M. Koch, A. Butz & J. Schlichter (Hrsg.): Mensch und Computer 2014 Workshopband,
München: Oldenbourg Wissenschaftsverlag, 2014, S. 205-210.

Demonstrator für ein handgestenbasiertes Interaktionskonzept im Automobil

Thomas Kopinski, Stefan Geisler, Uwe Handmann

Institut Informatik, Hochschule Ruhr West

Zusammenfassung

Handgesten im Automobil haben das Potenzial einer Kombination von gut sichtbaren Displays nahe der Windschutzscheibe und einer als intuitiv empfundenen Gestensteuerung, wie sie berührungsgesteuert von Smartphones aber auch berührungslos von einigen Fernsehgeräten bekannt ist. Bei entsprechender Positionierung der Sensoren können so die Augen auf der Straße und die Hände am Lenkrad oder zumindest sehr nahe dazu verbleiben. Der hier beschriebene frühe Demonstrator zeigt die Machbarkeit dieser Technologie mit einem neuartigen Erkennungsverfahren.

1 Einleitung

Berührungslose Handgesten sind eine vielversprechende Option zur Bedienung von Infotainmentsystemen im Fahrzeug. Anders als Touchscreens im Greifbereich kann das zugeordnete Display näher an der Sichtlinie zum Verkehr positioniert werden. Somit wird die Ablenkung vom Verkehrsgeschehen bei der Gestensteuerung verringert. Verschiedene Forschungsprojekte hierzu wurden in der Vergangenheit durchgeführt. Einen Überblick bietet (Pickering et al. 2007). Einige ausgewählte Arbeiten sollen auch hier kurz erwähnt werden.

Bereits 2000 wurden Arbeiten von BMW in Zusammenarbeit mit der RWTH Aachen publiziert (Akyol et al., 2000). Eine Nah-Infrarotkamera wurde über der mittleren Armlehne positioniert. Mit dem System konnten sechs Handgesten erkannt werden. In einem von Audi unterstützten Projekt mit „Time-of-Flight"-Kameras konnten zwölf unterschiedliche statische Handgesten erkannt werden. (Kollorz et al. 2008)

Neben den technischen Aspekten darf die Akzeptanz durch die Nutzerinnen und Nutzer nicht vernachlässigt werden. In (Zobl et al. 2001). Wird festgestellt, dass hauptsächlich dynamische Gesten und ein beschränktes Alphabet statischer Gesten benutzt wurden. Zudem wird auf die Kulturabhängigkeit einzelner Gesten verwiesen. Die Verringerung der Ablenkung

von berührungslosen Gesten bei der Bedienung von Sekundärtasks in Verbindung mit einem Head-up-Display wurde in (Alpern, M., & Minardo, K. 2003) nachgewiesen.

Audi hat bereits 2012 auf der CES einen Demonstrator vorgestellt. Google hält zudem ein US Patent für Handgestenerkennung im Fahrzeug mit einer in der Decke verbauten Kamera (Hobbs, N. K., & Chi, L-Y. 2013).

Bereits auf dem 2. Workshop Automotive HMI wurden unsere ersten Ansätze zur Gestensteuerung im Fahrzeug vorgestellt (Kopinski et al. 2013). Auf diesem Workshop soll nun ein Live-Demonstrator präsentiert werden.

2 Interaktionskonzept für ein Infotainmentsystem

Moderne Fahrzeuge verfügen über ein umfangreiches Infotainmentsystem. Über eingebaute Funktionen oder verbundene Mobilgeräte werden diese noch erweitert, von der einfachen Freisprechanlage bis hin zu Apps für soziale Netzwerke. Die Gefahr der Ablenkung ist hoch.

