

# Herausforderungen menschengerechter Forschung mit Body-Tracking-Sensoren in Langzeit-Feldstudien

Julian Fietkau

julian.fietkau@unibw.de

Universität der Bundeswehr München  
Neubiberg, Deutschland

Jan Schwarzer

jan.schwarzer@haw-hamburg.de

Hochschule für Angewandte Wissenschaften Hamburg  
Hamburg, Deutschland

## Zusammenfassung

In der Mensch-Computer-Interaktion sind Beobachtungsstudien ein bewährtes Mittel, um Erkenntnisse darüber zu gewinnen, wie Menschen sich in der Interaktion mit digitalen Systemen verhalten. Um solche Beobachtungen über längere Zeiträume durchführen und reichhaltigere Beobachtungsdaten sammeln zu können, werden mitunter optische Sensoren eingesetzt, die bestimmte Aspekte des Verhaltens ohne kontinuierliche menschliche Betreuung aufzeichnen können. Wie können solche Sensoren menschenwürdig und unter Wahrung der Privatsphäre eingesetzt werden und welche Fallstricke sind dabei zu beachten? Wir berichten aus dem Projekt *Untersuchung des Honey-pot-Effekts an (halb-)öffentlichen interaktiven Ambient Displays in Langzeitfeldstudien* und reflektieren in diesem Beitrag Aspekte der Inklusion und des respektvollen Umgangs mit Versuchspersonen.

## CCS Concepts

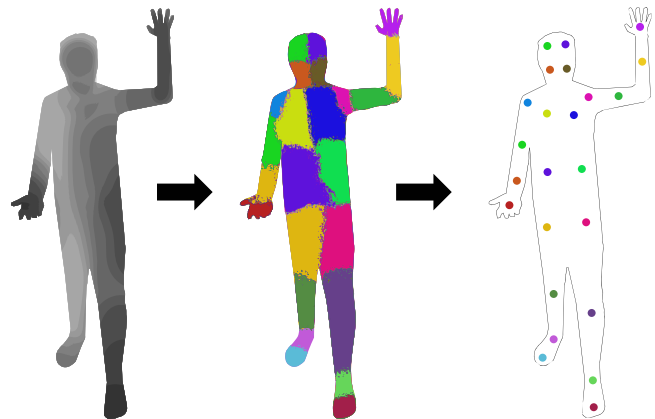
• **Human-centered computing** → *Empirical studies in ubiquitous and mobile computing*; • **Hardware** → *Sensor applications and deployments*; • **Social and professional topics** → *People with disabilities*.

## Stichwörter

Feldstudien, Langzeitstudien, sensorbasierte Empirie, quantitative Empirie, Body Tracking, Kinect, Barrierefreiheit, Inklusion

## 1 Hintergrund

Das Projekt *Untersuchung des Honey-pot-Effekts an (halb-)öffentlichen interaktiven Ambient Displays in Langzeitfeldstudien* (kurz: *HoPE*) befasst sich seit 2021 mit der Analyse des sozialen Verhaltens von Personen vor interaktiven Wandbildschirmen. Zur detaillierten Untersuchung des Verhaltens werden, neben qualitativen Methoden wie Interviews und Beobachtungen, optische Body-Tracking-Sensoren (Tiefenkameras) eingesetzt und sich dabei auf weitestgehend anonymisierte Körpermodelle konzentriert (siehe Abbildungen 1 und 2). Die im Projekt eingesetzten optischen Sensoren (Microsoft Kinect, Stereolabs ZED 2) arbeiten grob wie folgt: Zunächst wird ein Bild der Umgebung durch die Kamera-Sensoren digitalisiert. Das Bildmaterial wird dann allerdings nicht unmittelbar aufgezeichnet, sondern es wird in Echtzeit ein Erkennungsverfahren



**Abbildung 1: Die Verarbeitungskette des Kinect v2-Sensors: Aus einem Tiefenbild werden identifizierte Körperregionen und daraus bestimmte Körperpunkte abgeleitet. (Darstellung nach Shotton et al. [7])**

angewandt, welches Personen mit ihren Körperhaltungen im Bild markiert. Während der Kinect v2-Sensor dazu auf einen Random Forest-Algorithmus setzt, baut die ZED 2 auf ein neuronales Netz.

