

Auswertungsverfahren von Body-Tracking-Daten

Bachelorarbeit

Fabian Schaaf

1196508

Erstprüfer: Prof. Dr. Michael Koch
Zweitprüfer: Prof. Dr. Gunnar Teege
Betreuer: Julian Fietkau

Universität der Bundeswehr München
Fakultät für Informatik

Inhaltsverzeichnis

1	Einleitung	1
2	Erklärung der Methoden	2
2.1	Manuelle Auswertung	2
2.2	Algorithmus	3
2.3	Machine Learning	4
3	Literatur-Review	8
3.1	Review Eingrenzung	8
3.2	Vorgehensweise Suche	8
4	Aktuelle Anwendungsbereiche der Verfahren	11
4.1	Verwendung der manuellen Auswertung	11
4.2	Verwendung eines Algorithmus	11
4.3	Verwendung von Machine Learning	16
5	Vergleich der Auswertungsverfahren	20
5.1	Vorteile und Nachteile der Verfahren	20
5.1.1	Manuelle Auswertung	20
5.1.2	Algorithmus	20
5.1.3	Machine Learning	21
5.2	Vergleich der Verfahren	22
5.3	Optimaler Anwendungsbereich eines Verfahrens	24
6	Fazit	26
	Literatur	27
	Abbildungsverzeichnis	30

1 Einleitung

Die Verarbeitung von Body-Tracking-Daten ist ein wichtiger Bestandteil für viele Aspekte des heutigen Lebens. Dabei erstreckt sich das Feld der Anwendungen von Werbung zu sportlichen Auswertungen über die Überwachungs- und Sicherheitssysteme bis hin zu Videospiele. Hierzu werden diverse Auswertungsverfahren benutzt, welche für den jeweiligen Anwendungsbereich optimiert wurden. Durch die Weiterentwicklung und Verbreitung von Kameras werden auch die Auswertungsverfahren immer weiterentwickelt und sind stark verbreitet. Dadurch wurden noch mehr Anwendungsbereiche geschaffen und noch mehr Auswertungsverfahren entwickelt.

Diese Arbeit beschäftigt sich mit den verschiedenen Auswertungsverfahren von Body-Tracking-Daten. Die diversen Verfahren werden durch ein Literatur-Review recherchiert, dann einzeln analysiert und zuletzt zusammenfassend verglichen. Weitere wichtige Aspekte der Verfahren werden ebenfalls dargestellt. Dadurch soll ein qualifizierter Überblick über benutzte Auswertungsverfahren erstellt werden und zwischen den Verfahren differenziert werden können. Im Verlauf der Arbeit soll auch ein optimaler Anwendungsbereich für ein Auswertungsverfahren genau dargestellt werden, um die zuvor analysierten Verfahren bestmöglich nutzen zu können. Die Arbeit soll also den aktuellen Stand der verschiedenen Auswertungsverfahren präsentieren und die Verfahren differenziert vergleichen können. Dafür wird besonders auf die Analyse der Auswertungsverfahren Wert gelegt und eine Vergleichstabelle für eine gute Übersicht angefertigt.

Durch die Strukturierung soll ein möglichst präziser Überblick über die aktuellen Auswertungsverfahren von Body-Tracking-Daten erzeugt werden und dadurch die verschiedenen Verfahren optimal für den jeweiligen Anwendungsbereich ausgewählt werden. Durch den Vergleich sollen Parallelen gezogen werden, um die Verfahren mit Verfahren aus anderen Bereichen zu verknüpfen und möglicherweise neue Anwendungsbereiche für schon benutzte Verfahren zu finden.

Bevor die verschiedenen Auswertungsmethoden analysiert und verglichen werden, werden zunächst die Kategorien der Auswertungsmethoden näher dargestellt und mit Anwendungsbeispielen erläutert. Dann werden durch ein Literatur-Review die Quellen der Verfahren eingegrenzt, um diese Verfahren dann analysieren zu können. Nach der Analyse werden die Verfahren verglichen und zuletzt ein Beispiel und die Herangehensweise für die optimale Nutzung dargestellt.

2 Erklärung der Methoden

2.1 Manuelle Auswertung

Die manuelle Auswertung wird im folgendem Kapitel näher erläutert und es werden Beispiele der Verwendung geliefert.

Die manuelle Auswertung ist ein Verfahren, welches ohne weitere Programme oder technischer Unterstützung funktioniert. Sie ist also unabhängig und geht nur von dem Menschen aus. Die manuelle Auswertung ist daher die älteste Methode und es gibt zahlreiche verschiedene Anwendungen.

Die Auswertung kann je nach auswertender Person auch variieren. Eine oder mehrere Personen beobachten ein Szenario und dokumentieren dann gegebene Situationen oder Beobachtungen und führen dadurch eine Auswertung durch. Es werden meistens mehrere manuelle Auswertungen von unterschiedlichen Menschen durchgeführt, wobei die auswertenden Personen oft auch in dem Fachgebiet belehrt sind. „Diese klassische und sehr sichere Methode eignet sich grundsätzlich für alle papierbasierenden Arten von Dokumenten. Sie wird heutzutage vornehmlich für schwer lesbare Belege, überwiegend per Hand ausgefüllte Formulare sowie Belegen mit viel Fließtext eingesetzt.“ (DESH, 2023).

Ein wichtiger Aspekt einer Auswertungsmethode ist auch die Fehlerquote. Diese kann bei der manuellen Auswertung stark variieren, denn die Fachkompetenz aber auch Konzentration oder Ermüdung sind wichtige Faktoren. Daher kann man von einer Fehlerquote von mindestens 5 Prozent ausgehen, welche sich aber auch durch Qualitätssicherung verbessern lässt. (inserve, 2023).

Bei neueren Datenauswertungen werden die Daten von Menschen erfasst und dann, teilweise auch zur Weiterverarbeitung, in einer Datenbank oder ähnliches abgespeichert. Ein Aufgabenfeld von solchen Datenbanken dient beispielsweise dem Vergleich beziehungsweise der Kontrolle von anderweitig erhobenen Daten. Die manuelle Auswertung dient also auch oft zur Überwachung oder Kontrolle von anderen Datenauswertungen.

Alltägliche, beziehungsweise die meist genutzten Szenarien für eine manuelle Auswertung, sind beispielsweise Umfragen von Kunden oder Passanten, Beobachtung und damit verbundene Messungen, Dokumentation von Zählerständen oder Datensätzen und auch Dokumentation in Form von Texten wie Berichte.

Die manuelle Datenerfassung und Datenauswertung wird sehr oft durch automatische Datenauswertung ersetzt, da es berechenbarer und oft auch finanziell sinnvoller ist. Es ist also eher die

Was ist ein Algorithmus?

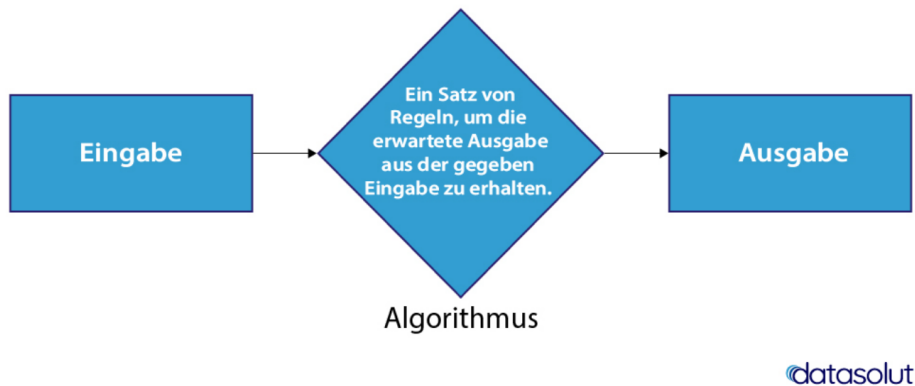


Abbildung 1: Funktionsweise Algorithmus (Wuttke, 2021)

letzte Alternative für schwierig verarbeitbare Prozesse. Sie ist nicht mehr die aktuelle Methode und wird in Zukunft auch noch weiter abnehmen. Im Bezug zu den Auswertungsverfahren bei Body-Tracking-Daten wird die manuelle Auswertung als unterstützende und kontrollierende Methode verwendet.

2.2 Algorithmus

Nachdem die manuelle Auswertung näher erläutert wurde, wird nun der Algorithmus erklärt und reale Anwendungsbeispiele genannt.

Ein Algorithmus ist ein Verfahren in der Informatik oder Mathematik. „Grundsätzlich handelt es sich bei einem Algorithmus um eine formal festgelegte Vorgehensweise, nach der eine definierte Aufgabe gemäß einem strukturierten Schema gelöst wird.“ (Wuttke, 2021). Es gibt also einen Eingabewert der durch ein festgelegtes Verfahren einen entsprechenden Ausgabewert ausgibt, siehe Abbildung 1. Man kann einen Algorithmus als eine automatisierte Anleitung bezeichnen, in der die einzelnen Schritte beschrieben sind, die die Daten durchlaufen um zum Ergebnis zu kommen.

Es kann ein einfacher Schritt sein oder eine komplexe Rechnungsfolge oder auch aufeinanderfolgende Prozesse. Der Komplexität ist also keine Grenze gesetzt solange der Algorithmus seine Eigenschaften beibehält.

Folgende Eigenschaften sind für einen Algorithmus in der Informatik definiert:

Determiniertheit: Bei gleichen Voraussetzungen muss ein Algorithmus immer das gleiche Ergebnis liefern.

2 Erklärung der Methoden

Determinismus: Es besteht in jedem Schritt nur eine Möglichkeit, die richtig ist. Nur mit dieser richtigen Möglichkeit kann der Algorithmus zum Folgeschritt der Problemlösung übergehen. Der Ablauf ist also eindeutig festgelegt.

Ausführbarkeit: Jeder einzelne Schritt muss zwingend ausführbar sein.

Eindeutigkeit: In einem Algorithmus darf es keine widersprüchlichen Bedingungen oder Beschreibungen geben.

Terminierung: Der Algorithmus muss nach endlich vielen Schritten zu einem Ergebnis kommen und das Lösungsverfahren abschließen.

Finitheit (Endlichkeit): Der Algorithmus selbst verfügt über eine begrenzte Anzahl an Zeichen und hat somit eine endliche Beschreibung. Zumal ist der von ihm belegte Speicherplatz in jedem Schritt begrenzt.

