der Bundeswehr
Universität München

Kategorisierung von Kinect-Bewegungsdaten mittels deterministischer Algorithmen

Bachelorarbeit von
Laurenz Fuchs
1196307

Draft vom 2. März 2022

Universität der Bundeswehr München
Fakultät für Informatik

der Bundeswehr
Universität München

Kategorisierung von Kinect-Bewegungsdaten mittels deterministischer Algorithmen

Bachelorarbeit von
Laurenz Fuchs
1196307

Erstprüfer:	Prof. Dr. Michael Koch
Zweitprüfer:	Prof. Dr. Gunnar Teege
Betreuer:	M. Sc. Julian Fietkau
Abgabetermin:	31. Mai 2022

Universität der Bundeswehr München
Fakultät für Informatik

Abstract

Hier steht eine kurze Zusammenfassung der Arbeit. Sie darf nicht länger als eine Seite sein, sollte aber mindestens eine halbe Seite umfassen.

Inhaltsverzeichnis

1 Einleitung und Motivation

1.1 Einleitung

1.2 Motivation

Von September 2021 bis August 2024 forschen die Hochschule für angewandte Wissenschaften in Hamburg und die Universität der Bundeswehr München im Rahmen des *HoPE-Projekts* zusammen an Effekten rund um die Steigerung der Aufmerksamkeit bei der Nutzung von großen, interaktiven Wandbildschirmen, sogenannten *Ambient Displays*. Dieser Bereich weist noch einen grundsätzlichen Forschungsbedarf auf. Im Projekt wird unter anderem der *Honeypot-Effekt* erforscht (UniBw, 2021). Dieser beschreibt in der Mensch-Computer-Interaktion (HCI) wie Menschen die mit einem System interagieren weitere Passanten anregen die Interaktion zu beobachten oder sogar an ihr teilzuhaben (Wouters et al., 2016). Der *Honeypot-Effekt* soll bei der Nutzung von *Ambient Displays* im öffentlichen und halböffentlichen Raum in Langzeit-Feldstudien analysiert werden. Dabei soll auch der Aspekt der Datenerhebung und -analyse weiter ausgebaut werden. Hierzu muss ein methodisches Rahmenwerk entwickelt werden, welches eine auf Sensordaten-basierende, automatische und zeitlich uneingeschränkte Evaluation von *Ambient Displays* ermöglicht (UniBw, 2021). Die Sensordaten werden von Body-Tracking-Kameras bereitgestellt. Konkret wurden im die Wandbildschirme im vorliegenden Datensatz mit Kinect v2 Kameras ausgestattet. Diese zeichnen die Interaktion von Nutzern mit den Displays auf, wodurch das Nutzerverhalten zu einem späteren Zeitpunkt ausgewertet werden kann. Besonders interessant ist, welche Muster sich in den Daten erkennen lassen. Lässt sich eine Menge von Kategorien identifizieren, die etwas über das Verhalten von Menschen vor Wandbildschirmen aussagt? Ist die Einteilung in diese Kategorien automatisierbar? Aufgrund der Größe des vorliegenden Datensatzes ist eine manuelle Durchsicht zur Beantwortung der Fragen nicht möglich. In dieser Bachelorarbeit wird versucht, dieses Problem mithilfe von deterministischen Algorithmen zu lösen. Wesentliches Ziel ist die Implementierung eines Systems zur Kategorisierung der vorliegenden Kinect-Bewegungsdaten. Dieses System wird detailiert beschrieben und anschließend evaluiert. Dabei soll geklärt werden, welches Verfahren sich am Besten für den genannten Sachverhalt eignet. Außerdem wird die Frage beantwortet, welches Vorwissen benötigt wird, um das Verfahren einsetzen zu können (z.B. die Vorgabe von konkreten Templates) und welche Datenpunkte zur Analyse interessant sind (z.B. Laufpfade oder Engaged-Werte).

1.3 Kategorien der Interaktion mit Wandbildschirmen

Interaktive digitale Medien sind in der Öffentlichkeit immer präsenter. Deshalb wird es für Wandbildschirme immer schwieriger die Aufmerksamkeit von vorbeigehenden Menschen zu erregen und sie zur Interaktion zu animieren. Diese Herausforderungen können nicht einfach durch verbesserte Hardware oder attraktivere Displays gelöst werden. Stattdessen muss ein besseres Verständnis zwischen Menschen und deren Nutzung von Technologie geschaffen werden (Wouters et al., 2016). *Ambient Displays* sind große, interaktive Bildschirme im (halb-) öffentlichen Raum, mit denen Nutzer interagieren können. Es handelt sich meist um ästhetisch ansprechende Displays die Personen mit Informationen versorgen (Mankoff et al., 2003). Für ein verbessertes Verständnis der Interaktion von Personen mit solchen Wandbildschirmen kann eine Kategorisierung erfolgen. Dazu existieren verschiedenste *Audience Behaviour-Interaktionsmodelle*, wovon im Folgenden zwei näher beschrieben werden.

