



Titel:

**Erkennung und Bewertung von Körperhaltungen und
Bewegungen anhand von Gelenkwinkeln mit Einsatz
von Motion-Capture-Sensorik**

Von der

Fakultät für Informatik, Wirtschafts- und Rechtswissenschaften der
Carl von Ossietzky Universität Oldenburg zur Erlangung des
Grades und Titels eines

Doktors der Ingenieurwissenschaften (**Dr.-Ing.**)

angenommene Dissertation

von

Jan P. Vox

geboren am

18.10.1986 in Haselünne

Gutachter: Prof. Dr.-Ing. Andreas Hein

Weiterer Gutachter: Prof. Dr.-Ing. Frank Wallhoff

Tag der Disputation: 08.07.2022

Erklärung

Ich erkläre, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Jan P. Vox

Oldenburg, den 22.03.2022

Abstract

Diseases of the musculoskeletal system are one of the most frequent causes of incapacity to work. Therefore, it is very important to ensure proper ergonomic posture during physically strenuous and one-sided movements. The analysis of movements enables the identification of critical postures. Usually, the trained eye and knowledge of experts (ergonomists or medical professionals) is necessary to evaluate the physical risk of movement sequences. A data-driven motion analysis is possible by means of motion capture sensors. The gold standard for motion capture is an optical and marker-based method. Such motion capture sensor technology is generally expensive to purchase and time-consuming to set up, but it is characterized by high accuracy and robustness. User-friendly and inexpensive motion capture sensor technology is available from the consumer electronics sector, but is often less accurate and robust compared to above mentioned motion capture systems.

This thesis investigates the application of alternative motion capture sensor technology such as RGB-D Camera (Depth Sensor) and Lighthouse Tracking System for the analysis of movements and postures based on joint angles. Potential areas of application for the automatic evaluation of movements include the support of movement training or the ergonomic analysis of movements in the workplace. In both cases, the quality of the analysis is determined by and relies on the accuracy of the derived joint angles.

The first research question addresses the evaluation of sensor accuracies of the above mentioned alternative motion capture technologies in comparison to the gold standard. In addition to the results of joint angle accuracy, the limitations of motion capture sensor technology in terms of application are being addressed. The second research question investigates algorithms for motion detection and temporal segmentation of movements. The duration of holding a particular posture as well as the frequency of repetitions provide information about the physiological intensity and for ergonomic risk assessment. The third research question deals with methods of automatic evaluation of movements based on ergonomic criteria. Overall, this thesis describes the development and evaluation of a prototypical framework for the automatic recognition and evaluation of body postures and movements using motion capture sensor technology.

Zusammenfassung

Erkrankungen des Muskel-Skelett-Systems sind eine der häufigsten Ursachen für Arbeitsunfähigkeit. Umso mehr ist bei körperlich anstrengenden und einseitigen Bewegungen auf eine ergonomische Körperhaltung zu achten. Die Analyse von Bewegungen ermöglicht die Identifikation kritischer Körperhaltungen. In der Regel ist ein geschulter Blick von Expert*innen (Ergonom*innen oder medizinisches Fachpersonal) notwendig, um das körperliche Risiko von Bewegungsabläufen zu bewerten. Eine datengetriebene Bewegungsanalyse ist mittels Sensorik für die Erfassung von Bewegungsabläufen, sogenannter Motion-Capture-Sensorik, möglich. Der Goldstandard für die Bewegungserfassung ist ein optisches und markerbasiertes Verfahren. Derartige Motion-Capture-Sensorik ist generell kostspielig in der Anschaffung und zeitintensiv im Messaufbau, zeichnet sich jedoch durch eine hohe Genauigkeit und Robustheit aus. Anwenderfreundliche und günstige Motion-Capture-Sensorik findet sich im Bereich der Unterhaltungselektronik, der es jedoch im Vergleich zu dem oben genannten System oft an Genauigkeit und Robustheit mangelt.

Diese Arbeit untersucht die Applikation alternativer Motion-Capture-Sensorik, wie RGB-Tiefenkameras und Lighthouse Tracking Systeme, für die Analyse von Bewegungen und Körperhaltungen anhand von Gelenkwinkeln. Anwendungsfelder für die automatische Bewertung sind beispielsweise die Unterstützung von Bewegungstraining oder die Ergonomieanalyse von Bewegungen am Arbeitsplatz. In beiden Fällen ist die Betrachtung von Gelenkwinkeln maßgebend für die Analyse.

In der ersten Fragestellung wird die Sensorgenauigkeit der oben genannten Motion-Capture-Sensorik anhand von Vergleichsstudien mit dem Goldstandard evaluiert. Darin werden neben Ergebnissen zur Gelenkwinkelgenauigkeit auch Einschränkungen in der Anwendbarkeit der Motion-Capture-Sensorik untersucht. Die zweite Fragestellung untersucht Algorithmen für die Erkennung und zeitliche Segmentierung von Bewegungen. Die Dauer in einer Körperhaltung sowie die Häufigkeit von Wiederholungen geben Aufschluss über die physiologische Intensität und sind relevant für die ergonomische Risikobewertung. Die dritte Fragestellung befasst sich mit Verfahren der automatischen Bewertung von Bewegungen anhand von ergonomischen Kriterien. Insgesamt beschreibt diese Arbeit die Entwicklung und technische Evaluation eines prototypischen Frameworks für die automatische Erkennung und Bewertung von Körperhaltungen und Bewegungen mit Einsatz von Motion-Capture-Sensorik.

Danksagung

An dieser Stelle möchte ich allen danken, die mich während der Zeit als Doktorand unterstützt haben. Zunächst möchte ich mich bei dem Erst- und dem Zweitgutachter für die langjährige wissenschaftliche Betreuung bedanken. Prof. Dr.-Ing. Andreas Hein danke ich für beeindruckend treffende und zielführende Rückfragen und Anregungen, die mir immer eine große Hilfe waren, um den Fokus der Arbeit zu schärfen und die Motivation aufrechtzuerhalten. Prof. Dr.-Ing. Frank Wallhoff gilt großer Dank dafür, dass mir diese Chance und der Weg zur Promotion ermöglicht wurde. Die Freiheit etwas Neues ausprobieren zu dürfen und inhaltlich zu durchdringen habe ich sehr geschätzt.

Weiter möchte ich allen Mitgliedern der Prüfungskommission danken, die die Durchführung des Promotionsverfahrens ermöglichten. Prof. Dr.techn. Susanne Boll-Westermann (Vorsitz) danke ich für die interessierten und motivierenden Worte bei der Einführung in die Disputation, die mir in dem Moment großen Rückhalt gegeben haben. Ebenfalls danke ich Dr.-Ing. Vanessa Cobus für das hilfreiche Feedback und die Bereitschaft, sich diesem Thema anzunehmen.

Vielen Dank an Johannes Hurka und Jannik Fleßner für das unzählige Korrekturlesen kurz vor Deadline, die vielen Diskussionen, die Kaffeepausen und insbesondere auch die freundschaftliche Unterstützung. Ebenso danke ich Dr.-Ing. Insa Wolf für die "komplexen" Quaternion-Diskussionen und die Motivation, all das durchzuziehen. Apl. Prof. Dr. Helge Bormann danke ich für die organisatorische Begleitung des Promotionsprojekts (Jade2Pro).

Ebenfalls möchte ich allen Kolleg*innen der Abteilung TGM (Jade Hochschule) für den wertvollen Austausch und die Inspiration sowie dem IAPG für die Zurverfügungstellung von Messtechnik und Support danken. Vielen Dank auch allen Kolleg*innen am Fraunhofer IDMT, insbesondere der Gruppe MNT, für die konstruktive Kritik, Hilfsbereitschaft und den Ansporn. Danke auch an alle Kolleg*innen der Abteilung AMT (Universität Oldenburg) und des Forschungsbereichs Gesundheit (OFFIS) für den hilfreichen fachlichen Austausch und das Feedback.

Besonderer Dank gilt meiner Familie und Freunden für willkommene Ablenkungen und emotionale Unterstützung. Zuletzt möchte ich meiner Verlobten Verena Müller für das unentwegte Zuhören und Diskutieren, den fachlichen Input, das akribische Korrekturlesen und allgemein für die unbeschreiblich große Unterstützung danken. Danke!

Inhaltsverzeichnis

Abbildungsverzeichnis	XII
Tabellenverzeichnis	XV
Abkürzungsverzeichnis	XVII
1. Einleitung	1
1.1. Problemstellung	1
1.2. Lösungsansatz und Herausforderung	3
1.3. Zielsetzung	4
1.4. Forschungsfragen und Hypothese	6
1.5. Publikationen	7
1.6. Gliederung der Dissertation	9
2. Grundlagen	11
2.1. Beschreibung Motion-Capture-Daten	11
2.1.1. Kartesische Koordinaten von Gelenkpunkten	12
2.1.2. Räumliche Drehung anhand von Quaternionen	14
2.2. Motion Tracking mit einer RGB-Tiefenkamera	17
2.2.1. Funktionsweise RGB-Tiefenkamera	18
2.2.2. Skeleton Tracking in RGB- und Tiefenbildern	19
2.3. Motion Tracking mit einem Lighthouse Tracking System	21
2.3.1. Funktionsweise Lighthouse Tracking	23
2.3.2. Inverse Kinematik für einen menschlichen Avatar	24
2.4. Algorithmen für die Erkennung von Körperhaltungen	26
2.4.1. Support Vector Machine	27
2.4.2. Artificial Neural Network	30
2.5. Klinischer Standard für die Gelenkwinkelmessung	33
2.6. Bewegungsbewertung auf Basis von Gelenkwinkeln	34
2.6.1. DGOU Nachbehandlungsempfehlungen	35

2.6.2.	RULA	36
2.6.3.	REBA	37
2.6.4.	IFA Ampelschema	38
2.6.5.	OWAS	39
3.	Stand der Technik	42
3.1.	Sensorgenauigkeit von Motion-Capture-Sensorik	42
3.1.1.	Sensorgenauigkeit von RGB-Tiefenkameras	42
3.1.2.	Sensorgenauigkeit von Lighthouse Tracking Systemen	45
3.2.	Bewegungserkennung auf Basis von MoCap-Daten	48
3.3.	Bewegungsbewertung mit Einsatz von MoCap-Sensorik	54
4.	Eigener Ansatz	59
4.1.	Fragestellung I - Sensorgenauigkeit	59
4.1.1.	Relevanz der Sensorgenauigkeit für die Bewegungsbewertung	59
4.1.2.	Ermittlung der Sensorgenauigkeit einer RGB-Tiefenkamera	60
4.1.3.	Ermittlung der Sensorgenauigkeit eines Lighthouse Tracking Systems	61
4.2.	Fragestellung II - Bewegungserkennung	62
4.2.1.	Relevanz der Bewegungserkennung für die Bewegungsbewertung	62
4.2.2.	Ansatz für die Entwicklung einer Bewegungserkennung	62
4.3.	Fragestellung III - Bewegungsbewertung	64
4.3.1.	Relevanz einer automatischen Bewegungsbewertung	65
4.3.2.	Ansatz für die Realisierung einer automatischen Bewegungsbewertung	65
4.4.	Globale Darstellung des eigenen Beitrags	66
5.	Realisierung	69
5.1.	Transformation und Normalisierung von MoCap-Daten	70
5.2.	Berechnung von physiologischen Gelenkwinkeln	72
5.2.1.	Kalibrierung und Setzen der Neutral-Null-Stellung	78
5.2.2.	Offline- und Online-Filterung	79
5.2.3.	Plausibilitätsprüfung anhand der RoM	80
5.2.4.	Überprüfung der Berechnungsmethode	80
5.3.	Sensorrauschen von der Kinect v2 in statischen Haltungen	81
5.3.1.	Beschreibung der Sensorik	81

5.3.2.	Experimenteller Aufbau	82
5.3.3.	Methodik	82
5.3.4.	Ergebnisse	84
5.4.	Sensorgenauigkeit des HTC Vive Tracking Systems	85
5.4.1.	Beschreibung der Sensorik	86
5.4.2.	Experimenteller Aufbau und Datenerhebung	87
5.4.3.	Datenverarbeitung	90
5.4.4.	Methodik	90
5.4.5.	Ergebnisse	92
5.5.	Entwicklung eines Modells für die Bewegungserkennung	98
5.5.1.	Datenerhebung	99
5.5.2.	Methodik	99
5.5.3.	Ergebnisse	104
5.6.	Algorithmus für die Segmentierung von Bewegungssequenzen	105
5.7.	Automatische Bewegungsbewertung	109
5.7.1.	Digitalisierung von Bewertungsschemata	109
5.7.2.	Framework für die automatische Bewegungsbewertung	110
5.7.3.	Anwendung der automatischen Bewegungsbewertung auf mobilen Endgeräten	112
6.	Technische Evaluation	114
6.1.	Sensorgenauigkeit in kritischen Körperhaltungen	114
6.2.	Erkennung von kritischen Körperhaltungen	117
7.	Diskussion	119
7.1.	Fragestellung I - Sensorgenauigkeit	119
7.1.1.	Optische und markerbasierte Systeme als technische Referenz	120
7.1.2.	RGB-Tiefenkamera	120
7.1.3.	Lighthouse Tracking System	121
7.2.	Fragestellung II - Bewegungserkennung	122
7.3.	Fragestellung III - Bewegungsbewertung	124
7.4.	Personenbezogene Informationen in MoCap-Daten	125
8.	Fazit und Ausblick	126
A.	Anhang	129
A.1.	Benutzeroberflächen der eigens entwickelten Applikationen	129

A.2. Statistik der analysierten Gelenkwinkel von Vive und Qualisys	131
A.3. Weitere Publikationen	133
Literaturverzeichnis	134

Abbildungsverzeichnis

1.1. Bevölkerungsvorausberechnung für Deutschland von 2020 - 2050 . . .	2
1.2. Anwendungsfälle für eine automatische Bewegungsanalyse	5
2.1. Darstellung eines Samples aus MoCap-Daten mit 24 Gelenkpunkten und den Rohdaten einer RGB-Tiefenkamera als Punktwolke	13
2.2. Räumliche Drehung von Objekten bzw. Gelenkpunkten	16
2.3. Prinzipielle Datenverarbeitung bei einer RGB-Tiefenkamera als Motion- Capture-Sensorik	17
2.4. Wesentliche Komponenten einer RGB-Tiefenkamera am Beispiel der Kinect v2	18
2.5. Vereinfachte Darstellung des Skeleton Trackings im <i>Azure Kinect Bo- dy Tracking SDK</i>	20
2.6. Prinzipielles Verfahren eines Lighthouse Tracking Systems für die Erfassung von MoCap-Daten	22
2.7. Vergleich zwischen realer und virtueller Körperhaltung	22
2.8. Lighthouse Tracking System <i>HTC Vive</i> mit Basisstation, HMD, Con- troller und Trackern an Oberarmen und Hüfte	23
2.9. Beschreibung einer IK am Beispiel der Gelenkkette <i>Shoulder - Elbow - Hand</i>	24
2.10. Exemplarische Darstellung verschiedener Kernel von SVM	29
2.11. Exemplarisches ANN in einer 2-4-5-4-2 Architektur für die Klassi- fizierung von <i>lunge left</i> und <i>lunge right</i> anhand von <i>hip left flexion</i> und <i>hip right flexion</i>	30
2.12. Signalverarbeitung eines Neurons	31
2.13. Typische Aktivierungsfunktionen eines künstlichen Neurons	31
2.14. Gelenkwinkelmessung der Flexion des linken Kniegelenks mit einem Standard-Goniometer	33

2.15. Bsp. für die RoM von Extension nach Flexion ($40^\circ/0^\circ/170^\circ$) im Schultergelenk; Bewertung der physischen Belastung bei Flexion und Extension	36
4.1. Globale Darstellung des eigenen Ansatzes	68
5.1. Darstellung der Verarbeitungsschritte von MoCap-Daten für die Umsetzung eines prototypischen Systems zur automatischen Bewegungsbewertung auf Basis von Dauer und Label einer erkannten Bewegung oder Körperhaltung	70
5.2. Skelettmodell mit 25 Gelenkpunkten als Grundlage für die Vereinheitlichung von Positionen und Benennungen der Gelenkpunkte . . .	71
5.3. Körperebenen und MoCap-Daten einer aufrecht stehenden Person .	74
5.4. Berechnung von physiologischen Gelenkwinkeln aus MoCap-Daten mit Bezug zu Körperebenen und aus drei kartesischen Koordinaten .	77
5.5. Darstellung von verketteten Quaternionen und der Winkel <i>Yaw, Pitch, Roll</i> am Beispiel der Gelenkpunkte <i>Elbow - Wrist - Hand</i>	78
5.6. Experimenteller Aufbau der Kinect v2 vor einer real großen menschlichen Plastikfigur und neun umringenden Kameras des Qualisys Oqus Systems	83
5.7. Vergleich zwischen der Kinect v2 und dem Qualisys als Referenzsystem auf Basis berechneter Gelenkwinkel in einer statischen Haltung .	83
5.8. Experimenteller Aufbau für den technischen Vergleich zwischen der MoCap-Sensorik von Qualisys und Vive	88
5.9. Concordance Correlation Coefficient für alle untersuchten Gelenkwinkel und Tracker-Konfigurationen	93
5.10. Untersuchung der Abweichung zwischen Vive und Qualisys als Referenz auf Basis von modifizierten Bland-Altman-Plots	94
5.11. Violin-Plots für die <i>head</i> - und <i>neck</i> -Winkel bei der 3-Tracker-Konfiguration	96
5.12. Violin-Plots für alle untersuchten Gelenkwinkel bei der 6-Tracker-Konfiguration	97
5.13. Vereinfachte Darstellung der Verarbeitungsschritte und Modellentwicklung für die frameweise Erkennung von Körperhaltungen	101
5.14. Vorverarbeitung der MoCap-Daten für die Modellentwicklung	102
5.15. Konfusionsmatrix mit normalisierten Genauigkeiten $[0, 1]$ für die frameweise Erkennung auf Basis von SVMs	105

5.16. Konfusionsmatrix mit normalisierten Genauigkeiten $[0, 1]$ für die frameweise Erkennung auf Basis eines ANN	106
5.17. Algorithmus für die zeitliche Segmentierung von Bewegungssequenzen	107
5.18. Automatische Bewegungsbewertung, Erkennung und Segmentierung am Beispiel von Kniebeugen (<i>fullSquat</i>)	111
5.19. Testung der automatischen Bewegungsbewertung auf mobilen Endgeräten	113
6.1. Violin-Plots für alle untersuchten Gelenkwinkel bei der 6-Tracker-Konfiguration in kritischen Gelenkwinkelbereichen	116
6.2. Darstellung des Sub-Datensatzes aus den Bewegungen (<i>diagoArmL-LegR</i> , <i>fullSquat</i> , <i>jumpingJack</i> , <i>sideLungeLegL</i> , <i>sideLungeLegR</i>) mit Einzeichnung eines <i>threshold</i> für Haltungen mit einem Score ≥ 6	117
6.3. Konfusionsmatrizen mit normalisierten Genauigkeiten $[0, 1]$ für die frameweise Erkennung von fünf Ganzkörperbewegungen mit einem Score ≥ 6	118
7.1. Ellenbogenflexion von (a) 60° bis (d) 90° in 10° Schritten gemessen mit einem Standard-Goniometer	119
A.1. Eigens entwickelte Software für die Berechnung und Analyse von Gelenkwinkeln in MoCap-Daten verschiedener Sensortypen	129
A.2. Eigens entwickelte Software für die manuelle Annotation von MoCap-Daten	130
A.3. Eigens entwickelte Software-Benutzeroberfläche für die Anwendung der Bewegungsbewertung mit der Kinect v2 in einem mobilen Endgerät	131

Tabellenverzeichnis

2.1. Unterschiede bei der Notation einer Quaternion zwischen Hamilton und JPL	15
2.2. Typische Eigenschaften von RGB-Tiefenkameras am Beispiel der Kinect v2 und Azure	19
2.3. Auflistung von relevanten Gelenkwinkeln für eine Bewegungsbewertung; Freiheitsgrad und RoM nach Neutral-Null-Methode	35
2.4. Beispiele für Gelenktraumata mit entsprechender Behandlungsempfehlung zur Einhaltung von Regeln (z.B. Begrenzung der RoM) über einen gewissen Zeitraum	37
2.5. Bewegungsbewertung RULA; tabellarische Auflistung des Scorings für kritische Winkelbereiche von Gelenken	38
2.6. Erweiterung von RULA um die Bewegungsbewertung REBA	38
2.7. Bewertung der physischen Belastung nach IFA Ampelschema für die Gelenke Head, Neck, Chest Spine, Lumbar Spine, Shoulder und Knee	40
2.8. Bewertung der physischen Belastung nach IFA Ampelschema für die Gelenke Elbow und Hand	41
3.1. Offene Datensätze mit Bewegungsinformationen präsentiert als Bewegungssequenzen in verschiedenen Anwendungsfeldern	50
5.1. Auflistung von Gelenkpunkten, Freiheitsgraden und angewendetem Rechenweg für die Bestimmung des jeweiligen Gelenkwinkels	73
5.2. Für die Berechnung von Gelenkwinkeln relevante Fixpunkte und Körperebenen	76
5.3. Statistische Auswertung der Vergleichsmessung zwischen Kinect v2 und dem Qualisys auf Basis berechneter Gelenkwinkel	84
5.4. Beschreibung der Bewegung und zugeordneter Gelenkwinkel	89
5.5. Beschreibung der Bewegungsübungen (Klassen) in den MoCap-Daten	100

5.6. Berechnungszeiten für die Erkennen-Modelle (SVM, ANN) auf verschiedenen Endgeräten 104

6.1. Untersuchte Gelenkwinkel mit zugehörigen Regeln nach IFA Ampelschema für die Auswahl von kritischen Winkeln ($\geq \bullet$) 115

A.1. Statistik der analysierten Gelenkwinkel von Vive und Qualisys . . . 132

Abkürzungsverzeichnis

ANN Artificial Neural Network

CCC Concordance Correlation Coefficient

CNN Convolutional Neural Network

DGOU Deutsche Gesellschaft für Orthopädie und Unfallchirurgie

DGUV Deutsche Gesetzliche Unfallversicherung

DL Deep Learning

HMD Head Mounted Display

IFA Institut für Arbeitsschutz

IK Inverse Kinematik

IMU Inertial Measurement Unit

IPR Inter-Percentile-Range

IR Infrarot

JPL Jet Propulsion Laboratory

LSTM Long Short-Term Memory

LoA Limits of Agreement

ML Machine Learning

MSS Muskel-Skelett-System

MoCap Motion-Capture

RAM Random Access Memory

- RBF** Radial Basis Function
- REBA** Rapid Entire Body Assessment
- RNN** Recurrent Neural Network
- ReLU** Rectified Linear Unit
- RMSE** Root Mean Square Error
- RoM** Range of Motion
- RULA** Rapid Upper Limb Assessment
- ToF** Time-of-Flight
- OWAS** Ovako Working Posture Assessment
- SDK** Software Development Kit
- SVM** Support Vector Machine
- VR** Virtual Reality

1. Einleitung

1.1. Problemstellung

Der Mensch verfügt über ein komplexes Muskel-Skelett-System (MSS) bestehend aus ca. 650 Muskeln, 200 Knochen, mehr als 100 Gelenken sowie Bändern, Sehnen, Knorpel und Bindegewebe. Dieses System ermöglicht die Ausführung von komplexen Bewegungen wie Greifen, Heben und Laufen, sowie allgemein die Ausübung von Kräften auf die Umgebung. Ein gesundes MSS ist grundlegend für Mobilität und Eigenständigkeit. Überlastungen und unergonomische Bewegungsabläufe, aber auch Bewegungsmangel, können das MSS beeinträchtigen [DGUV, 2018].

Funktionseinschränkungen und Erkrankungen des MSS, wie der Verschleiß von Gelenken (Arthrose), gehen häufig mit chronischen Schmerzen einher und haben in der Regel gravierende Auswirkungen auf die Lebensqualität und die Arbeitsfähigkeit des betroffenen Menschen [KKH, 2008, Liebers et al., 2013]. Im Jahr 2019 gab es im Durchschnitt 12 Krankheitstage je Fall unter den Pflicht- und freiwilligen Mitgliedern der Gesetzlichen Krankenkasse, die zu einer Arbeitsunfähigkeit führten. Nahezu ein Viertel der Krankheitstage (23.9 %) von Erwerbstätigen im Alter zwischen 15 - 65 Jahren sind auf Erkrankungen des MSS und des Bindegewebes zurückzuführen. Im Vergleich zu anderen Diagnosegruppen verursachen Erkrankungen des MSS die meisten Krankheitstage und somit die höchsten Kosten durch Arbeitsunfähigkeit in Deutschland. Bei Betrachtung der Krankheitstage in Abhängigkeit des Alters zeigt sich, dass mit zunehmendem Alter das Risiko einer Arbeitsunfähigkeit infolge von Krankheiten des MSS steigt. Bei Erwerbstätigen jünger als 45 Jahre liegt das Risiko bei 18.4 %, bei über 45-Jährigen bei 27.4 % [Brenscheidt et al., 2021, S.47; BMAS, 2020, S.136-137].

Eine Zunahme der Fallzahlen von degenerativen Erkrankungen ist, bedingt durch den demografischen Wandel einhergehend mit einer älter werdenden Gesellschaft, bereits in den kommenden zwei Jahrzehnten zu erwarten [Hämel and Schaeffer,

2013]. Die Bevölkerungsvorausberechnung¹ für Deutschland zeigt einen Zuwachs von Personen älter als 67 von 19 % auf 27 % bezogen auf die Gesamtbevölkerung in Deutschland in den Jahren von 2020 bis 2040. Gleichzeitig sinkt der Anteil von Personen in der Altersgruppe zwischen 20 - 66 Jahren. Dieser Trend ist in Abbildung 1.1 dargestellt.