Die Hersteller der beiden von uns eingesetzten kommerziellen Sensoren legen die exakte Funktionsweise ihrer jeweiligen Erkennungsverfahren nicht offen. Jedoch können wir anhand von Studien (z. B. Shotton et al. [7]) ableiten, wie der grundsätzliche Ablauf aussieht. Am Beispiel der Kinect v2 möchten wir im Folgenden die Funktionsweise genauer schildern. Es ist davon auszugehen, dass das Training des in der Kinect v2 integrierten Random Forest-Algorithmus – analog zu Shotton et al. [7] – mit mehreren hunderttausend künstlich erstellten Bildern vollzogen wurde, um ihre Funktionalität zu gewährleisten. Die Verarbeitungskette der Kinect v2 lässt sich gedanklich dreiteilen (siehe Abbildung 1):

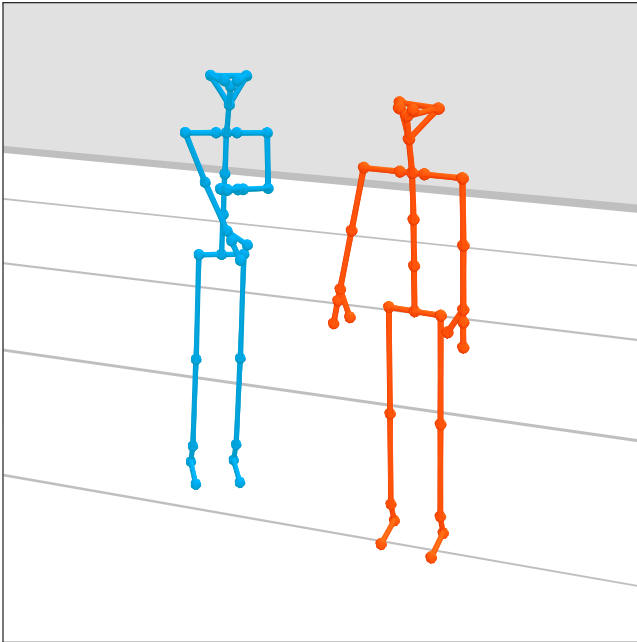
- (1) Der Kinect-Sensor erhebt zuallererst Tiefenbilder über Infrarot, in welchen jeder Pixel Tiefeninformationen bis auf wenige Zentimeter genau enthält. Die Vorteile von Tiefenbildern liegen u. a. darin mit schlechten Lichtverhältnissen umgehen zu können, farb-, textur- und Körperform-invariant zu sein sowie in der Fähigkeit, realistische Abbilder von Personen synthetisieren zu können.
- (2) Mithilfe von Klassifikationsalgorithmen werden im nächsten Schritt probabilistische Pixel-basierte Körperregionen ermittelt. Einige dieser Teile sind so definiert, dass sie bestimmte Körperpunkte direkt lokalisieren, während andere die Lücken füllen oder in Kombination verwendet werden können, um andere Gelenke vorherzusagen.



This work is licensed under a Creative Commons Attribution 4.0 International License. *Mensch und Computer 2024 – Workshopband, Gesellschaft für Informatik e.V., 01.–04. September 2024, Karlsruhe, Deutschland*

© 2024 Copyright held by the owner/author(s). Publication rights licensed to GI.

<https://doi.org/10.18420/muc2024-mci-ws14-170>



**Abbildung 2: Darstellung exemplarischer Body-Tracking-Daten von zwei Personen als abstrakte 3D-Körpermodelle in unserer Visualisierungssoftware PoseViz [1].**

- (3) Im letzten Schritt werden die konkreten Positionen von Körperpunkten in dreidimensionalen Koordinaten vorgeschlagen. Die zuvor ermittelten Pixel-basierten Informationen bzgl. der Körperregionen müssen nun über alle Pixel zusammengefasst werden, um verlässliche Vorschläge für die Positionen der Körperpunkte machen zu können. Für die Kinect v2 ergeben sich auf Basis dieses Prozedere insgesamt 26 individuelle Körperpunkte.