(Sandmann, 2022).

Eine genaue Definition für einen Algorithmus beinhaltet eine Turingmaschine. Diese Definition besagt, dass ein Algorithmus ein Verfahren ist, welches eine Turingmaschine ausführen kann oder selbst eine Turingmaschine ist (Höpping, 2023).

Häufige Anwendungen findet ein Algorithmus in alltäglichen Situationen. Vom Ampelverkehr bis hin zum Autofahren sind Algorithmen sehr weit verbreitet und prägen moderne Methoden im Alltag. Auch Simulationen zum Gebäudebau oder Wettervorhersagen basieren auf Algorithmen. Zusammenfassend werden also viele Berechnungen und Verfahren heutzutage durch Algorithmen gelöst, weshalb Sie auch essentiell für die heutige Zeit sind. Im Bezug zu den Auswertungsverfahren bei Body-Tracking-Daten werden Algorithmen zum Beispiel zur Berechnung von Bewegungen und erwarteten Positionen der Körperteile verwendet.

2.3 Machine Learning

Abschließend wird das Machine Learning erklärt und verschiedene Aufgabenbereiche von Machine Learning gezeigt.

Machine Learning gehört zur künstlichen Intelligenz und befasst sich mit dem Lernpotenzial einer Maschine. In Deutschland wird es auch maschinelles Lernen genannt. Das besondere beim Machine Learning ist die Herangehensweise des Programmierens. Es wird nicht die genaue Problemlösung programmiert und muss daher auch nicht für den Menschen ersichtlich sein. Nur das Ziel muss bekannt sein und man braucht geeignete Übungsdatensätze. Der genaue Lösungsansatz und Lösungsweg wird ohne Programmieren erstellt, sondern antrainiert. Dies grenzt Machine Learning sehr von den anderen Algorithmen ab (Rebala et al., 2019).

„Maschinelles Lernen kann automatisiert Wissen generieren, Algorithmen trainieren, Zusammenhänge identifizieren und unbekannte Muster erkennen. Diese identifizierten Muster und Zusammenhänge lassen sich auf einem neuen, unbekanntem Datensatz anwenden, um so

2 Erklärung der Methoden

Vorhersagen zu treffen und Prozesse zu optimieren.“ (Wuttke, 2022). Dies wird in Abbildung 2 gezeigt.

Dieses selbstständige Lernen ist der Kern von Machine Learning und dessen Identifikationsmerkmal. Das Training wird von einem Menschen übernommen und Datensätze werden verarbeitet. Es gibt jedoch verschiedene Ansätze von Machine Learning.

Es gibt ein überwachtes Lernen bei dem vorbereitete Datensätze zur Anlernung benutzt werden und die Prognosen und erkannten Muster geprüft werden. Dies wird teilweise oft wiederholt, bis man eine genaue Prognose bekommt und die Zusammenhänge gut erfasst wurden. Dann wendet man diese Methode auf unbekannte, neue Datensätze an, um genaue Prognosen und Zusammenhänge von dem unbekanntem Datensatz zu ermitteln. Dies wird oft für Prognosen von Verbrauch oder auch am Aktienmarkt benutzt (Wuttke, 2022).

Unüberwachtes Lernen gibt zuerst auch bekannte Datensätze ein, jedoch nicht um das Machine Learning zu verbessern, sondern um neue Ansätze, versteckte Muster oder neue Zusammenhänge von bereits bekannten Daten zu bekommen. Es wird also nicht auf eine genaue Ausgabe angelernt, sondern zur Überprüfung von bekannten Datensätzen genutzt (Wuttke, 2022). Diese Methode ist meistens bei Clusteranalysen in Gebrauch.

Zusätzlich gibt es auch eine Mischung und zwar das teilüberwachte Lernen. Dort werden sowohl bekannte als auch unbekannte Datensätze ausgewertet und beschleunigen durch die vorbereiteten Daten den Lernprozess ohne eine bestimmte Lösung zu forcieren. Dies wird oft bei Bild- oder Objekterkennung benutzt.

Zuletzt gibt es auch noch das verstärkende Lernen. Diese Methode benötigt keine vorgefertigten Datensätze, sondern lernt durch Feedback die richtige Aktion in der richtigen Situation auszuführen. Aufgrund des Feedbacks lernt der Algorithmus also selbstständig eine Strategie oder Lösung für ein Problem oder eine Situation zu finden. Zum Anlernen muss man also nicht den Weg kennen sondern nur das gewünschte Endergebnis und alles Weitere wird dem Machine Learning überlassen. Dies ist daher auch eine vielversprechende Methode für sehr komplexe Situationen, wie zum Beispiel autonomes Fahren oder Robotik (Wuttke, 2022). Im Bezug zu den Auswertungsverfahren bei Body-Tracking-Daten wird Machine Learning zur Verfolgung und Erkennung markanter Punkte benutzt, um dadurch die Bewegungen sehr genau antizipieren zu können.

Eine Übersicht für die verschiedenen Methoden und Anwendungsbeispielen zeigt Abbildung 3.

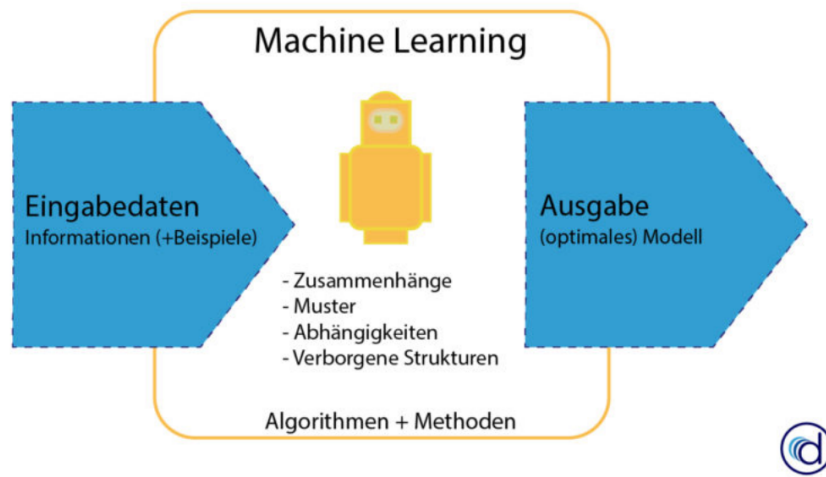


Abbildung 2: Funktionsweise Machine Learning (Wuttke, 2022)

2 Erklärung der Methoden

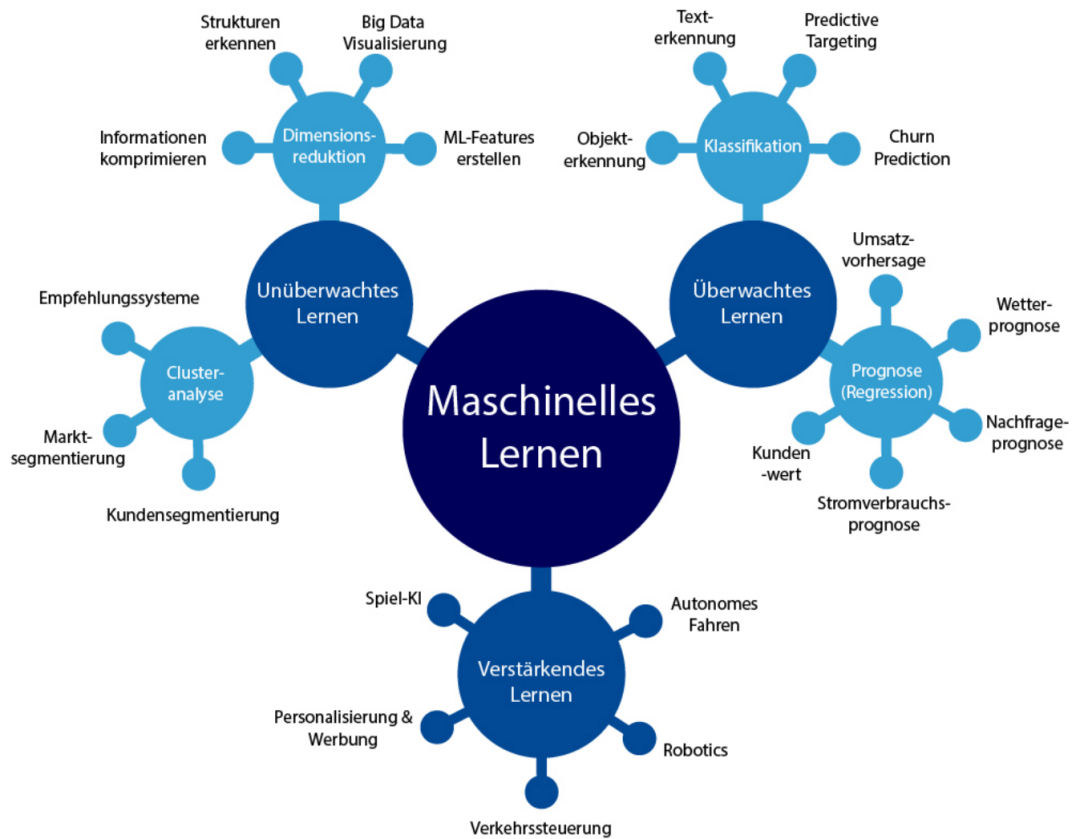


Abbildung 3: Maschinelles Lernen im Überblick: Anwendungsbeispiele nach Arten (Wuttke, 2022)

3 Literatur-Review

3.1 Review Eingrenzung

Im folgendem Kapitel werden nötige Anforderungen an die Paper näher artikuliert und dadurch die Literatur dieser Arbeit eingegrenzt.

Die Bachelorarbeit dient zur Erfassung und Auswertung der aktuell genutzten Auswertungsverfahren für Body-Tracking-Daten in 3D. Die Auswertungsverfahren bei Body-Tracking-Daten mit 3D-Kameras sind noch nicht gut etabliert und deshalb werden mehrere Verfahren in der Praxis genutzt. Um die verschiedenen Verfahren erfassen und analysieren zu können, braucht man erst einen Überblick über die Anwendungen. Dafür ist ein Literatur-Review sehr nützlich.