Das *Audience Funnel Modell* beschreibt, wie Menschen sich um ein großes öffentliches Display versammeln und von Beobachtern zu Interagierenden mit dem System, und anschließend wieder zu Beobachtern werden. Menschen neigen dazu verschiedene Phasen der Interaktivität zu durchlaufen, bevor sie direkt mit dem System interagieren (Wouters et al., 2016; Mai and Hußmann, 2018). Die einzelnen Phasen des *Audience Funnel* werden in Abbildung 1.1 gezeigt. Eine der Aufgaben eines Wandbildschirms ist also Aufmerksamkeit auf sich zu ziehen und den Nutzer zu motivieren mit dem System zu interagieren (Mai and Hußmann, 2018). Mai and Hußmann (2018) verweisen darauf, dass *Ambient Displays* in der Öffentlichkeit nicht undbedingt der zentrale Punkt der Aufmerksamkeit sind, da sie eigene intrinsische Ziele verfolgen. Die Herausforderung für Entwickler ist es die Systeme so zu gestalten, dass sie Aufmerksamkeit erregen, sich aber gleichzeitig nicht gezwungen in den Mittelpunkt stellen.

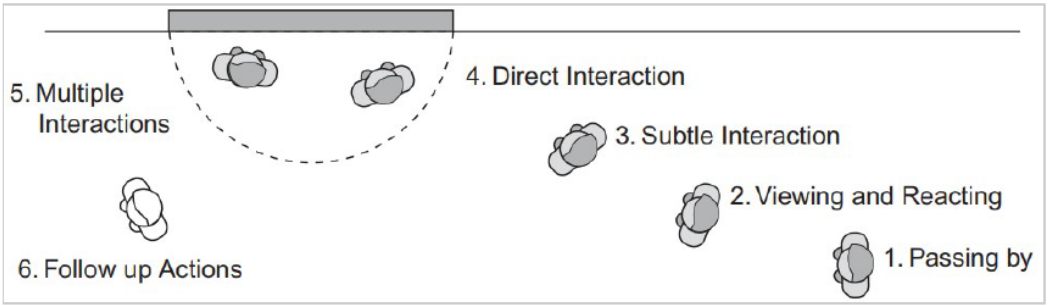

Abbildung 1.1: Audience Funnel Framework. Abbildung aus Mai and Hußmann (2018).

Ein zweites Modell wird durch den bereits erwähnten *Honeypot-Effekt* beschrieben. Dieser Effekt ist ein Einfluss des sozialen Lernens. Er zeigt, dass Individuen undabhängig von Belohnungen, Bestrafungen oder sozialem Wettbewerb von der reinen Präsenz oder den Ak-

tivitäten anderer beeinflusst werden. In der HCI wird dies meist erkennbar, indem Passanten sich einem System nähern und überlegen, ob sie mit ihm interagieren sollen, nachdem sie anderen Menschen dabei zugesehen haben (Wouters et al., 2016).

Eine Einordnung der Bewegungsdaten in solche Kategorien und eine weiterführende Analyse der Bewegungen kann Aufschluss über das Verhalten von Menschen vor interaktiven Bildschirmen geben. Wie bereits erwähnt ist eine Kategorisierung bei einer großen Datenmenge nicht manuell realisierbar. Im Folgenden werden die Grundlagen der verwendeten Sensorik und die Struktur des Datensatzes beschrieben. Aufbauend darauf werden Überlegungen angestellt, wie eine Implementierung zur Automatisierung der Kategorisierung aussehen kann.