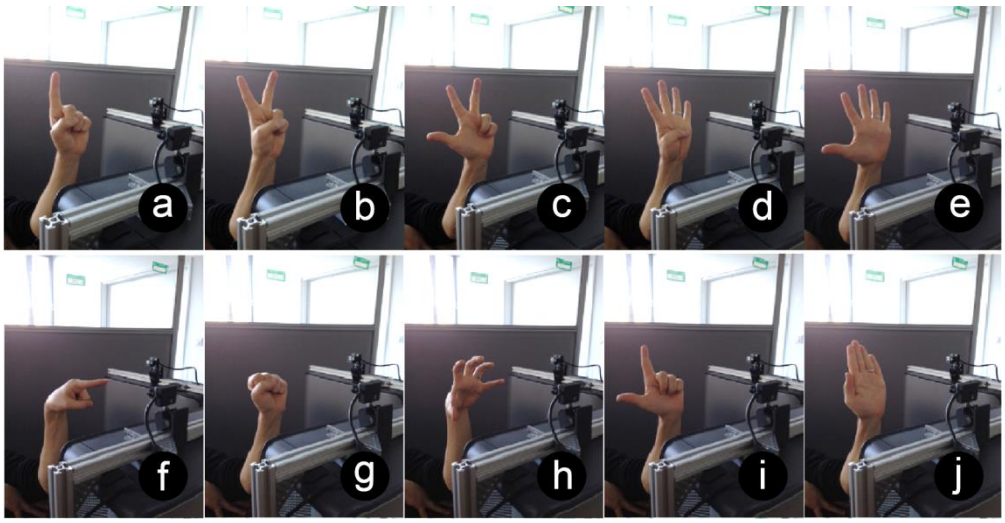

Abbildung 1: Das Gestenalphabet

Die meisten Automobilhersteller versuchen die Ablenkung durch den Einsatz von Spracherkennung und Sprachausgabe zu verringern. Jedoch haben viele Menschen Hemmungen, mit einer Spracherkennung zu arbeiten, insbesondere wenn sich andere Menschen mit im Auto befinden. Im Folgenden soll dargelegt werden, wie berührungslose Gesten bestehende Interaktionskonzepte ergänzen können. Dies wird anhand einzelner Use-Cases dargestellt.

Die eingelernte Datenbank besteht aus 10 statischen Handgesten (s. Abb. 1). Diese Auswahl bildet gut anwendbare Use Cases ab - wir können die Gesten im Fahrzeug sinnvoll benutzen

– und stellen zugleich hinreichend komplexe Fälle zur Ermittlung der Klassifikationsgüte dar.

Mit den oben gezeigten Gesten wurde beispielhaft die Interaktion der folgenden Use Cases angedacht. Abbildung 2 zeigt eine Mock-up Demonstration von zwei Beispielen.

- Eingehender Anruf
  - Geste F: Gespräch annehmen
  - Geste J: Anruf ablehnen
- Eingehende SMS, Twitter/Facebook-Nachricht o.ä.
  - Geste F: Nachricht vorlesen
  - Geste J: Nachricht ignorieren
- Aktive Verkehrsnachrichten u.Ä.
  - Geste J: Abbrechen der Nachricht, zurück zur vorhergehenden Audioquelle
- Favoritenauswahl
  - Gesten A-E für Favoriten 1-5

  Favoriten können sein: Radiosender, Navigationsziele, Telefonkontakte. Dies wird durch den Systemkontext festgelegt.
- Auswahl einer von maximal fünf Optionen auf dem Bildschirm (z.B. Alternative Routen im Navigationssystem, unklare Sprachkommandos, …):
  - Gesten A-E: Option 1-5
  - Geste J: Abbruch des Dialogs

Diese Liste lässt sich für verschiedene Bildschirme erweitern bzw. konkretisieren. In nächsten Entwicklungsschritten können dynamische Gesten mit in die Konzeptdefinition einbezogen werden. So werden z.B. die bekannten Wischgesten zum Wechsel zwischen Seiten oder zum Scrollen durch Listen ermöglicht.

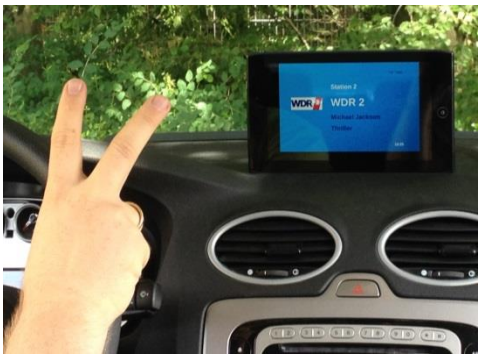

Abbildung 2: Mock-up Demonstration zur Auswahl eines Radio-Presets und Ablehnen eines eingehenden Anrufs.

Als nächste Erweiterung ist die Verbesserung der Zeigegeste geplant. Durch die Tiefensensorik kann das Ziel des Zeigefingers nach ersten Tests hinreichend gut erkannt werden. Dies ermöglicht zwar nicht die Auswahl kleiner Schaltflächen oder von Listenelementen, aber auf großen Displays können somit einzelne Regionen aktiviert werden, wodurch dann

die Aktivierung einer Hauptfunktion, z.B. Wechsel von Navigation auf Radio, realisiert wird. Mit einer Zahlgeste kann dann im nächsten Schritt ein Radiosender aus den Favoriten gewählt werden. Die Hand muss dabei nicht zum Bildschirm bewegt werden.