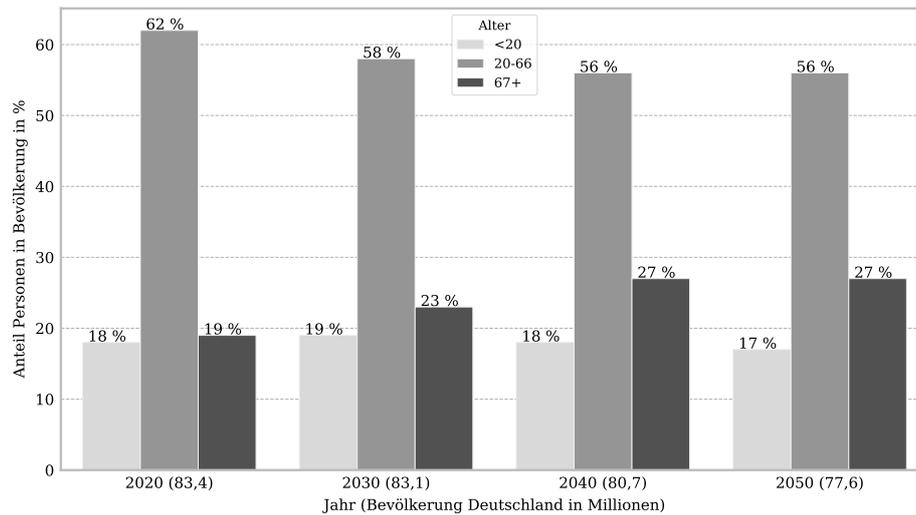


Abbildung 1.1.: Bevölkerungsvorausberechnung für Deutschland von 2020 - 2050 unterteilt in Altersklassen; Daten erhoben von [Destatis, 2021]

Dieser Wandel stellt sowohl das Gesundheitssystem als auch Unternehmen vor große Herausforderungen. Zum einen steigt in einer alternden Gesellschaft der Bedarf an Pflege und Krankenversorgung. Zum anderen ist durch den Anstieg des Durchschnittsalters in den Betrieben, einhergehend mit überdurchschnittlich vielen Krankheitstagen und Früh-Verrentungen, mit einem Fachkräftemangel am Arbeitsmarkt zu rechnen. Dieser Umstand rückt Krankheitsprävention und Gesundheitsförderung stark in den Fokus. In Unternehmen sichert die Gesundheit von Mitarbeiter*innen die Aufrechterhaltung des Betriebs. Gleichzeitig sorgt im Allgemeinen eine gesündere Gesellschaft für eine Entlastung des Gesundheitssystems [Amstutz et al., 2018, Gellert et al., 2018, Kraft, 2019].

¹Stand aus der 14. koordinierten Bevölkerungsvorausberechnung des statistischen Bundesamts (Destatis). Die sich aktuell entwickelnden demografischen Prozesse (Fertilität, Mortalität und Migration) sind in dieser Statistik nicht berücksichtigt.

1.2. Lösungsansatz und Herausforderung

Als Lösungsansatz bieten sich u.a. Präventionsmaßnahmen zur Vermeidung von Erkrankungen des MSS an. Prävention durch Bewegungstraining ist eine Möglichkeit, um das MSS zu kräftigen und die Verletzungsanfälligkeit zu reduzieren [Wessinghage and Morsch, 2013]. Auch die Optimierung von Heilungsprozessen nach einer Erkrankung kann zu einer Entlastung des Gesundheitssystems beitragen. Durch gezieltes Bewegungstraining kann die Rehabilitation nach Erkrankung des MSS beschleunigt werden [Fehr and Slesina, 2004]. Insbesondere kann durch weiterführende Maßnahmen im häuslichen Umfeld ein nachhaltiger Rehabilitationserfolg erzielt werden [Sabo et al., 2016].

Um arbeitsbedingten Erkrankungen des MSS vorzubeugen, ist es notwendig, Arbeitsplätze ergonomisch zu gestalten [Amell and Kumar, 2001, Kjellberg et al., 2016]. Teil der Ergonomiebewertung ist die visuelle Beobachtung von Körperhaltungen und Bewegungen bei Arbeitsprozessen [Punnett and Wegman, 2004]. Gängige Praxis ist die Beobachtung und Bewertung von Bewegungsabläufen durch Ergonomieexpert*innen [Lowe et al., 2019]. Insbesondere die Analyse von Gelenkwinkeln gibt Aufschluss, ob eine Körperhaltung oder Gelenkbewegung in ungünstigen Winkelbereichen ausgeführt wird und ein Risiko darstellen könnte [IFA, 2015].

Sowohl Krankheitsprävention durch Bewegungstraining, als auch die Ergonomiebewertung von Körperhaltungen und Bewegungen am Arbeitsplatz können durch den Einsatz von Sensorik und Datenverarbeitung unterstützt werden [Capecci et al., 2016, Ellegast et al., 2010]. Als technischer Goldstandard gilt ein optisches Verfahren basierend auf digitalen Hochgeschwindigkeitskameras und passiv reflektierenden Markern für die Aufzeichnung von Motion-Capture (MoCap)-Daten². Dieses Verfahren zeichnet sich durch eine hohe Positionsgenauigkeit im Millimeterbereich aus, wird aber aufgrund hoher Anschaffungskosten und eines aufwendigen Messaufbaus³ vorwiegend unter Laborbedingungen eingesetzt. Anwendungsfelder sind Sportwissenschaften sowie Forschung zu Biomechanik und Bewegungsergonomie [Göpfert et al., 2009, Liu et al., 2007, Windolf et al., 2008].

Leicht einsetzbare und kostengünstige MoCap-Sensorik findet sich im Bereich der

²Positionen und Quaternionen von Gelenkpunkten im menschlichen Skelett-Modell als Zeitreihe.

³Kalibrierung Mehrkamera-System; anatomisch korrekte Positionierung von Markern.

Fitness- und Unterhaltungselektronik (Consumer-Elektronik). Diese Systeme basieren in der Regel auf Verfahren der optischen Bildverarbeitung, Trägheitssensoren oder elektromagnetischen Sensoren, zeigen jedoch Einbußen in der Genauigkeit und Robustheit gegenüber dem Goldstandard [van der Kruk and Reijne, 2018].

Die Entwicklung und Anpassung von Sensorik und Software für eine automatische Bewegungsanalyse ist aufgrund von komplexen, nicht linearen und hochdynamischen Bewegungen sowie auftretenden Sensorungenauigkeiten weiterhin eine große Herausforderung für Wissenschaft und Forschung [Liu et al., 2016, Vox et al., 2021].

1.3. Zielsetzung

Im Rahmen dieser Arbeit wird die Verwendbarkeit von MoCap-Sensorik aus dem Consumer-Segment für die automatische Erkennung und Bewertung von Körperhaltungen und Bewegungen anhand von Gelenkwinkeln untersucht. Zielsetzung ist die Umsetzung und Evaluation einer anwenderfreundlichen und mobil einsetzbaren Bewegungsanalyse zur Identifikation von ungünstigen und damit potenziell schädlichen Gelenkbewegungen. Ziel der Forschung ist die Entwicklung eines technischen Ansatzes, um einen Beitrag zur Reduzierung von Erkrankungen des MSS zu leisten. Vor diesem Hintergrund wurden zwei Anwendungsfälle identifiziert und anhand dieser die Anforderungen an Sensorik und Software untersucht.

Im ersten Anwendungsfall soll Bewegungstraining im häuslichen Umfeld, als Teil von Maßnahmen zur Krankheitsprävention, durch eine automatische Bewegungsanalyse unterstützt werden (siehe Abbildung 1.2a). Für diesen Fall wird eine markerlos einsetzbare MoCap-Sensorik, basierend auf einer RGB-Tiefenkamera wie der Microsoft Kinect v2, näher untersucht.

Im zweiten Anwendungsfall geht es um die Risikobewertung von Körperhaltungen und Bewegungen bei der Planung von Arbeitsplätzen in Virtual Reality (VR) (siehe Abbildung 1.2b). Für dieses Szenario wird die für VR-Anwendungen ausgelegte MoCap-Sensorik *HTC Vive*, basierend auf dem *Lighthouse* Verfahren, evaluiert. Damit ist ein Infrarot-Laser basierendes Tracking System bestehend aus festen Basisstationen (sog. *Lighthouses*) und tragbaren Trackern gemeint.

In beiden Fällen kann anhand von Gelenkwinkeln eine Bewertung von kritischen Körperhaltungen und Bewegungen erfolgen. Zudem ist die Ausführungsdauer einer



(a) Bewegungsanalyse via Kinect v2 zur Unterstützung von Bewegungstraining im häuslichen Umfeld [Vox et al., 2016, Abb. 1]



(b) MoCap-Daten des HTC Vive Systems projiziert auf einen Avatar am Arbeitsplatz in VR [Vox et al., 2021, Fig. 1]

Abbildung 1.2.: Anwendungsfälle für eine automatische Bewegungsanalyse; exemplarisch eingezeichnete Gelenkwinkel

kritischen Körperhaltung relevant für die Risikoabschätzung [DGOU, 2021, McAtamney and Nigel Corlett, 1993].

Auf Basis der Anwendungsfälle wurden folgende allgemeine und technische Anforderungen an Sensorik und Software konkretisiert:

Verfügbarkeit

Der Einsatz von am Markt verfügbarer Hardware aus dem Bereich der Unterhaltungselektronik wird untersucht, um eine flächendeckende Einsetzbarkeit zu erreichen.

Leicht einsetzbar

Zur Minimierung von Barrieren sollten die Systemkomponenten anwenderfreundlich und ohne Begleitung von Expert*innen installier- und nutzbar sein.

Live Feedback

Das System sollte in der Lage sein, die Nutzer*innen aktiv während der Ausführung von Bewegungen zu unterstützen bzw. zu korrigieren. Aus diesem Grund wird an einer echtzeitfähigen⁴ und auf mobilen Plattformen einsetzbaren Verarbeitung von MoCap-Daten geforscht.

⁴Aufgrund von Spezifikationen der genutzten Modelle zur Schätzung von Gelenkpositionen (Skeleton Tracking) wird eine *weiche Echtzeit* angenommen.

Individuelle Bewegungsanalyse

Die Analyse und die Erkennung von Körperhaltungen und Bewegungen sollte an den jeweiligen Anwendungsfall anpassbar sein. Das bedeutet, dass dem System eine Körperhaltung oder Bewegung vorgemacht wird, welche anschließend automatisch erkannt und analysiert wird. Im Anwendungsfall des Bewegungstrainings würde durch Trainer*innen eine individuelle Trainingsbewegung vorgemacht werden, die dann mit dem System erkannt und analysiert werden kann. Auch in der Ergonomieanalyse könnten dadurch individuelle Körperhaltungen identifiziert werden. Technische Anforderungen sind, dass der jeweilige Erkennungs-Algorithmus mit wenigen Trainingsdaten trainiert werden kann und dass das Training auf einem mobilen Endgerät möglich ist.

Die Anforderungen an ein System für die automatische Bewegungsanalyse wurden in einer eigens durchgeführten Befragung [Vox et al., 2016] von Expert*innen aus dem medizinischen Bereich (Mediziner*innen, Physiotherapeut*innen) erhoben. Die Relevanz der Zielsetzung wird in Kapitel 4 *Eigener Ansatz*, nach der Darstellung des Stands der Technik, diskutiert.

1.4. Forschungsfragen und Hypothese

Der Schwerpunkt dieser Arbeit ist anhand der Zielsetzung auf drei Forschungsfragen und eine allgemeine Hypothese verteilt. Um die Struktur dieser Arbeit darzustellen, werden die Forschungsfragen und die Hypothese bereits an dieser Stelle eingeführt. Der Aufbau der folgenden Kapitel ist entsprechend der drei Fragestellungen gegliedert.

Fragestellung I - Sensorgenauigkeit

Bietet Motion-Capture-Sensorik (RGB-Tiefenkamera, Lighthouse Verfahren) eine ausreichende Gelenkwinkelgenauigkeit für eine Bewegungsbewertung?

Zunächst wird die Genauigkeit von entsprechender MoCap-Sensorik für die genannten Anwendungsfälle untersucht. Insbesondere die Genauigkeit bei der Messung von Gelenkwinkeln wird evaluiert. Dafür wird die jeweilige Sensorik anhand von technischen Vergleichsstudien mit einer optischen und markerbasierten MoCap-Sensorik evaluiert. Die Anforderungen an die Genauigkeit werden anhand der Anforderungen in den Bewertungskriterien beantwortet.

Fragestellung II - Bewegungserkennung

Mit welcher Genauigkeit können Körperhaltungen und Bewegungen auf Basis von Motion-Capture-Daten erkannt werden?

In dieser Fragestellung werden Ansätze für die Segmentierung von Bewegungen und die Erkennung von Körperhaltungen untersucht. Die Ausführungsdauer einer Körperhaltung und die Benennung der Körperhaltung oder Bewegung sind relevante Parameter für die Bewegungsbewertung. Insbesondere die Methoden Artificial Neural Network (ANN) und Support Vector Machine (SVM) werden genauer untersucht.

Fragestellung III - Bewegungsbewertung

Ist eine Bewegungsbewertung auf Basis physiologischer Gelenkwinkel und erkannter Körperhaltungen mit Einsatz von Motion-Capture-Sensorik in einem mobilen System möglich?

Zur Beantwortung der Fragestellung werden Schemata für die Bewertung von Körperhaltungen anhand von Gelenkwinkeln untersucht und in ein prototypisches System für die automatische Bewertung von Bewegungen mit Einsatz von MoCap-Sensorik implementiert und auf den mobilen Endgeräten getestet.

Hypothese

„Motion-Capture-Sensorik im Consumer-Bereich bietet eine ausreichende Genauigkeit für die Erkennung und Bewertung von Bewegungen und Körperhaltungen anhand von Gelenkwinkeln.“

1.5. Publikationen

Teile der vorliegenden Arbeit wurden bereits im Peer-Review-Verfahren begutachtet sowie in Fachzeitschriften und Konferenzbänden veröffentlicht. Ergebnisse aus publizierten Beiträgen werden an entsprechender Stelle in dieser Arbeit referenziert. Im Folgenden werden die Veröffentlichungen in chronologischer Reihenfolge aufgelistet:

- Vox, J. P., Franz, S., and Wallhoff, F. (2016). Adaptive Bewegungsanalyse von physiotherapeutischen Übungen für eine optimierte Mensch-Roboter-

Trainingsinteraktion. *Kognitive Systeme*, 2016(1), <https://doi.org/10.17185/dupublico/42386>, (Artikel, Workshop Präsentation)

- Vox, J. P. and Wallhoff, F. (2016). Evaluation of motion tracking methods for therapeutic assistance in everyday living environments. In *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Baden-Baden, Germany*, pages 96–101, <https://doi.org/10.1109/MFI.2016.7849473>, (Konferenzpaper, Präsentation)
- Vox, J. P. and Wallhoff, F. (2017b). Recognition of human motion exercises using skeleton data and SVM for rehabilitative purposes. In *2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, Australia*, pages 266–269, <https://doi.org/10.1109/LSC.2017.8268194>, (Konferenzpaper, Präsentation, best paper finalist)
- Vox, J. P. and Wallhoff, F. (2018a). Artificial Increase of 3D-Skeleton-Data for Human Motion Recognition using supervised SVM and NN. In *2018 IEEE International Conference on Intelligent Systems (IS), Funchal, Portugal*, pages 602–607, <https://doi.org/10.1109/IS.2018.8710572>, (Konferenzpaper, Präsentation)
- Vox, J. P. and Wallhoff, F. (2018b). Human Motion Recognition Using 3D-Skeleton-Data and Neural Networks. In Bramer, M. and Petridis, M., editors, *Artificial Intelligence XXXV, SGAI International Conference on Artificial Intelligence, Cambridge, England*, pages 204–209, Cham. Springer International Publishing, https://doi.org/10.1007/978-3-030-04191-5_19, (Konferenzpaper, Poster)
- Vox, J. P. and Wallhoff, F. (2018c). Preprocessing and Normalization of 3D-Skeleton-Data for Human Motion Recognition. In *2018 IEEE Life Sciences Conference (LSC), Montreal, QC, Canada*, pages 279–282, <https://doi.org/10.1109/LSC.2018.8572153>, (Konferenzpaper, Präsentation)
- Vox, J. P. and Wallhoff, F. (2019). A Framework for the Analysis of Biomechanical Loading Using Human Motion Tracking. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), Los Angeles, CA, USA*, pages 39–46, <https://doi.org/10.1109/IRI.2019.00020>, (Konferenzpaper, Präsentation)

- Vox, J. P., Weber, A., Wolf, K. I., Izdebski, K., Schüler, T., König, P., Wallhoff, F., and Friemert, D. (2021). An Evaluation of Motion Trackers with Virtual Reality Sensor Technology in Comparison to a Marker-Based Motion Capture System Based on Joint Angles for Ergonomic Risk Assessment. *Sensors*, 21(9), <https://doi.org/10.3390/s21093145>, (Artikel)

Weitere Publikationen, auch in Co-Autorenschaft, sind im Anhang A.3 aufgelistet.

1.6. Gliederung der Dissertation

Die vorliegende Arbeit ist wie folgt strukturiert: In Kapitel 2 *Grundlagen* werden die Funktionsprinzipien der untersuchten Methoden und Sensorik vorgestellt. Dieses Kapitel ist bewusst vor den Stand der Technik gesetzt, da die Grundlagen für das Verständnis zu den referenzierten Arbeiten relevant sind.

In Kapitel 3 *Stand der Technik* werden verwandte Arbeiten zu den drei Forschungsfragen vorgestellt und der aktuelle Stand der Technik jeweils zusammengefasst.

In Kapitel 4 *Eigener Ansatz* werden Zusammenhänge der drei Fragestellungen erläutert, der Stand der Technik diskutiert und der eigene Beitrag in dieser Arbeit herausgestellt.

In Kapitel 5 *Realisierung* werden konkrete Lösungsansätze und die Umsetzung von Studien zur Beantwortung der Forschungsfragen sowie Ergebnisse präsentiert.

In Kapitel 6 *Technische Evaluation* werden die Genauigkeit eines Lighthouse Tracking Systems und die Performance der entwickelten Modelle für die Bewegungserkennung auf Sub-Datensätzen mit kritischen Körperhaltungen getestet.

In Kapitel 7 *Diskussion* werden die Ergebnisse diskutiert und das entwickelte Gesamtsystem bewertet.

In Kapitel 8 *Fazit und Ausblick* wird die Arbeit zusammengefasst und ein Ausblick zu den einzelnen Fragestellungen gegeben.

Der prinzipielle Aufbau der Hauptkapitel ist entsprechend der drei Forschungsfragen in die Abschnitte **I. Sensorgenauigkeit**, **II. Bewegungserkennung** und **III.**

Bewegungsbewertung unterteilt.

Im Folgenden wird bei der Nennung von Gelenkpunkten und Gelenkwinkeln auf englische Begriffe (wie z.B. Elbow, Knee, Hip) zurückgegriffen, um eine Einheitlichkeit der Begriffe in der im Rahmen dieser Arbeit entwickelten Software und den Beschreibungen in dieser Dissertation zu schaffen.

2. Grundlagen

In diesem Kapitel wird zunächst das Datenmodell beschrieben, welches als Ausgangslage für die Verarbeitung von MoCap-Daten in dieser Arbeit verwendet wird. Des Weiteren werden die Grundlagen und Funktionsprinzipien von verwendeter und untersuchter Methodik und Sensorik eingeführt.

2.1. Beschreibung Motion-Capture-Daten

Ausgangslage für die Datenverarbeitung in dieser Arbeit sind MoCap-Daten, welche in der Regel nicht den Rohdaten der jeweiligen Sensorik entsprechen. Erst nach der Vorverarbeitung durch ein entsprechendes sensornahes Verfahren zur Bestimmung von Gelenkpositionen (Skeleton Tracking, inverse Kinematik) können MoCap-Daten bereitgestellt werden. Die Verfahren für die Bereitstellung von MoCap-Daten mit einer RGB-Tiefenkamera und nach dem Lighthouse Verfahren werden in den Abschnitten 2.2 und 2.3 erläutert. In verwandten Arbeiten werden für MoCap-Daten u.a. die Bezeichnungen *Skeleton Data* oder *Bewegungsdaten* verwendet. In dieser Arbeit sind *MoCap-Daten* wie folgt definiert.

MoCap-Daten repräsentieren Körperhaltungen und Bewegungen anhand von kartesischen Koordinaten (Positionen) und Quaternionen als Zeitreihe. Ein Sample in den Zeitreihendaten besteht aus ca. 25 Gelenk- oder Skelettpunkten¹, die sich zu einem abstrakten menschlichen Skelettmodell zusammensetzen (Skelettierung). Die Anzahl der Gelenkpunkte variiert gemäß dem verwendeten Skelettmodell von Sensorik und Software für die Bereitstellung von MoCap-Daten.

Ein Gelenkpunkt, z.B. Knie, Schulter oder Ellenbogen, wird durch ein 7-Tupel $P_{m,n} = (x, y, z, w, xi, yj, zk)$, bestehend aus der Position (x, y, z) sowie dem Quaternion (w, xi, yj, zk) , beschrieben. In dieser Arbeit sind MoCap-Daten als multidimensionales Array mit n -Gelenkpunkten und m -Samples über die Zeit defi-

¹Im Weiteren werden Gelenk- oder Skelettpunkte nur durch den Begriff *Gelenkpunkte* beschrieben.

niert (siehe A in Gleichung 2.1).

$$A = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1n} \\ P_{21} & P_{22} & \cdots & P_{2n} \\ \vdots & \vdots & & \vdots \\ P_{m1} & P_{m2} & \cdots & P_{mn} \end{bmatrix} \quad (2.1)$$

Die Gelenkpunkte P sind in der Regel anhand einer festgelegten Skelett-Hierarchie miteinander verknüpft. Durch jeden Knochen ist ein Gelenk mit einem übergeordneten Gelenk verbunden. Beispielsweise bezieht sich das Handgelenk auf das Ellenbogengelenk und dieses wiederum bezieht sich auf das Schultergelenk. Ursprung ist generell ein Gelenk im Körperzentrum. Gelenkpunkte (Positionen und Quaternionen) in einer zusammenhängenden Skelett-Hierarchie sowie die Rohdaten einer RGB-Tiefenkamera als Punktwolke sind in Abbildung 2.1 in einem dreidimensionalen Koordinatensystem dargestellt.

Die räumliche Drehung jedes Gelenkpunktes ist anhand der Quaternionrepräsentation durch die Ausrichtung der X- (rot), Y- (grün) und Z-Achse (blau) gezeigt (siehe auch Abschnitt 2.1.2). Zur Veranschaulichung sind auch die Rohdaten als Punktwolke, aufgenommen durch eine RGB-Tiefenkamera, präsentiert.

In den untersuchten MoCap-Daten sind die Zahlenwerte in der Regel mit einer Auflösung von 64 bit als Fließkommazahl dargestellt. Bei 25 Gelenkpunkten, jeweils repräsentiert durch sieben Zahlenwerte, entspricht das einer Datenmenge von 1,4 *KB* per Sample. Bei einer angenommenen Abtastrate von 30 Hz läge eine Aufnahme von 60 Sekunden bei 2,52 *MB* (ohne Rohdaten).

MoCap-Daten sind i.d.R. nicht standardisiert. Die Achsenausrichtung, Rotationsrichtung sowie Anzahl und Positionierung der Gelenkpunkte im Skelettmodell sind vorwiegend durch die Sensortechnologie, Hersteller oder das verwendete Software Development Kit (SDK) für das Skeleton Tracking vorgegeben. Für eine unabhängige Datenverarbeitung werden in Abschnitt 5.1 Prozessschritte für die Vereinheitlichung (Transformation und Normalisierung) von MoCap-Daten vorgestellt.

2.1.1. Kartesische Koordinaten von Gelenkpunkten

Die kartesischen Koordinaten werden vorwiegend in Meter angegeben. Ursprung des Koordinatensystems ist bei einem RGB-Tiefenkamera System allgemein die räumli-

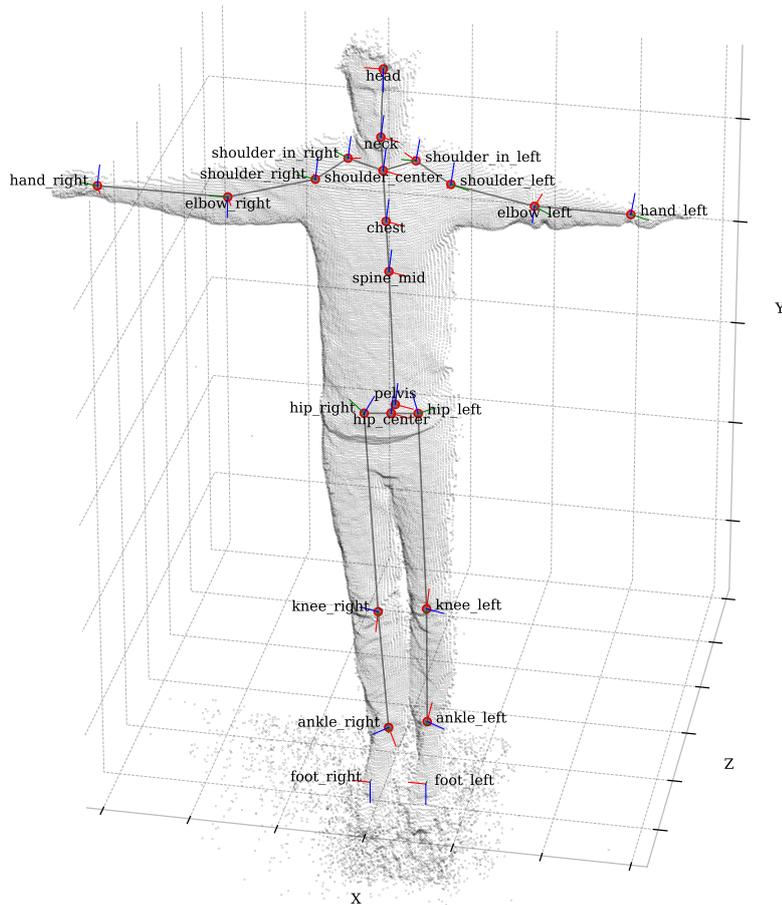


Abbildung 2.1.: Darstellung eines Samples aus MoCap-Daten mit 24 Gelenkpunkten und den Rohdaten einer RGB-Tiefenkamera als Punktwolke

che Position der Kamera. Bei tragbarer Sensorik liegt der Ursprung gewöhnlich auf einem Gelenkpunkt oder einer übergeordneten Position mit einer fixen Zuordnung zu den Gelenkpunkten.

Anhand der Koordinaten können Rückschlüsse auf anthropometrische Informationen wie Körpergröße und Gliedmaßenlängen gezogen werden. In Abschnitt 7.4 *Per-*

sonenbezogene Informationen in MoCap-Daten werden Methoden für die Anonymisierung von Merkmalen in MoCap-Daten diskutiert.

2.1.2. Räumliche Drehung anhand von Quaternionen

Quaternionen sind in dieser Arbeit insbesondere für die Berechnung von Rotationswinkeln von Gelenken relevant. In den MoCap-Daten repräsentieren die Quaternionen die räumliche Drehung von Gelenkpunkten.