Durch Sortierung und Filterung werden die am meisten benutzten Methoden gefunden und man hat einen Überblick der in wissenschaftlichen Experimenten benutzten Verfahren. Um geeignete Paper zu erhalten, welche Body-Tracking-Daten auswerten, muss das Paper zunächst zum Beispiel durch wissenschaftliche Zeitungen, Bücher oder Konferenzen veröffentlicht werden. Ein weiteres Kriterium ist die Erwähnung der benutzten Methode zur Auswertung und nicht nur die Annahme, welche Methode benutzt wird.

Nur Paper mit klar ersichtlichen Auswertungsverfahren werden berücksichtigt. Ob die Auswertung im Fokus des Papers liegt ist dabei irrelevant, denn nur die Benutzung eines Verfahrens qualifiziert das Paper als Quelle. Dies bedeutet, dass jedes Paper, welches mit Body-Tracking-Daten arbeitet und diese mit einem bekannten Verfahren auswertet, eine geeignete Quelle ist. Dabei wird auch auf die Art und Größe der Auswertung geachtet, um diese später als Vergleichswerte einbeziehen zu können. Die Datenerhebungen, welche mit zusätzlichen tragbaren Systemen erfolgt sind, werden auch nicht berücksichtigt, da die Auswertungsverfahren sehr individuell und nicht so universell verwendbar sind wie die Auswertungsmethoden, welche nur Kameras zur Datenerhebung benutzen.

3.2 Vorgehensweise Suche

Gemäß der Eingrenzung folgt nun die Suche. Dabei wird der Verlauf der Suche und die vorgenommenen Anpassungen detailliert dargestellt.

Die erste Suche bei den Fachdatenbanken Google Scholar, Springer und IEEE mit den Stichworten „Body Tracking“ bringt kaum relevante Ergebnisse mit sich, da keine 3D Daten verarbeitet werden. Es werden hauptsächlich über 10 Jahre alte Quellen gezeigt, die sich mit

2D Body-Tracking-Daten beschäftigen. Deshalb ist die neue Suche „3D Body Tracking“ viel erfolgreicher. Es gibt sehr viel relevante und geeignete Literatur. Dies bildet die Grundlage der Literatur und soll nun im weiteren genauer gefiltert werden, um einzelne Aspekte weiter zu beleuchten und die Entwicklung der Verfahren besser wider zu spiegeln.

Ab diesem Punkt gibt es viele relevante Quellen, die für die Bachelorarbeit auch benutzt werden. Man kann die ersten Informationen und Daten sammeln und die ersten Auswertungsverfahren sind klar geschildert. Im Verlauf der Recherche wird ein Filter des Veröffentlichungszeitraumes hinzugefügt und angepasst. Der Filter berücksichtigt erstmal alle Veröffentlichungen bis 2000. Dies wird dann von 2000 bis 2004 angepasst. In diesen Intervallen, die 5 Jahre andauern, wird die Suche bis zu jetzigen Veröffentlichungen fortgeführt und dadurch sollen verschiedene Methoden und die Entwicklung der Verfahren besser erfasst werden. Wenn eine neuere Quelle eine ähnliche Methode verwendet und optimiert hat, ersetzt sie die ältere Quelle.

Der letzte Filter berücksichtigt nur Quellen aus den letzten 3 Jahren, um aktuellere Ergebnisse zu filtern. Dadurch soll ein aktuelleres Bild der aktuell benutzten Auswertungsverfahren erzeugt werden und gegenwärtige Quellen hervorbringen. Nach dem Einsetzen des neuen Filters gibt es sowohl Quellen die Algorithmen benutzen, als auch die ersten Experimente mit Deep Machine Learning. Es fallen immer wieder Quellen durch die Kriterien aufgrund einer nicht klar ersichtlichen Methode oder keiner realen Anwendung. Dies ist jedoch nur ein Bruchteil, da die meisten Paper Experimente durchführen und die Auswertung beschreiben. Es gibt auch Paper wie Zihajehzadeh et al. (2015), die tragbare inertielle Messeinheiten verwenden und daher nicht qualifiziert sind. Dazu zählt auch Drobnjakovic et al. (2018), da zusätzlich zu einer Kamera auch inertielle Messeinheiten benutzt werden. Auch Feigl et al. (2018) benutzt einen Sensor, der am Kopf der Person angebracht wird und daher nicht weiter berücksichtigt wird. Die in Liu et al. (2022) benutzten tragbaren Elektromyographie-Sensoren qualifizieren sich auch nicht für die Arbeit. Choutas et al. (2022) benutzt eine moderne Methode des Machine Learning, verwendet jedoch als Input nur ein 2D Foto und wird daher nicht berücksichtigt. Insgesamt wurde aus dem Intervall 2000 bis 2004 keine Quelle verwendet. Von 2005 bis 2009 wurden 2 Quellen verwendet. Im nächsten Intervall von 2010 bis 2014 gab es 3 relevante Quellen. Das Intervall 2015 bis 2019 lieferte auch 3 Quellen. Im letzten Intervall von 2020 bis 2023 wurden 2 Quellen verwendet.

Durch die bisherige Suche hat man einen Überblick über den Verlauf der Auswertungsverfahren und die Präsenz der einzelnen Methoden in der Literatur. Des Weiteren kann man die Erkenntnisse auch in zeitliche Phasen unterteilen, um die Entwicklung zu verdeutlichen. Nachdem man ein gutes und großes Bild der Literatur zu 3D Body-Tracking-Daten erzeugen konnte, werden im Nachhinein die einzelnen Methoden genau recherchiert. Dazu werden die Methoden an das „3D Body Tracking“ Suchkriterium angehängt.

Eine Suche der Methoden in den Fachdatenbanken auf deutsch ergab keine relevanten Quellen, weshalb sich die Suche nun auf ausschließlich englische Suchbegriffe beschränkt.

Die manuelle Auswertung hat in der Literatur keine eigenständigen Auswertungen, sondern

wird nur zusätzlich benutzt, weshalb die Suche nur wenige Ergebnisse liefert und sehr beschränkt ist.

Algorithmen werden sehr oft benutzt und man findet auch überschneidende Algorithmen, welche auch genauer recherchiert werden. Die einzelnen Algorithmen wie Particle Filters und Frameworks liefern auch viele Ergebnisse und sind sehr verbreitet in der gesuchten Literatur. Die Quellen bauen auch aufeinander auf und es lässt sich einfach ein großer Überblick bilden.

Die Suche nach Machine Learning bietet einige aktuelle Literatur, welche auch auf sich aufbaut und eine Weiterentwicklung des Verfahrens suggeriert. Eine Suche nach den einzelnen Modellen zeigt viele ähnliche Quellen auf.

Abschließend wurden weitere Daten, abgesehen von der Auswertungsart, kaum berücksichtigt und angegeben, da viele Projekte neue Methoden vorgestellt haben und nur kleine abschließende Experimente durchgeführt haben. Die Auswertungsverfahren der neu vorgestellten Methoden werden also stark in den Vordergrund der Arbeit rücken und die Rahmenbedingungen der daraus folgernden Auswertung vernachlässigt.

Dadurch ist sowohl die Literatur des gesamten Prozesses, als auch die Literatur der Auswertungsverfahren, detailliert erforscht worden. Abschließend werden die Begriffe ohne den Zusammenhang mit Body-Tracking-Daten recherchiert um die Methoden von Grund auf erklären zu können. Diese Literatursuche ist bei den Verfahren sehr simpel und man kann schnell die Methoden und ihre Funktionsweisen erklären. Die durch das Literatur-Review erbrachten Kenntnisse werden in den nachfolgenden Kapiteln benutzt und bieten die wissenschaftliche Grundlage der Literaturarbeit.

4 Aktuelle Anwendungsbereiche der Verfahren

4.1 Verwendung der manuellen Auswertung

Im folgendem Kapitel werden Anwendungsbereiche der manuellen Auswertung, aus der vorher definierten Literatur, vorgestellt.

Das in Abschnitt 2.1 erklärte Verfahren der manuellen Auswertung wird in der Praxis nicht oft benutzt. Falls das Verfahren angewendet wird, wird es selten als einzige Auswertungsmethode benutzt. Es wird oft als ergänzendes oder überwachendes Verfahren benutzt und kommt daher nur sekundär zum Einsatz. Im folgendem Text werden Anwendungsbeispiele vorgestellt.

In Alt et al. (2021) geht es um ein Projekt, welches Kinect-Kameras für die Erkennung und Auswertung von Skelett-Daten eingesetzt hat, um zu erkennen, zu welchem Zeitpunkt eine Person von einem öffentlichem Display weggehen wird.

Dazu wird in Alt et al. (2021) eine manuelle Auswertung mit mehreren Beteiligten vorgenommen. Die Beteiligten bekommen das aufgezeichnete Videomaterial und müssen den Zeitpunkt des Verlassens vorhersagen und dann dokumentieren (Alt et al., 2021). Diese manuelle Auswertung dient zur Testung der menschlichen Fähigkeiten der Vorhersage und benutzt später diese Erkenntnisse um Machine Learning anzulernen (Alt et al., 2021). Die manuelle Auswertung dient also sowohl zur Einordnung der Effektivität des Machine Learning im Vergleich zu dem Menschen, als auch für die Anlehnung von Machine Learning und hat daher eine unterstützende Funktion dem Machine Learning gegenüber (Alt et al., 2021).

In Hesse et al. (2020) wird ein Verfahren vorgestellt, um ein 3D Skinned Multi-Infant Linear body model (SMIL) für Säuglinge anzulehren und dadurch genug Bewegungsdaten der Säuglinge zu sammeln, um neurologische Entwicklungsstörungen früh zu erkennen.

Hesse et al. (2020) benutzt keine manuelle Auswertung, jedoch ist die Auswertung von einer manuellen Intervention stark abhängig, da die Bilder zur Anlehre manuell angepasst und erzeugt werden. Darüber hinaus werden auch die fehlerhaften Posen der Säuglinge manuell aussortiert, da es dafür keinen geeigneten Algorithmus gibt (Hesse et al., 2020). Also wird auch hier die manuelle Auswertung als zusätzliches Verfahren zur Kontrolle und Unterstützung benutzt.

4.2 Verwendung eines Algorithmus

Nach den Verwendungsbeispielen der manuellen Auswertung, wird nun der Algorithmus in den Fokus gesetzt. Deshalb werden die in der Literatur benutzten Algorithmen näher betrachtet

und analysiert.