1.4 Spezifikation der Kinect

Der Xbox 360 Kinect Sensor war eine Revolution im Bereich der erschwinglichen 3D-Erkennungssensorik. Ursprünglich war er für die Videospiel-Industrie gedacht. Er wurde aber schon bald von Wissenschaftlern verwendet. Später folgten weitere Iterationen der Kinect (Tölgyessy et al., 2021). Im vorliegenden Datensatz des *HoPE-Projekts* kam die Kinect v2 für Xbox One zum Einsatz. Diese Sensorik stellt Farbbilder einer Rot-Grün-Blau (RGB) Kamera, Tiefenbilder einer Tiefenkamera und Audiodateien von verschiedenen Mikrofonen zur Verfügung (Microsoft, 2014). Besonders die Tiefenkamera hilft zuverlässige Ergebnisse bei der Erkennung von Menschen vor *Ambient Displays* zu erzielen. Li et al. (2014) fassen es wie folgt zusammen. Die kompakte Größe, die Benutzungsfreundlichkeit, die stark vereinfachte Hintergrund-Subtraktion im Vergleich zu anderer Sensorik, sowie die hohe Genauigkeit und die hohe Bildrate machen Tiefenkameras zu einer attraktiven Lösung für ein breites Spektrum an Anwendungen. Die Kinect v2 verwendet dabei den Ansatz der kontinuierlichen Wellenintensitätsmodulation, der häufig bei Time-of-Flight (ToF)-Tiefenkameras zum Einsatz kommt. Dabei wird das Licht einer Lichtquelle von Objekten im Sichfeld der Kamera zurückgestreut und die Phasenverzögerung zwischen dem emittierten und dem reflektierten Licht gemessen. Diese Phasendifferenz wird für jedes Pixel im Bildfeld in einen Entfernungswert umgerechnet (Tölgyessy et al., 2021). Der Sensor kann Tiefenbilder mit einer Auflösung von 512 x 424 Pixeln und gewöhnliche Farbbilder mit 1920 x 1080 Pixeln aufnehmen (Marin et al., 2019). Bei der Kinect v2 können bis zu sechs Personen erfasst werden. Dabei wird die Lage von 25 Skelettpunkten, sowie verschiedene Gesichtsattribute erfasst (Microsoft, 2014). Abbildung 1.2 zeigt eine Übersicht dieser Punkte.

1.5 Struktur des vorliegenden Datensatzes

Zur Evaluation des implementierten System dient der Kinect-Datensatz des *HoPE-Projekts*. Hierfür wurden im Jahr 2017 für 18 Wochen zwei Kinect v2 Systeme an einem *Ambient*

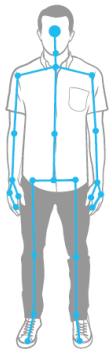

Abbildung 1.2: Skelettpunkte der Kinect v2. Abbildung aus Microsoft (2014).

Display angebracht. Der Versuchsaufbau kann Abbildung 1.3 entnommen werden. Eine Auswertung des Datensatzes ergab, dass dieser 97.626 Datenpunkte beinhaltet (Temiz, 2022). Jeder Datenpunkt enthält mehrere sogenannte Frames. Dabei handelt es sich um Momentaufnahmen der Kinect-Sensorik. Zu jedem Datenpunkt liegen verschiedene Dateien vor, die jeweils den Zeitstempel in Kombination mit einem geeigneten Postfix als Dateinamen tragen:

- timestamp.txt
- timestamp_bodies.txt
- timestamp_summary.txt
- timestamp.xef

Die Datei *timestamp.txt* enthält folgende Attributwerte:

1. Zeitstempel: Zeitpunkt der Aufnahme
2. KinectId
3. RecordId: Identifikationsnummer des Records
4. BodyIndex: Identifikationsnummer der Person
5. BodyCount: Anzahl der erfassten Personen
6. Happy: Person ist glücklich
7. Engaged: Person zeigt Interesse
8. WearingGlasses: Person trägt eine Brille
9. LeftEyeClosed: Person hat das linke Auge geschlossen

10. RightEyeClosed: Person hat das rechte Auge geschlossen
11. MouthOpen: Person hat den Mund geöffnet
12. MouthMoved: Person bewegt den Mund
13. LookingAway: Person schaut nicht zum Kinect-Sensor
14. Body.HandLeftState: Zustand der linken Hand
15. Body.HandRightState: Zusand der rechten Hand
16. X: x-Koordinate des Skelettpunkts SpineShoulder
17. Y: y-Koordinate des Skelettpunkts SpineShoulder
18. Z: z-Koordinate des Skelettpunkts SpineShoulder
19. Distance: Distanz zwischen Kinect-Sensor und Skelettpunkt SpineShoulder

Diese Attribute werden in der Textdatei durch drei Rauten ('###') voneinander abgegrenzt. Die Datei *timestamp_bodies.txt* enthält für jeden Frame die Position aller 25 Skelettpunkte. *timestamp_summary.txt* zeigt eine kurze Zusammenfassung des Datenpunkts, welche gut von Menschen lesbar ist und *timestamp.xef* kann genutzt werden, um den Datenpunkt in der Anwendung Kinect Studio zu visualisieren. Letzlich sind die Einträge sogenannte *Time Series Data*. Bei diesem Datentyp handelt es sich um geordnete Sequenzen von Datenpunkten, die über eine gewisse Zeit hinweg aufgenommen werden. Oft in regelmäßigen Abständen (Ali et al., 2019). Jeder Datenpunkt beinhaltet durchschnittlich 355 Frames, wodurch eine Gesamt-Frameanzahl von 34.687.630 entsteht (Temiz, 2022). An der Größe zeigt sich erneut die Notwendigkeit eines Software-Tools zur Auswertung.