Räumliche Drehungen können ebenfalls anhand von Drehwinkeln (Yaw ψ , Pitch θ , Roll φ) um die Z-, Y-, X-Achsen² angegeben werden (siehe Abbildung 2.2a). Diese Repräsentation birgt jedoch die Gefahr des sogenannten *Gimbal Lock*. Dieser tritt ein, wenn zwei Achsen parallel zueinander gedreht sind, was zu einem Verlust eines Freiheitsgrades führt [Perumal, 2011]. Dieser Zustand kann bei der Berechnung der Drehwinkel zu einer Teilung durch 0 führen und muss verhindert werden.

Zur Lösung des Problems können Drehungen anhand von Quaternionen abgebildet werden. Wie in Gleichung 2.2 gezeigt, wird eine Quaternion durch einen Realteil w und den Imaginärteil xi, yj, zk beschrieben.

$$q = \underbrace{w}_{\text{Real}} + \underbrace{xi + yj + zk}_{\text{Imaginary}} \quad (2.2)$$

Demnach ist eine Quaternion eine komplexe Zahl mit den imaginären Anteilen i, j, k . Folgende Gleichung 2.3 zeigt die Beziehung zwischen den Imaginärteilen, die nach Hamilton, dem Erfinder der Quaternionen, aufgestellt wurde [Shuster, 2008].

$$i^2 = j^2 = k^2 = ijk = -1 \quad (2.3)$$

Verschiedene Notationen

Nach Notation des NASA Jet Propulsion Laboratory (JPL) gilt $ijk = 1$ für die komplexen Zahlen [Trawny and Roumeliotis, 2005]. Hamilton und JPL unterscheiden sich auch darin, ob der Drehsinn einer Rotation mit der Rechte- oder Linke-Hand-Regel beschrieben wird, sowie in der Reihenfolge der vier Zahlenwerte. Hamilton stellt den Realteil w nach vorne, nach JPL wird dieser an letzter Stelle angegeben.

²Nach [DIN ISO 8855, 2013], ZYX-Konvention, wird zuerst um die Z- (Yaw), dann Y- (Pitch) und anschließend um die X-Achse (Roll) gedreht.

Die Unterschiede zwischen beiden Notationen sind in Tabelle 2.1 aufgelistet. Beide Notationen können in MoCap-Daten verschiedener Hersteller vorkommen [Solà, 2017].

Tabelle 2.1.: Unterschiede bei der Notation einer Quaternion zwischen Hamilton und JPL

Hamilton	JPL
$ijk = -1$	$ijk = 1$
(w, xi, yj, zk)	(xi, yj, zk, w)
Rechte-Hand-Regel	Linke-Hand-Regel

Quaternionen werden in der Regel als Einheitsquaternion angegeben. Das bedeutet, dass der Betrag einer Quaternion $|q| = \sqrt{w^2 + xi^2 + yj^2 + zk^2} = 1$ ergibt. Werden Quaternionen nicht als Einheitsquaternion angegeben, ist eine Normalisierung notwendig (siehe Abschnitt 5.1).

Beschreibung anhand exemplarischer Zahlenwerte

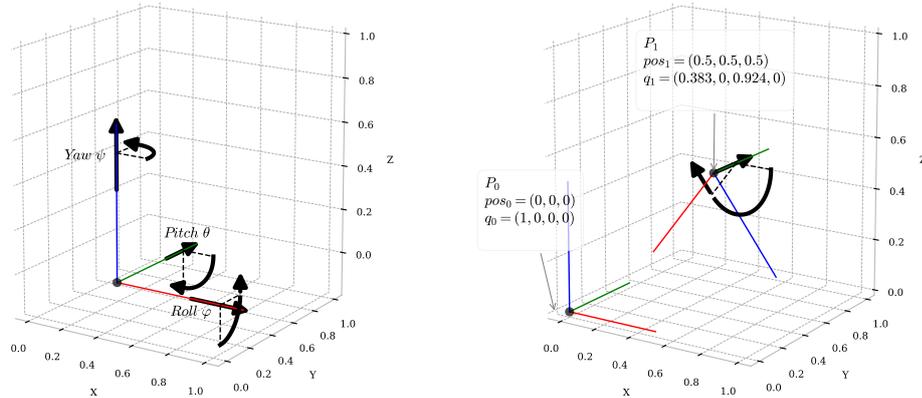
Im Folgenden wird die räumliche Drehung, beschrieben durch Quaternionen, anhand exemplarischer Zahlenwerte erläutert.

Ein Beispiel für eine räumliche Drehung in den Punkten P_0 und P_1 , mit Angabe der Einheitsquaternionen q_0 und q_1 , ist in Abbildung 2.2b dargestellt. Eine Beweisführung mit den realen Zahlenwerten der Quaternionen q_0 und q_1 , dass es sich um Einheitsquaternionen handelt, ist in Gleichung 2.4 gezeigt.

$$|q_0| = 1 = \sqrt{1^2 + 0^2 + 0^2 + 0^2}, \quad |q_1| = 1 = \sqrt{0.383^2 + 0^2 + 0.924^2 + 0^2} \quad (2.4)$$

P_0 beschreibt einen Punkt im Ursprung des Koordinatensystems ohne Drehung. Der Imaginärteil $(0, 0, 0)$ von q_0 , welcher einen Vektor beschreibt, zeigt in keine Richtung. Somit ist keine Drehung um diesen Vektor möglich. In P_1 ist eine Drehung von 135° um die Y-Achse (grün) zu sehen. Diese Drehung wird anhand der Quaternion $q_1 = (0.383, 0, 0.924, 0)$ angegeben. Der Imaginärteil $(0, 0.924, 0)$ von q_1 ist durch einen Vektor dargestellt, welcher in diesem Fall in positiver Richtung auf der grünen Y-Achse liegt. Anhand des Realteils $w = 0.383$ wird um den Vektor gedreht. Gleichung 2.5 zeigt die Transformation von Realteil w zu einem Drehwinkel α in Grad. In diesem Fall entspricht q_1 einem Pitch von 135° .

$$\alpha = 2 \arccos w \cdot 180/\pi, \quad 135^\circ = 2 \arccos 0.383 \cdot 180/\pi \quad (2.5)$$



(a) Yaw-Pitch-Roll: Drehung anhand Drehwinkel um die Z-, Y-, und X-Achse

(b) Quaternion: Drehung in P_1 beschrieben durch den Vektor im Imaginärteil $(0, 0.925, 0)$ und Drehung um den Vektor, definiert durch Realteil (0.381) ; entspricht Drehung von 135° um die Y-Achse; keine Drehung in P_0

Abbildung 2.2.: Räumliche Drehung von Objekten bzw. Gelenkpunkten; Vergleich zwischen den Beschreibungen nach Yaw-Pitch-Roll und Quaternion

In einer Skelett-Hierarchie orientiert sich der Ursprung einer Quaternion an einer festgelegten übergeordneten Quaternion. Angenommen q_0 beschreibt die Drehung im Schultergelenk und q_1 das Ellenbogengelenk. Definiert die Skelett-Hierarchie, dass das Schultergelenk dem Ellenbogen übergeordnet ist, dann orientiert sich q_1 unmittelbar an der Ausrichtung von q_0 . Sind Quaternionen nicht in einer Skelett-Hierarchie angeordnet, wird die Drehung zu einem globalen Ursprung angegeben.

Die Interpretation von Quaternionen ist in der Regel nicht intuitiv möglich. Die Drehung eines Gelenkes wird vornehmlich als physiologischer Gelenkwinkel angegeben. Dafür ist u.a. die Transformation von Quaternionen zu Yaw-Pitch-Roll erforderlich. Die Berechnung von physiologischen Gelenkwinkeln aus MoCap-Daten ist in Abschnitt 5.2 erläutert.

2.2. Motion Tracking mit einer RGB-Tiefenkamera

In diesem Abschnitt wird zunächst das Funktionsprinzip einer RGB-Tiefenkamera und anschließend das Verfahren für die Bereitstellung von MoCap-Daten aus den Rohdaten (RGB-Bilder, Punktwolke) beschrieben. Als RGB-Tiefenkamera wird ein Kameraverbund aus einer RGB-Kamera und einer Tiefenkamera in einem Sensor verstanden. Erst durch die Vorverarbeitung mittels Skeleton Tracking werden die Gelenkpunkte in den Rohdaten erkannt und anhand einer Skelett-Hierarchie zusammengesetzt. Das prinzipielle Schema für die Erfassung von MoCap-Daten mit einer RGB-Tiefenkamera ist in Abbildung 2.3 dargestellt.

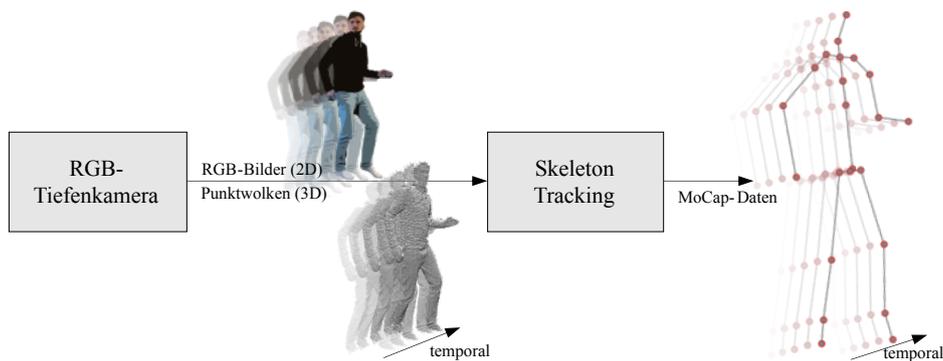


Abbildung 2.3.: Prinzipielle Datenverarbeitung bei einer RGB-Tiefenkamera als Motion-Capture-Sensorik

In dieser Arbeit wurde hauptsächlich mit der RGB-Tiefenkamera Microsoft Kinect v2 gearbeitet. Nachfolger der Kinect v2 ist die Microsoft Azure, welche 2019 veröffentlicht wurde. Referenzen und Erklärungen beziehen sich vorwiegend auf die Sensorik von Microsoft und das Time-of-Flight (ToF) Prinzip. Vergleichbare Tiefenkamera-Systeme und -Technologien wie Intel Realsense (ToF), Stereolabs ZED (Stereoscopy), Orbbec Astra (Structured-light) oder Asus Xtion (Structured-light) eignen sich ebenfalls für die Erfassung von Tiefeninformationen [D’Antonio et al., 2020, Yeung et al., 2021].

Ebenso verweisen die Erklärungen für das Skeleton Tracking auf die von Microsoft entwickelten Verfahren. Das Skeleton Tracking mit der Kinect v2 basiert auf dem Kinect SDK 2.0. Mit Einführung der Azure wurde das *Azure Kinect Body Tracking SDK* vorgestellt. Rudimentär können die gleichen Ansätze auch in anderen Skeleton Tracking Verfahren, wie OpenPose [Cao et al., 2021], DeepPose [Toshev and Szegedy, 2014], BlazePose [Bazarevsky et al., 2020], PoseNet [Papandreou et al.,

2018], MoveNet [Bajpai and Joshi, 2021], VNect [Mehta et al., 2017], Nuitrack [Li et al., 2019] oder DepthHRNet [Büker et al., 2021] angenommen werden.

Im Wesentlichen unterscheiden sich die Modelle für das Skeleton Tracking in den Eingabedaten (RGB- und/oder Tiefenbilder) und in der Anzahl der ausgegebenen Gelenkpunkte sowie der verwendeten Skelettmodelle. Des Weiteren finden sich Unterschiede in der Methodik für die Erkennung von Gelenkpunkten sowie in der Grundlage und Menge der Daten, mit welchen das jeweilige Modell trainiert wurde. Eine Evaluierung der Genauigkeiten zwischen verschiedenen Systemen ist in Abschnitt 3.1.1, im Kapitel Stand der Technik, eingeführt.

2.2.1. Funktionsweise RGB-Tiefenkamera

Eine RGB-Tiefenkamera besteht im Wesentlichen aus einer RGB-Kamera, einer Infrarot (IR)-Kamera und einem IR-Emitter. Durch die RGB-Kamera können 2D-Farbbilder mit den Farbinformationen für Rot, Grün und Blau (RGB) aufgenommen werden. Tiefeninformationen werden bei der Kinect v2 und der Azure nach dem ToF Prinzip gewonnen. Der IR-Emitter sendet amplitudenmodulierte Lichtimpulse aus, welche von Objekten oder Strukturen zur IR-Kamera reflektiert werden. Anhand der Phasenverschiebung zwischen emittiertem und reflektiertem Licht kann die *time of flight* des Lichtes in einen Entfernungswert oder Tiefe für jeden Pixel umgerechnet werden [Tölgyessy et al., 2021a].

In Abbildung 2.1 ist die Tiefe per Pixel als Punktwolke am Beispiel einer aufgenommenen Person in T-Position dargestellt. In Abbildung 2.4 ist der prinzipielle Aufbau einer RGB-Tiefenkamera am Beispiel der Kinect v2 dargestellt.

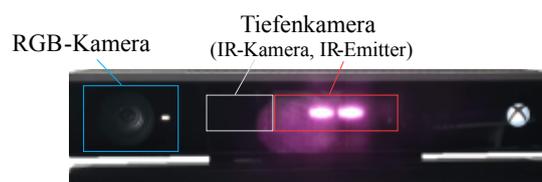


Abbildung 2.4.: Wesentliche Komponenten einer RGB-Tiefenkamera am Beispiel der Kinect v2

In Tabelle 2.2 sind typische Eigenschaften aktueller RGB-Tiefenkameras am Beispiel der Kinect v2 und der Azure aufgelistet. Die Azure ermöglicht verschiedene

Modi (breites oder enges Sichtfeld) mit unterschiedlichen Auflösungen bei der Tiefenbilderfassung.

Tabelle 2.2.: Typische Eigenschaften von RGB-Tiefenkameras am Beispiel der Kinect v2 und Azure [Tölgyessy et al., 2021a, Tab. 1]

Eigenschaft	Kinect v2	Azure
Auflösung RGB-Kamera	1920 × 1080 px (30 Hz)	3840 × 2160 px (30 Hz)
Auflösung Tiefenkamera	512 × 424 px (30 Hz)	640 × 576 px (30 Hz) 320 × 288 px (30 Hz) 1024 × 1024 px (15 Hz) 512 × 512 px (30 Hz)
Messbereich	0,5 – 4,5 m	0,5 – 5,5 m

Technologiebedingte Einschränkungen

RGB-Tiefenkameras sind aufgrund der Technologie nur in bestimmten Umgebungen und Anwendungsfällen einsetzbar. Die Nutzung ist überwiegend in beschatteten Umgebungen oder zu Tageszeiten mit wenig Tageslicht möglich. Einfallendes Sonnenlicht (Infrarot) kann die IR-Kamera stören. Ebenfalls kann durch Spiegelungen oder stark strukturierten Hintergründen die Funktionalität eingeschränkt sein. Ist die Reflexion von IR-Licht zu der IR-Kamera nicht ausreichend oder nur fehlerhaft möglich, kann keine Tiefeninformation gewonnen werden.

Die Erfassung von Personen und Gliedmaßen ist nur im Sichtfeld der Kamera möglich. Bei verdeckten Gliedmaßen kann keine exakte Position ermittelt werden. Auch locker getragene Kleidung kann die Sicht auf Gliedmaßen einschränken. Ebenfalls kann das Tracking durch den Kamerablickwinkel beeinflusst sein. Das ist weniger eine technologische Einschränkung, sondern auf das Skeleton Tracking zurückzuführen. Die Modelle für das Skeleton Tracking sind in der Regel mit frontal aufgenommenen Bildern trainiert [Yeung et al., 2021].

2.2.2. Skeleton Tracking in RGB- und Tiefenbildern

Skeleton Tracking beschreibt in dieser Arbeit das Verfahren für die Bereitstellung von MoCap-Daten aus RGB-Bildern und/oder Punktwolken. Dabei wird in der Regel in einzelnen Bildern (Frames) mittels Mustererkennung nach Gelenkpunkten

gesucht, welche dann in einer Skelett-Hierarchie zusammengeführt werden [Munea et al., 2020]. Um eine Echtzeit-Interaktion mit einer Benutzerschnittstelle zu ermöglichen, wird im Folgenden von echtzeitfähigem Skeleton Tracking ausgegangen.

Lange galt die Arbeit von [Shotton et al., 2013], basierend auf einem Random Forest Algorithmus, als Grundlage für die Erkennung von Gelenkpunkten. Dieses Verfahren ist in dem Kinect SDK 2.0 implementiert und in der Lage, 25 Gelenkpunkte von multiplen Nutzern gleichzeitig zu tracken. Shotton et al. trainierten das Modell mit einer großen Datenbank, mit 500.000 gelabelten Segmenten in Bildern, von unterschiedlichen Aktivitäten (Tanzen, Treten, Laufen, etc.). Bei dieser Erkennung wird nur die Tiefenbild-Information verwendet.

Im Jahr 2019 präsentierte [Liu, 2019] von Microsoft Research das Verfahren für das Skeleton Tracking in dem Azure Kinect Body Tracking SDK. Darin wurde die Erfassung um Gesichtspunkte wie Mund, Nase, Ohren und Positionen am Rücken erweitert und umfasst insgesamt 32 Punkte. Eine vereinfachte Darstellung der Verarbeitungsschritte ist in Abbildung 2.5 wiedergegeben.

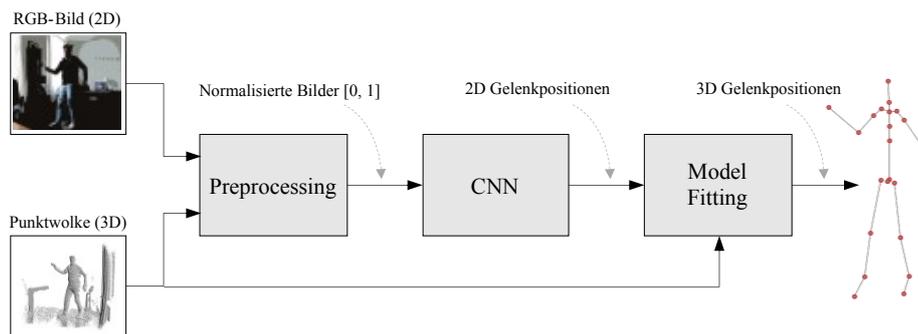


Abbildung 2.5.: Vereinfachte Darstellung des Skeleton Trackings im *Azure Kinect Body Tracking SDK*

Das vorgestellte Skeleton Tracking verarbeitet RGB-Bilder und Punktwolken. Im Preprocessing werden die Bilder in einen normalisierten Wertebereich $[0, 1]$ skaliert. Kern der Erkennung ist ein Convolutional Neural Network (CNN). Das CNN ist darauf trainiert, die Silhouette einer Person zu erkennen, in Segmente (Körperregionen) zu zerlegen, und Gelenkpositionen als 2D-Koordinaten auszugeben. Im Model Fitting werden die 2D-Gelenkpositionen durch die Tiefeninformation in der Punktwolke zu 3D-Koordinaten erweitert. Anschließend können die Gelenkpositio-

nen anhand einer festgelegten Skelett-Hierarchie zusammengeführt und ausgegeben werden.

An der Weiterentwicklung und Optimierung des Skeleton Trackings wird weitergearbeitet. In [Lin et al., 2021a, Lin et al., 2021b] veröffentlichte Microsoft Verfahren zur Verbesserung der Schätzung von verdeckten Gliedmaßen. Darin wurde eine Rekonstruktion von Körperhaltungen, ebenfalls auf dem Prinzip von CNNs, vorgestellt.

2.3. Motion Tracking mit einem Lighthouse Tracking System

Das Lighthouse Verfahren ist ein von Valve entwickeltes VR-Tracking-Verfahren, welches in den Systemen Valve Index und HTC Vive verwendet wird. VR-Tracking Systeme werden eingesetzt, um die Bewegungen einer Person aus der Ich-Perspektive in Echtzeit in eine Simulation zu integrieren. Die Technologie und Funktionsweise divergiert zwischen Herstellern und Versionen jeweiliger Systeme. Das HTC Vive ist eines der am weitesten verbreiteten Systeme, welches auf Basis der räumlichen Positionen eine hohe Genauigkeit aufweist. Aus diesem Grund wird in dieser Arbeit das HTC Vive näher betrachtet [Ikbal et al., 2021, Bauer et al., 2021a].

Das System besteht aus einem Head Mounted Display (HMD), zwei Controllern, einer variablen Anzahl von Trackern und den Basisstationen (*lighthouses*). In Abbildung 2.6 ist das prinzipielle Verfahren für die Erfassung von MoCap-Daten mittels Lighthouse Tracking System dargestellt.

Die Datenschnittstelle ist in der Regel für die Integration in eine 3D-Engine wie Unreal oder Unity ausgelegt. Anhand der erfassten Lage und Positionen von Trackern kann die Bewegung auf virtuelle Gliedmaßen oder Objekte transportiert werden. Ein menschlicher Avatar besitzt jedoch, wie ein realer Mensch, deutlich mehr Freiheitsgrade als bspw. die Tracking-Positionen der Füße, Hüfte, Hände, Oberarme, Brust und Kopf beschreiben könnten. Um einen menschlichen Avatar in VR natürlich abbilden zu können, ist die Berechnung weiterer Gelenkpunkte notwendig. Zu diesem Zweck wird eine Inverse Kinematik (IK) eingesetzt. Anhand der IK können MoCap-Daten aus den Tracking-Positionen berechnet und auf einen Avatar projiziert werden (siehe Abbildung 2.7).

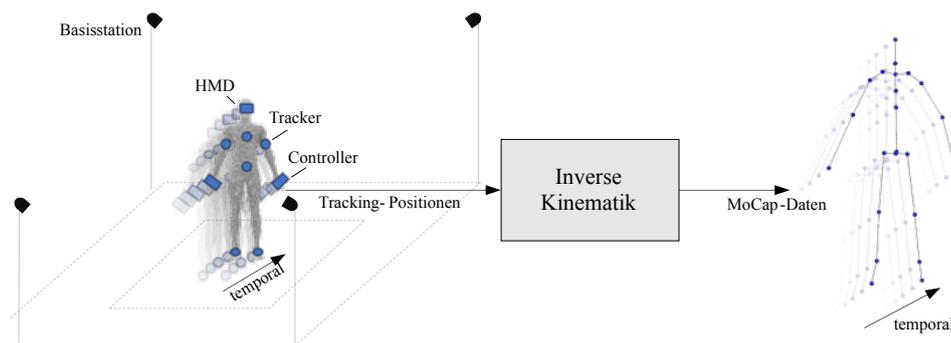


Abbildung 2.6.: Prinzipielles Verfahren eines Lighthouse Tracking Systems für die Erfassung von MoCap-Daten



Abbildung 2.7.: Vergleich zwischen realer und virtueller Körperhaltung; Tracking-Positionen von HMD, Controller und Trackern werden durch eine inverse Kinematik erweitert und auf einen Avatar projiziert

In dieser Arbeit werden die aus der IK berechneten MoCap-Daten ebenfalls verwendet, um Bewegungen anhand ergonomischer Kriterien zu bewerten und zu analysieren. Für diesen Fall ist es wichtig, dass die berechneten Gelenkpunkte möglichst realitätsnah abgebildet werden. Typische IK Anbieter für VR-Anwendungen sind *Final IK* oder *IKinema*. Die Genauigkeit von MoCap-Daten mit dem HTC Vive und Final IK wird in Kapitel 5.4 evaluiert.

2.3.1. Funktionsweise Lighthouse Tracking

Zu den Basiskomponenten eines Lighthouse Tracking Systems zählen Basisstation, HMD, Controller und Tracker. Tracker werden gewöhnlich an den Gliedmaßen und am Rumpf befestigt. Die Controller sind sowohl Eingabegeräte als auch Tracker für die Hände. In Abbildung 2.8 sind die Geräte eines Lighthouse Tracking Systems am Beispiel des HTC Vive dargestellt.

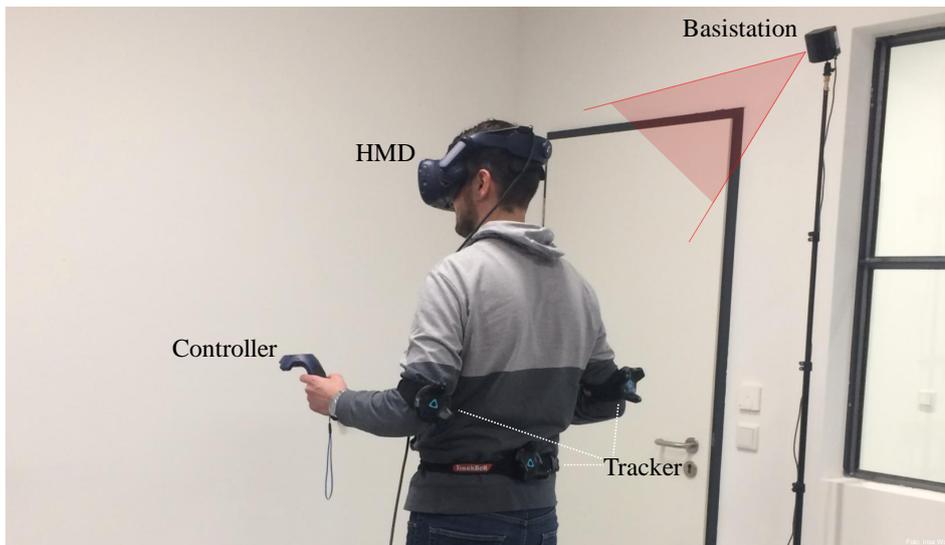


Abbildung 2.8.: Lighthouse Tracking System *HTC Vive* mit Basisstation, HMD, Controller und Trackern an Oberarmen und Hüfte

Kern des Lighthouse Trackings sind Licht emittierende Basisstationen. Jede Basisstation enthält ein IR-Array und zwei Laser der Klasse 1, die mit bis zu 60 Hz einen Raum in horizontaler und vertikaler Richtung scannen. Das IR-Array sendet diffuse IR-Impulse aus, welche von Photodioden (Rezeptoren) auf dem HMD, Controller oder den Trackern detektiert werden. Wird ebenfalls ein Laserstrahl von den Rezeptoren detektiert, kann anhand der Zeit zwischen IR-Impuls und Laserstrahl eine Entfernung zwischen Basisstation und Rezeptor ermittelt werden. Das HTC Vive System ist am HMD mit 31, an den Controllern jeweils mit 24 und an den Trackern mit 22 Rezeptoren bestückt. Laut Hersteller müssen wenigstens fünf der Rezeptoren im Sichtfeld zur Basisstation liegen, um anhand von Trigonometrie die räumliche Position zu ermitteln. Um bei Sichtverlust zu den Basisstationen das Tracking aufrechtzuerhalten, ist ebenfalls eine Inertial Measurement Unit (IMU) verbaut. Anhand von gemessenen Beschleunigungen und der Lage kann der Fortgang eines Positionsverlaufes geschätzt werden, bis der Sichtkontakt zu einer Basis-

station wieder hergestellt ist. Ein typischer Messraum mit zwei Basisstationen wird mit $4,5\text{ m} \times 4,5\text{ m}$ angegeben [Sitole et al., 2020, Bauer et al., 2021b].