Der in Abschnitt 2.2 erklärte Algorithmus ist das häufigst verwendete Auswertungsverfahren. Es gibt verschiedene Verwendungsmöglichkeiten für Algorithmen, weshalb er sehr universell verwendbar ist. Im Folgendem werden verschiedene Ausprägungen von benutzten Algorithmen beispielhaft vorgestellt.

In Knoop et al. (2006) wird ein Tracking-System zur 3D-Verfolgung von Bewegungen eines menschlichen Körpers vorgestellt. Dazu wird ein 3D Körpermodell und der iterative closest point (ICP) Algorithmus verwendet.

Der ICP Algorithmus berechnet von zwei Sätzen von Punkten, welche aus unterschiedlichen Koordinatensystemen stammen, die Translation und Rotation, um das erste Koordinatensystem in das zweite Koordinatensystem zu transformieren (Knoop et al., 2006). Bei Body-Tracking entspricht der erste Satz den Punkten des Sensors und der zweite Satz den Punkten auf dem Körper. (Knoop et al., 2006)

Der erste Satz Daten ist als $P = \{\vec{p}_i\}$ und der zweite Satz ist als $X = \{\vec{x}_i\}$ beschrieben (Knoop et al., 2006). Sie haben dieselbe Größe und jeder Punkt \vec{p}_i entspricht dem Punkt \vec{x}_i (Knoop et al., 2006). Da die Sensordaten immer mit Rauschen verfälscht sind, gibt es keine exakte Lösung, weshalb das Problem in Knoop et al. (2006) stattdessen in die Minimierung einer Summe quadrierter Entfernungen umgewandelt wird:

$$f(\mathbf{R}, \vec{t}) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{R}(\vec{x}_i) + \vec{t} - \vec{p}_i\|^2 \quad (1)$$

(Knoop et al., 2006)

Es gibt verschiedene Varianten des Partikelfilters. Die einfachste Variante ist der in Abbildung 4 erklärte generische Partikelfilter (Salih & Malik, 2011).

Darauf aufbauend gibt es den Kondensationsalgorithmus. Dieser Filter zielt darauf ab, die vorgeschlagene Verteilung in zwei Schritten zu verbessern:

Der erste Schritt ist die deterministische Abweichung, bei der der Prior basierend auf den akkumulierten Gewichten für jedes Partikel neu erfasst wird (Salih & Malik, 2011). Dies trägt dazu bei, das Entartungsphänomen zu verhindern, indem Partikel mit hohem Gewicht gehalten und diejenigen mit geringerem Gewicht entfernt werden, bevor ein neuer Satz von Partikeln erzeugt wird (Salih & Malik, 2011).

Der zweite Schritt ist die stochastische Diffusion, bei der durch einen Zufallsprozess aus dem Ergebnis der deterministischen Abweichung neue Teilchen generiert werden (Salih & Malik, 2011). Der Rest dieses Algorithmus ähnelt der allgemeinen Partikelfiltertechnik (Salih & Malik, 2011).

Darüber hinaus gibt es auch noch den Hilfspartikelfilter. Der Hilfspartikelfilter versucht, das Beobachtungsmodell des allgemeinen Partikelfilters zu verbessern, indem er Hilfgewichte verwendet, die vor dem Ziehen neuer Partikel berechnet werden (Salih & Malik, 2011). Diese

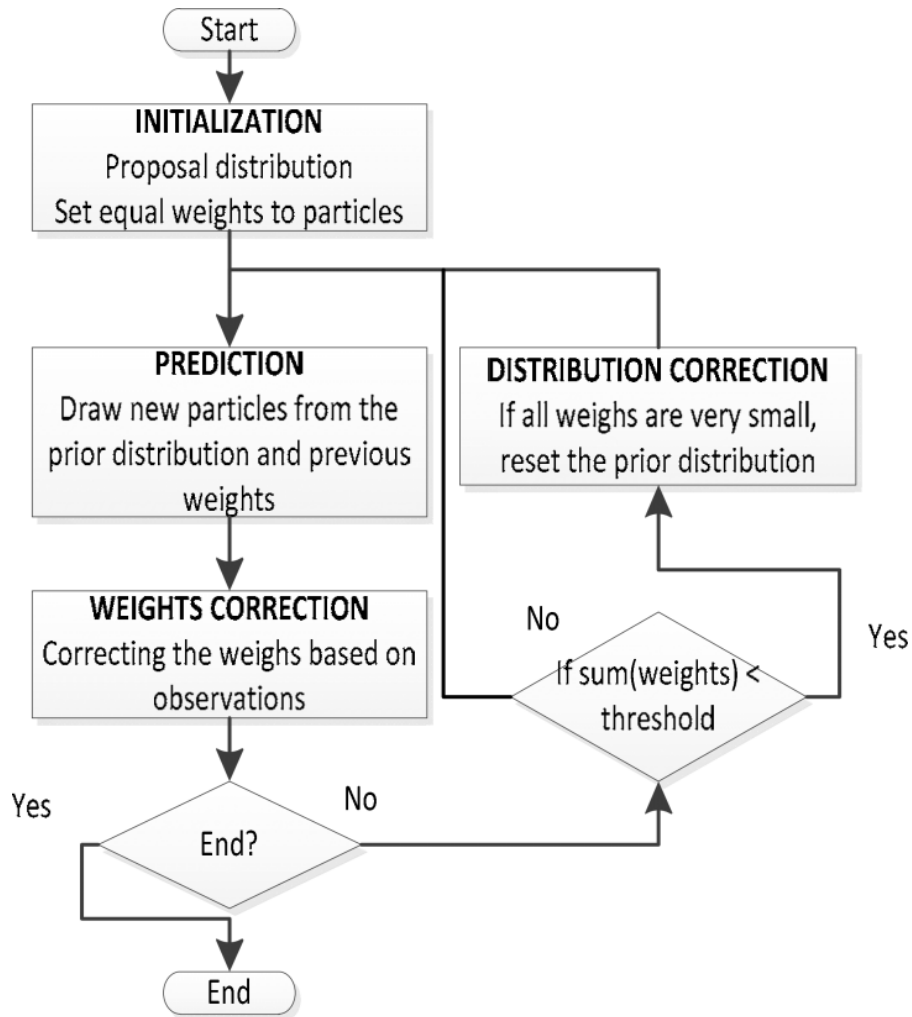


Abbildung 4: Funktionsweise generischer Partikelfilter (Salih & Malik, 2011)

Gewichte werden später für das Resampling der Prioritätsverteilung verwendet, um genauere Ergebnisse zu erhalten (Salih & Malik, 2011). Schließlich werden die Gewichte aus der neuen Wahrscheinlichkeit der neuen Schätzungen und Messungen berechnet (Salih & Malik, 2011).

Auch in Sungmin Kim et al. (2006) wird ein Partikelfilter benutzt. Er wird verwendet, um bei einer bewegten Kamera die Position und Bewegung von Personen unter Verwendung von Monte-Carlo-Sampling zu prognostizieren.

Lin et al. (2010) stellt eine Schnittstelle für die Multimediasteuerung unter Verwendung eines 3D-modellbasierten Gestenverfolgungsalgorithmus vor. Der Partikelfilter, der auf dem Bayes'schen Framework basiert, ist eine nützliche Technik für die 3D-Körperbewegungsverfolgung, da er mehrere Vorhersagen für komplexe menschliche Bewegungen liefert (Lin et al., 2010). In Lin et al. (2010) ist die posteriore Bayes'sche Formulierung des Partikelfilters definiert als:

$$p(x_t|Z_t) \propto p(z_t|x_t) \cdot p(x_t|Z_{t-1}) \quad (2)$$

Dabei zeigt bei Lin et al. (2010) x_t den Zustandsvektor zum Zeitpunkt t an und Z_t stellt die Beobachtung dar. Der Verlauf der Beobachtungen von 1 bis t wird als $Z_t = \{z_1, \dots, z_t\}$ angegeben (Lin et al., 2010). Die pdf $p(x_t|Z_{t-1})$ ist die Vorhersagewahrscheinlichkeitsverteilung zum Zeitpunkt $t-1$ und kann ausgedrückt werden als:

$$p(x_t|Z_{t-1}) = \int p(x_t|x_{t-1}) \cdot p(x_{t-1}|Z_{t-1})d_{x_{t-1}} \quad (3)$$

(Lin et al., 2010)

In Lin et al. (2010) wird der Partikelfilter weiter optimiert. Dafür wird er mit anderen Verfahren ergänzt. Zuerst wird durch Adaboost-Gesichtserkennung die 3D-Gesichtsposition ermittelt. Die Schulterposition kann durch die Kopfposition abgeleitet werden. Durch Verfolgung der Handfläche können dann die gesamten Bewegungsdaten der Arme abgeschätzt werden (Lin et al., 2010).

In Torres et al. (2015) wird eine Person mit einer Kinect Kamera aufgezeichnet, um durch eine Auswertung der Bewegungsdaten medizinische Diagnosen vorherzusagen. Dabei wird die NUI API (Natural User Interface Application Programming Interface) gestartet, die die Schnittstelle für Skeletal Tracking enthält (Torres et al., 2015). Diese Anwendungen ermöglichen es, die Daten für die individuelle Erkennung einer Person und die Verfolgung ihrer skelettalen Figur zu erhalten (Torres et al., 2015). Die Aufzeichnung und Verarbeitung der Daten zu einem Skelett wird in Torres et al. (2015) durch Microsoft Kinect durchgeführt. Die erhaltenen Koordinatensätze wurden mit Hilfe einer Matlab-Routine strukturiert und verarbeitet (Torres et al., 2015). Die erhaltenen Punkte wurden in 10 verschiedenen Zweigen aus drei verschiedenen Matrizen gruppiert, die mit den Koordinaten x , y und z aus jedem Bild verbunden sind (Torres et al., 2015).