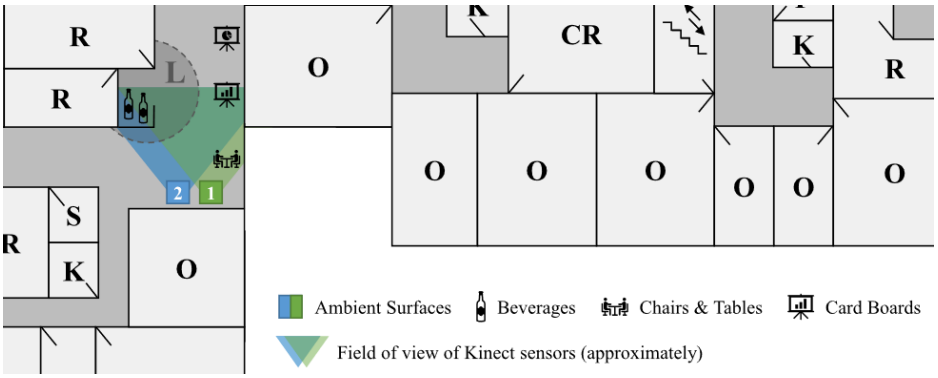

Abbildung 1.3: Kinect-Setup des Datensatzes. Abbildung von Jan Schwarzer ToDo

2 Mustererkennung in Bewegungsdaten mittels deterministischer Algorithmen

2.1 Dynamic Time Warping Algorithmus

2.2 Related Work Analyse

2.2.1 DTW

2.2.2 K-Means

2.3 Vergleich der Algorithmen

3 Konzeption

3.1 Anforderungsanalyse

3.2 Programmablauf

3.3 Teilsysteme

4 Implementierung

4.1 Grundlagen

4.1.1 Entwicklungsumgebung

4.1.2 Programmiersprache

4.2 Aufbau

4.3 Herausforderungen

4.3.1 Fehler in den Daten

4.4 Codebeschreibung

4.4.1 Verwendete Konzepte

4.4.2 Komplexität

5 Evaluation

5.1 Auswertung des Datensatzes

5.2 Übertragbarkeit

6 Fazit

Abkürzungsverzeichnis

HCI	Mensch-Computer-Interaktion
RGB	Rot-Grün-Blau
ToF	Time-of-Flight

Abbildungsverzeichnis

Literaturverzeichnis

Ali, M., Alqahtani, A., Jones, M. W., and Xie, X. (2019). Clustering and classification for time series data in visual analytics: A survey. *IEEE Access*, pages 181314–181338. Conference Name: IEEE Access.

Li, L. et al. (2014). Time-of-flight camera–an introduction. *Technical white paper*. SLOA190B.

Mai, C. and Hußmann, H. (2018). The audience funnel for head-mounted displays in public environments. In *2018 IEEE 4th Workshop on Everyday Virtual Reality (WEVR)*, page 5.

Mankoff, J., Dey, A. K., Hsieh, G., Kientz, J., Lederer, S., and Ames, M. (2003). Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 169–176.

Marin, G., Agresti, G., Minto, L., and Zanuttigh, P. (2019). A multi-camera dataset for depth estimation in an indoor scenario. *Data in Brief*, page 104619.

Microsoft (2014). Human interface guidelines v2.0.

Temiz, J. (2022). Konzeption und implementierung eines datenanalyse-werkzeugs für body-tracking-kameras. Bachelor's thesis, Universität der Bundeswehr München.

Tölgyessy, M., Dekan, M., Chovanec, L., and Hubinský, P. (2021). Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors*, page 413.

UniBw (2021). Honeypot-effekt an interaktiven ambient displays (HoPE) — inf2.

Wouters, N., Downs, J., Harrop, M., Cox, T., Oliveira, E., Webber, S., Vetere, F., and Vande Moere, A. (2016). Uncovering the honeypot effect: How audiences engage with public interactive systems. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 5–16.

Hiermit versichere ich, die vorliegende Arbeit selbständig und ohne fremde Hilfe verfasst, die Zitate ordnungsgemäß gekennzeichnet und keine anderen, als die im Literatur/Schriftenverzeichnis angegebenen Quellen und Hilfsmittel benutzt zu haben.

Ferner habe ich vom Merkblatt über die Verwendung von studentischen Abschlussarbeiten Kenntnis genommen und räume das einfache Nutzungsrecht an meiner Bachelorarbeit der Universität der Bundeswehr München ein.

München, den 31. Mai 2022

.....

(Unterschrift)