Die Basisstationen benötigen ausschließlich Strom und sind nicht direkt mit dem System verbunden. Die Lokalisation der Positionen von HMD, Controller und Tracker wird auf den Geräten selbst bestimmt. Dieses Verfahren wird auch *inside-out* (from inside looking outside) genannt. Die Geräte sind in der Lage, ihre Positionen zu bestimmen und brauchen keine zusätzliche Kommunikationsschnittstelle zu den Basisstationen [Ikbal et al., 2021].

2.3.2. Inverse Kinematik für einen menschlichen Avatar

Die meisten Animationen in VR-Anwendungen können durch die Information der Positionen (x, y, z) und der Lagewinkel (ψ, θ, φ) von HMD, Controller oder Tracker erzeugt werden. Um Bewegungen auf einen menschlichen Avatar zu projizieren, ist die Berechnung zusätzlicher Gelenkstellungen und -positionen erforderlich. Wie bereits in Abschnitt 2.1 beschrieben, sind MoCap-Daten anhand einer festgelegten Skelett-Hierarchie aufgebaut. Dadurch sind bestimmte Gelenkketten wie *Shoulder - Elbow - Hand* fest definiert.

In einer *Inversen Kinematik* wird jede Gelenkstellung von einem entgegengesetzten Standpunkt aus betrachtet. Sind bspw. die Position und Ausrichtung der Hand bekannt, kann rückwärts anhand einer Gelenkkette gearbeitet werden, um die Gelenke von Ellenbogen und Schulter auszurichten. Eine prinzipielle Beschreibung einer IK, am Beispiel der Gelenkkette *Shoulder - Elbow - Hand*, mit einem Controller als Startposition der Hand, ist in Abbildung 2.9 dargestellt.

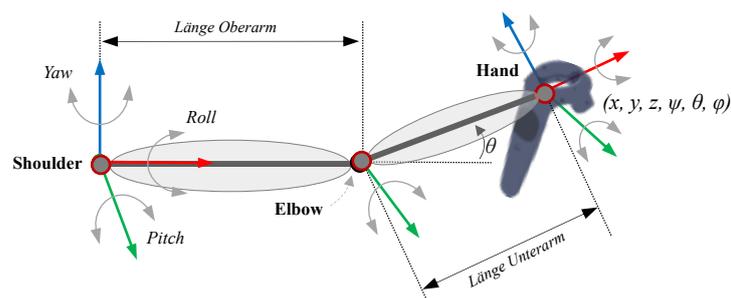


Abbildung 2.9.: Beschreibung einer IK am Beispiel der Gelenkkette *Shoulder - Elbow - Hand*; Eingabe der Startposition an der Hand via Controller

Neben den Startpositionen sind weitere Parameter in der IK notwendig, um eine Gelenkstellung möglichst genau zu berechnen. Zum einen ist das System auf die in jedem Gelenk physiologisch mögliche Bewegungsamplitude, die Range of Motion (RoM), entlang von Bewegungsachsen (den Freiheitsgraden) zu begrenzen. Zum Beispiel ist die Bewegung im Ellenbogengelenk zwischen Ober- und Unterarm nach der Neutral-Null-Methode [Ryf and Weymann, 1995] nur zwischen $10^\circ/0^\circ/150^\circ$ (von Extension nach Flexion) möglich. Die Kombinationen der möglichen Gelenkstellungen für die IK werden durch die Begrenzung anhand der RoM und der Freiheitsgrade für jedes Gelenk eingeschränkt [Chen et al., 2018].

Zum anderen ist die Angabe von Gliedmaßenlängen, wie die Länge des Ober- oder Unterarms, wesentlich für die Berechnung. Dafür wird in VR-Anwendungen zunächst ein Kalibrierungsschritt durchgeführt, bei dem die Controller eingesetzt werden, um den Körper zu vermessen. Typischerweise müssen gewisse Körperhaltungen wie die T-Position eingenommen werden, um anhand dessen die Gliedmaßenlängen zu konstruieren [Caserman et al., 2019].

Anhand der Kalibrierungsschritte ist jedoch nur eine Abschätzung der Körpermaße möglich. Werden die MoCap-Daten eines Avatars auch für eine Bewegungsanalyse verwendet, ist es sinnvoll, die Körpermaße so exakt wie möglich in die Berechnungsgrundlage der IK einzugeben [Vox et al., 2021]. In [Jürgens et al., 1998] sind die Messpunkte von anthropometrischen Daten nach [DIN EN ISO 7250-1, 2017] angegeben und können als Grundlage für eine manuelle Bestimmung der Körpermaße herangezogen werden.

In die IK können eine variable Anzahl von Tracking-Positionen eingegeben werden, um Gelenkstellungen zu berechnen. Die minimale Konfiguration ist durch die Positionen der Hände (Controller) und des Kopfes (HMD) beschrieben. Dabei ist kein Tracking der unteren Gliedmaße möglich. Erst mit zusätzlichen Trackern an den Füßen, Oberarmen und Torso ist eine Berechnung von allen Gelenken eines Avatars durchführbar. Zwar mögen die berechneten Gelenkstellungen für das VR-Erlebnis ausreichen, dennoch ist nicht sichergestellt, dass diese auch den realen Gelenkstellungen entsprechen. Ebenfalls kann es durch Messartefakte oder das Verdrutschen von Trackern zu einer Verzerrung zwischen den realen und virtuellen Körperhaltungen kommen. Eine Evaluation der Genauigkeit des HTC Vive Systems in Kombination mit Final IK ist in Abschnitt 5.4 durchgeführt.

2.4. Algorithmen für die Erkennung von Körperhaltungen

Algorithmen für die Erkennung von Körperhaltungen anhand von MoCap-Daten sind in der Wissenschaft breit untersucht (siehe Abschnitt 3.2 in Stand der Technik). Generell wird zwischen Ansätzen aus dem Machine Learning (ML) und Deep Learning (DL) unterschieden. Mit beiden Ansätzen wird ein Computer befähigt, aus Daten zu *Lernen*, ohne das Wissen explizit zu programmieren. Damit unterscheiden sich ML- und DL-Ansätze wesentlich von Expertensystemen, in denen mit *Wenn-Dann-Regeln* differenziert wird. Des Weiteren wird zwischen überwachtem und unüberwachtem Lernen (Training) der jeweiligen Modelle unterschieden. Bei überwachtem Training ist die Antwort für das Modell bekannt. Das heißt, zu jedem Datensatz liegt eine eindeutige Beschreibung (Label) vor. Bei unüberwachtem Training werden die jeweiligen Modelle für Clustering herangezogen, um zwischen unbekanntem Mustern in Daten automatisch zu differenzieren.

Das ML beschreibt generell Algorithmen, in denen mathematische Methoden sowie statistische Verfahren eingesetzt werden, um Wissen automatisch zu erlernen. Als ML-Modell wird ein Algorithmus verstanden, welcher ein automatisch erworbenes Wissen repräsentiert. Generell werden ML-Modelle eingesetzt, um Daten automatisch zu interpretieren bzw. klassifizieren zu können (bspw. Mustererkennung in Bild-Daten). Generell gelten ML-Ansätze als robust und können auch in sicherheitskritischen Systemen eingesetzt werden [Nusser et al., 2009].

Das Preprocessing (Vorverarbeitung) und Feature Selection erfordert bei ML-Ansätzen besondere Aufmerksamkeit. Als *Feature* wird ein individuell messbarer Wert (Merkmal) verstanden, welcher einem Objekt in einem Datensatz zugeordnet ist. Gewöhnlich werden Objekte durch n-dimensionale Features beschrieben. Unter *Feature Selection* wird eine Auswahl der n-dimensionalen Features mit den besten Diskriminationsfähigkeiten verstanden. Gute Genauigkeiten können erzielt werden, wenn in den Features klare Abgrenzungen entlang der Klassen gefunden werden können und die Datenlage mit wenigen Störgrößen oder Rauschen behaftet ist.

Unter DL werden Ansätze verstanden, welche im Wesentlichen auf einem Artificial Neural Network (ANN) basieren. Ein ANN folgt einer mathematischen Logik, welche dem biologischen Vorbild der Neuronen nachempfunden ist. Das Grundprinzip wird in der Wissenschaft auch als *Multilayer Perceptron* (MLP)-Ansatz be-

schrieben [Gardner and Dorling, 1998]. Als *Network* werden über Schichten (Layer) miteinander verbundene Neuronen verstanden. Mit *Deep* ist dabei ein Netzaufbau gemeint, welcher zwischen Eingabe- und Ausgabebayere eine gewisse *Tiefe* aufweist (siehe Abschnitt 2.4.2). Generell ist DL für Aufgaben der Computer Vision ausgelegt, in denen nach Mustern in großen Pixel-Matrizen gesucht wird. Erweiterungen von ANN sind Convolutional Neural Networks (CNN), Recurrent Neural Network (RNN) oder auch Long Short-Term Memory (LSTM) Networks. Unterschiede liegen in der Vorverarbeitung (Berechnung von Feature Maps bei CNN), Aufbau (rückgekoppelte Neuronverbindungen bei RNN) oder Verfahren für eine Optimierung der Wissensrepräsentation und des Trainings von Netzen³.

Prinzipiell eignen sich DL-Ansätze für alle komplexen Probleme, in denen Zusammenhänge in einer Datenbasis vorliegen und Muster nicht einfach erkennbar sind. Das Training der Modelle erfordert in der Regel eine große Menge an Daten. Das Preprocessing und das Feature Extraction ist weniger eine Herausforderung, da die Modelle eigenständig erlernen, welche Merkmale für die Differenzierung relevant sind. Insbesondere bei CNN ist bei großen Datenmengen eine hohe Rechenperformance erforderlich, um Modelle zu berechnen.

Im Folgenden werden eine Support Vector Machine (SVM), ein Ansatz aus dem ML-Bereich und das Grundprinzip von ANN vorgestellt. Die Ansätze werden anhand von konkreten Daten für die Erkennung von Körperhaltungen in Zeitreihendaten exemplarisch erläutert.

2.4.1. Support Vector Machine

Eine SVM ist ein Algorithmus für die Klassifikation von unbekanntem Objekten in bekannten Kategorien. Das Grundprinzip besteht darin, die Daten in Klassen einzuordnen und eine Grenze zwischen den Klassen zu berechnen. Dabei wird versucht, die Grenze so zu setzen, dass ein möglichst breiter Rand (Margin) zwischen den Klassen frei bleibt, um die Unterscheidbarkeit zu optimieren. Als optimal gilt die Trennung, wenn der größtmögliche Margin zwischen zwei Klassen gefunden wurde. Jedes Objekt in den Daten wird durch einen n-dimensionalen Feature-Vektor beschrieben. Die Trennungsgrenze wird anhand ihrer nächstgelegenen Vektoren bestimmt, weswegen diese auch *Stützvektoren* genannt werden [Hearst et al., 1998].

³Bspw. bei LSTM: Optimierung des Gradientenabstiegsverfahrens durch Erinnerung an *tieferen Täler* bei Backpropagation [Hochreiter and Schmidhuber, 1997].

Beschreibung anhand eines Minimalbeispiels für die Haltungserkennung

Im Folgenden werden weitere Eigenschaften von SVMs anhand eines Minimalbeispiels für die Haltungserkennung beschrieben. Das Beispiel beschreibt ein Problem, in dem zwischen zwei Klassen unterschieden werden soll. Die erste Klasse stellt einen Ausfallschritt des linken Beins nach vorne *class 1: lunge left* dar. Die zweite Klasse zeigt einen Ausfallschritt mit dem rechten Bein nach vorne *class 2: lunge right*. Als Features werden Zeitreihendaten (timeseries) von den Gelenkwinkeln der linken *feature 1: hip left flexion* und der rechten *feature 2: hip right flexion* (Flexion in den Hüftgelenken) eingesetzt. Jeder Messzeitpunkt (Frame) beschreibt eine Körperhaltung. Jede Körperhaltung ist anhand der beiden Features beschrieben und einer Klasse zugeordnet. Eine Klasse beschreibt eine Bewegung bestehend aus einzelnen Körperhaltungen. Ziel der Klassifizierung ist es, anhand von zwei Gelenkwinkeln eine Körperhaltung bzw. -bewegung zu erkennen. In Abbildung 2.10a sind die Zeitreihendaten beider Features und beider Klassen dargestellt. Ein Beispiel für eine klare Trennung zwischen den Klassen kann in Sekunde 02 gezeigt werden. Hier ist *feature 1: hip left flexion* über 80° und *feature 2: hip right flexion* ca. 30° . Diesen Zustand gibt es nur in *class 1: lunge left* und nicht in *class 2: lunge right*.

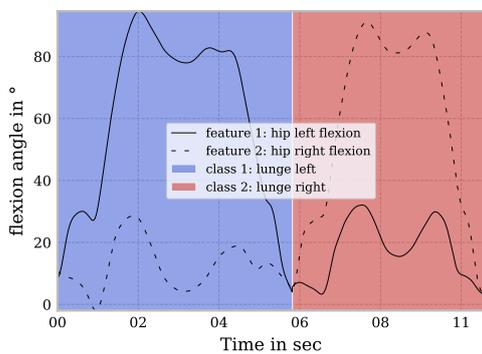
Um Klassen mit einer SVM voneinander trennen zu können, werden Stützvektoren oder auch Trennungsvektoren berechnet. Die Idee dabei ist, die Stützvektoren in weitere Dimensionen (Hyperebenen) zu überführen, um zusätzliche Ebenen für die Trennung zu generieren. Dafür können verschiedene Kernel in einer SVM eingesetzt werden.

In Abbildung 2.10b ist die Trennung durch einen *Linear Kernel* dargestellt. Die Rohdaten der beiden Features sind anhand der blau- und rot-schwarzen Punkte präsentiert. Die blaue und rote Fläche beschreibt die Zuordnung zu einer jeweiligen Klasse. In diesem Fall können die Daten durch eine lineare Trennung nahezu vollständig voneinander abgegrenzt werden. Ein unbekanntes Objekt, mit den Winkeln von *hip left flexion* und *hip right flexion*, würde anhand der Koordinate und der darunter liegenden Klasse zugeordnet werden können.

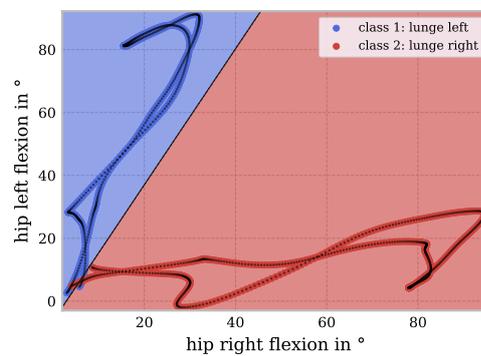
Ist eine lineare Trennung zwischen den Daten nicht möglich, kann eine Trennung auf Basis eines Polynomial-Kernels berechnet werden, um eine fließende Grenze zu erreichen (siehe Abbildung 2.10c). Mit dem Radial Basis Function (RBF)-Kernel ist es möglich, ganze Bereiche zu umschließen, um diese von anderen Klassen ab-

zugrenzen (siehe Abbildung 2.10d). Dies ist in diesem Beispiel aber nicht notwendig.

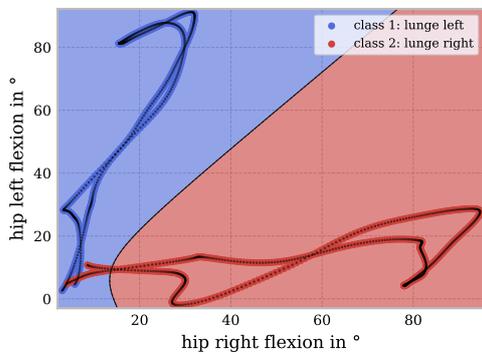
Für eine Klassifizierung auf Basis von zwei- oder dreidimensionalen Features ist eine Visualisierung der Grenzen noch möglich. In den meisten Fällen sind die Objekte in den Daten jedoch durch eine höhere Anzahl an Features beschrieben. Der Berechnungsaufwand von SVM-Modellen steigt deutlich mit der Anzahl der Features. Dies ist auch ein Grund, warum eine sorgfältige Feature Selection ratsam ist [Noble, 2006].



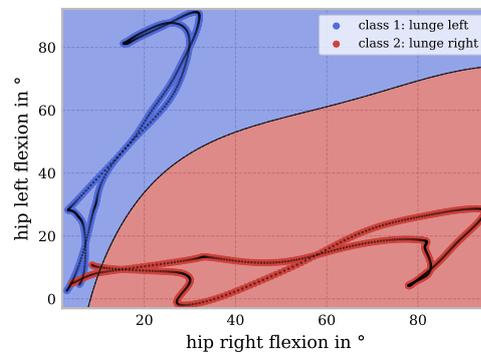
(a) Timeseries



(b) Linear Kernel



(c) Polynomial-Kernel



(d) RBF-Kernel

Abbildung 2.10.: Exemplarische Darstellung verschiedener Kernel von SVM; Beispiel eines Zwei-Klassen-Problems für die Unterscheidung von Körperhaltungen in Zeitreihendaten

2.4.2. Artificial Neural Network

Ein ANN ist ein Algorithmus, in dem die Signalverarbeitung des menschlichen Gehirns nachempfunden ist. Das Modell eines einfachen ANN besteht im Wesentlichen aus drei Layern: *Input Layer*, *Hidden Layer* und *Output Layer*. Jeder Layer besteht aus n -Neuronen. Jedes Neuron ist anhand von gerichteten Graphen mit dem nächsten Layer verbunden. Ein exemplarisches ANN in einer 2-4-5-4-2 Architektur (zwei grüne Neuronen im *Input Layer*, drei blaue Layer als 4-5-4 im *Hidden Layer*, zwei rote Neuronen im *Output Layer*) für die Klassifizierung zwischen zwei Bewegungen (*lunge left*, *lunge right*) anhand von zwei Gelenkwinkeln (*hip left flexion*, *hip right flexion*) ist in Abbildung 2.11 dargestellt. Die Anzahl der Neuronen im *Input Layer* entspricht der Anzahl der genutzten Features. Die Anzahl der Neuronen im *Output Layer* entspricht der Anzahl der zu erkennenden Klassen. Der *Hidden Layer* besteht wiederum aus n -Layern mit jeweils n -Neuronen.

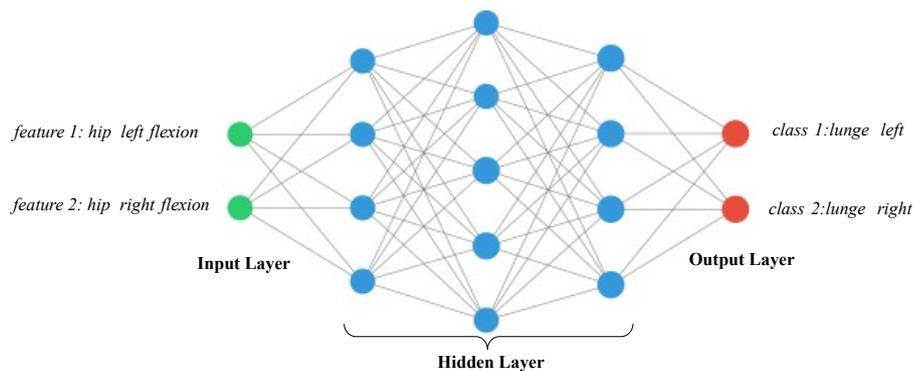


Abbildung 2.11.: Exemplarisches ANN in einer 2-4-5-4-2 Architektur für die Klassifizierung von *lunge left* und *lunge right* anhand von *hip left flexion* und *hip right flexion*

In diesem Beispiel ist ein *fully connected Network* dargestellt. Das bedeutet, dass jedes Neuron eines Layers mit jedem Neuron des nächsten Layers vollständig verbunden ist. Man spricht auch von einem *feed forward Network*, da nur zwischen den Layern Verbindungen, aber keine Querverbindungen innerhalb eines Layers bestehen. Kern jedes Neurons ist eine Aktivierungsfunktion $f(x)$. Die Verbindungen der Neurone sind mit Gewichten w behaftet. Werden an den grünen Neuronen im *Input Layer* zwei Werte eingegeben, erfolgt zunächst eine Wichtung, bis sie zu den blauen Neuronen in der ersten Schicht des *Hidden Layers* gelangen. An jedem Neuron kommt somit eine gewichtete Summe an. Diese Summe ist die Eingabe für die Aktivierungsfunktion jedes Neurons. Wird ein erlernter Schwellwert überschritten,

feuert das Neuron, d.h. es gibt das Ergebnis der Aktivierungsfunktion, wiederum gewichtet, an die verbundenen Neuronen in der nächsten Schicht weiter. Zusammengefasst: ein Neuron summiert die Eingabewerte, verarbeitet die Summe mithilfe der Aktivierungsfunktion $f(x)$ und gibt das Ergebnis y an die nächste Schicht weiter. Der prinzipielle Aufbau eines Neurons ist in Abbildung 2.12 dargestellt [Rashid, 2016].

Die Auswahl einer passenden Aktivierungsfunktion für die Neuronen in einem Layer ist wichtig, um ein effizientes ANN zu entwickeln. Häufig wird eine Rectified Linear Unit (ReLU) Funktion oder eine Sigmoidfunktion eingesetzt. Der Verlauf beider Funktionen ist in Abbildung 2.13 a und b dargestellt [Sharma et al., 2020].

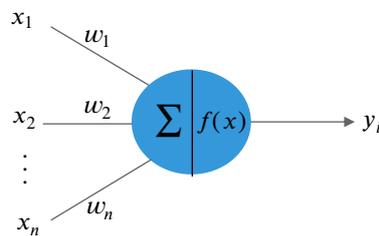
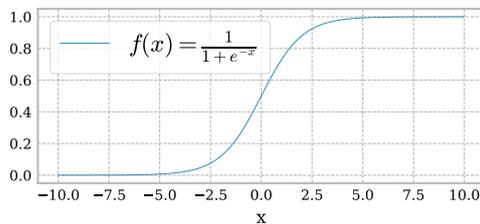
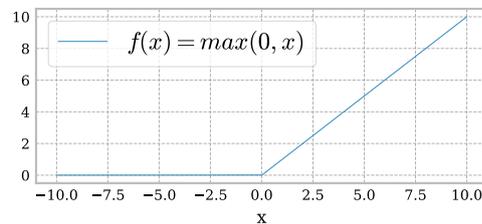


Abbildung 2.12.: Signalverarbeitung eines Neurons; Gewichtung der Eingaben x_i anhand der Gewichte w_i ; Summierung der gewichteten Eingaben; Berechnung der Ausgabe y_i anhand der Aktivierungsfunktion $f(x)$



(a) Sigmoidfunktion



(b) Rectified Linear Unit (ReLU) Funktion

Abbildung 2.13.: Typische Aktivierungsfunktionen eines künstlichen Neurons

Im untrainierten Zustand sind alle Gewichte an den Verbindungen zufällig gewichtet. Für die Anpassung der Gewichte eignen sich sog. Gradientenabstiegsverfahren, wie das *Backpropagation* [Leung and Haykin, 1991], um ein ANN zu trainieren. Wird ein ANN überwacht trainiert, findet eine Art Belohnung und Bestrafung statt. Überwachtes Trainieren bedeutet, dass die Ergebnisse, die das ANN ausgeben

soll, bekannt sind. Belohnt werde die Gewichte an den Neuronen, die korrekterweise aktiviert wurden und somit das Ergebnis im *Output Layer* richtig beeinflusst haben. Bestraft werden die Gewichte an Neuronen, die dafür gesorgt haben, dass das Ergebnis falsch war. Bestrafung und Belohnung beschreibt somit eine Anpassung (Training) der Gewichte an den Neuronen, um den Schwellwert für die Aktivierung zu optimieren und somit den Fehler zu minimieren.

Anhand einer definierten *Learning Rate* wird festgelegt, wie stark die Gewichte jeweils angepasst werden. Bei einem richtigen Ergebnis werden die Gewichte nach oben angepasst, bei einem Falschen nach unten. Als *Backpropagation* wird somit eine Korrektur der Gewichte zur Minimierung des Fehlers verstanden. Dabei wird ausgehend vom Ergebnis *zurück* auf die Gewichte geschaut und diese entsprechend des Fehlers angepasst [Rashid, 2016].

Das einmalige Durchlaufen eines Datensatzes in den Trainingsdaten wird *Epoche* genannt. Gewöhnlich durchläuft ein Datensatz ein ANN in mehreren Iterationen, um die Anpassung der Gewichte entsprechend zu verfeinern. Wird ein ANN mit zu vielen Iterationen oder Daten trainiert, kann es zu einer Zunahme der Fehler-rate kommen. Dieses Phänomen wird *Overfitting* genannt. *Overfitting* kann auch entstehen, wenn die Datenlage für das Training *unbalanciert*⁴ vorliegt. *Overfitting* kann anhand einer zunehmenden Fehlerrate während des Trainings erkannt werden.

Das gesamte Wissen eines ANNs wird durch Training erlernt und in den Gewichten verankert. Ein fertiges Modell besteht aus angepassten bzw. trainierten Gewichten. Werden im ANN aus Abb. 2.11 unbekannte Gelenkwinkel für *feature 1: hip left flexion* und *feature 2: hip right flexion* im Input Layer ANN eingegeben, wird eine unbekannte Anzahl an gewichteten Verbindungen und feuernden Neuronen durchlaufen. Im *Output Layer* wird entweder das Neuron für *class 1: lunge left* oder *class 2: lunge right* höher aktiviert. Ist in den Neuronen im *Output Layer* eine Sigmoidfunktion definiert, wird für jede Klasse ein Wert zwischen [0,1] angegeben. Die Eingabe am Input Layer ist der Klasse des Output Layers mit dem höchsten Wert zuzuordnen.

Das Finden eines passenden Designs für eine ANN Architektur und die Auswahl und Vorbereitung der Trainingsdaten sind die größten Herausforderungen bei der

⁴In unbalancierten Daten sind bestimmte Zustände häufiger präsentiert als andere.

Modellentwicklung [Vox and Wallhoff, 2018b].

2.5. Klinischer Standard für die Gelenkwinkelmessung

Die physiologische Beweglichkeit von Gelenken ist anhand der Freiheitsgrade und der Range of Motion (RoM) für jedes Gelenk definiert. Ausgangslage für die Messung des Bewegungsausmaßes ist die Neutral-Null-Stellung [Ryf and Weymann, 1995]. Diese ist beschrieben als aufrechte Körperhaltung mit hängenden Armen parallel zum Rumpf, Handflächen zum Körper zeigend, einem hüftbreiten Stand mit gestreckten Kniegelenken und den Fußspitzen nach vorne gerichtet. In dieser Stellung sind alle Gelenkwinkel auf 0° .