In Tran und Trivedi (2009) wird ein Multiview-Ansatz zur Oberkörperposenverfolgung unter

Verwendung von 3D-Bewegungen der Extremitäten (Kopf und Hände) vorgestellt, der XMOB (eXtremity Movement OBServation) upper body pose tracker genannt wird. Bei diesem Ansatz wird das hochdimensionale Suchproblem der 3D-Positionsverfolgung des Oberkörpers in zwei Teilprobleme unterteilt. Zunächst werden die 3D-Bewegungen von Kopf und Händen anhand von Multiview-Eingaben verfolgt (Tran & Trivedi, 2009). Basierend auf dem Wissen über das Oberkörpermodell werden diese 3D-Extremitätenbewegungen dann verwendet, um die gesamte Oberkörperbewegung als inverse Kinematik vorherzusagen (Tran & Trivedi, 2009). Unter Verwendung geometrischer Einschränkungen der Oberkörpergelenke und Extremitäten bestimmt XMOB bei jedem Bild eine Reihe von Hypothesen für mögliche innere Gelenkpositionen (Tran & Trivedi, 2009). Durch Beobachtung der Bewegungen der Extremitäten über einen bestimmten Zeitraum hinweg wählt XMOB die Hypothesenfolge aus, die die Gelenkverschiebung minimiert (Tran & Trivedi, 2009).

In Tang et al. (2018) wird eine neue Methode zur Rekonstruktion von Videoszenen vorgestellt, um bei einem Überwachungssystem mit überlappenden Kameraaufzeichnungen mehrere menschliche Objekte zu verfolgen und ihre 3D-Position zu schätzen. In jeder Einzelansicht wird zunächst ein Objektdetektor verwendet, um die erkannten Bounding Boxes für jedes Bild zu erhalten (Tang et al., 2018). Dann wird einen Kalman-Filter-basierten Ansatz benutzt, um sie zu Tracklets zusammenzufügen (Tang et al., 2018). Konkret wird jede Trajektorie fragmentiert, wenn sie einen Bildrand verlässt, verdeckt ist oder eine Kalman-Vorhersage der 3D-Position hat, die 1 Meter von der nächstgelegenen Position entfernt ist (Tang et al., 2018). Alle Kameras werden auf der Grundlage einer Reihe von 2D-Tracklets selbst kalibriert und anhand einiger gemeinsamer Referenzpunkte in ein globales Koordinatensystem umgewandelt (Tang et al., 2018). Die Hierarchie der 3D-Positionsschätzung ist geordnet nach Rumpfschätzung, Schätzung der oberen Gliedmaßen und Schätzung der unteren Gliedmaßen (Tang et al., 2018). Ausgehend von den Geometrieinformationen beim Multi-View-Tracking kann man die optimale Ansicht für jedes Element für jeden Zeitpunkt bestimmen und aus dieser Ansicht eine 3D-Positionsschätzung machen (Tang et al., 2018). Diese vorgeschlagene Methode wird mit dem EPFL-Benchmark und dem Human3.6M-Benchmark evaluiert und demonstriert (Tang et al., 2018). Dabei ist die Methode von Tang et al. (2018) effektiver als die schon bekannten Methoden.

In Bogo et al. (2015) wird ein Verfahren vorgestellt um die 3D-Geometrie und das Aussehen des menschlichen Körpers anhand einer monokularen RGB-D-Sequenz eines sich frei vor dem Sensor bewegenden Benutzers genau abzuschätzen. Die Daten werden mit einer Kinect One erhoben (Bogo et al., 2015).

Stufe 1 unterteilt die ursprüngliche Sequenz der Länge n in kurze Intervalle von $n'=3$ aufeinanderfolgenden Bildern und schätzt die Körperform und -haltung in jedem Intervall von grob bis fein (Bogo et al., 2015). Es wird für jedes Bild in der Sequenz, beginnend mit dem ersten Bild und fortlaufend mit sich überlappenden Intervallen, minimiert wobei jedes Intervall mit den für das vorherige optimierten Werten initialisiert wird (Bogo et al., 2015). Dies ergibt eine Körperform t und drei Schätzungen der Pose bei fast jedem Bild. Um eine einzige Körperform

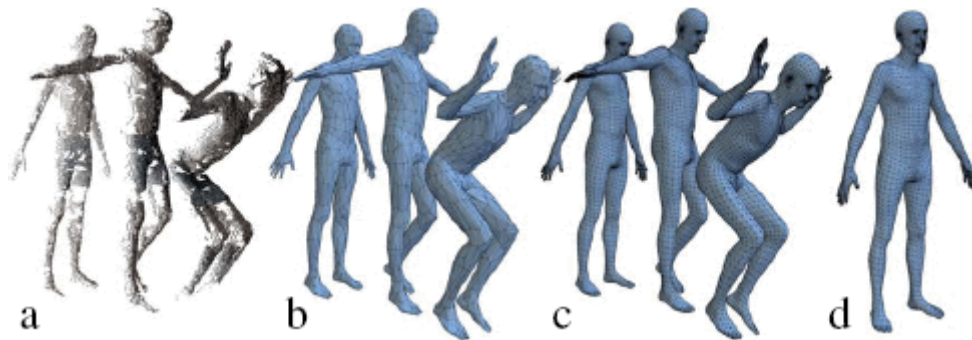


Abbildung 5: Stufe 1: Drei Eingabepunktwolken (a) und die entsprechenden Modelle mit niedriger (b) und hoher Auflösung (c), die nach der Optimierung erhalten wurden. Gezeigt ist auch das Endergebnis von Stufe 1 - eine konsistente hochauflösende Form (d) (Bogo et al., 2015)

aus Abbildung 5 auszugeben, mitteln wir die Formkoeffizienten der hochauflösenden Modelle . In ähnlicher Weise werden die drei geschätzten Posen bei jedem Bild gemittelt (Bogo et al., 2015). Dies funktioniert gut, da die Schätzungen dazu neigen, sehr ähnlich zu sein.

Stufe 2 ist dann die erscheinungsbildbasierte Optimierung (Bogo et al., 2015). Ausgehend von der anfänglichen Vermutung von oben löst man nun eine detailliertere Körperform, die nicht mehr auf den PCA-Unterraum beschränkt ist (Bogo et al., 2015). S sei der Vektor der gesuchten Körperformverformungen und um S zu berechnen, wird direkt die Scheitelpunktpositionen eines frei deformierenden Netzes, das Alignment T^t genannt wird, optimiert. Alignments werden in Richtung des Modells reguliert, aber ihre Scheitelpunkte können davon abweichen, um die Daten besser anzupassen (Bogo et al., 2015). Die Scheitelpunkte von T^t zusammen mit den Modellparametern werden optimiert. Zur Initialisierung des Aussehens U wird die in Abbildung 5 geschätzte Form und Pose verwendet. Dann wird die Farbe aus allen Frames auf Basis der einzelnen Texel gemischt und jeden Anteil entsprechend dem Winkel zwischen Oberflächennormale und Blickrichtung gewichtet (Bogo et al., 2015). Dies funktioniert gut, außer beim Gesicht, das viele hochfrequente Details aufweist. In Stufe 1 werden möglicherweise keine präzisen Kopfposen erzeugt, weil die Auflösung des Modells gering ist, was zu unscharfen Gesichtsdetails führt (Bogo et al., 2015). Um dies zu beheben, wird ein durchschnittliches Gesicht pro Geschlecht verwendet, das aus einer Trainingsmenge in der Gesichtsregion von U berechnet wird (Bogo et al., 2015).

4.3 Verwendung von Machine Learning

Abschließend werden Verwendungsbeispiele von Machine Learning, aus der Literatur, benannt und analysiert.

Das in Abschnitt 2.3 vorgestellte Maschine Learning ist eine neuere Auswertungsmethode und wird daher in aktuellen Papern häufiger verwendet als in älteren Papern. Im folgendem

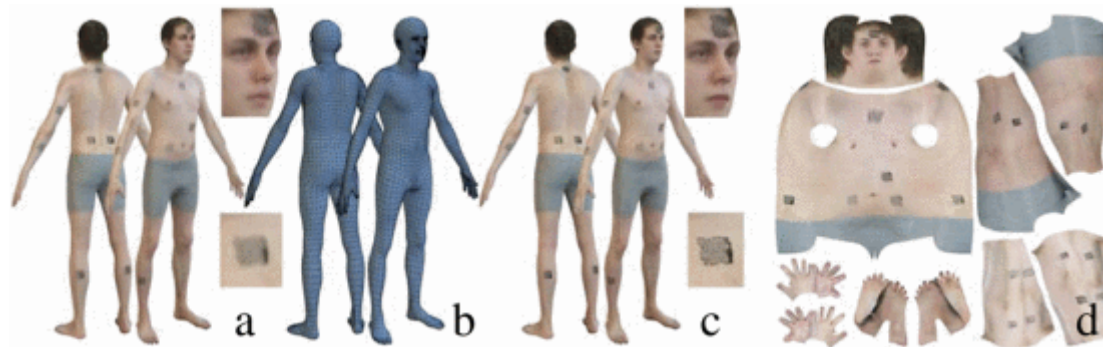


Abbildung 6: Stufe 2 : Ausgabeform S (b) und Texturkarte U (d). Zum Vergleich: S wird mit U vor der Optimierung (a) und nach der Optimierung (c) gerendert. (Bogo et al., 2015)

Text werden verschieden ausgeprägte Anwendungsbeispiele von Machine Learning beispielhaft vorgestellt.

Anjum et al. (2014) stellt einen neuen und effizienten Algorithmus zur Erkennung komplexer menschlicher Aktivitäten vor, welche Tiefenvideos benutzt, die von einer einzelnen Microsoft Kinect-Kamera aufgenommen wurden. Der Algorithmus basiert auf Robot Operating System (ROS) und benutzt eine modifizierte Version, um offline Videoaufnahmen benutzen zu können (Anjum et al., 2014). Nach einer Psi-Pose zur Kalibrierung werden die 3D-Positionen und Rotationsquaternionen von 15 Gelenken oder Körperteilen in Bezug auf einen festen Rahmen zur Orientierung gezeigt (Anjum et al., 2014). Die gezeigten Skelettgelenke umfassen Füße, Knie, Schultern, Hände, Ellbogen, Hüfte, Kopf, Hals und Rumpf (Anjum et al., 2014).