Die auf diese Weise analytisch ermittelten Körperpunkte werden mit ihren Positionen im Raum aufgezeichnet. Als Resultat liefert die Sensorik also vergleichsweise genaue Daten zur Position, Körperhaltung, Blickrichtung usw. der Personen im räumlichen Bereich vor den Bildschirmen (siehe Abbildung 2). Zwar ist es auf Basis von abstrakten Körpermodellen nicht möglich Verhalten realen Personen zuzuordnen, jedoch können bspw. Schlussfolgerungen mit Blick auf ein Gruppenverhalten gezogen werden. Dies wirft unvermeidbar neue ethische Fragen auf, denen sich die HCI- und CSCW-Feldforschung zunehmend annehmen muss.

Wir erheben seit mehreren Jahren nahezu durchgehend Body-Tracking-Daten mit optischen Sensoren und werten sie qualitativ sowie quantitativ aus. In anderen Veröffentlichungen präsentieren wir das Forschungskonzept [2, 3], die Verarbeitung der Daten [1], die dafür aufgebaute technische Infrastruktur [5] sowie erste empirische Ergebnisse [6]. Bereits publizierte Aspekte des Forschungsprojekts (einschließlich seiner Kernerkenntnisse) wiederholen wir an dieser Stelle bewusst nicht. Den Rahmen dieses Beitrags möchten wir stattdessen nutzen, um die nicht beabsichtigten Konsequenzen des Einsatzes solcher Sensoren darzustellen und dabei einen

Schwerpunkt auf die Inklusion, die Menschengerechtigkeit und den respektvollen Umgang mit Studienteilnehmenden zu legen.

Dies tun wir vor dem Hintergrund, dass der Einsatz vergleichbarer Sensorik für empirische Langzeitstudien aufgrund von steigender Datenqualität und fallenden Anschaffungspreisen an Attraktivität gewinnt und wir mit einer stärkeren zukünftigen Verbreitung solcher Datenerhebungen rechnen.

## 2 Ungleichbehandlung durch normative Körpermodelle

Die Funktionsweise von Body-Tracking-Verfahren besteht darin, Kamera-Bildmaterial anhand eines menschlichen Körpermodells zu interpretieren. Jenes Körpermodell entscheidet darüber, welche der visuellen Strukturen im Kamerabild als Menschen gelten und welche nicht. Dieser deskriptiv intendierte Prozess hat unweigerlich einen normativen Anteil: Die impliziten Annahmen des Körpermodells können die Erkennung von untypischen menschlichen Körpern dauerhaft und auf tückische Weise erschweren.

Die Körpermodelle gehen davon aus, dass ein Mensch aus einem Kopf, einem Rumpf, zwei Armen und zwei Beinen besteht, welche auf eine bestimmte Weise angeordnet und innerhalb gewisser Toleranzbereiche proportioniert sind [4, 8]. Tritt eine Person in den Kamerabereich, die bspw. keinen rechten Arm hat, so wird die Body-Tracking-Software diese Person deshalb weniger gut als Mensch erkennen. In Fällen, in denen das Körpermodell trotzdem „anschlägt“, wird es einen rechten Arm stillschweigend dazu erfinden und potenziell unbemerkt in den Body-Tracking-Daten ablegen. Da das Verfahren darauf basiert, verdeckte und anderweitig schlecht erkennbare Körperteile anhand des Körpermodells abzuleiten, sind solche Verfälschungen unvermeidlich. Zudem werden Personen, die Hilfsmittel wie Rollstühle oder Blindenstöcke verwenden, ebenfalls schlechter auf das Norm-Körpermodell abgebildet und werden deshalb weniger gut erkannt.

Würde das Körpermodell flexibler konzipiert, könnten sich allerdings in der Folge die Fehlerraten bei der Erkennung von Menschen mit verbreiteten Körpertypen erhöhen, von den Entwicklungskosten (moderne Body-Tracking-Verfahren verwenden Machine Learning und sind auf umfangreiche Trainingsdatensätze angewiesen) noch gar nicht zu reden. So ergibt sich für die Herstellerfirmen der fehlgeleitete ökonomische Anreiz, Menschen mit seltenen Körpertypen zu vernachlässigen. Dies kann Betroffene in Body-Tracking-Aufzeichnungen unsichtbar machen und zu einem Domino-Effekt der Vernachlässigung führen.