Klinischer Standard für die Bestimmung der RoM ist das Goniometer [Gajdosik and Bohannon, 1987], welches in verschiedenen Arten und Größen existiert. Eine Messung der Flexion im Kniegelenk mit einem Standard-Goniometer (Armlänge: 20 cm, Winkelschrittweite: 2°) ist in Abbildung 2.14 dargestellt.



Abbildung 2.14.: Gelenkwinkelmessung der Flexion des linken Kniegelenks mit einem Standard-Goniometer [Vox and Wallhoff, 2016, Fig. 1]

Dabei wird der stationäre Arm des Goniometers mittig und parallel zum Oberschenkel und der bewegliche Arm parallel zum Wadenbein platziert. Das Zentrum des Goniometers liegt mittig auf Höhe des Kniegelenkspaltes. Geringe Abweichungen durch die Platzierung des Goniometers sind aus medizinischer Sicht tolerierbar.

Die höchste Genauigkeit bei der Messung der Gelenkbeweglichkeit ist mit dem Röntgenbild zu erreichen [Gogia et al., 1987].

Die RoM ist für jeden Freiheitsgrad des jeweiligen Gelenkes definiert und wird ausgehend von der Neutral-Null-Stellung angegeben. Als Freiheitsgrade werden die möglichen Bewegungsrichtungen verstanden. Pro Freiheitsgrad sind zwei Bewegungsrichtungen möglich, welche im Weiteren als Flexion (Beugung) und Extension (Streckung), Lateralflexion (seitliche Beugung), Abduktion (Abspreizen) und Adduktion (Heranziehen), Radialduktion (Abwinkeln der Hand auf Seite des Daumens) und Ulnarduktion (Abwinkeln der Hand auf Seite des kleinen Fingers), Pronation (Einwärtsdrehung z.B. des Unterarms) und Supination (Auswärtsdrehung z.B. des Unterarms) und Außen- und Innenrotation angegeben werden. Kombinationsbewegungen im unteren Sprunggelenk werden als Eversion (Heben der Außenseite des Fußes) und Inversion (Heben der Innenseite des Fußes) bezeichnet. Die RoM ist in der Literatur nicht übereinstimmend definiert. Die Benennung der Bewegungsrichtungen und -ausmaße (RoM) können divergieren. In Tabelle 2.3 sind die Freiheitsgrade und RoM für ausgewählte Gelenkwinkel, mit Angabe der Quelle, aufgelistet. Die Auswahl bezieht sich auf Gelenkwinkel, welche typischerweise für die Bewegungsbewertung (siehe folgender Abschnitt) herangezogen werden.

2.6. Bewegungsbewertung auf Basis von Gelenkwinkeln

In der Therapie von Gelenkerkrankungen und -verletzungen, besonders nach operativen Eingriffen, kann eine temporäre Limitation des Bewegungsausmaßes sowie eine Reduktion der Belastung nötig sein, um den Heilungsprozess zu unterstützen. In solchen Fällen sind u.a. durch die Deutsche Gesellschaft für Orthopädie und Unfallchirurgie (DGOU) sogenannte Nachbehandlungsempfehlungen [DGOU, 2021] definiert. Ähnliche Schemata bestehen für die Bewertung der physischen Belastung von Bewegungen am Arbeitsplatz oder im Alltag. Zu diesem Zweck sind u.a. durch das Institut für Arbeitsschutz (IFA) Schwellwerte definiert, in denen Gelenkwinkelbereiche anhand ergonomischer Kriterien durch ein Ampelschema bewertet sind [IFA, 2015]. In Abbildung 2.15a,b,c ist die RoM der Extension und Flexion im Schultergelenk dargestellt. In Abbildung 2.15d ist ein Beispiel für die Risikobewertung bei Extension und Flexion der Schulter nach [IFA, 2015] Ampelschema gezeigt.

In den weiteren Abschnitten werden typische Schemata für die Bewertung von

Tabelle 2.3.: Auflistung von relevanten Gelenkwinkeln für eine Bewegungsbewertung; Freiheitsgrad und RoM nach Neutral-Null-Methode

Gelenk	Freiheitsgrad	RoM	Quelle
Neck	Extension, Flexion Lateralflexion rechts, links Rotation rechts, links	45°/0°/45° 45°/0°/45° 70°/0°/70°	[Pjontek et al., 2013]
Chest Spine	Extension, Flexion Lateralflexion rechts, links	15°/0°/20° 10°/0°/10°	[Mannerfelt et al., 2005]
Lumbar Spine	Extension, Flexion Lateralflexion rechts, links Rotation rechts, links	35°/0°/45° 30°/0°/30° 70°/0°/70°	[Mannerfelt et al., 2005]
Shoulder	Extension, Flexion Adduktion, Abduktion Rotation, angelegter Arm Rotation, abduzierter Arm	40°/0°/170° 40°/0°/180° 60°/0°/95° 70°/0°/60°	[von Salis-Soglio, 2015]
Elbow	Extension, Flexion Pronation, Supination	10°/0°/150° 90°/0°/90°	[von Salis-Soglio, 2015]
Hand	Extension, Flexion Radialduktion, Ulnarduktion	60°/0°/60° 30°/0°/40°	[Pjontek et al., 2013]
Hip	Extension, Flexion Adduktion, Abduktion Rotation innen, außen	20°/0°/140° 45°/0°/30° 50°/0°/45°	[von Salis-Soglio, 2015]
Knee	Extension, Flexion Rotation innen, außen	10°/0°/150° 30°/0°/10°	[von Salis-Soglio, 2015]
Foot	Extension, Flexion Eversion, Inversion	30°/0°/50° 10°/0°/20°	[Pjontek et al., 2013]

Bewegungen auf Basis kritischer Gelenkwinkelbereiche eingeführt. Die Angaben von Regeln für eine schwellwertbasierte Bewegungsbewertung beziehen sich auf entsprechende Winkelbereiche in der RoM aus Tabelle 2.3. Die Benennung der Gelenke und Freiheitsgrade ist für alle Schemata vergleichbar dargestellt und kann aus diesem Grund von der Wortwahl in der angegebenen Literatur abweichen. Für die Benennung der Gelenke werden die englischen Begriffe genutzt, welche auch in Abbildung 2.1 bei der Beschreibung der MoCap-Daten verwendet werden. Diese finden sich auch in dem Kapitel 4 Eigener Ansatz und in der Systementwicklung wieder.

2.6.1. DGOU Nachbehandlungsempfehlungen

In den [DGOU, 2021] Nachbehandlungsempfehlungen sind Behandlungsziele und Maßnahmen nach einem Trauma oder einer Operation von Gelenken aufgeführt. Es werden u.a. Angaben zur Begrenzung des Bewegungsmaßes für einen gewissen Zeitraum gemacht und diese im Therapieverlauf erweitert, bis das volle Bewe-

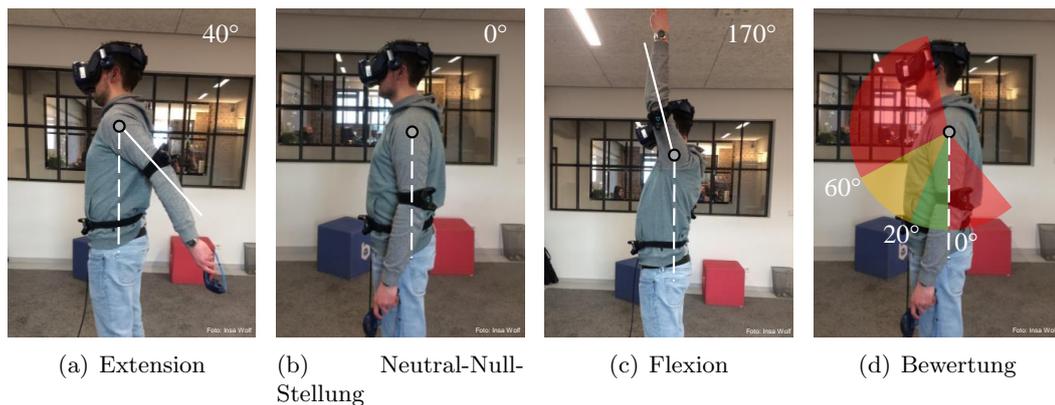


Abbildung 2.15.: (a,b,c) Bsp. für die RoM von Extension nach Flexion ($40^\circ/0^\circ/170^\circ$) im Schultergelenk; (d) Bewertung der physischen Belastung bei Flexion und Extension nach [IFA, 2015] Ampel-schema

gungsausmaß freigegeben ist. Dabei ist ausdrücklich zu erwähnen, dass es sich um Empfehlungen handelt und individuelle Anpassungen der Maßnahmen durch medizinisches Fachpersonal an den Schweregrad des Traumas vorbehalten sind.

In den Maßnahmen ist eine Bewertung von Gelenkwinkelbewegungen auf Basis von Regeln mit Gelenkwinkelgrenzen definiert. Für bestimmte Therapien ist innerhalb des Behandlungszeitraums eine Begrenzung der RoM einzuhalten und nicht zu überschreiten. In Tabelle 2.4 sind beispielhafte Traumata betroffener Gelenke mit Angabe des Zeitraums sowie der Regel der Bewegungsbegrenzung eines Freiheitsgrades dargestellt.

2.6.2. RULA

Das Rapid Upper Limb Assessment (RULA) von [McAtamney and Nigel Corlett, 1993] ist für die ergonomische Bewertung von Arbeitsprozessen entwickelt worden. Darin wird anhand von 15 Schritten ein Score für die Ergonomiebewertung berechnet. In Tabelle 2.5 sind die Schritte aufgelistet, in denen kritische Gelenkwinkelbereiche anhand von Schwellwerten definiert sind.

Wird ein Schwellwert in einem Winkelbereich überschritten, wird der Wert in Spalte *Score* in einer Bewertungstabelle eingetragen. Damit wird dokumentiert, wenn

Tabelle 2.4.: Beispiele für Gelenktraumata mit entsprechender Behandlungsempfehlung zur Einhaltung von Regeln (z.B. Begrenzung der RoM) über einen gewissen Zeitraum [DGOU, 2021]; α ist der Winkel in eine Bewegungsrichtung des Freiheitsgrades des genannten Gelenks

Gelenk: Trauma	Zeit nach Unfall/OP	Freiheitsgrad	Regel
Knee: Patellafraktur	bis 3. Woche	Flexion	$\alpha \leq 30^\circ$
	bis 6. Woche		$\alpha \leq 60^\circ$
Knee: Vorderes Kreuzband, operativ	bis 2. Woche	Flexion	$\alpha \leq 90^\circ$
Shoulder: Instabilität, hintere, operativ	bis 6. Woche	Flexion	$\alpha \leq 90^\circ$
	bis 6. Woche	Abduktion	$\alpha \leq 90^\circ$
	bis 12. Woche	Rotation, innen	$\alpha \leq 20^\circ$
		Rotation, außen	$\alpha \leq 40^\circ$

Gliedmaßen in ungünstigen Haltungen stehen und einen entsprechend höheren Score aufweisen. Des Weiteren werden Punkte vergeben, wenn eine Haltung statisch und länger als 10 min auftritt oder eine Bewegung wiederholt durchgeführt wird. Ebenfalls wird erfasst, ob eine zusätzliche Belastung durch Gewichte besteht.

Die Bewertung erfolgt anhand von Zuordnungen in Tabellen, wodurch sich aus den einzelnen Zuständen (bspw. Flexion im Schultergelenk $> 90^\circ$, Arm ist abduziert, Arm in Innenrotation, Hand rotiert, Oberkörper rotiert, Oberkörper seitlich, Füße in festem Stand) ein Gesamtscore ermitteln lässt. Ein Score von 1 bis 2 bedeutet, dass keine Notwendigkeit für Maßnahmen besteht und die Bewegungen in einem ergonomisch unbedenklichen Rahmen ausgeführt werden. Ein Score von 3 bis 4 weist darauf hin, dass in Zukunft Maßnahmen zur Verbesserung der Ergonomie zu treffen sind. Ein Score von 5 bis 6 zeigt an, dass in Kürze entlastende Maßnahmen getroffen werden sollten. Bei einem Score größer als 6 müssten sofort Maßnahmen zur Entlastung eingeleitet werden.

2.6.3. REBA

Das Rapid Entire Body Assessment (REBA) nach [Hignett and McAtamney, 2000] ist im Wesentlichen mit RULA vergleichbar, priorisiert jedoch die unteren Gliedmaßen. Bezogen auf die Bewertung in den Gelenkwinkelbereichen sind in REBA zusätzlich Regeln zur Bewertung der Flexion im Kniegelenk abgebildet. Die Erweiterung gegenüber RULA ist in Tabelle 2.6 dargestellt.

Der REBA Score ist mit RULA vergleichbar und wird ebenfalls durch eine tabel-

Tabelle 2.5.: Bewegungsbewertung RULA von [McAtamney and Nigel Corlett, 1993]; tabellarische Auflistung des Scorings für kritische Winkelbereiche von Gelenken

Gelenk	Freiheitsgrad	Regel	Score
Shoulder	Extension, Flexion	$\alpha \leq 20^\circ$	+1
	Extension	$\alpha > 20^\circ$	+2
	Flexion	$20^\circ < \alpha \leq 45^\circ$	+2
	Flexion	$45^\circ < \alpha \leq 90^\circ$	+3
	Flexion	$\alpha > 90^\circ$	+4
Elbow	Flexion	$60^\circ \leq \alpha \leq 100^\circ$	+1
	Flexion	$\alpha < 60^\circ$	+2
	Flexion	$\alpha > 100^\circ$	+2
	Rotation	$\alpha > 0^\circ$	+1
Hand	Flexion	$\alpha = 0^\circ$	+1
	Extension, Flexion	$\alpha \leq 15^\circ$	+2
	Extension, Flexion	$\alpha > 15^\circ$	+3
	Radialduktion, Ulnarduktion	$\alpha > 0^\circ$	+1
Neck	Flexion	$\alpha \leq 10^\circ$	+1
	Flexion	$10^\circ < \alpha \leq 20^\circ$	+2
	Flexion	$\alpha > 20^\circ$	+3
	Extension	$\alpha > 0^\circ$	+4
Lumbar Spine	Flexion	$\alpha = 0^\circ$	+1
	Flexion	$\alpha \leq 20^\circ$	+2
	Flexion	$20^\circ < \alpha \leq 60^\circ$	+3
	Flexion	$\alpha > 60^\circ$	+4

Tabelle 2.6.: Erweiterung von RULA um die Bewegungsbewertung REBA nach [Hignett and McAtamney, 2000]

Gelenk	Freiheitsgrad	Regel	Score
Knee	Flexion	$30^\circ < \alpha \leq 60^\circ$	+1
	Flexion	$\alpha > 60^\circ$	+2

larische Kategorisierung erhoben. Ein Score von 1 beschreibt, dass kein Risiko für Erkrankungen des MSS besteht. Von 2 bis 3 besteht ein leichtes Risiko. Zwischen 4 bis 7 ein mittleres Risiko, welches eine Veränderung der Situation in Zukunft erfordert. Ein Score von 8 - 10 signalisiert ein erhöhtes Risiko und ein Score größer als 11 ein sehr hohes Risiko.

2.6.4. IFA Ampelschema

Das [IFA, 2015] Beurteilungsverfahren für die Bewertung von physischen Belastungen ist in Anlehnung an Normen (ISO, DIN EN) der Bewegungsergonomie und auf

Basis von veröffentlichten wissenschaftlichen Bewertungskriterien entworfen. Das Bewertungsschema ist in drei Teile gegliedert und ermöglicht die Bewertung einzelner Gelenke. In Teil 1 werden Winkelbereiche verschiedener Gelenke aufgelistet und nach einem Ampelschema in drei Zustände kategorisiert. Die Kategorien werden als akzeptabel (grün), bedingt akzeptabel (gelb) oder nicht akzeptabel (rot) eingestuft. Teil 2 berücksichtigt auch Kräfte und Momente in den Bereichen der Lendenwirbelsäule, des Schultergelenks und der Hand. In Teil 3 sind Methoden für die Erfassung manueller Arbeitsprozesse abgebildet. In den Tabellen 2.7 und 2.8 sind die Bewertungen aus Teil 2 für die Gelenke Head, Neck, Chest Spine, Lumbar Spine, Shoulder, Elbow, Hand und Knee aufgelistet. Die Winkelbereiche der Zustände grün, gelb, rot sind als Regeln für den Winkel α dargestellt.

2.6.5. OWAS

Das Ovako Working Posture Assessment (OWAS) nach [Karhu et al., 1977] ist ein Schema, welches für die Ergonomiebewertung von Mitarbeiter*innen im finnischen Stahlwerk *Ovako Oy* entwickelt wurde. Die OWAS-Methode basiert nicht auf einer Bewertung anhand von Gelenkwinkeln. Die Methode ist dennoch ein etabliertes Instrument für die Ergonomiebewertung mit direktem Themenbezug zu den anderen Methoden und wird aus diesem Grund hier genannt [Kong et al., 2018].

Das Schema basiert auf einer Kategorisierung des Risikos für die Entwicklung einer Erkrankung des MSS anhand von typischen Körperhaltungen aufgeteilt in die drei Kategorien: Rücken, Arme und Beine. Jeder Kategorie sind Körperhaltungen (bspw. Rücken gebeugt, Arme über Schulter, Beine gestreckt) mit entsprechenden Risikoklassen zugeordnet. Für jede Klasse sind Haltungen als Foto repräsentiert und mit einem Code (1 bis 4) versehen. Anhand des Codes wird die Dringlichkeit von Maßnahmen zur Vermeidung der entsprechenden Körperhaltung klassifiziert. Ein Code von 1 zeigt ein geringes oder kein Risiko an. Der Code 2 und 3 signalisiert ein moderates und ein erhöhtes Risiko mit Handlungsbedarf. Bei Code 4 besteht ein erhöhtes Risiko und es sollten sofort Gegenmaßnahmen eingeleitet werden.

Durch eine manuelle Beobachtung wird für jede Kategorie eine Körperhaltung ausgewählt, welche einer Haltung in den Fotos am nächsten kommt. Anschließend wird der Code aus den Kategorien Rücken, Arme und Beine als dreistelliger OWAS-Code zusammengefasst. Ansätze für eine automatische Bewertung von Körperhaltungen mit der OWAS-Methode auf Basis von MoCap-Daten werden in aktuellen Forschungsarbeiten behandelt [Lins, 2021].

Tabelle 2.7.: Bewertung der physischen Belastung nach [IFA, 2015] Ampelschema für die Gelenke Head, Neck, Chest Spine, Lumbar Spine, Shoulder und Knee

Gelenk	Freiheitsgrad	Regel	Bewertung
Head	Extension	$\alpha > 0^\circ$	• rot
	Flexion	$\alpha \leq 25^\circ$	• grün
	Flexion	$25^\circ < \alpha \leq 85^\circ$	• gelb
	Flexion	$\alpha > 85^\circ$	• rot
Neck	Lateralflexion rechts, links	$\alpha \leq 10^\circ$	• grün
	Lateralflexion rechts, links	$\alpha > 10^\circ$	• rot
	Rotation rechts, links	$\alpha \leq 45^\circ$	• grün
	Rotation rechts, links	$\alpha > 45^\circ$	• rot
	Flexion	$\alpha \leq 25^\circ$	• grün
	Flexion	$\alpha > 25^\circ$	• rot
Chest Spine	Extension	$\alpha > 0^\circ$	• rot
	Flexion	$\alpha \leq 20^\circ$	• grün
	Flexion	$20^\circ < \alpha \leq 40^\circ$	• gelb
	Flexion	$\alpha > 40^\circ$	• rot
Lumbar Spine	Extension	$\alpha > 0^\circ$	• rot
	Flexion	$\alpha \leq 20^\circ$	• grün
	Flexion	$20^\circ < \alpha \leq 60^\circ$	• gelb
	Flexion	$\alpha > 60^\circ$	• rot
	Lateralflexion rechts, links	$\alpha \leq 10^\circ$	• grün
	Lateralflexion rechts, links	$10^\circ < \alpha \leq 20^\circ$	• gelb
	Lateralflexion rechts, links	$\alpha > 20^\circ$	• rot
	Rotation rechts, links	$\alpha \leq 10^\circ$	• grün
	Rotation rechts, links	$10^\circ < \alpha \leq 20^\circ$	• gelb
	Rotation rechts, links	$\alpha > 20^\circ$	• rot
Shoulder	Adduktion	$\alpha > 0^\circ$	• rot
	Abduktion	$\alpha \leq 20^\circ$	• grün
	Abduktion	$20^\circ < \alpha \leq 60^\circ$	• gelb
	Abduktion	$\alpha > 60^\circ$	• rot
	Extension	$\alpha > 0^\circ$	• rot
	Flexion	$\alpha \leq 20^\circ$	• grün
	Flexion	$20^\circ < \alpha \leq 60^\circ$	• gelb
	Flexion	$\alpha > 60^\circ$	• rot
	Rotation, innen	$\alpha \leq 30^\circ$	• grün
	Rotation, innen	$30^\circ < \alpha \leq 60^\circ$	• gelb
	Rotation, innen	$\alpha > 60^\circ$	• rot
	Rotation, außen	$\alpha \leq 15^\circ$	• grün
	Rotation, außen	$15^\circ < \alpha \leq 30^\circ$	• gelb
	Rotation, außen	$\alpha > 30^\circ$	• rot
Knee, Sitz	Extension	$\alpha > 0^\circ$	• rot
	Flexion	$\alpha \leq 45^\circ$	• gelb
	Flexion	$45^\circ < \alpha \leq 90^\circ$	• grün
	Flexion	$\alpha > 90^\circ$	• rot
Knee, Stand	Extension	$\alpha > 0^\circ$	• rot
	Flexion	$\alpha \leq 5^\circ$	• grün
	Flexion	$5^\circ < \alpha \leq 10^\circ$	• gelb
	Flexion	$\alpha > 10^\circ$	• rot

Tabelle 2.8.: Bewertung der physischen Belastung nach [IFA, 2015] Ampelschema für die Gelenke Elbow und Hand

Gelenk	Freiheitsgrad	Regel	Bewertung
Elbow	Extension	$\alpha > 0^\circ$	● rot
	Flexion	$\alpha \leq 60^\circ$	● rot
	Flexion	$60^\circ < \alpha \leq 100^\circ$	● grün
	Flexion	$\alpha > 100^\circ$	● rot
	Pronation	$\alpha \leq 20^\circ$	● grün
	Pronation	$20^\circ < \alpha \leq 40^\circ$	● gelb
	Pronation	$\alpha > 40^\circ$	● rot
	Supination	$\alpha \leq 30^\circ$	● grün
	Supination	$30^\circ < \alpha \leq 55^\circ$	● gelb
Hand	Supination	$\alpha > 55^\circ$	● rot
	Extension	$\alpha \leq 25^\circ$	● grün
	Extension	$25^\circ < \alpha \leq 50^\circ$	● gelb
	Extension	$\alpha > 50^\circ$	● rot
	Flexion	$\alpha \leq 20^\circ$	● grün
	Flexion	$20^\circ < \alpha \leq 45^\circ$	● gelb
	Flexion	$\alpha > 45^\circ$	● rot
	Radialduktion	$\alpha \leq 10^\circ$	● grün
	Radialduktion	$10^\circ < \alpha \leq 25^\circ$	● gelb
	Radialduktion	$\alpha > 25^\circ$	● rot
	Ulnarduktion	$\alpha \leq 10^\circ$	● grün
Ulnarduktion	$10^\circ < \alpha \leq 15^\circ$	● gelb	
Ulnarduktion	$\alpha > 15^\circ$	● rot	

3. Stand der Technik

In diesem Kapitel werden aktuelle Arbeiten zu den behandelten Forschungsfragen vorgestellt. Dabei werden insbesondere Veröffentlichungen präsentiert, in denen die eingeführten Sensorprinzipien und Methodiken aus Kapitel 2 *Grundlagen* angewendet wurden.

3.1. Sensorgenauigkeit von Motion-Capture-Sensorik

Wie in den Abschnitten 2.2 *Motion Tracking mit einer RGB-Tiefenkamera* und 2.3 *Motion Tracking mit einem Lighthouse Tracking System* erläutert, hängt die Genauigkeit der Bereitstellung der MoCap-Daten sowohl von der Sensor-Hardware als auch von der Software (Skeleton Tracking, Inverse Kinematik) ab. In dieser Arbeit wird eine Evaluierung des Gesamtsystems vorgestellt. Bei optischen und markerbasierten Systemen (Qualisys, Vicon) liegt der ermittelte Fehler bei $< 1 \text{ mm}$ für die jeweilige Markerposition [Liu et al., 2007, Windolf et al., 2008]. Diese Systeme werden generell in technischen Vergleichsstudien für die Evaluierung der Genauigkeit von MoCap-Sensorik eingesetzt.

Schwerpunkt der vorliegenden Arbeit ist insbesondere die Evaluation von Gelenkwinkelmessungen. Dabei ist die Genauigkeit in der Regel nicht nur von einer Gelenkposition in den MoCap-Daten abhängig, sondern auch von den benachbarten Gelenkpositionen (siehe auch 5.2 *Berechnung von physiologischen Gelenkwinkeln*). Im Folgenden werden u.a. technische Vergleichsstudien vorgestellt, in denen auf Basis von Gelenkpositionen und/oder Gelenkwinkeln eine Evaluierung der jeweiligen Sensorik durchgeführt wurde.

3.1.1. Sensorgenauigkeit von RGB-Tiefenkameras

In der Arbeit von [Yeung et al., 2021] wurden die RGB-Tiefenkamera Systeme Kinect v2, Azure und Orbbec Astra mit dem Vicon System als Goldstandard evaluiert. Dabei wurden die Systeme in fünf verschiedenen Kamerablickwinkeln ($0^\circ / 22.5^\circ /$

45° / 67.5° / 90°) auf eine laufende Person gerichtet. Das Experiment wurde mit zehn Proband*innen durchgeführt, welche auf einem Laufband in drei Geschwindigkeitsstufen normales Gehen durchführten. Die Evaluierung der Genauigkeit ist auf Basis von berechneten Gelenkwinkeln von Extension/Flexion (Hip-ExFlex) und Adduktion/Abduktion der Hüftgelenke (Hip-AdAb); Extension/Flexion der Kniegelenke (Knee) und Extension/Flexion der Sprunggelenke (Foot) erfolgt. Für den Vergleich wurde der Root Mean Square Error (RMSE) zwischen Vicon und dem RGB-Tiefenkamera System berechnet. Bei der Kinect v2 konnte mit frontalem Blickwinkel (0°) ein RMSE von [Hip-ExFlex 3.9°, Hip-AdAb 7.9°, Knee 11.4°, Foot 17.4°] identifiziert werden. Bei der Kinect v2 wurde festgestellt, dass der Fehler mit zunehmendem Blickwinkel ansteigt. Die Azure zeigte im frontalen Blickwinkel etwas größere RMSEs von [Hip-ExFlex 4.0°, Hip-AdAb 9.6°, Knee 16.1°, Foot 13.3°]. In den weiteren Blickwinkeln performte die Azure jedoch am genauesten. Bei der Orbbec Astra wurde im Gesamten ein höherer Fehler gegenüber Kinect v2 und Azure ermittelt.