Durch die richtige Kombinierung der Gelenke, kann man jede Bewegung oder Aktivität erkennen (Anjum et al., 2014). Die Konstruktion von Merkmalsvektoren ist der wichtigste Schritt dieser Arbeit, bei dem entschieden wird, welche Gelenke oder Körperteile für jede Aktivität verfolgt werden sollen und wie sie in einem mathematischen Modell für die Konstruktion von Trainings- und Testdaten für eine Support Vector Machine (SVM) angeordnet werden (Anjum et al., 2014). Damit der Algorithmus in Echtzeit funktioniert, können nicht alle Gelenke verfolgt und erfasst werden (Anjum et al., 2014). Daher werden Merkmalsvektoren von bestimmten Gelenkpositionen in Abhängigkeit zueinander erstellt (Anjum et al., 2014). Die Gelenke mit den deutlichsten Bewegungen werden für die Konstruktion des Vektors benutzt (Anjum et al., 2014). Die Auswahl der Gelenke muss manuell vorbestimmt werden, jedoch muss man dies nur einmal machen und der Algorithmus erkennt die zu folgenden Gelenke danach automatisch (Anjum et al., 2014).

Für die SVM gilt nun die Aufgabe aus jedem Merkmalsvektor einer beliebigen Aktivität, die korrekte Bezeichnung der Aktivität zu benennen (Anjum et al., 2014). Es wird dafür eine multi-class SVM benutzt um mehrere Klassifizierungen vornehmen zu können (Anjum et al., 2014). Dabei wird jeder Merkmalsatz mit allen Proben verglichen und dann die Bezeichnung

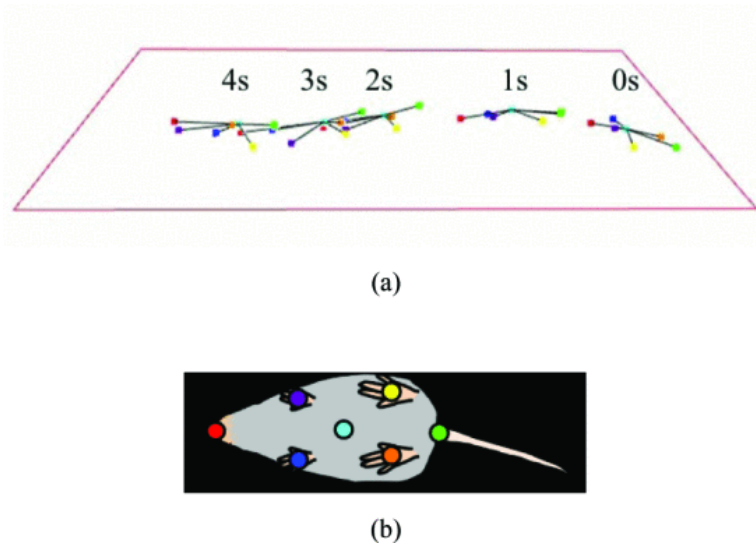


Abbildung 7: Ergebnisse der Methode (a) zeigt die Ergebnisse der 3 Dimensionalisierung der Körperteilverfolgung bei der Maus. (b) zeigt die Farbe der Punkte, die den einzelnen Körperteilen der Maus entsprechen (Tsuruda et al., 2021)

mit höchster Übereinstimmung zugewiesen (Anjum et al., 2014). Von dem Datensatz, der 200 Videos enthält, werden 120 Videos für das Training benutzt und die übrigen 80 Videos zur Testung des Verfahrens eingesetzt (Anjum et al., 2014). Dabei wurde zum Anlernen ein „linear Gaussian kernel“ benutzt (Anjum et al., 2014).

In Tsuruda et al. (2021) wird eine Methode zur markerlosen 3D-Positionsbestimmung von Mäusen mit einer einzigen Tiefenkamera präsentiert. Mithilfe von Hintergrundsubtraktion und den Tiefendaten konnte der Gewichtsschwerpunkt der Maus berechnet werden (Tsuruda et al., 2021). Die Gliedmaßen, die Nase und der Schwanzansatz wurden mithilfe von Deep Learning durch DeepLabCut verfolgt (Tsuruda et al., 2021). DeepLabCut verfügt über mehrere Netzwerke mit unterschiedlichen Schichttiefen, die mit über einer Million Bildern im ImageNet vortrainiert wurden (Tsuruda et al., 2021). In dieser Studie wurde ResNet-50, ein neuronales Netz mit einer Tiefe von 50 Schichten und der Struktur eines Residualnetzes, verwendet (Tsuruda et al., 2021). Die Position der Körperteile auf dem Videobild wurde mit den von der RGB-D-Kamera erhaltenen Tiefeninformationen kombiniert und durch die inverse perspektivische Transformation in 3D umgewandelt (Tsuruda et al., 2021). Die Daten wurden gelöscht, wenn sie entweder außerhalb des definierten Bereichs liegen, indem sich die Maus aufhalten kann, oder die Wahrscheinlichkeit der Schätzung durch DeepLabCut unter dem Schwellenwert lag. Danach wurden die Daten aus den 3D-Positionen des vorherigen und nachfolgendem Bilds ergänzt (Tsuruda et al., 2021). Bei Frames mit ungenauen Daten wird angenommen, dass sich die Körperteile linear konstant weiterbewegen. Die Ergebnisse sind in Abbildung 7 dargestellt.

In Alt et al. (2021) geht es um ein Projekt, welches Kinect-Kameras für die Erkennung und Auswertung von Skelett-Daten eingesetzt hat, um zu erkennen, zu welchem Zeitpunkt eine

Person von einem öffentlichem Display weggehen wird. Aus aufgenommenem Videomaterial wurden markante Merkmale erfasst, um den Zeitpunkt genauer vorhersagen zu können (Alt et al., 2021). Um das Konzept zu testen wurde ein Prototyp entwickelt, um die markanten Merkmale zu erkennen, dadurch eine Vorhersage zu treffen und die Informationen dann an eine App weiterzuleiten. (Alt et al., 2021). Es wurde eine Kinect Kamera benutzt und aufgrund der Skelettdaten wurden die Merkmale berechnet (Alt et al., 2021). Die Merkmalerkennung läuft permanent im Hintergrund mit und kann daher die Merkmale in Echtzeit erfassen (Alt et al., 2021). Dabei wurde in Alt et al. (2021) ein einfacher Entscheidungsbaum anstelle einer SVM verwendet, um die Entscheidungsfindung der trainierten Maschine im Vergleich zur menschlichen Entscheidungsfindung zu verstehen (Alt et al., 2021).

5 Vergleich der Auswertungsverfahren

5.1 Vorteile und Nachteile der Verfahren

5.1.1 Manuelle Auswertung

Vorteile und Nachteile der manuellen Auswertung werden im folgendem Kapitel dargestellt.

Die manuelle Auswertung hat Vorteile, welche meistens für das Machine Learning benutzt werden. Das Anlernen kann besonders effektiv sein, wenn die Person die Daten optimal vorbereitet und sortiert. Die involvierte Person kann also einen sehr positiven Effekt auf die Auswertung haben, wenn die Person bestens dafür ausgebildet ist. Dies kann sowohl Vorteil als auch Nachteil sein, denn die Person kann ebenso einen negativen Effekt haben. Diese unberechenbare Varianz ist insgesamt auch ein Nachteil der manuellen Auswertung. Die Fehlerquote ist abhängig vom Menschen und daher nicht planbar. Auch wenn es positive Tendenzen gibt, ist die nicht mögliche Planbarkeit einer der Hauptgründe, weshalb die manuelle Auswertung meistens nur eine unterstützende Funktion hat. Die begrenzte Komplexität ist in der heutigen Zeit auch ein limitierender Faktor.

Darüber hinaus ist ein großer negativer Aspekt der Zeitaufwand. Der Zeitaufwand für die Verarbeitung von großen Datensätzen ist nicht vergleichbar mit dem Zeitaufwand der anderen Methoden. Darüber hinaus ist der Zeitaufwand bei der manuellen Auswertung auch gleich dem Arbeitsaufwand und macht damit diese Methode kaum einsetzbar.

Die manuelle Auswertung hat zusammenfassend klare Stärken und Schwächen. Da nicht alle Schwächen ausgleichbar sind, ist die Methode nicht für eine alleinige Auswertung geeignet. Die Rolle der unterstützenden oder kontrollierenden Methode kann die manuelle Auswertung jedoch wahrnehmen, da dort die Nachteile weniger relevant sind.

5.1.2 Algorithmus

Die Algorithmen haben viele Vorteile sowie Nachteile, weshalb diese Aspekte für die unterschiedlichen Algorithmen differenziert dargestellt werden.

Der Algorithmus ist generell universell einsetzbar. Aufgrund der zahlreichen verschiedenen Algorithmen ist ein Algorithmus immer verwendbar. Daher werden die Algorithmen, welche in der Arbeit schon näher vorgestellt wurden, einzeln betrachtet.

Zuerst wurde der ICP Algorithmus von Knoop et al. (2006) vorgestellt. Dieser ist durch viele Annäherungen und Wiederholungen am Ende sehr genau. Dadurch ist der ICP Algorithmus

aber auch sehr zeitaufwändig, was sein größter Nachteil ist (Knoop et al., 2006).

Der bei Salih und Malik (2011) vorgestellte Partikelfilter wird als Nächstes betrachtet. Der generische Partikelfilter hat jedoch Probleme mit Beleuchtungsschwankungen durch Schatten von anderen Objekten (Salih & Malik, 2011). Darüber hinaus schneidet der generische Partikelfilter, bei der von Salih und Malik (2011) vorgenommenen Testung, aufgrund der hohen Fehlerquote schlecht ab. Der Kondensationsalgorithmus schneidet bei den Ergebnissen besser ab, ist jedoch auch zeitaufwändiger. Der zuletzt getestete Hilfspartikelfilter scheint, laut der Testung, die beste Wahl zu sein (Salih & Malik, 2011). Die Partikelfilter sind jedoch allesamt sehr zeitaufwändig.

Der in Lin et al. (2010) optimierte Partikelfilter spezialisiert sich auf die Gestenerkennung. Der Algorithmus ist schneller als die anderen Partikelfilter und kann daher in Echtzeit angewendet werden. Darüber hinaus ist er genauer als andere Verfahren, welche nur auf das Tracking spezialisiert sind (Lin et al., 2010). Nur die Gestenerkennung ohne ein Körpermodell besitzt eine höhere Fehlerquote und ist daher nicht optimal verwendbar. Für die Gestenerkennung des Oberkörpers mit erkanntem Körpermodell ist dieses Verfahren sehr optimal.

Der bei Torres et al. (2015) benutzte Algorithmus ist genau und kann ohne ein Setup verwendet werden. Negative Aspekte wurden nicht erwähnt.

