite vorgestellte System kann das Skelett der Insassen mit Hilfe von Tiefenkameras automatisch ableiten. Dieses System stellt den ersten Schritt zur Umsetzung einer automatischen Klassifikation der mit der DPAM-Methode identifizierten Nebentätigkeiten dar. Wir zeigen die Funktionalität des Systems qualitativ auf Videodaten aus dem Fahrzeug.

- [1] Diederichs, F. (in press) Entwicklung von verhaltensbasierten Verfahren zur Erkennung von Fahrerintention für die Prädiktion von Fahrmanövern. Diss. Universität Stuttgart
- [2] SAE. Automated Driving. Abgerufen am 15.09.2016, von http://www.sae.org/misc/pdfs/automated_driving.pdf
- [3] Buchholz, B., et al.: "PATH: A work sampling-based approach to ergonomic job analysis for construction and other non-repetitive work." *Applied Ergonomics*, 27 (1996): 177–187.
- [4] Ohn-Bar, E., et al.: "Head, eye, and hand patterns for driver activity recognition." *International Conference on Pattern Recognition* (2014).
- [5] Shotton, J., et al.: "Real-time human pose recognition in parts from single depth images." *Communications of the ACM* 56.1 (2013): 116-124.
- [6] van Wyk, Paula M., et al.: "Determining the optimal size for posture categories used in video-based posture assessment methods." *Ergonomics* 52.8 (2009): 921-930.