Von [Ma et al., 2019] konnten ähnliche Beobachtungen mit gemessenen RMSEs für [Hip-ExFlex 5.3°, Hip-AdAb 6.6°, Knee 7.9°, Foot > 17.5°] bei frontalem Blickwinkel mit der Kinect v2 ermittelt werden. In [Xu et al., 2015] wurde der RMSE in unterschiedlichen Ganggeschwindigkeiten berechnet. Die Ergebnisse [Hip-ExFlex 11.7°, Knee 28.6°] zeigten bei einer Ganggeschwindigkeit von 1.07 m/s (langsames Gehen) einen vergleichsweise höheren RMSE.

In [Albert et al., 2020] wurden die Kinect v2 und Azure ebenfalls mit dem Vicon System mittels Durchführung von Laufbewegungen auf einem Laufband evaluiert. Hierbei wurden Positionen von Gelenkpunkten der oberen und unteren Gliedmaßen auf Basis der Pearson Korrelation [Benesty et al., 2009] und Gangparametern (Schrittweite, Schrittgeschwindigkeit) ausgewertet. Dabei konnte gezeigt werden, dass die Azure eine signifikant höhere Genauigkeit gegenüber der Kinect v2 bei der Bewertung von räumlichen Gangparametern aufweist. Im Vergleich zu den zeitlichen Parametern konnten keine auffälligen Unterschiede festgestellt werden. Die Evaluierung der linken oberen Gliedmaße¹ anhand der Pearson Korrelation zeigte jedoch, dass die Kinect v2 [Lumbar Spine 0.86 ± 0.10, Chest Spine 0.88 ± 0.07, Shoulder left 0.92 ± 0.04, Elbow left 0.62 ± 0.22] gegenüber der Azure [Lumbar Spine 0.67 ± 0.12, Chest Spine 0.75 ± 0.13, Shoulder left 0.66 ± 0.11, Elbow left

¹Für die rechte obere Gliedmaße können ähnliche Werte angenommen werden.

0.62 ± 0.13] eine bessere Übereinstimmung in vertikalen Bewegungen aufweist.

In der Untersuchung von [Tölgyessy et al., 2021b] wurde die Genauigkeit von Gelenkpositionen des gesamten Körpers für die Kinect v1, Kinect v2 und Azure evaluiert. In dem Experiment wurde ein menschlicher Körper als Plastikfigur (mit Kleidung) an einem Endeffektor eines Industrieroboters befestigt und linear bewegt. Alle Gliedmaßen der Plastikfigur waren statisch, ohne Bewegung. Evaluiert wurden die Systeme auf Basis der Standardabweichung von einzelnen Gelenkpositionen im Raum. Die Plastikfigur wurde in 400 mm auseinander liegenden Positionen linear von den Systemen wegbewegt. Es wurde untersucht, wie reproduzierbar die Gelenkpositionen im dreidimensionalen Raum bestimmt werden können. Wesentliche Unterschiede zeigen sich zwischen der Kinect v1 und den nachfolgenden Systemen. Bei der Kinect v1 wurde allgemein ein größeres Rauschen in Positionen der äußeren Gliedmaßen (Hände, Füße) festgestellt. Die Standardabweichung von Kinect v2 und Azure zeigte keine wesentlichen Unterschiede. In den Grenzbereichen des Messraums, bei Entfernungen $> 3.5 m$, war die Reproduzierbarkeit der gemessenen Gelenkpositionen von der Azure jedoch deutlich genauer gegenüber der Kinect v2 und Kinect v1. Untermauert werden die Ergebnisse durch eine daran anknüpfende Studie, ebenfalls durchgeführt von [Tölgyessy et al., 2021a]. Darin wurde u.a. die Genauigkeit der Erfassung von Strukturen auf Basis der Tiefeninformation (Punktwolken) evaluiert. Hierbei zeigte sich ebenfalls eine höhere Genauigkeit der Azure gegenüber den vorangegangenen Versionen.

In der Arbeit von [Ivorra et al., 2021] wurden u.a. das Skeleton Tracking *OpenPose* und *Azure Kinect Body Tracking* mit Rohdaten der Azure evaluiert. Die Autor*innen nannten als mögliche Anwendungsszenarien die Vermessung des Bewegungsausmaßes sowie allgemein die technische Unterstützung in der Therapie. Als Referenz wurde das markerbasierte System *OptiTrack* eingesetzt. Evaluiert wurde auf Basis von Gelenkwinkeln der oberen Gliedmaßen (Extension, Flexion, Abduktion, Adduktion und Rotation der Schultergelenke; Extension, Flexion der Ellenbogengelenke). Während der Messung wurden typische Armbewegungen in der RoM der untersuchten Gelenkwinkel durchgeführt. In den Ergebnissen erreichte das *Azure Kinect Body Tracking* einen RMSE von 10° und *OpenPose* einen RMSE von 8° , zusammengefasst über alle Gelenkwinkel. Demzufolge konnte eine etwas bessere Genauigkeit mit *OpenPose* erreicht werden. In der Arbeit von [Ota et al., 2020] wurde *OpenPose* anhand von Gelenkwinkeln bei der Ausführung von bilateralen

Ausfallschritten validiert. Die Autor*innen gaben an, dass *OpenPose* eine verlässliche Methode für die Bewegungsanalyse sei.

In [Wiedemann et al., 2015] wurde die Kinect v2 mit dem Vicon System auf Basis von Gelenkwinkeln der oberen und unteren Gliedmaßen evaluiert. In der Studie mussten die Proband*innen 16 verschiedene statische Körperhaltungen im Sitzen und Stehen einnehmen. Dabei konnte ein absoluter median Fehler (vergleichbar mit RMSE) über alle Proband*innen und Haltungen von unter 10° per Winkel ermittelt werden. Für nur stehende Körperhaltungen liegt der Fehler bei 7.6° . Im Vergleich zu der Studie von [Ivorra et al., 2021], worin ein RMSE von 10° (*Azure Kinect Body Tracking*) ermittelt wurde, zeigt die Kinect v2 einen kleinen Vorteil in der Genauigkeit. Diese Beobachtung wurde auch in der oben genannten Studie von [Yeung et al., 2021], beim Vergleich zwischen der Kinect v2 und Azure mit frontalem Blickwinkel gemacht.

Zusammenfassung

Die RGB-Tiefenkamera Systeme Kinect v1, Kinect v2 und Azure, sowie die dazugehörige Software für das Skeleton Tracking, sind in der Wissenschaft umfassend evaluiert. Insbesondere ist die Studienlage zu einer Bewegungsanalyse von Laufbewegungen sehr breit. Ein Vergleich zu anderen RGB-Tiefenkamera Systemen wie Intel Realsense, Stereolabs ZED, Orbbec Astra oder Asus Xtion, sowie ein Vergleich der Software für das Skeleton Tracking ist weniger umfassend evaluiert [Clark et al., 2018].

Insgesamt erreichen die Kinect v2 und Azure, auf Basis der oben genannten Studien, Messgenauigkeiten mit einem RMSE um $\approx 10^\circ$ gemittelt über alle Gelenkwinkel im Vergleich zu einem optischen und markerbasierten System. Bei ungünstigen Kamerablickwinkeln und verdeckten Gliedmaßen muss mit einem RMSE bis zu $\approx 30^\circ$ gerechnet werden [Yeung et al., 2021].

3.1.2. Sensorgenauigkeit von Lighthouse Tracking Systemen

In der Arbeit von [Niehorster et al., 2017] wurde die Genauigkeit des HTC Vive Systems anhand von statischen Positionen und Ausrichtungen (Lagewinkel) untersucht. Dabei wurde das HMD auf Rasterlinien (Rasterpunktfehler 1.7 ± 0.9 cm), die auf dem Boden gezeichnet wurden, platziert und die gemessene Position mit der

Rasterposition verglichen (manueller Platzierungsfehler $< 2\text{ cm}$). Dabei konnte ein $\text{RMSE} < 0.08\text{ mm}$ zwischen erfasster und tatsächlicher Position ermittelt werden. Ebenfalls wurde ein $\text{RMSE} < 0,011^\circ$ zwischen erfasster und tatsächlicher Drehung gemessen.

Ein ähnliches Verfahren wurde in der Arbeit von [Lockett, 2018] angewandt, jedoch unter Verwendung der *PhaseSpace* Kamera (Positionierungsfehler $< 1\text{ cm}$) als Referenz anstelle von Rasterlinien. Dabei konnte ein mittlerer Abstand zwischen der wahren Position und der gemessenen Position $< 1\text{ cm}$ ermittelt werden. Die Autor*innen berichteten auch über Probleme mit der Kalibrierung.

In [Borges et al., 2018] wurde die Genauigkeit des HTC Vive auf Basis von Positionen in statischen Haltungen und dynamischen Bewegungen der Tracker getestet. Die Autor*innen zeigten, dass die Genauigkeit im Submillimeterbereich für statische Trackerpositionen liegt. Für dynamische Bewegungen konnte eine Genauigkeit von Millimetern bis 5 cm gemessen werden. Des Weiteren wurde festgestellt, dass je nach Ausrichtung des Trackers zur Basisstation die Genauigkeit auf bis zu 80 cm abnehmen kann.

In der Evaluierungsstudie von [Spitzley and Karduna, 2019] wurde ein HTC Vive Tracker auf einer starren Struktur befestigt und kontrolliert bewegt und gedreht. Dabei konnte ein $\text{RMSE} < 3.0\text{ mm}$ bei Positionen und ein $\text{RMSE} < 0.4^\circ$ bei den Lagewinkeln ermittelt werden. Die Autor*innen kamen zu dem Schluss, dass das HTC Vive System das Potenzial birgt, zuverlässige kinematische Daten zu erfassen.

In [Lubetzky et al., 2019] wurde die Genauigkeit von dynamischen Kopfbewegungen anhand der Intra-Class Correlation (ICC) bewertet. In der Studie wurden die Positionen der Kopfbewegungen zwischen dem HTC Vive und dem markerbasierten Qualisys System (Referenz) verglichen. Die Autor*innen ermittelten einen ICC zwischen 0.9 und 0.99 , was auf eine sehr gute Übereinstimmung zwischen den Systemen hinweist. Eine schwächere Übereinstimmung wurde für die Lagewinkel (Gier- und Nickbewegungen) festgestellt.

In der Studie von [van der Veen et al., 2019] wurde ein Vergleich zwischen dem Vicon System und dem HTC Vive durchgeführt. In dem Experiment wurde ein Industrieroboter zur reproduzierbaren Simulation von Rumpfbewegungen eingesetzt.

In einem zweiten Experiment wurden Rumpfbewegungen von sieben Proband*innen gemessen. Die optischen Marker wurden direkt an den Vive HMD, Controller und Trackern angebracht. Evaluiert wurde auf Basis von räumlichen Positionen und Lagewinkeln. Im Vergleich der simulierten Bewegungsdaten konnten keine signifikanten Unterschiede zwischen Vicon und HTC Vive festgestellt werden. Die Ergebnisse zeigen einen RMSE von $0.68 \pm 0.32 \text{ cm}$ auf Basis der Positionen und einen RMSE von $1.64 \pm 0.18^\circ$ bei den Lagewinkeln. Auch im Vergleich der Daten von Rumpfbewegungen der Studienteilnehmer*innen wurden keine wesentlichen Unterschiede festgestellt. Die Autor*innen kamen zu dem Schluss, dass das HTC Vive System zur genauen Verfolgung von Gelenkbewegungen für klinische und Forschungszwecke verwendet werden kann.

Die Autor*innen [Ameler et al., 2019] bewerteten die Genauigkeit des *SteamVR*-Tracking Systems bei der Verfolgung der Position und Ausrichtung einer Ultraschallsonde (medizinisches Gerät) unter Laborbedingungen. Sie verglichen das *SteamVR* System mit dem markerbasierten System *OptiTrack* und bewerteten auf Basis des RMSE. Dabei konnte eine Genauigkeit bei der Positionsverfolgung mit einem RMSE von $< 1 \text{ mm}$ festgestellt werden. Für Drehbewegungen konnte ein RMSE von $< 1^\circ$ ermittelt werden.

Die Autor*innen [Sitole et al., 2020] untersuchten Lighthouse Tracking für Anwendungen in Biomechanik und Robotik. Sie montierten einen HTC Vive Tracker an einen Endeffektor eines Industrieroboters und verglichen die Positionen des Trackers mit den vom Endeffektor ermittelten Positionen. Es wurde gezeigt, dass die Genauigkeit bei Positionen im Millimeterbereich und bei den Lagewinkeln unter 1° liegt. Die Autor*innen beschrieben das Auftreten von Trackingverlusten und gaben an, dass eine Genauigkeit, bei denen kein Trackingverlust auftrat (93.4%), bis zu Submillimeter für Positionen und Subgrade für Lagewinkel möglich ist. Die Machbarkeit der Messung von Gelenkwinkeln an einem künstlichen Arm wurde ebenfalls untersucht. Die Autor*innen kamen zu dem Ergebnis, dass die Quantifizierung von Gelenkwinkeln mit dem Lighthouse Tracking-Verfahren möglich ist.

In [Bauer et al., 2021a] wurde das HTC Vive System mit einem Laser Tracking System in einem statischen Szenario verglichen. Die Autor*innen zeigten, dass das HTC Vive eine Genauigkeit im unteren Millimeterbereich bei der Positionserfassung aufweist. Allerdings beschrieben die Autor*innen auch systematische Abweichungen

im Zentimeterbereich, wenn mehrere Basisstationen eingesetzt werden. Des Weiteren wurde eine Abweichung von 0.4° bei den Lagewinkeln festgestellt.

In [Ikbal et al., 2021] wurde die Genauigkeit von räumlichen Positionen aller drei Komponenten (HMD, Controller und Tracker) des HTC Vive Systems durch eine Reihe von angepassten standardisierten Tests der American Society for Testing and Materials (ASTM) bewertet. Die gemessene Genauigkeit in statischen Positionen und Lagewinkeln konnte mit einem durchschnittlichen Fehler von 3 mm und 0.5° ermittelt werden. Bei bewegten Positionen mit einer durchschnittlichen Geschwindigkeit von 900 mm/s konnten ein Positionsfehler von weniger als 10 mm und ein Fehler von 2° bei Messung der Drehung ermittelt werden. Jedoch berichteten die Autor*innen ein hohes Aufkommen von Trackingverlusten bei 900 mm/s , weshalb von den Positionsdaten nur [Tracker 18.70 %, Controller 44.38 %, HMD 17.53 %] und bei den Lagewinkeln nur [Tracker 61.53 %, Controller 81.26 %, HMD 91.45 %] der Daten ausgewertet werden konnten. Bei einer Verringerung der Geschwindigkeit auf $< 135\text{ mm/s}$ sank der Trackingverlust bei allen Geräten auf $< 0.5\%$. Der Tracker zeigte die geringste Genauigkeit, gefolgt vom Controller. Die beste Genauigkeit und der geringste Trackingverlust wurde für das HMD festgestellt. Das HMD zeigte auch bei höhere Bewegungsgeschwindigkeiten die besten Genauigkeiten.

Zusammenfassung

In den vorgestellten Arbeiten wurde die Genauigkeit von räumlichen Positionen und Lagewinkeln von statischen Positionen und dynamischen Bewegungen, gemessen mit Lighthouse Tracking Systemen, untersucht. Insgesamt wurden Genauigkeiten von Submillimeter bis zu 1 cm und Subgrad bis 2° für statische Bedingungen von HMD, Tracker und Controller erhoben. Bei den dynamischen Bewegungen schwankten die Ergebnisse aufgrund methodischer Unterschiede (Referenzsystem, Bewegungsumfang, -geschwindigkeit) von Submillimetern bis zu einigen Zentimetern. Einige Studien beschrieben Probleme bei der Kalibrierung und das Auftreten von Trackingverlusten.

3.2. Bewegungserkennung auf Basis von MoCap-Daten

Die Entwicklung und Evaluierung von Algorithmen und Modellen für die Erkennung von Bewegungen ist ein breit untersuchtes Forschungsgebiet. Insbesondere nach Einführung der Microsoft Kinect v1 im Jahre 2010 ist ein deutlicher Anstieg

von wissenschaftlichen Arbeiten in diesem Themenfeld zu verzeichnen. In diesem Kapitel werden Arbeiten für die Erkennung von Bewegungen anhand von MoCap-Daten (Definition wie in Abschnitt 2.1) vorgestellt. Die Modellentwicklung basiert dabei zum Teil auf offenen Datensätzen. Ein wesentlicher Unterschied besteht in der Anzahl der Bewegungen bzw. Klassen in den Daten. Je mehr Klassen, desto größer das Klassifizierungsproblem. In Tabelle 3.1 sind Datensätze aufgelistet, welche in den vorgestellten Studien referenziert sind. Die Tabelle ermöglicht einen Vergleich der Klassenanzahl und eine Beschreibung der Klassen. Weitere Informationen zu den Datensätzen werden in der Beschreibung zu den referenzierten Studien angegeben.

In der Arbeit von [Patsadu et al., 2012] wurde ein Vergleich der Methoden ANN, SVM, Naive Bayes und Decision Tree für die Erkennung von Sitzen, Stehen und Liegen untersucht. Datengrundlage waren normalisierte MoCap-Daten von 20 Gelenkpositionen, aufgenommen mit der Kinect v1. Das ANN wurde in einer 20-60-3 Architektur mit 60 Neuronen im Hidden Layer aufgebaut. Die SVM wurde mit einem Polynomial-Kernel eingesetzt. Die einzelnen Verfahren erzielten dabei Genauigkeiten von [ANN 100 %, SVM 99.75 %, Naive Bayes 81.94 %, Decision Tree 93.13 %]. Die Autor*innen konnten grundsätzlich nachweisen, dass NN und SVM für die Klassifizierung der Körperhaltungen Sitzen, Stehen und Liegen perfekt geeignet sind.

Die Autor*innen [Pham et al., 2014] untersuchten Dynamic Time Warping (DTW) für die Erkennung von sechs verschiedenen Arm- und Handbewegungen (beidhändiges hohes und niedriges Winken, Klatschen, beidhändig Grüßen). Die Methode wurde auf Basis von eigens erhobenen Datensätzen mit MoCap-Daten entsprechender Bewegungen und dem MSR-Action3D evaluiert. Als Features wurden relative Winkel zwischen Gelenkpositionen berechnet und in verschiedenen Konfigurationen untersucht. Die Autor*innen zeigten Genauigkeiten von maximal [Pham Data 93.91 %, MSR-Action3D 85.82 %] und beschrieben zudem einen erhöhten Rechenaufwand für komplexere Repräsentationen von Features.

In der Arbeit von [Cho and Chen, 2014] wurde eine Bewegungserkennung auf Basis eines Deep ANN entworfen und gegen eine Extreme Learning Machine (ELM) nach [Huang et al., 2006] und eine SVM mit RBF-Kernel evaluiert. Trainiert und getestet wurden die Modelle mit der MoCap-Databse HDM05 [Müller et al., 2007], aufgenommen mit einem optischen und markerbasierten System bei 120 Hz, wel-

Tabelle 3.1.: Offene Datensätze mit Bewegungsinformationen präsentiert als Bewegungssequenzen in verschiedenen Anwendungsfeldern; Kinetics Datensatz enthält Videos; alle weiteren Datensätze enthalten MoCap-Daten

Bezeichnung	Klassen	Beschreibung	Quelle
HDM05	130	Bewegungen wie Tanzen, Tritte und Schläge, Hinsetzen, Aufstehen, verschiedene Armbewegungen, u.w.	[Müller et al., 2007]
MSR-Action3D	20	Fangen, Klatschen, Tennis Schwung, Golf Schwung, seitliches Fußbaustrecken, u.w.	[Li et al., 2010]
UT-Kinect	10	Laufen, Hinsetzen, Aufstehen, Aufheben, u.w.	[Xia et al., 2012]
SBU-Kinect	8	Interaktion von zwei Personen, wie sich nähern, sich entfernen, Schubsen, Treten, Schlagen, Austauschen von Gegenständen, Umarmen und Händeschütteln	[Yun et al., 2012]
CAD-60	12	Aktivitäten im häuslichen Umfeld (Küche, Wohnzimmer, etc.), wie Zähne putzen, Telefonieren, Wasser trinken, Kochen (umrühren) u.w.	[Sung et al., 2012]
Berkeley MHAD	11	Springen, Hampelmänner, mit einer Hand und zwei Händen winken, Klatschen, Werfen, Hinsetzen, Aufstehen, u.w.	[Ofli et al., 2013]
Florence3D	9	Trinken, Hinsetzen, Aufstehen, Klatschen, Winken, u.w.	[Seidenari et al., 2013]
KARD	18	10 Gesten wie hohes und tiefes Winken, Klatschen, Bein nach vorne oder zur Seite bewegen, u.w. und 8 Aktivitäten wie Hinsetzen, Aufstehen, Laufen, Trinken, u.w.	[Gaglio et al., 2015]
UTD-MHAD	27	Klatschen, Ausfallschritt, Kniebeuge, Laufen, Joggen, Luftboxen, u.w.	[Chen et al., 2015]
NTU-RGB-D	60	40 tägliche Handlungen (Trinken, Essen, Lesen, u.w.), 9 gesundheitsbezogene Handlungen (Niesen, Taumeln, Hinfallen, u.w.) und 11 gegenseitige Handlungen (Schlagen, Treten, Umarmen, u.w.)	[Shahroudy et al., 2016]
Kinetics	400	Videos mit Inhalten wie Zeichnen, Trinken, Lachen, Umarmen, Küssen, Händeschütteln, Geschenk öffnen, Rasenmähen, Geschirrspülen, u.w.	[Kay et al., 2017]

ches 130 Klassen beinhaltet. Die Autor*innen bildeten Unterkategorien und teilten die Bewegungssequenzen in 65 Klassen auf. Als Features wurden relative Gelenkpositionen, temporale Differenzen zwischen Gelenkpositionen und normalisierte Trajektorien eingesetzt. Die Autor*innen erzielten eine Genauigkeit bei der Evaluierung auf Basis einzelner Frames von [ANN 87.42 %, ELM 74.28 %, SVM 87.06 %]. Die Evaluierung auf Basis ganzer Bewegungssequenzen zeigte eine etwas höhere Genauigkeit von [ANN 94.86 %, ELM 92.76 %, SVM 95.12 %]. Die Erkennung mit SVM zeigt in beiden Fällen leicht bessere Ergebnisse.

In der Veröffentlichung von [Du et al., 2015] wurde ein Modell für die Bewegungserkennung in MoCap-Daten basierend auf einem RNN-Ansatz gebaut. Evaluiert wurde das Modell auf dem MSR-Action3D Datensatz, den Berkeley MHAD Daten und der MoCap Database HDM05. Die Autor*innen teilten die Gelenkpositionen in den MoCap-Daten in fünf Körperregionen (linkes, rechtes Bein; linker, rechter Arm; Rücken mit Hals und Kopf) auf und trainierten für jede Region eigene RNNs. Mit diesem Prinzip wurden zunächst separate Bewegungen einzelner Gliedmaßen verarbeitet. Für die endgültige Erkennung wurde ein künstliches Perzeptron eingesetzt, in welchem die Ergebnisse der einzelnen RNNs von den Körperregionen zusammengeführt sind. Mit dieser Methode konnten Genauigkeiten von [MSR-Action3D 94.49 %, Berkeley MHAD 100 %, HDM05 96.92 %] für die jeweiligen Datensätze erzielt werden.

In der Studie von [Laraba et al., 2015] wurde ein Erkenner auf Basis eines Hidden Markov Models (HMM) für die Erkennung von vier verschiedenen Tanzschritten in MoCap-Daten untersucht. Für die Evaluierung und Modellentwicklung wurden MoCap-Daten von Tanzbewegungen, aufgenommen mit der Kinect v2 und dem optischen und markerbasierten System Qualisys, eingesetzt. Die Autor*innen verwendeten Distanzen zwischen Gelenkpositionen von Rumpf und unteren Gliedmaßen als Features. Bei Verwendung von Daten der Kinect v2 für das Training konnte eine Erkennungsgenauigkeit von 72.41 % erreicht werden. Bei einer Anpassung der Modelle mit den Qualisys Daten steigerte sich die Genauigkeit auf 84.48 %.

Die Autor*innen [Gaglio et al., 2015] kombinierten K-Means, SVM und HMM und evaluierten diese Methode auf dem Cornell Activity Dataset (CAD-60) und einem selbst erhobenen Datensatz, dem Kinect Activity Recognition Dataset (KARD). In der Modellbeschreibung wird SVM für die Erkennung von einzelnen (frameweise)

Körperhaltungen eingesetzt und durch HMM erweitert, um im zeitlichen Verlauf auch ganze Bewegungssequenzen zu erkennen. Mit dieser Methode konnten Genauigkeiten von [CAD-60 77.3 %, KARD 84.4 %] erreicht werden.

In der Untersuchung von [Cippitelli et al., 2016] wurde ein Bewegungserkennung auf Basis von SVM mit einem RBF-Kernel in einer *one vs. one* Strategie entwickelt. Evaluiert wurde die Methode auf Basis des KARD, CAD-60, MSR-Action3D, Florence3D und dem UT-Kinect Datensatz. Die Autor*innen nutzten Distanzen von Gelenkpositionen zu einer übergeordneten Position in der Körpermitte als Features. Für jede Aktivität wurden die charakteristisch wichtigsten Körperhaltungen einzelner Frames ausgewählt und entsprechend gelabelt. Mit diesem Vorgehen wurden Genauigkeiten von [KARD 98.7 %, CAD-60 93.9 %, MSR-Action3D 81.2 %, Florence3D 91 %, UT-Kinect 96.7 %] erreicht.