Für uns als Konsumenten von Body-Tracking-Technologie wäre es an dieser Stelle einfach zu sagen, die Hersteller müssten die Körpermodelle verbessern. Obwohl wir eine stärkere Einbeziehung von Menschen mit seltenen Körpertypen in die Entwicklungs- und Testprozesse ausdrücklich begrüßen würden, sind wir nicht zuversichtlich, dass dieses Problem durch iterative Verbesserungen der Software gelöst werden kann. Vielmehr müssen die unterliegenden Daten, auf denen das Tracking basiert, grundlegend um verschiedenartige Körpertypen erweitert und auf Diskriminierungsfreiheit geprüft werden. Der Ressourcenaufwand hinter einer solchen Maßnahme wäre beträchtlich, weshalb wir eine Chance eher durch legislative Regulierung der Technologie sehen als durch bloße Produktkonkurrenz.

### 3 Abschreckung durch Sensor-Präsenz

Technisch handelt es sich bei Body-Tracking-Sensoren um Videokameras mit Spezialsoftware. Obwohl wir die Videobilder nicht aufzeichnen, ist dies für Personen im Kamerabereich prinzipiell nicht überprüfbar. Insofern läuft ein Feldeinsatz solcher Sensoren Gefahr, allein durch ihre Präsenz das Verhalten der Personen zu beeinflussen (vgl. Hawthorne-Effekt), indem sie z. B. den Kamerabereich meiden oder sich weniger intensiv bewegen. Es ist nach einiger Zeit mit einem Gewöhnungseffekt zu rechnen. Wir begegnen dieser Herausforderung mit Transparenz im Forschungsprozess und Aufklärung über die technischen Möglichkeiten.

### 4 Deanonymisierung

Eine der Voraussetzungen für menschengerechte Empirie ist der sorgfältige Umgang mit personenbezogenen Daten. Body-Tracking-Sensoren abstrahieren recht erfolgreich vom Aussehen einzelner Personen und werden deshalb von einigen Forschenden als intrinsisch anonymisiert betrachtet. Bestimmte Merkmale (z. B. Körpergröße) lassen sich aus den Daten allerdings mit Einschränkungen rekonstruieren. Dies ist insbesondere dann der Fall, wenn bspw. qualitative Beobachtungsdaten mit Body-Tracking-Daten trianguliert werden. Weiterhin könnten Body-Tracking-Daten, welche mit präzisen Zeitstempeln versehen sind, mit externen Datenquellen (z. B. Labortagebuch oder Dienstplan) verbunden werden, um Personen zu deanonymisieren und Bewegungsprofile zu erstellen.

Zur Anonymisierung der Körperproportionen gibt es Verfahren, welche die Längen der Gliedmaßen normieren und so die damit verbundenen individuellen Eigenschaften in den Aufzeichnungsdaten verstecken [9]. Eine Möglichkeit ist die Berechnung der Winkel aller Gelenke einer Person aus den Positionsdaten gefolgt von einer Anwendung dieser Gelenkwinkel auf ein Körpermodell in Normgröße. So bleiben Positionen und Bewegungsabläufe erhalten, aber die Körperproportionen sind nicht mehr rekonstruierbar. Der größte Nachteil solcher Verfahren ist, dass durch die einhergehenden Änderungen von Hand- und Kopfpositionen die Präzision in der Erkennung von Blickrichtungen und Zeigegesten verschlechtert wird. Da individuell charakteristische Bewegungen weiterhin in den Daten enthalten sind, gewährleisten auch diese Verfahren keine vollständige Anonymisierung. Dennoch sind Normalisierungsverfahren für Body-Tracking-Daten ein probates Mittel, um die Gefahr der Deanonymisierung veröffentlichter Datensätze zu senken.

Der insgesamten Gefahr können wir als Experimentatoren allerdings nur begegnen, indem wir auf die Veröffentlichung von umfassenden Rohdaten verzichten und lediglich statistisch aggregierte Datensätze verfügbar machen. Beispielhafte Rohdaten-Auszüge können zusätzlich ohne exakte Angabe des Aufnahmezeitpunkts veröffentlicht werden. Je nach Sensor kann eine weitere Maßnahme das bewusste Entfernen etwaiger personenbezogener Merkmale in den Rohdaten sein.