3 Gestenerkennung mittels ToF-Sensoren

Der Ansatz zeigt, gegenüber anderer Verfahren, dass Gestenerkennung datengetrieben - hier: basierend auf reinen Tiefendaten – und unter wechselnden Lichteinwirkungen zu robusten Ergebnissen gebracht werden kann.

Wir benutzen hierbei zwei ToF-Sensoren (Time-of-Flight) die eine vergleichsweise niedrige Tiefenauflösung von 165 x 120px haben dafür jedoch eine Framerate von bis zu 90 fps besitzen. Aufgrund der ToF-Prinzips, nach dem das ausgesandte Licht mit einer vordefinierten Frequenz moduliert wird, um es von Störquellen unterscheiden zu können, sind wir gezwungen, die Aufnahmen der Kameras alternierend durchzuführen, um Interferenzen zu vermeiden. Das führt automatisch zu einer Reduktion der Framerate, zugleich allerdings auch zu der Möglichkeit, die Sensoren in Außenbereichen einsetzen zu können.

Wir verwenden zur Erkennung der Handgesten neuronale Netze, die wir für unsere Bedürfnisse entsprechend anpassen und konfigurieren. Um hinreichend gute Ergebnisse zu erzielen, wurde eine große Datenbank aufgenommen auf deren Basis die neuronalen Netze trainiert worden sind.

Um die gewünschte Robustheit in unserem Erkennungsverfahren zu erlangen, haben wir eine Datenbank aus 10 Gesten erstellt und zu jeder Geste und Person 2000 Samples aufgenommen. Während der Aufnahme wurden die Hände in einem VOI (Volume of Interest) bewegt und rotiert, so dass ein Mindestmaß an Variation in den aufgenommenen Daten vorhanden ist.

Die aufgenommenen Punktwolken werden mit Hilfe auf die Bedürfnisse zugeschnittener Algorithmen in Histogramme überführt (Kopinski et al. 2014), welche die Eingabe an das neuronale Netz bilden. Da wir Inputs beider Kameras fusionieren, ergibt sich insgesamt eine Datenbank von 400000 Samples, die wir als Grundlage für unser Erkennungsverfahren verwenden können.

Die Tatsache, dass wir Erkennungsraten bis zu 95% haben zeigt, dass ein einfacher, echtzeitfähiger, aber dennoch ausreichend robuster Deskriptor genügt, um ein Gestenerkennungssystem aufzubauen.

4 Gestenerkennung im Fahrzeug

Das Livedemonstrationssystem ist wie in Abb. 3 aufgebaut. Die Sensoren sind in ca. 25cm Abstand zueinander im zentral-rechten Teil der Mittelkonsole montiert, so dass der Sichtwinkel zur Hand ca. 30° beträgt. Das VOI umfasst einen Kubus mit den Maßen

30cmx40cmx25cm. Die Kameras nehmen alternierend Aufnahmen von der Szene, entfernen den Hintergrund und liefern die aufgenommene Punktwolke an das System.

Einige der oben gezeigten Gesten sind schwer zu disambiguieren (z.B. a vs. g oder d vs. e) weshalb es zu Fehlklassifikationen kommt. Dieses Problem können wir dadurch beheben, indem wir eine einfache zeitliche Mittelung durchführen und eine Geste als solche erkennen, wenn mind. fünf der letzten sieben Klassifikationen für den Kandidaten sprechen.

Mit Hilfe des genannten Verfahrens, erreichen wir eine sehr hohe Klassifikationsrate (>95%), wenn die Testperson mit in den Trainingsdaten enthalten ist. Sofern die genannte Testperson nicht in die Trainingsdaten eingeschlossen wird, reduziert sich die Klassifikationsgüte um 5-15%.

Eine genaue Abschätzung des Klassifikationsfehlers ist unter Live-Bedingungen schwer umsetzbar, jedoch haben wir eine als ausreichend empfindbare Performanz erzielt. Erste Rückmeldungen der Testpersonen lassen eine gute Nutzerakzeptanz erwarten.