In der Arbeit von [Zhang et al., 2017] wurde ein LSTM Netz für die Erkennung von Bewegungen in MoCap-Daten entwickelt. Die Autor*innen verwendeten geometrische Features in Form von Distanzen zwischen einzelnen Gelenkpunkten. Evaluiert wurde auf Basis von vier verfügbaren Datensätzen: SBU-Kinect Daten (8 Klassen), UT-Kinect Datensatz (10 Klassen), Berkeley MHAD Daten (11 Klassen) und NTU-RGB-D (40 Klassen). Der NTU-RGB-D ist im Vergleich der umfangreichste Datensatz. Die Autor*innen erzielten Genauigkeiten von [SBU-Kinect 99.02 %, UT-Kinect 95.96 %, Berkeley MHAD 100 %, NTU-RGB-D 82.39 %].

In der Publikation von [Yan et al., 2018] wurde eine Erkennung auf dem NTU-RGB-D und einem Subset des Kinetics Datensatzes von 30 verschiedenen Aktivitäten entwickelt. Für die Videos im Kinetics Datensatz wurde zunächst *OpenPose* eingesetzt, um Gelenkpositionen zu identifizieren. Auf Basis der Gelenkpositionen wurden räumliche und zeitliche Features (Distanzen zwischen zeitlich versetzten Gelenkpositionen) abgeleitet und für das Training von CNNs für die beiden Datensätze eingesetzt. Die Autor*innen evaluierten auch verschiedene Ansätze für das Feature Extraction. Insgesamt wurden Genauigkeiten von [NTU-RGB-D \approx 80 %, Kinetics \approx 70 %] erzielt.

In der Arbeit von [Pham et al., 2019] wurde ein CNN eingesetzt, um Bewegungen in MoCap-Daten zu erkennen. Die Autor*innen bildeten eine Feature Matrix mit Positionen von Gelenken über die Zeit und repräsentierten diese als Bilddar-

stellung (*spatial* \times *temporal*). Zudem verzichteten sie bewusst auf Anpassungen der Gelenkpositionen zu normalisierten Features. Evaluiert wurde u.a. auf Basis des KARD mit unterschiedlichen Netzarchitekturen verschiedener *Tiefen* im *Dense Layer*². Insgesamt wurden Genauigkeiten für die Erkennung von $> 97.5\%$ erzielt. In der Arbeit von [Gesnouin et al., 2020] wurde ebenfalls das Verfahren angewendet, eine Bewegungssequenz in MoCap-Daten als Bild zu repräsentieren und diese als Eingabe für ein CNN zu nutzen. Die Autor*innen setzten das CNN-Modell ein, um Bewegungsvorhersagen von Fußgänger*innen im Verkehr zu treffen und erreichten eine Genauigkeit von 94.4% .

Auch in der Studie von [Tasnim et al., 2021] wurden anhand der MoCap-Daten *spatial* \times *temporal* Feature Matrizen gebildet. Die Autor*innen wandten *Transfer Learning*³ auf die Modelle MobileNetV2 [Sandler et al., 2018], DenseNet121 [Huang et al., 2016], und ResNet18 [Feichtenhofer et al., 2016] an. Evaluiert und trainiert wurde auf dem UTD-MHAD Datensatz und dem MSR-Action3D Datensatz. Dabei konnten Genauigkeiten von 98.93% , 99.65% und 98.80% für UTD-MHAD und 96.00% , 98.75% und 97.08% für MSR-Action3D mit den vortrainierten Netzen MobileNetV2, DenseNet121 und ResNet18 erreicht werden.

In der Arbeit von [Yadav et al., 2021] wurde eine CNN-LSTM Kombination für die Erkennung von sieben verschiedenen Bewegungen (Sitzen, Stehen, Bücken, schnelles Gehen, langsames Gehen, Liegen und Sturzaktivitäten) entwickelt. Im Preprocessing wurden geometrische und zeitliche Features (relative Gelenkpositionen, Differenzen zwischen Gelenken, Gelenkwinkel, Winkelgeschwindigkeiten) berechnet. Das CNN-LSTM-Modell erreichte eine Genauigkeit von 98.89% . Zum Vergleich wurde auch ein CNN und ein LSTM Network in nicht kombinierter Form mit denselben Daten evaluiert. Dabei wurden etwas geringere Genauigkeiten von CNN 93.89% und von LSTM Network 92.75% erzielt.

Zusammenfassung

In den vorgestellten Arbeiten sind verschiedene Methoden für die Erkennung von Bewegungen in MoCap-Daten evaluiert worden. Insgesamt werden Erkennungsgenauigkeiten im Bereich von $\approx 90\%$ erreicht. In den meisten Veröffentlichungen wird angegeben, dass die ermittelte Genauigkeit ausreichend für den jeweiligen Anwen-

²Layer in einer CNN Architektur.

³Retraining vorhandener Modelle auf weitere oder andere Muster.

dungsfall sei. Neben einer Untersuchung differenzierter Methoden liegen wesentliche Unterschiede im Preprocessing und Feature Extraction, der genutzten Datenbasis und den Verfahren für das Training und das Testen der jeweiligen Modelle.

Die vorgestellten Arbeiten zeigen die Entwicklung und Evaluierung von Erkennungsmethoden, zum Teil anhand von offenen Datensätzen (Tabelle 3.1). Eine Prävalenz, dass die Modelle auf bestimmten Datensätzen bzw. Typen von Klassen (Ganzkörperbewegungen oder Aktivitäten wie Laufen oder Bewegungen einzelner Gliedmaßen) besser oder schlechter arbeiten, konnte nicht festgestellt werden.

3.3. Bewegungsbewertung mit Einsatz von MoCap-Sensorik

Es gibt eine Vielzahl von MoCap-Sensorik mit zugehöriger Software für die Bewegungsanalyse. In diesem Kapitel wird untersucht, welche Möglichkeiten etablierte Systeme bieten, MoCap-Daten automatisch zu interpretieren und zu analysieren. Dabei werden insbesondere Arbeiten vorgestellt, in denen die dargestellten Bewertungsmethoden aus dem Abschnitt 2.6 *Bewegungsbewertung auf Basis von Gelenkwinkeln* angewendet wurden.

Stand der Technik für die Erfassung von MoCap-Daten sind optische und markerbasierte Systeme, wie Vicon und Qualisys. Diese erlauben eine Aufzeichnung von mehr als 100 Markern mit bis zu 2000 Frames per Second (FPS). Die zugehörige Software, z.B. der *Qualisys Track Manager* (QTM) [Qualisys, 2022], ermöglicht eine Echtzeit- und Ex-post-Beobachtung der rohen Markerkoordinaten, sowie eine Simulation und Visualisierung von Körpermodellen. Des Weiteren ist eine Ableitung von biomechanischen Parametern und Gelenkwinkeln möglich. Eine automatische Bewertung von MoCap-Daten ist in der Software nicht vorgesehen [Richards, 1999, Göpfert et al., 2009].

Als tragbare MoCap-Sensorik, welche unabhängig von Blickwinkeln oder Messräumen ist, gilt der Sensoranzug von Xsens [Roetenberg et al., 2009] als hoch präzise. Damit ist die Erfassung von Gelenkpositionen mit Hilfe von Trägheitssensoren möglich. Die Software ermöglicht ebenfalls die Beobachtung und die Aufzeichnung von MoCap-Daten, jedoch keine automatische Auswertung. [Paulich et al., 2018]

Für die Interpretation von MoCap-Daten, unabhängig von der Sensorik, kann die

Analysesoftware *OpenSim* [Delp et al., 2007] eingesetzt werden. *OpenSim* ermöglicht eine Auswertung von biomechanischen Belastungen und die Simulation von Knochenbewegungen und Muskelkontraktionen. Somit können Kräfte auf einzelne Gelenke bei bestimmten Bewegungen ermittelt werden. Normalerweise wird dieses Tool für die Offline-Analyse und Simulation verwendet und bietet eine gute Möglichkeit, Bewegungsparameter für bestimmte Zwecke (z.B. Sport, Medizin) zu ermitteln.

Systeme für Bewegungstraining

Die Kinetisense-Software [Kinetisense, 2021] ist für eine Bewegungsanalyse mit der Kinect v2 oder Azure entwickelt worden. Mit der Software sind Gelenkwinkelmessungen, die allgemeine Bestimmung der RoM und Balance-Tests möglich. Des Weiteren kann die Software insbesondere für Ganganalysen und die Ermittlung von Gangparametern genutzt werden. Darüber hinaus können Haltungparameter und Gelenkwinkel live berechnet und visualisiert werden. Laut Hersteller kann Kinetisense sowohl für therapeutische Anwendungen als auch für präventives Bewegungstraining eingesetzt werden. Das System bietet den Nutzer*innen aktive Unterstützung bei der korrekten Ausführung der jeweiligen Bewegung. Wie bereits in Kapitel 2.2 *Motion Tracking mit einer RGB-Tiefenkamera* beschrieben, unterliegt das System den technologiebedingten Einschränkungen einer RGB-Tiefenkamera, weshalb eine Nutzung vorwiegend in aufrechten Körperhaltungen möglich ist.

Das icura-System [icura, 2021] ist ein Bewegungstrainingssystem, das auf tragbarer Inertialsensorik basiert. Das System besteht aus anlegbaren Sensoren, welche an Armen und Beinen platziert werden. Empfang und Auswertung der Sensordaten ist über ein Smartphone und eine dazugehörige App realisiert. Das System ist in der Lage, Bewegungsübungen zu erkennen und den Nutzer*innen gezielte Hilfestellungen für die korrekte Ausführung von Bewegungen zur Verfügung zu stellen. Darüber hinaus können Trainingsstatistiken erstellt und motivierende Anreize, in Form von vorgegebenen Bewegungszielen, präsentiert werden.

Systeme für die Ergonomiebewertung

Im Bereich der ergonomischen Arbeitsplatzbewertung hat sich die proprietäre Bewegungsanalysesoftware *Winkel Daten Analyse* (WIDAAN) etabliert, die von der IFA und der Deutsche Gesetzliche Unfallversicherung (DGUV) entwickelt wurde [El-

legast et al., 2009, Ellegast et al., 2010]. Um unabhängig von individuellen Körpergrößen vergleichbare Ergebnisse zu erhalten, werden die MoCap-Daten auf ein Körpermodell, dem sogenannten "Dortmunder", abgebildet [Jäger et al., 2001]. Durch diesen Schritt werden die individuellen anthropometrischen Maße einer jeweiligen Person vereinheitlicht. Auf Basis des Dortmunder Körpermodells können auch biomechanische Kräfte in der Lendenwirbelsäule erhoben werden. Des Weiteren bietet WIDAAN die Möglichkeit, Gelenkwinkel zu berechnen und eine automatische Haltungsanalyse nach der OWAS-Methode auszugeben. Prinzipiell ist WIDAAN für Analysen zusammen mit dem CUELA-Anzug⁴ (tragbare MoCap-Sensorik) entwickelt, dennoch können auch MoCap-Daten anderer Systeme geladen und ausgewertet werden [Weber, 2011].

Eine weitere Software für die Ergonomiebewertung ist das *Dynamicus-Tool* des Instituts für Mechatronik e.V. Chemnitz [IFM, 2021]. Damit ist es möglich, MoCap-Daten nach dem *European Assembly Worksheet* (EWAS) [Schaub et al., 2013] und nach RULA automatisch auszuwerten.

Die Autor*innen [Manghisi et al., 2017] entwickelten das *K2Rula* System, eine halb-automatische RULA-Bewertungssoftware mit der Kinect v2. Die Software eignet sich, um ungünstige Körperhaltungen in Echtzeit zu identifizieren und zu bewerten. Die Autor*innen validierten das System in zwei Experimenten, in denen 15 Körperhaltungen wie Bücken, Greifen, Aufheben u.w. in frontalem Blickwinkel zu der Kinect v2 durchgeführt wurden. Im ersten Experiment wurde eine Vergleichsmessung mit einem optischen und markerbasierten System untersucht. Verglichen wurden die Systeme auf Basis der berechneten RULA-Scores beider Systeme. Die Autor*innen erreichten eine statistisch perfekte Übereinstimmung nach der [Lendis and Koch, 1977] Skala von $k = 0.87$. Im zweiten Experiment wurde das System gegen RULA-Expert*innen (menschliche Beobachter*innen) evaluiert und die Übereinstimmung auf Basis des RULA-Gesamtscores verglichen. Dabei wurde ein etwas geringerer Übereinstimmungindex von $k = 0.84$ ermittelt. Insgesamt zeigten die Autor*innen, dass ein System für die Bewertung von Bewegungen nach dem RULA-Score mit der Kinect v2 anwendbar ist.

In der Arbeit von [Plantard et al., 2017] wurde die Kinect v2 ebenfalls gegen das

⁴Computer-unterstützte Erfassung und Langzeit-Analyse von Belastungen des Muskel-Skelett-Systems (CUELA)

Vicon System und gegen menschliche Beobachter*innen evaluiert. In der Untersuchung wurden typische Bewegungen für das Anheben und Ablegen eines Kartons unter Laborbedingungen und realen Arbeitsbedingungen untersucht. In der ersten Untersuchung wurde die Kinect v2, auf Basis der für RULA relevanten Gelenkwinkel, mit dem Vicon System als Referenz verglichen. Dabei konnte ein mittlerer RMSE von 7.7° über alle untersuchten Gelenkwinkel ermittelt werden. In der zweiten Untersuchung wurde die Kinect v2 anhand des RULA-Gesamtscores gegen RULA-Expert*innen verglichen. Dabei konnten Übereinstimmungen von $k = 0.46$ bis $k = 0.66$ für verschiedene Bewegungen erzielt werden.

Die Autor*innen [Li et al., 2020] entwickelten ein System für die automatische Risikobewertung von Körperhaltungen in RGB-Bildern. Im ersten Schritt wurde *OpenPose* eingesetzt, um Gelenkpositionen in den RGB-Bildern zu identifizieren. Im zweiten Schritt wurde ein ANN trainiert, um direkt anhand der Gelenkpositionen einen RULA-Score zu schätzen. Mit diesem Verfahren wird eine Berechnung von Gelenkwinkeln für die Ermittlung des RULA-Scores ausgelassen. Trainiert und getestet wurde das Verfahren mit zwei Datensätzen [Xu et al., 2011, Ionescu et al., 2014], welche typische Hebebewegungen von Kisten sowie Alltagsbewegungen wie Winken, Laufen u.w. beinhalten. Die Autor*innen erreichten eine Genauigkeit von 93 % für die korrekte Ermittlung des RULA-Scores der jeweiligen Körperhaltungen. Sie sagten aus, dass das vorgeschlagene Verfahren für Online-Anwendungen bei 29 FPS für gängige Rechnerplattformen geeignet sei.

Zusammenfassung

Die vorgestellten Arbeiten zeigen Systeme für die automatische Bewegungsanalyse. Mit den Systemen von Vicon, Qualisys und der Software OpenSim können Analysen auf Expertenniveau durchgeführt werden. Damit sind detaillierte Auswertungen von biomechanischen Belastungen möglich. Systeme für das Bewegungstraining gestalten sich i.d.R. als anwenderfreundlich und bieten eine intuitiv interpretierbare Hilfestellung bei der Ausführung von Bewegungen. Diese Systeme sind in den meisten Fällen an eine bestimmte Sensorik gebunden. Des Weiteren existieren eine Vielzahl von Systemen für die Risikobewertung an Arbeitsplätzen [Maczewska et al., 2020]. In einem Großteil der Forschungsarbeiten wurden der Einsatz von RGB-Tiefenkameras und eine Bewegungsbewertung nach RULA oder OWAS umgesetzt. In den Arbeiten von [Manghisi et al., 2017, Plantard et al., 2017] wurde die automatische Bewegungsbewertung nach RULA gegen menschliche Beobachter*innen

evaluiert. Dabei konnten hohe Übereinstimmungen (Interrater-Reliabilität) gezeigt werden.

Eine Interpretation des untersuchten Stands der Technik und die daraus resultierenden Schlüsse für den eigenen Ansatz und die Realisierung einer automatischen Bewegungsbewertung mit Einsatz von MoCap-Sensorik wird im folgenden Kapitel wiedergegeben.

4. Eigener Ansatz

In diesem Kapitel wird der eigene Ansatz für die Entwicklung und Evaluation von Algorithmen zur automatischen Bewegungserkennung und -bewertung auf Basis von Gelenkwinkeln aus MoCap-Daten vorgestellt. Richtungweisend für die Auswahl von Methoden sind Forschungsergebnisse aus Kapitel 3 *Stand der Technik*. Die folgenden Abschnitte, zu den drei Forschungsfragen

I. Sensorgenauigkeit,

II. Bewegungserkennung und

III. Bewegungsbewertung,

beginnen jeweils mit der Relevanz der jeweiligen Fragestellung. Anschließend wird der erhobene Forschungsstand diskutiert und der eigene Beitrag dieser Dissertation herausgestellt. Des Weiteren wird in Abschnitt 4.4 eine globale Darstellung des eigenen Ansatzes präsentiert.

4.1. Fragestellung I - Sensorgenauigkeit

Die Fragestellung I - Sensorgenauigkeit hat eine hohe Bedeutsamkeit für die Fragestellung II - Bewegungsbewertung. Im Folgenden wird zunächst der Zusammenhang und die Relevanz diskutiert und anschließend der eigene Ansatz für die Ermittlung der Sensorgenauigkeit von einer RGB-Tiefenkamera und einem Lighthouse Tracking System dargestellt.

4.1.1. Relevanz der Sensorgenauigkeit für die Bewegungsbewertung

In Abschnitt 2.6 *Bewegungsbewertung auf Basis von Gelenkwinkeln* sind Schemata für die Bewertung von Bewegungen dargestellt. Das Schema der DGOU für Nachbehandlungsempfehlungen sowie Schemata für die ergonomische Bewegungsbewertung wie RULA, REBA und das IFA Ampelschema zeigen eine Bewegungsbewertung anhand von kritischen Gelenkwinkeln. In Abschnitt 3.3 *Bewegungsbewertung*

mit Einsatz von MoCap-Sensorik sind Systeme für die automatische Bewegungsbewertung gezeigt. Ersichtlich ist, dass eine automatische Bewegungsbewertung mit Einsatz von MoCap-Sensorik möglich ist. Um kritische Grenzwinkel anhand der Schemata richtig zu bewerten, ist eine gewisse Genauigkeit bei der Erfassung von Gelenkwinkeln relevant. Aus diesem Grund ist es wichtig, den Fehler der jeweiligen MoCap-Sensorik zu kennen, um die Ergebnisse der Bewegungsbewertung richtig beurteilen zu können. Die Bedeutsamkeit der Genauigkeit bei der Gelenkwinkel-messung wurde auch in eigens durchgeführten Befragungen [Vox et al., 2016, Vox and Wallhoff, 2016] an Expert*innen aus dem medizinischen und physiotherapeutischen Bereich bekräftigt.

In Bezug auf die in Kapitel 1 *Einleitung* beschriebenen Anwendungsfälle werden in dieser Arbeit sowohl eine RGB-Tiefenkamera als auch ein Lighthouse Tracking System näher untersucht. Beide Systeme stammen aus dem Consumer-Segment und zeichnen sich durch eine hohe Marktverfügbarkeit und eine benutzerfreundliche Einsetzbarkeit aus.

4.1.2. Ermittlung der Sensorgenauigkeit einer RGB-Tiefenkamera

Die Eignung von Systemen auf Basis einer RGB-Tiefenkamera für die Bewegungsanalyse ist bereits breit untersucht. In den Grundlagen, in Abschnitt 2.2 *Motion Tracking mit einer RGB-Tiefenkamera* wurde erklärt, dass bei der Erfassung von MoCap-Daten, neben der RGB-Tiefenkamera, auch die Software für das Skeleton Tracking Einfluss auf die Genauigkeit von Gelenkpunkten nimmt. Im Kapitel zum Stand der Technik unter Abschnitt 3.1.1 *Sensorgenauigkeit von RGB-Tiefenkameras* konnte gezeigt werden, dass in einem Großteil der referenzierten Arbeiten die RGB-Tiefenkamera Systeme von Microsoft (Kinect v2, Azure) mit entsprechender Software für das Skeleton Tracking untersucht wurden.

Generell sind die Kinect v2 und Azure bereits sehr gut evaluiert. In Vergleichsstudien zwischen Kinect v2, Azure und optischen markerbasierten Systemen wurde insgesamt eine mittlere Genauigkeit für die Gelenkwinkel-messung mit einem RMSE von $\approx 10^\circ$ ermittelt. Des Weiteren wurde festgestellt, dass in ungünstigen Blickwinkeln der Kamera und bei der Verdeckung von Gliedmaßen mit unplausiblen Tracking-Positionen gerechnet werden muss [Wiedemann et al., 2015, Yeung et al., 2021].

In Ergänzung zu den Ergebnissen aus dem Stand der Technik, wird in dieser Arbeit die Kinect v2 unter statischen Bedingungen untersucht. Ziel ist es, in einer technischen Vergleichsstudie zwischen der Kinect v2 und dem Qualisys das **allgemeine Sensorrauschen zu ermitteln und die Kinect v2 auf systematische Fehler zu untersuchen**. Die Durchführung und Ergebnisse dieser Studie sind in Abschnitt 5.3 *Sensorrauschen von der Kinect v2 in statischen Haltungen* beschrieben und in [Vox and Wallhoff, 2016] publiziert.

4.1.3. Ermittlung der Sensorgenauigkeit eines Lighthouse Tracking Systems

In den referenzierten Arbeiten, in Abschnitt 3.1.2, wurden Genauigkeiten von HMD, Controller und Tracker auf Basis von Positionen und Drehungen erhoben. Dabei konnten in statischen und dynamischen Szenarien hohe Genauigkeiten ermittelt werden. In der überwiegenden Anzahl der Arbeiten wurde eine technische Vergleichsstudie zwischen einem Lighthouse Tracking System und einem Referenzsystem, anhand von Bewegungen eines Roboter Endeffektors, durchgeführt [van der Veen et al., 2019, Sitole et al., 2020].

Ansatz in dieser Arbeit ist die Untersuchung der **Genauigkeit bei der Erfassung von menschlichen Bewegungen**. Wie in Abschnitt 2.3 *Motion Tracking mit einem Lighthouse Tracking System* beschrieben, wird für die Berechnung von MoCap-Daten aus Lighthouse-Tracker-Positionen eine IK eingesetzt. Die Genauigkeit hängt also nicht nur von der räumlichen Trackingleistung ab, sondern auch von der Berechnung der IK.

Des Weiteren ist bei der Befestigung von HMD, Trackern und Controllern an menschlichen Gliedmaßen mit Sensorverrutschen und Okklusion (ein Körperteil verdeckt die optische Verbindung zur Basisstation) bei der Bewegungsausführung zu rechnen. Nach aktuellem Stand der Recherche konnte keine Evaluierungsstudie eines Lighthouse Tracking Systems in Kombination mit einer IK, auf Basis von MoCap-Daten bzw. Gelenkwinkeln, gefunden werden.

Aus diesem Grund ist das Ziel in dieser Arbeit, eine **technische Evaluierung zwischen dem HTC Vive in Kombination mit Final IK und dem Qualisys durchzuführen** und den Fehler bei der Gelenkwinkelmessung zu identifizieren. Die Durchführung einer Vergleichsstudie zwischen dem HTC Vive System in Kombination mit Final IK und

dem Qualisys System ist in Abschnitt 5.4 *Sensorgenauigkeit des HTC Vive Tracking Systems* beschrieben. Die Ergebnisse aus dieser Untersuchung sind bereits in [Vox et al., 2021] publiziert.

Die Frage, ob die ermittelte Sensorgenauigkeit für eine Bewegungsbewertung anhand von kritischen Gelenkwinkeln ausreicht, wird in Kapitel 6 *Technische Evaluation*, Abschnitt 6.1 *Sensorgenauigkeit in kritischen Körperhaltungen* beantwortet.

4.2. Fragestellung II - Bewegungserkennung

Die Fragestellung II - Bewegungserkennung hat ebenfalls einen engen Zusammenhang zu Fragestellung III - Bewegungsbewertung. Im Folgenden wird zunächst die Relevanz der Bewegungserkennung für die Bewegungsbewertung diskutiert und anschließend der eigene Ansatz für die Entwicklung einer Bewegungserkennung erläutert.

4.2.1. Relevanz der Bewegungserkennung für die Bewegungsbewertung

Neben der Einschätzung von kritischen Gelenkwinkeln sind auch der Bewegungstyp, die Ausführungsdauer und die Anzahl von Wiederholungen relevant für die Bewegungsbewertung. Diese Parameter ermöglichen es, einzelne Bewegungen voneinander differenziert zu bewerten. Im Anwendungsfall von Bewegungstraining kann dadurch eine automatische Trainingsstatistik erstellt werden und bietet die Möglichkeit, bestimmte Grenzwinkel für individuelle Bewegungen festzulegen [Vox et al., 2016]. Das gleiche Verfahren kann für die Bewertung von Bewegungssequenzen am Arbeitsplatz hilfreich sein. Sollen aus aufeinander folgenden Bewegungssequenzen am Arbeitsplatz einzelne Abläufe bewertet werden, ist es wichtig, diese zu erkennen und voneinander zu differenzieren.

4.2.2. Ansatz für die Entwicklung einer Bewegungserkennung

Bei der Entwicklung von Modellen für die Bewegungserkennung sind einige Aspekte zu berücksichtigen. Wie in Abschnitt 2.4 *Algorithmen für die Erkennung von Körperhaltungen* beschrieben, sind sowohl die Datengrundlage und Feature Extraction als auch die Auswahl und Parametrisierung geeigneter Methoden entscheidend für eine effiziente Erkennung.

In referenzierten Arbeiten (Abschnitt 3.2) konnten generell zwei Arten von Feature Extraction identifiziert werden: *Handcrafted* Features und MoCap-Daten in Rohform. Bei den *Handcrafted* Features werden normalisierte Gelenkpositionen, Distanzen zwischen Gelenkpunkten oder -winkeln eingesetzt. Diese wurden häufiger in Veröffentlichungen bis 2019 forciert. Dabei zeigen die Methoden SVM und ANN vielversprechende Genauigkeiten von $> 90\%$ [Patsadu et al., 2012, Cho and Chen, 2014, Gaglio et al., 2015].

In jüngeren Arbeiten werden zunehmend Modelle auf Basis von CNNs, mit Genauigkeiten von $> 95\%$, für die Bewegungserkennung eingesetzt [Pham et al., 2019, Tasnim et al., 2021, Yadav et al., 2021]. In diesem Fall werden die MoCap-Daten nahezu in Rohform als Features verwendet. Lediglich eine Transformation der MoCap-Daten zu einer Pixelmatrix (i.d.R. *spatial* \times *temporal*) ist notwendig [Kale and Patil, 2016, Nanni et al., 2017]. Anzumerken ist, dass der Rechenaufwand bei Verwendung von CNNs vergleichsweise hoch ist und i.d.R. eine Grafikerunterstützung für das Training erforderlich ist.