Der in Tran und Trivedi (2009) vorgestellte Algorithmus braucht ein bestimmtes Setup, um die Daten optimal verarbeiten zu können. Weitere negative Aspekte werden leider nicht genannt. Die geschätzten Körperbewegungen sind durchaus genau, weshalb der Algorithmus dort seine optimale Anwendung findet.

In Tang et al. (2018) wird ein Algorithmus vorgestellt, welcher auch bei schlechten Aufnahmebedingungen gute Ergebnisse liefert. Der Algorithmus ist in seinem Anwendungsbereich der beste in allen Kategorien, benötigt jedoch sich überschneidende Kamerabilder ohne Lücken.

Zuletzt wurde der Algorithmus von Bogo et al. (2015) vorgestellt. Dieser Algorithmus bestimmt die Körperform und Pose einer Person sehr erfolgreich. Es gibt nur wenige Nachteile. Ein Nachteil ist der Zeitaufwand der Rekonstruktion und ein weiterer Nachteil ist die manuelle Auswahl des Geschlechts. Diese manuelle Intervention ist notwendig. Ansonsten ist der Algorithmus automatisiert und für die Rekonstruktion bestens geeignet, da der zeitliche Aspekt meistens nicht relevant ist.

5.1.3 Machine Learning

Abschließend werden die Vorteile und Nachteile von Machine Learning erläutert.

Fast alle Nachteile des Machine Learnings beziehen sich auf das Anlernen. Besonders die Datensätze können einen großen Nachteil des Verfahrens bilden. Besagte Datensätze, welche zum Anlernen des Machine Learnings benötigt werden, müssen für die jeweilige Verwendung individuell vorbereitet werden. Hierbei ist ein negativer Aspekt die Suche nach geeigneten Datensätzen. Des Weiteren ist die Erstellung solcher Datensätze sehr Zeit- und Arbeitsaufwändig. Wenn daher keine vorbereiteten, geeigneten Datensätze vorhanden sind, erschwert dies den Einsatz von

Machine Learning drastisch. Die Lerndatensätze sind ein maßgeblicher Faktor der erfolgreichen Verwendung von Machine Learning. Falls der Aufgabenbereich jedoch gleich bleibt, kann man die Datensätze wiederverwenden, weshalb die arbeitsaufwändige Erstellung nicht oft durchgeführt werden muss. In Themenbereichen, in denen aktuell viele Experimente auf dem Gebiet durchgeführt werden, gibt es häufig sehr gute Trainingsdatensätze, weshalb der Aspekt nicht immer so negativ ist.

Ein positiver Aspekt ist die Geschwindigkeit und, bei guter Anlernung, die Genauigkeit des Machine Learnings. Darüber hinaus wird das Machine Learning immer effektiver, ohne von Personen optimiert zu werden. Die Verarbeitung von großen Datenmengen ist beim Machine Learning normal und ist auch bestens dafür geeignet. Ein negativer Aspekt der Datenmengen fällt erst bei einer Echtzeitanwendung ins Gewicht. Bei einer solchen Anwendung des Verfahrens müssen die Datenmengen überschaubar sein, um diese in Echtzeit mitzurechnen. Dadurch kann man zum Beispiel nicht alle Skelettgelenke einer Person tracken, sondern nur einige Ausgewählte (Anjum et al., 2014). Trotzdem ist das Verfahren sehr genau.

Das bei Alt et al. (2021) benutzte Verfahren kann aufgrund der Simplizität in Echtzeit angewendet werden. Der einzige negative Aspekt ist die Ungenauigkeit, wenn eine Person nicht vollständig aufgenommen werden kann.

Bei Tsuruda et al. (2021) ist ein großer Nachteil die Fehlerquote der Nasenverfolgung. Jedoch wird die Maus als Ganzes sehr erfolgreich verfolgt und antizipiert.

Zusammenfassend ist die Vorbereitung des Machine Learning teilweise sehr zeitaufwändig, jedoch umso präziser und effektiver in der späteren Anwendung und daher ein, für viele Anwendungsbereiche, sinnvolles Auswertungsverfahren.

5.2 Vergleich der Verfahren

Die folgende Tabelle soll einen Überblick über die verschiedenen vorgestellten Methoden verschaffen. Dabei sollen auch die Besonderheiten der einzelnen Verfahren herausgestellt werden. Die manuelle Auswertung wurde, aufgrund der fehlenden Relevanz und Beispiele, nicht beachtet. Die technischen Voraussetzungen werden dabei nur angegeben, wenn sie auch in der Quelle explizit angegeben sind.

5 Vergleich der Auswertungsverfahren

Von	Verfahren	Ziel	Input	Nachteile	Vorteile	Technische Voraussetzungen
Knoop et al. (2006)	iterative closest point algorithm	3D-Verfolgung von Bewegungen	time-of-flight Kamera mit Stereokameraaufsatz	Zeitaufwand	Genauigkeit	Pentium 4 mit 3.2 GHz
Salih und Malik (2011)	Partikelfilter	3D-Verfolgung von Objekten	2D Videokamera	Zeitaufwand	Genauigkeit	nicht erwähnt
Lin et al. (2010)	Partikelfilter	Oberkörper-Gestenverfolgung	Stereokamera	Ungenau ohne Körpermodell	Echtzeit-Anwendung und sehr genau	Intel Core2 Quad CPU 2.66GHz
Torres et al. (2015)	Microsoft Skelett Tracking	Erfassung der Bewegung und Haltung des Körpers	Kinect Kamera	nicht erwähnt	Genauigkeit	nicht erwähnt
Tran und Trivedi (2009)	XMOB Algorithmus	Oberkörper-Gestenverfolgung	Zwei 2D-Kameras	Setup benötigt	Genauigkeit	Pentium(R) D CPU 2.8 GHz
Tang et al. (2018)	hierarchische Modellierung	Video-Rekonstruktion für 3D-Tracking	mehrere 2D-Kameras	benötigt überlappende Bildaufnahmen	gute Rekonstruktion und genaues Tracking	nicht erwähnt
Bogo et al. (2015)	Delta Algorithmus	Schätzung des 3D-Erscheinungsbilds des menschlichen Körpers	Kinect Kamera	Zeitaufwand und manuelle Intervention	Genauigkeit	nicht erwähnt

Von	Verfahren	Ziel	Input	Nachteile	Vorteile	Technische Voraussetzungen
Anjum et al. (2014)	Robot Operating System skeleton tracking	Erkennung komplexer menschlicher Aktivitäten	Kinect Kamera	Arbeitsaufwendig	Echtzeit-Anwendung und sehr genau	nicht erwähnt
Tsuruda et al. (2021)	DeepLabCut	Schätzung der 3D-Position einer Maus	Kinect Kamera	Setup notwendig und Nasenverfolgung fehlerhaft	Positionsbestimmung sehr genau	nicht erwähnt
Alt et al. (2021)	Entscheidungsbaum	Zeitpunkt des Verlassens erkennen	Kinect Kamera	Ungenau bei unvollständiger Aufnahme einer Person oder mehreren Personen gleichzeitig	Echtzeit-Anwendung und Genauigkeit	nicht erwähnt

5.3 Optimaler Anwendungsbereich eines Verfahrens

Im folgendem Kapitel wird der optimale Anwendungsbereich für ein Auswertungsverfahren näher beschrieben und dabei ein Bezug auf die vorher herausgestellten Aspekte der verschiedenen Methoden hergestellt.

Da die meisten vorgestellten Verfahren extra für deren Anwendungsbereiche optimiert wurden, kann man anhand der in Abschnitt 5.2 gezeigten Tabelle, die optimalen Anwendungsbereiche gut entnehmen. Jedoch gibt es auch Ziele die sich überschneiden oder auch ähnliche Prozesse haben. Wichtig dabei ist erstmal der verwendete Input, um das Verfahren benutzen zu können. Aus dem Input und dem Ziel kann man entsprechende Verfahren herausfiltern, welche für den gewünschten Anwendungsbereich optimal sind. Zuletzt wird auf die Vor- und Nachteile geachtet, um zum Beispiel besonders wichtige Aspekte berücksichtigen zu können. Dabei kann man beispielsweise zwischen Echtzeit-Anwendungen oder Anwendungen unterscheiden, bei denen der Zeitaufwand nicht relevant ist und die Genauigkeit mehr im Vordergrund steht.

Als Beispiel kann man die Verfahren von Tran und Trivedi (2009) und Lin et al. (2010) vergleichen. Beide sind für die Gestenverfolgung von Oberkörpern vorgesehen. Der Input

unterscheidet sich jedoch schon erheblich. Der Input ist sowohl relevant für den Aufbau des Projekts als auch für die Finanzierung. Danach erkennt man an den Nachteilen, dass Tran und Trivedi (2009) ein Setup braucht und daher nur für bestimmte, aufgesetzte Experimente nutzbar ist. Bei Lin et al. (2010) wird kein bestimmtes Setup benötigt und ist daher bei realen Anwendungen und Projekten verwendbar. Zuletzt kann man auch die technischen Voraussetzungen vergleichen, um sich ein Bild über die Finanzierung zu machen.

Als Letztes kann man auch die verwendeten Methoden abändern. Dies benötigt einen höheren Aufwand, kann sich aber bei einem ähnlichen Anwendungsbereich anbieten. Hierbei bleiben viele Aspekte der Methode gleich, kann jedoch einen anderen Output liefern. Der Hauptfaktor bleibt der negative Aspekt, da die meisten Verfahren durchaus genau arbeiten können. Wenn man die negativen Aspekte ausgleichen kann, ist ein Verfahren meist optimal für den gewünschten Anwendungsbereich geeignet. Der Ausgleich kann durch weitere Optimierung der Software oder Hardware sein, oder auch durch Optimierung des Aufbaus des Experiments.

Zusammenfassend ist die Nutzung der positiven Aspekte und der bestmögliche Ausgleich der negativen Aspekte, zusammen mit dem passenden Input, der Kern der optimalen Verwendung eines Auswertungsverfahrens von Body-Tracking-Daten.