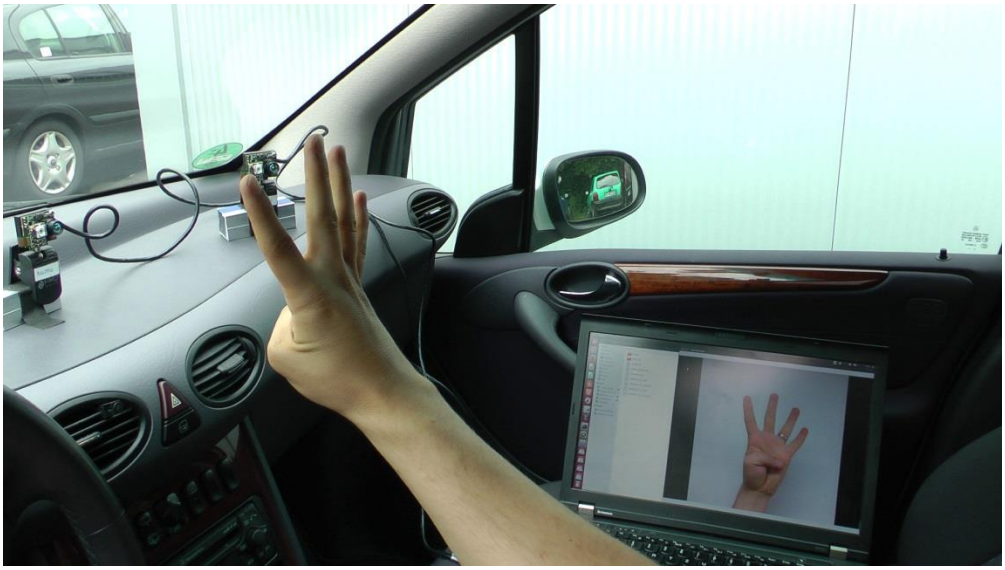

Abbildung 3: Gestenerkennung im Fahrzeug (Testsystem)

5 Zusammenfassung und Ausblick

Gestenbedienung im Fahrzeug bildet nach unserer Einschätzung eine neue, zusätzliche Eingabemodalität im Fahrzeug. Bei entsprechender Positionierung der Kameras können so die Hände nahe des Lenkrads verbleiben, dennoch besteht die Möglichkeit, von Smartphones bekannte Gesten zu verwenden.

Der entwickelte Algorithmus basiert auf Tiefenkameras. Die Erkennung erfolgt in Echtzeit und ist robust für den Einsatz im Automobil. Derzeit können zehn unterschiedliche statische Gesten erkannt werden.

Die Gestenerkennung erlaubt die Ergänzung bestehender Interaktionskonzepte, ein Ersatz traditioneller Eingabeformen ist jedoch nicht möglich. Bereits mit den derzeit implementierten statischen Gesten sind erste Funktionen umsetzbar.

In nächsten Schritten ist geplant, verschiedene Kamerapositionen zu evaluieren. Weiterhin soll das Interaktionskonzept ausgebaut werden, um es in Benutzertests validieren zu können.

Literaturverzeichnis

Akyol, S., Canzler, U., Bengler, K., & Hahn, W. (2000). Gesture Control for Use in Automobiles. In MVA (IAPR Workshop on Machine Vision Applications) (pp. 349-352).

Alpern, M., & Minardo, K. (2003). Developing a car gesture interface for use as a secondary task. In CHI'03 extended abstracts on Human factors in computing systems. pp. 932-933.

Hobbs, N. K., & Chi, L-Y., (2013). Gesture-Based Automotive Control, U.S. Patent Application No. 13/437,730. Washington, DC: U.S. Patent and Trademark Office.

Kopinski, T., Gepperth, A., Geisler, S., & Handmann, U. (2014). Neural network based data fusion for hand pose recognition with multiple ToF sensors. ICANN 2014, to be published.

Kopinski, T., Geisler, S. & Handmann, U., (2013). Contactless Interaction for Automotive Applications. In: Boll, S., Maaß, S. & Malaka, R. (Hrsg.), Mensch & Computer 2013 – Workshopband. München: Oldenbourg Verlag. (S. 87-94).

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 Automotive Electronics, 3rd Institution of Engineering and Technology Conference on Automotive Electronics, pp. 1-15.

Zobl, M., Geiger, M., Bengler, K., & Lang, M. (2001). A usability study on hand gesture controlled operation of in-car devices. In Abridged Proceedings, HCI 2001, 9th Int. Conference on Human Machine Interaction, pp. 166–168.

Kontaktinformationen

Thomas Kopinski, E-Mail: thomas.kopinski@hs-ruhrwest.de
Prof. Dr. Stefan Geisler, E-Mail: stefan.geisler@hs-ruhrwest.de
Prof. Dr.-Ing. Uwe Handmann, E-Mail: uwe.handmann@hs-ruhrwest.de