### 5 Fazit

Body-Tracking-Technologie ist nach wie vor weit entfernt von der menschlichen Fähigkeit zur visuellen Erkennung anderer Menschen. Auf diesem Weg erhobene Daten müssen entsprechend kritisch bewertet und eingeordnet werden. Bei der Durchführung von Langzeitstudien mit Body-Tracking-Sensorik empfehlen wir, die Qualität und Konsistenz der Erkennung durch Pilotstudien möglichst mit Menschen verschiedener Körpertypen zu testen und die Annahmen des verwendeten Körpermodells bei der Auswertung ausführlich zu berücksichtigen. Weiterhin sollte die Auswertung und Publikation von Body-Tracking-Daten der Gefahr der Deanonymisierung Rechnung tragen. Nach Möglichkeit sollten Stakeholder aus dem örtlichen Umfeld der Feldstudien in die Planung eingebunden werden, damit eine Vertrauensbasis aufgebaut werden kann.

### Danksagung

Dieses Forschungsprojekt wird gefördert durch die Deutsche Forschungsgemeinschaft (DFG) unter der Projektnummer 451069094.

### Literatur

- [1] Julian Fietkau. 2023. A New Software Toolset for Recording and Viewing Body Tracking Data. In *Mensch und Computer 2023 – Workshopband* (Rapperswil, Schweiz). Gesellschaft für Informatik e.V., Bonn, Deutschland, 4 Seiten. <https://doi.org/10.18420/muc2023-mci-ws13-334>
- [2] Michael Koch, Julian Fietkau, Susanne Draheim, Jan Schwarzer, Kai von Luck. 2023. Methods and Tools for (Semi-)Automated Evaluation in Long-Term In-the-Wild Deployment Studies. In *Mensch und Computer 2023 – Workshopband* (Rapperswil, Schweiz). Gesellschaft für Informatik e.V., Bonn, Deutschland, 5 Seiten. <https://doi.org/10.18420/muc2023-mci-ws13-116>
- [3] Michael Koch, Julian Fietkau, Laura Stojko. 2023. Setting up a long-term evaluation environment for interactive semi-public information displays. In *Mensch und Computer 2023 – Workshopband* (Rapperswil, Schweiz). Gesellschaft für Informatik e.V., Bonn, Deutschland, 5 Seiten. <https://doi.org/10.18420/muc2023-mci-ws13-356>
- [4] Microsoft. 2022. Azure Kinect body tracking joints. <https://learn.microsoft.com/en-us/azure/kinect-dk/body-joints> Abgerufen am 6. Juni 2024.
- [5] Christopher Rohde, Michael Koch, Laura Stojko. 2023. Using an Elastic Stack as a Base for Logging and Evaluation of Public Displays. In *Mensch und Computer 2023 – Workshopband* (Rapperswil, Schweiz). Gesellschaft für Informatik e.V., Bonn, Deutschland, 6 Seiten. <https://doi.org/10.18420/muc2023-mci-ws13-303>
- [6] Jan Schwarzer, Julian Fietkau, Laurenz Fuchs, Susanne Draheim, Kai von Luck, Michael Koch. 2023. Exploring Mobility Behavior Around Ambient Displays Using Clusters of Multi-dimensional Walking Trajectories. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 6 Seiten. <https://doi.org/10.1145/3544549.3585661>
- [7] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake. 2011. Real-time human pose recognition in parts from single depth images. In *CVPR 2011* (Colorado Springs, CO, USA). Institute of Electrical and Electronics Engineers, New York, NY, USA, 1297–1304. <https://doi.org/10.1109/CVPR.2011.5995316>
- [8] Stereolabs. 2021. Body Tracking Overview. <https://www.stereolabs.com/docs/body-tracking> Abgerufen am 6. Juni 2024.
- [9] Jan P. Vox Frank Wallhoff. 2018. Preprocessing and Normalization of 3D-Skeleton-Data for Human Motion Recognition. In *2018 IEEE Life Sciences Conference (LSC)*. Institute of Electrical and Electronics Engineers, New York, NY, USA, 279–282. <https://doi.org/10.1109/LSC.2018.8572153>