6 Fazit

Das Ziel dieser Literaturarbeit war die Darstellung und Erklärung verschiedener benutzter Auswertungsverfahren von Body-Tracking-Daten. Dazu wurden die verschiedenen Kategorien der Auswertungsverfahren genauer erklärt. Danach wurde ein Literatur-Review durchgeführt. Dabei wurden die Anforderungen der Quellen beschrieben und dadurch die Literatur eingegrenzt. Die Suche wurde dokumentiert und sowohl die weitere Sortierung als auch nicht qualifizierte Paper wurden dargestellt. Darauf basierend wurden die verwendeten Auswertungsverfahren vorgestellt und analysiert. Die Vorteile und Nachteile der analysierten Methoden wurden danach explizit dargestellt und näher erläutert. Zusammenfassend wurde dann eine Vergleichstabelle angefertigt, um die Methoden bestmöglich miteinander vergleichen zu können. Hierbei wurde auf wichtige Aspekte geachtet, um den Vergleich besser herstellen zu können. Diese Aspekte waren zum Beispiel der benutzte Input, das Ziel der Methode oder auch die technischen Voraussetzungen. Mithilfe dieser Tabelle wurde dann ein optimales Anwendungsbeispiel für ein Auswertungsverfahren näher erläutert und verschiedene Ansätze dafür dargestellt. Durch die gesammelten Informationen können verschiedene Auswertungsverfahren für Body-Tracking-Daten differenziert betrachtet werden und Unterschiede ermittelt werden, um das optimale Verfahren für einen bestimmten Anwendungsbereich ermitteln zu können.

Ein kritischer Aspekt der Arbeit ist die beispielhafte Analyse von Auswertungsverfahren, da viele Verfahren in jegliche Richtungen optimiert werden können und sich dadurch die Verwendung und Aufgabenbereiche verändern können. Dabei ist auch die optimale Anwendung veränderbar, wenn eine große Veränderung oder Optimierung vorgenommen wird. Dies wurde in der Arbeit nicht beispielhaft durchgeführt, sondern nur erwähnt. Das Potenzial der verschiedenen Auswertungsverfahren wird auch nicht analysiert und wäre eine aufwändige, aber informative, Arbeit über die möglichen Auswertungsverfahren der Zukunft.

Es bietet sich daher an, abseits der Begrenzungen dieser Arbeit, die Optimierungen näher zu betrachten und dadurch neue Anwendungsbereiche für Verfahren zu finden. Im Zuge dessen könnten auch neue Verfahren kombiniert werden und möglicherweise optimale Ergebnisse erzeugt werden. Darüberhinaus könnten auch Experimente mit verschiedenen Auswertungsverfahren durchgeführt werden, um die Unterschiede bestmöglich vergleichen zu können und neue Anwendungsbeispiele für die Methoden zu testen.

Literatur

- Alt, F., Buschek, D., Heuss, D. & Müller, J. (2021). Orbuculum - Predicting When Users Intend to Leave Large Public Displays. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 5(1). <https://doi.org/10.1145/3448075>
- Anjum, M. L., Ahmad, O., Rosa, S., Yin, J. & Bona, B. (2014). Skeleton Tracking Based Complex Human Activity Recognition Using Kinect Camera. In M. Beetz, B. Johnston & M.-A. Williams (Hrsg.), *Social Robotics* (S. 23–33). Springer International Publishing. ISBN: 978-3-319-11973-1.
- Bogo, F., Black, M. J., Loper, M. & Romero, J. (2015). Detailed Full-Body Reconstructions of Moving People from Monocular RGB-D Sequences. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2300–2308. <https://doi.org/10.1109/ICCV.2015.265>
- Choutas, V., Müller, L., Huang, C.-H. P., Tang, S., Tzionas, D. & Black, M. J. (2022). Accurate 3D Body Shape Regression using Metric and Semantic Attributes. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2708–2718. <https://doi.org/10.1109/CVPR52688.2022.00274>
- DESH. (2023). *Methoden der Datenerfassung*. Verfügbar 12. Januar 2023 unter <https://www.desh-datenservice.de/lp/methoden-der-datenerfassung.html>
- Drobnjakovic, F., Douangpaseuth, J. B., Gadea, C., Haider, M., Ionescu, D., Ionescu, B. & Poon, L. (2018). Fusing data from inertial measurement units and a 3D camera for body tracking. *2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 1–6. <https://doi.org/10.1109/I2MTC.2018.8409754>
- Feigl, T., Mutschler, C. & Philippsen, M. (2018). Head-to-Body-Pose Classification in No-Pose VR Tracking Systems. *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 1–2. <https://doi.org/10.1109/VR.2018.8446495>
- Hesse, N., Pujades, S., Black, M. J., Arens, M., Hofmann, U. G. & Schroeder, A. S. (2020). Learning and Tracking the 3D Body Shape of Freely Moving Infants from RGB-D sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10), 2540–2551. <https://doi.org/10.1109/TPAMI.2019.2917908>
- Höpping, I. (2023). *Präzisierung des Algorithmusbegriffs*. Verfügbar 12. Januar 2023 unter <https://info-wsf.de/praezisierung-des-algorithmusbegriffs/>

- inserve. (2023). *Manuelle Datenerfassung vs. automatisierte Datenerfassung (mit KI) Teil 2*. Verfügbar 12. Januar 2023 unter <https://www.inserve.de/2022/11/18/manuelle-datenerfassung-vs-automatisierte-datenerfassung-mit-ki/>
- Knoop, S., Vacek, S. & Dillmann, R. (2006). Sensor fusion for 3D human body tracking with an articulated 3D body model. *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*, 1686–1691. <https://doi.org/10.1109/ROBOT.2006.1641949>
- Lin, S.-Y., Lai, Y.-C., Chan, L.-W. & Hung, Y.-P. (2010). Real-Time 3D Model-Based Gesture Tracking for Multimedia Control. *2010 20th International Conference on Pattern Recognition*, 3822–3825. <https://doi.org/10.1109/ICPR.2010.931>
- Liu, Y., Zhang, S. & Gowda, M. (2022). A Practical System for 3D Hand Pose Tracking using EMG Wearables with Applications to Prosthetics and User Interfaces. *IEEE Internet of Things Journal*, 1. <https://doi.org/10.1109/JIOT.2022.3223600>
- Rebala, G., Ravi, A. & Churiwala, S. (2019). Machine Learning Definition and Basics. *An Introduction to Machine Learning* (S. 1–17). Springer International Publishing. ISBN: 978-3-030-15729-6. https://doi.org/10.1007/978-3-030-15729-6_1
- Salih, Y. & Malik, A. S. (2011). 3D Tracking using particle filters. *2011 IEEE International Instrumentation and Measurement Technology Conference*, 1–4. <https://doi.org/10.1109/IMTC.2011.5944040>
- Sandmann, L. (2022). *Was bedeutet Algorithmus?* Verfügbar 12. Januar 2023 unter <https://www.heise.de/tipps-tricks/Was-bedeutet-Algorithmus-6324069.html>
- Sungmin Kim, Chang-Beom Park & Seong-Whan Lee. (2006). Tracking 3D Human Body using Particle Filter in Moving Monocular Camera. *18th International Conference on Pattern Recognition (ICPR'06)*, 4, 805–808. <https://doi.org/10.1109/ICPR.2006.1130>
- Tang, Z., Gu, R. & Hwang, J.-N. (2018). Joint Multi-View People Tracking and Pose Estimation for 3D Scene Reconstruction. *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. <https://doi.org/10.1109/ICME.2018.8486576>
- Torres, R., Monica, Clotet, R., Gonzalez, R., Pirrone, J., Sagbay, G., Rivas, D. & Soto, A. (2015). Analysis of corporal movement alterations using a 3D body tracking sensor. *2015 IEEE International Symposium on Systems Engineering (ISSE)*, 202–205. <https://doi.org/10.1109/SysEng.2015.7302757>
- Tran, C. & Trivedi, M. M. (2009). Introducing “XMOB”: Extremity Movement Observation Framework for Upper Body Pose Tracking in 3D. *2009 11th IEEE International Symposium on Multimedia*, 446–447. <https://doi.org/10.1109/ISM.2009.89>

Literatur

- Tsuruda, Y., Akita, S., Yamanaka, K., Matsumoto, Y., Yamamoto, M., Sano, Y., Furuichi, T. & Takemura, H. (2021). 3D Body Parts Tracking of Mouse Based on RGB-D Video from Under an Open Field. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 7252–7255. <https://doi.org/10.1109/EMBC46164.2021.9630565>
- Wuttke, L. (2021). *Was ist ein Algorithmus?* Verfügbar 12. Januar 2023 unter <https://datasolut.com/wiki/was-ist-ein-algorithmus/>
- Wuttke, L. (2022). *Was ist Machine Learning?* Verfügbar 12. Januar 2023 unter <https://datasolut.com/was-ist-machine-learning/>
- Zihajehzadeh, S., Yoon, P. K., Kang, B.-S. & Park, E. J. (2015). UWB-Aided Inertial Motion Capture for Lower Body 3-D Dynamic Activity and Trajectory Tracking. *IEEE Transactions on Instrumentation and Measurement*, 64(12), 3577–3587. <https://doi.org/10.1109/TIM.2015.2459532>

Abbildungsverzeichnis

1	Funktionsweise Algorithmus (Wuttke, 2021)	3
2	Funktionsweise Machine Learning (Wuttke, 2022)	6
3	Maschinelles Lernen im Überblick: Anwendungsbeispiele nach Arten (Wuttke, 2022)	7
4	Funktionsweise generischer Partikelfilter (Salih & Malik, 2011)	13
5	Stufe 1: Drei Eingabepunktswolken (a) und die entsprechenden Modelle mit niedriger (b) und hoher Auflösung (c), die nach der Optimierung erhalten wurden. Gezeigt ist auch das Endergebnis von Stufe 1 - eine konsistente hochauflösende Form (d) (Bogo et al., 2015)	16
6	Stufe 2 : Ausgabeform S (b) und Texturkarte U (d). Zum Vergleich: S wird mit U vor der Optimierung (a) und nach der Optimierung (c) gerendert. (Bogo et al., 2015)	17
7	Ergebnisse der Methode (a) zeigt die Ergebnisse der 3 Dimensionalisierung der Körperteilverfolgung bei der Maus. (b) zeigt die Farbe der Punkte, die den einzelnen Körperteilen der Maus entsprechen (Tsuruda et al., 2021)	18

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst habe, dass keine anderen als die angegebenen Quellen und Hilfsmittel benutzt wurden und dass alle Zitate ordnungsgemäß gekennzeichnet worden sind.

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

Neubiberg, den 23.03.2023

.....
Fabian Schaaf