Ein weiterer Unterschied liegt im generellen Aufbau der Features. Dabei ist zwischen Features auf Basis einzelner Frames und Features von ganzen Bewegungssequenzen zu unterscheiden. Werden einzelne Frames genutzt, betrachtet das Modell eine Bewegung als eine Sequenz aus einzelnen Körperhaltungen [Cho and Chen, 2014]. In diesem Fall sind zeitliche Komponenten wie Bewegungsgeschwindigkeiten oder -beschleunigungen irrelevant. Wird ein Modell eingesetzt, um ganze Bewegungssequenzen (z.B. vollständige Kniebeuge) zu erkennen, werden in den Features ganze Zeiträume (Fenster) abgebildet [Yan et al., 2018]. In [Cho and Chen, 2014] konnte gezeigt werden, dass ein Modell für die Erkennung ganzer Bewegungssequenzen eine höhere Genauigkeit aufweist als ein Modell für die Erkennung einzelner Frames.

Ansatz in dieser Arbeit ist eine **Erkennung von Körperhaltungen in einzelnen Frames** für die Anwendung auf mobilen Endgeräten. Sowohl die Bewegungserkennung als auch die Bewegungsbewertung werden für eine frameweise Analyse entwickelt. Um auch Bewegungssequenzen differenziert voneinander zu analysieren, wird in Abschnitt 5.6 ein Algorithmus für die Segmentierung von Bewegungen vorgestellt.

In diesem Ansatz werden **handcrafted Features in Form von Gelenkwinkeln** eingesetzt. Gelenkwinkel werden auch für die Bewegungsbewertung benötigt, jedoch ist

die Motivation zwischen physiologischen Gelenkwinkeln für die Bewegungsbewertung und Features für die Bewegungserkennung gänzlich verschieden. Bei der Bewegungsbewertung wird versucht, möglichst individuelle Gelenkwinkel zu betrachten. Wohingegen bei Gelenkwinkeln als Features eine möglichst homogene Darstellung der Daten erreicht werden sollte. Der Vorteil von Gelenkwinkeln gegenüber Gelenkpositionen als Features ist, dass Informationen in den Winkeln unabhängig von räumlichen Distanzen oder individuellen anthropometrischen Parametern sind. Es kann die Annahme getroffen werden, dass normalisierte Gelenkwinkel die Informationen für die Erkennung einer Haltung annähernd personenunabhängig transportieren. Ziel der Normalisierung ist, die individuelle RoM zu vereinheitlichen und persönliche Bewegungsmerkmale auszuklammern. Die Normalisierung und Evaluierung von Feature-Kombinationen wird in Abschnitt 5.5.2 erläutert.

Anforderung an das System ist, dass die Auswertung auch auf mobilen Plattformen und in Echtzeit möglich ist. Ebenfalls soll es möglich sein, die Erkennung dynamisch auf neue Bewegungen und Körperhaltungen anzupassen. In diesem Fall würde das Training der jeweiligen Methode für die Erkennung auf einer mobilen Rechnerplattform durchgeführt werden. Die Methoden ANN und SVM konnten in den referenzierten Arbeiten hohe Genauigkeiten erzielen und zeigen sich gegenüber CNNs sowohl in der Entwicklung als auch in der Anwendung weniger rechenintensiv. Aus diesem Grund werden im Abschnitt 5.5 die **Methoden ANN und SVM untersucht**. Datengrundlage für die Entwicklung sind MoCap-Daten aus der *1. Bewegungsstudie*.

Die Realisierung und Evaluierung eines Modells für die Bewegungserkennung ist in Abschnitt 5.5 beschrieben. Teile der Entwicklungsschritte sind in den Arbeiten [Vox and Wallhoff, 2017b, Vox and Wallhoff, 2018a, Vox and Wallhoff, 2018b, Vox and Wallhoff, 2018c] publiziert. Die Frage, ob die ermittelten Erkennungsgenauigkeiten ausreichend für eine automatische Bewegungsbewertung sind, wird in Abschnitt 6.2 *Erkennung von kritischen Körperhaltungen* evaluiert.

4.3. Fragestellung III - Bewegungsbewertung

In Fragestellung III - Bewegungsbewertung wird die Anwendbarkeit einer automatischen Bewegungsbewertung mit Einsatz von MoCap-Sensorik in Gänze untersucht. Zur Beantwortung ist sowohl die Sensorgenauigkeit bei der Gelenkwinkelmessung

als auch eine Bewegungserkennung von Bedeutung. Im Folgenden wird zunächst die Relevanz dargestellt und anschließend der eigene Ansatz für eine technische Umsetzung erläutert.

4.3.1. Relevanz einer automatischen Bewegungsbewertung

Die Relevanz einer automatischen Bewegungsbewertung spiegelt sich in der globalen Motivation dieser Arbeit wieder. Wie in Kapitel 1.1 *Problemstellung* dargestellt sind Erkrankungen des Muskel-Skelett-Systems (MSS) häufig. Somit ist der Einsatz von Assistiven Technologien für die Identifikation und Vermeidung von potenziell schädlichen Bewegungen und Körperhaltungen vielversprechend zur Prävention und Reduktion der Krankheitslast. In einer eigens durchgeführten Befragung [Vox and Wallhoff, 2016] an Physiotherapeut*innen bzgl. Akzeptanzfaktoren an Systeme für eine automatische Bewegungsanalyse wurde sich für eine benutzerfreundliche und einfache Einsetzbarkeit sowie ein nachvollziehbares Systemverhalten, insbesondere bei der Gelenkwinkelmessung, ausgesprochen.

4.3.2. Ansatz für die Realisierung einer automatischen Bewegungsbewertung

In Abschnitt 3.3 *Bewegungsbewertung mit Einsatz von MoCap-Sensorik* wurden aktuelle Systeme für die Bewegungsbewertung vorgestellt und zwischen Systemen für

- Bewegungstraining und
- Ergonomiebewertung

differenziert. In den meisten Fällen sind die Systeme auf bestimmte Aufgaben und spezifische Sensorik zugeschnitten. Das bedeutet, das System (Software + Sensorik) eignet sich entweder für Bewegungstraining oder für die Ergonomiebewertung von Arbeitsbewegungen. Des Weiteren existieren Systeme, welche zwar eine detaillierte biomechanische Analyse unabhängig von der Sensorik ermöglichen, jedoch i.d.R. nur von Expert*innen eingesetzt werden können. Eine Betrachtung der Sensorgenauigkeit für die jeweilige Bewegungsbewertung wird in den meisten, im Stand der Technik untersuchten, Systemen vernachlässigt.

Ansatz in dieser Arbeit ist die **Realisierung eines mobil einsetzbaren Systems für die Bewegungsbewertung auf Grundlage von digitalisierten Bewertungsschemata**

(Abschnitt 5.7 *Automatische Bewegungsbewertung*). Sowohl die eingesetzte MoCap-Sensorik als auch der Kontext für die automatische Bewertung soll anpassbar sein. Es wird angestrebt, die Bewertungsschemata zu vereinheitlichen und austauschbar für den jeweiligen Kontext in einem Framework zu integrieren. Dadurch wird erreicht, dass das System sowohl für die Bewertung von Bewegungstraining als auch für die Ergonomiebewertung eingesetzt werden kann. Anforderungen an die Algorithmen sind, dass eine Bewertung von Bewegungen sowohl **offline als auch in Echtzeit** möglich ist. Des Weiteren wird diskutiert, welche Sensorgenauigkeiten erforderlich sind, um kritische Grenzwinkel anhand der Schemata richtig einschätzen zu können. Darüber hinaus wird die Systemperformance auf mobilen Endgeräten (Einplatinencomputer, Laptop) untersucht.

Grundlage für die Realisierung einer automatischen Bewegungsbewertung mit Einsatz von MoCap-Sensorik ist eine Berechnungsmethode für die Bestimmung von Gelenkwinkeln aus MoCap-Daten und digitalisierte Schemata für die Bewegungsbewertung. Neben der Erfassung von kritischen Gelenkwinkeln ist auch die Erkennung der jeweiligen Körperhaltung oder Bewegung relevant für die Bewertung. Darüber hinaus nimmt die Dauer in einer Körperhaltung oder Bewegung und die Anzahl von Wiederholungen Einfluss auf die Bewertung. Die Schritte für die **Implementierung der Schemata: DGOU Nachbehandlungsempfehlungen, RULA, REBA und das IFA Ampelschema** sind in Abschnitt 5.7.1 *Digitalisierung von Bewertungsschemata* beschrieben.

In Abschnitt 5.7 *Automatische Bewegungsbewertung* wird ein Ansatz für die **Integration der oben genannten Verfahren und Algorithmen in ein prototypisches System** vorgestellt. Teile der Implementierung und des Systems sind bereits in [Vox and Wallhoff, 2019] publiziert.

4.4. Globale Darstellung des eigenen Beitrags

Grundlage für die Evaluation der Sensorgenauigkeit und Entwicklung von Modellen für die Bewegungsbewertung sind MoCap-Daten. Im Zuge dieser Arbeit wurden zwei Bewegungsstudien durchgeführt. In der ersten Bewegungsstudie wurden MoCap-Daten mit der Kinect v2, mit dem Ziel der Entwicklung eines Modells für die Bewegungserkennung, erhoben. Dabei wurden 20 verschiedene Bewegungsmuster (Klassen) aufgezeichnet. Eine detaillierte Beschreibung der Datenerhebung ist

in Abschnitt 5.5.1 wiedergegeben.

In der zweiten Bewegungsstudie wurden MoCap-Daten mit dem Qualisys System und dem HTC Vive System aufgezeichnet. Ziel dieser Studie ist die Ermittlung der Sensorgenauigkeit durch einen Systemvergleich mit dem Qualisys als Referenzsystem. Bei dieser Studie wurden ebenfalls 20 verschiedene Bewegungen, explizit unter Ausnutzung der kompletten RoM untersuchter Gelenkwinkel, aufgezeichnet. Das experimentelle Design und die Durchführung der Studie ist in Abschnitt 5.4.2 dargestellt.

Die Berechnung von Gelenkwinkeln aus MoCap-Daten ist ein zentraler Prozessschritt dieser Arbeit. Die Realisierung einer Berechnungsgrundlage ist in Abschnitt 5.2 beschrieben. Sowohl für die Evaluierung der Sensorgenauigkeit als auch für die Bewegungsbewertung sind physiologische Gelenkwinkel, angegeben in der RoM (siehe Abschnitt 2.5), bedeutsam. Des Weiteren können einerseits Gelenkwinkel, andererseits Winkel zwischen physiologisch nicht benachbarten Gelenkpunkten als Features für die Bewegungserkennung eingesetzt werden.

Eine globale Darstellung des eigenen Ansatzes und der Zusammenhänge zwischen den Fragestellungen ist in Abbildung 4.1 dargestellt.

Abgrenzung

In Fragestellung I - Sensorgenauigkeit wird ausschließlich die Genauigkeit der MoCap-Sensorik (RGB-Tiefenkamera, Lighthouse Tracking System) untersucht. Nach Algorithmen und Verfahren für die Verbesserung der Trackingleistungen wird nicht geforscht.

In Fragestellung III - Bewegungsbewertung werden ausschließlich Schemata auf Basis von kritischen Gelenkwinkelgrenzen untersucht. Die Bewertung von biomechanischen Belastungen auf Gelenke, wie z.B. bei dem "Dortmunder" Körpermodell [Jäger et al., 2001], werden vernachlässigt.

Die Bewegungsanalyse bezieht sich auf moderate Bewegungsgeschwindigkeiten mit Winkelgeschwindigkeiten bis zu $100^\circ/s$. Analysen von Beschleunigungen und Geschwindigkeiten, wie in den Sportwissenschaften üblich [Haller, 2007], werden in dieser Arbeit nicht behandelt.

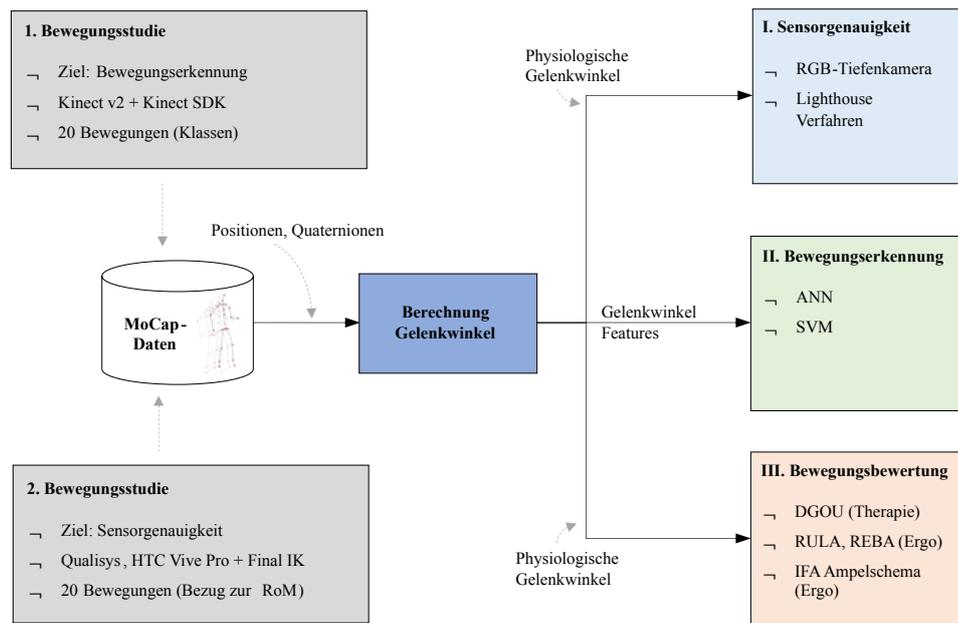


Abbildung 4.1.: Globale Darstellung des eigenen Ansatzes; Bewegungsstudien und Datengrundlage; Verbindung zwischen den Fragestellungen I. Sensorgenauigkeit, II. Bewegungserkennung und III. Bewegungsbewertung

5. Realisierung und Ergebnisse

In diesem Kapitel werden sowohl die Entwicklungsschritte für ein prototypisches System für die Bewegungsanalyse als auch Evaluierungsstudien zur Erhebung von Sensorgenauigkeiten vorgestellt. Zunächst werden allgemeine Verarbeitungsprozesse für die Transformation und Normalisierung von MoCap-Daten präsentiert (Abschnitt 5.1). Dies ist notwendig, um MoCap-Daten verschiedener Sensorsysteme zu vereinheitlichen und vergleichbar zu machen. Zum einen wird dadurch erreicht, dass in Studien zur Sensorgenauigkeit zwei Systeme miteinander verglichen werden können (Abschnitt 5.3 und 5.4). Zum anderen ist eine Vereinheitlichung des Datenmodells sinnvoll, um verschiedene Sensorsysteme an das gezeigte Gesamtsystem anbinden zu können.

Der nächste Schritt ist die Berechnung von Gelenkwinkeln aus MoCap-Daten (Abschnitt 5.2). Gelenkwinkel werden in dieser Arbeit als Features für die Bewegungserkennung (Abschnitt 5.5) und in Form von physiologischen Gelenkwinkeln für die Bewegungsbewertung (Abschnitt 5.7) sowie als Eingang für einen Algorithmus zur zeitlichen Segmentierung von Bewegungen (Abschnitt 5.6) eingesetzt. Des Weiteren wird in den Studien zur Sensorgenauigkeit die Genauigkeit bei Gelenkwinkelmessungen erhoben.

In Gänze wird ein prototypisches System für die Bewegungsbewertung auf Basis der erkannten Bewegung oder Körperhaltung (Label), der Dauer und einer Bewertung von kritischen Gelenkwinkeln vorgestellt (Abschnitt 5.7). Eingang der Verarbeitung sind entweder lokale MoCap-Daten oder ein Datenstream von entsprechender MoCap-Sensorik für die Live-Analyse. Ausgang des Systems ist eine eigens entwickelte Softwareoberfläche mit interpretierbaren graphischen Elementen für die Bewegungsbewertung.

In Abbildung 5.1 sind die Verarbeitungsschritte und einzelnen Module des Gesamtsystems dargestellt.

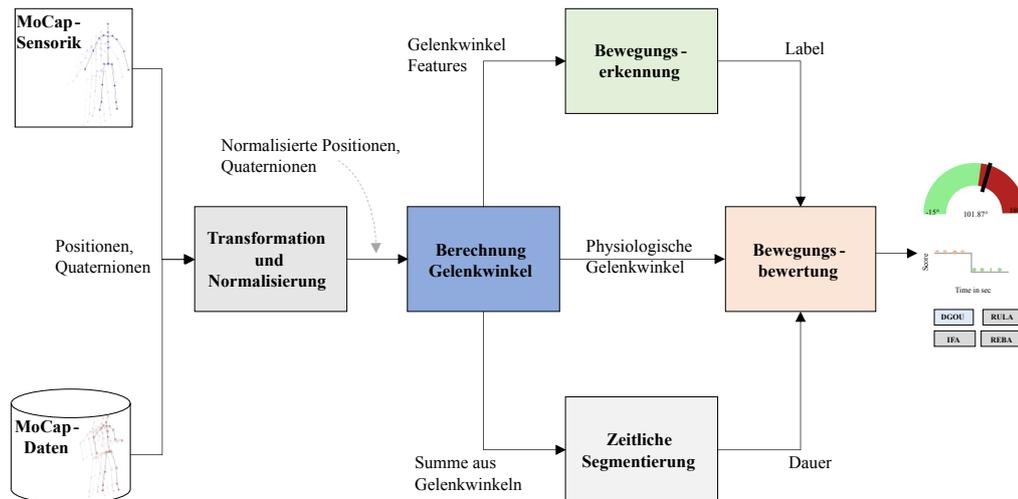


Abbildung 5.1.: Darstellung der Verarbeitungsschritte von MoCap-Daten für die Umsetzung eines prototypischen Systems zur automatischen Bewegungsbewertung auf Basis von Dauer und Label einer erkannten Bewegung oder Körperhaltung

5.1. Transformation und Normalisierung von MoCap-Daten

Die Transformation und Normalisierung von MoCap-Daten zu einem vereinheitlichten Datenmodell ist notwendig, um Daten verschiedener Anbieter von MoCap-Sensorik verarbeiten zu können. In dieser Arbeit wurden folgende vier Schritte identifiziert:

1. Zuordnung der Gelenkpunkte zu einem vereinheitlichten Skelettmodell
2. Skalierung der kartesischen Koordinaten
3. Notation der Quaternionen und Normalisierung
4. Festlegung eines lokalen oder globalen Ursprungs und Achsenausrichtung

(1.) Die Benennung und Positionierung von Gelenkpunkten ist bei MoCap-Daten verschiedener Anbieter in der Regel nicht einheitlich. Ansatz in dieser Arbeit ist ein händisches Verfahren für die Zuordnung zu einem vereinheitlichten Skelettmodell. Konkret werden dabei spezifische Namensgebungen und Positionierungen von Gelenkpunkten auf das in dieser Arbeit entwickelte Skelettmodell in Abbildung 5.2 gemappt. Die weitere Verarbeitung basiert somit auf 25 Gelenkpunkten. Ist eine Zuordnung von einzelnen Gelenkpunkten nicht möglich, d.h. es fehlen einzelne Punkte,

wirkt sich dies entsprechend auf die Berechnung von Gelenkwinkeln aus. Es können nur die Gelenkwinkel bestimmt werden, bei denen die erforderlichen Gelenkpunkte für die Berechnung vorhanden sind.

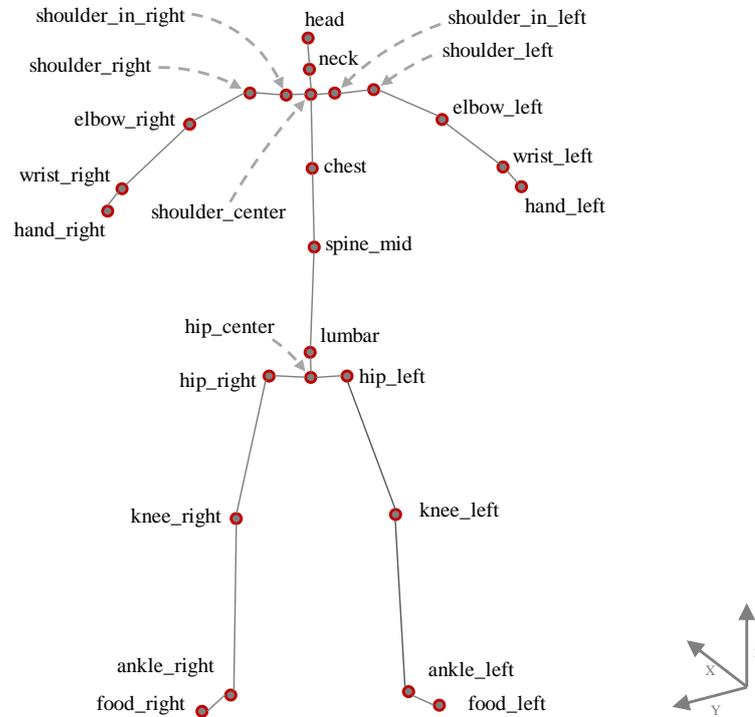


Abbildung 5.2.: Skelettmodell mit 25 Gelenkpunkten als Grundlage für die Vereinheitlichung von Positionen und Benennungen der Gelenkpunkte

(2.) Die kartesischen Koordinaten in MoCap-Daten werden in der Regel in Meter angegeben. Herstellerspezifisch können auch andere Längeneinheiten angegeben werden. Ansatz in dieser Arbeit ist, das jeweilige Format zunächst zu identifizieren und anschließend auf eine Angabe in Metern zu skalieren.

(3.) Wie in Abschnitt 2.1.2 *Räumliche Drehung anhand von Quaternionen* beschrieben, gibt es für Quaternionen verschiedene Notationen. In dieser Arbeit wird eine Notation nach Hamilton verwendet und andere Notationen entsprechend umformatiert. In der Berechnungsgrundlage für die Bestimmung von Gelenkwinkeln (Abschnitt 5.2) wird die Angabe als Einheitsquaternion angenommen. Liegen die Quaternionen in den MoCap-Daten nicht als Einheitsquaternionen vor, ist eine Normalisierung erforderlich. Dafür ist zunächst der Betrag z (Gleichung 5.1) zu bestimmen und anschließend werden die Werte des Quaternionen q durch den Betrag

z geteilt (Gleichung 5.2).

$$z = \sqrt{w^2 + xi^2 + yj^2 + zk^2} \quad (5.1)$$

$$q = \left(\frac{w}{z}, \frac{xi}{z}, \frac{yj}{z}, \frac{zk}{z} \right) \quad (5.2)$$

(4.) Die Anpassung der Achsenausrichtung auf ein einheitliches Koordinatensystem, sowie das Setzen des Ursprungs, ist für die Darstellung von MoCap-Daten und für die Bestimmung der Bewegungsrichtung (vor, zurück, oben, unten, links, rechts) relevant. In dieser Arbeit werden die kartesischen Koordinaten auf ein rechtshändiges Koordinatensystem transformiert und die lokale Position von *hip_center* als Ursprung (0,0,0) eingesetzt. Dadurch wirken die in der Software dargestellten Skelettmodelle ähnlich wie Marionetten, welche an der Hüfte (*hip_center*) hochgehalten werden. Das ist insbesondere für die Darstellung von Laufbewegungen sinnvoll, um den Koordinatenbereich nicht zu verlassen.

5.2. Verarbeitung und Berechnung von physiologischen Gelenkwinkeln aus MoCap-Daten

Im Zuge dieser Arbeit wurde ein eigens entwickeltes Tool für die Berechnung und Analyse von Gelenkwinkeln implementiert. Eine Darstellung der Software-Oberfläche ist in Abbildung A.1 im Anhang gezeigt. In der folgenden Beschreibung werden die implementierten Berechnungen für die Bestimmung von Gelenkwinkeln aus MoCap-Daten detailliert erläutert.

Physiologische Gelenkwinkel werden in Grad angegeben. Bei der Umsetzung wurde für jedes Gelenk mit den jeweiligen Freiheitsgraden eine individuelle Rechenvorschrift definiert. Im Allgemeinen kann jedoch zwischen den folgenden drei Berechnungen unterschieden werden:

1. Winkel mit Bezug zu einer Körperebene
2. Winkel berechnet aus drei kartesischen Koordinaten
3. Rotationswinkel berechnet aus der relativen Drehung zwischen zwei Quaternionen

In Tabelle 5.1 sind die untersuchten Gelenkpunkte und Freiheitsgrade aufgelistet. Des Weiteren sind Bezeichnungen der jeweiligen Gelenkwinkel und die jeweils angewendete Berechnung zugeordnet.

Tabelle 5.1.: Auflistung von Gelenkpunkten, Freiheitsgraden und angewendetem Rechenweg für die Bestimmung des jeweiligen Gelenkwinkels

Gelenk	Freiheitsgrad	Bez. Gelenkwinkel	Rechenweg
Neck	Extension, Flexion	neck_flex	(1.)
	Lateralflexion rechts, links	neck_latflex	(1.)
	Rotation rechts, links	head_rot	(3.)
Chest Spine	Extension, Flexion	chest_flex	(1.)
	Lateralflexion rechts, links	chest_latflex	(1.)
Lumbar Spine	Extension, Flexion	lumbar_flex	(1.)
	Lateralflexion rechts, links	lumbar_latflex	(1.)
	Rotation rechts, links	lumbar_rot	(3.)
Shoulder	Extension, Flexion	shoulder_(left, right)_flex	(1.)
	Adduktion, Abduktion	shoulder_(left, right)_elev	(1.)
	Rotation, angelegter Arm	shoulder_(left, right)_transrot	(1.)
	Rotation, abduzierter Arm	shoulder_(left, right)_sagrot	(1.)
Elbow	Extension, Flexion	elbow_(left, right)_flex	(2.)
	Pronation, Supination	elbow_(left, right)_rot	(3.)
Hand	Extension, Flexion	hand_(left, right)_flex	(1.)
	Radialduktion, Ulnarduktion	hand_(left, right)_azim	(1.)
Hip	Extension, Flexion	hip_(left, right)_flex	(1.)
	Adduktion, Abduktion	hip_(left, right)_elev	(1.)
	Rotation innen, außen	hip_(left, right)_rot	(3.)
Knee	Extension, Flexion	knee_(left, right)_flex	(2.)
	Rotation innen, außen	knee_(left, right)_rot	(3.)
Foot	Extension, Flexion	foot_(left, right)_flex	(2.)
	Eversion, Inversion	foot_(left, right)_vers	(1.)

(1.) Winkel mit Bezug zu einer Körperebene

In diesem Fall werden die Gelenkwinkel als Winkel zwischen einer Extremität und einer Körperebene berechnet [Doorenbosch et al., 2003]. Die Körperebenen sind diejenigen, die den Körper frontal, sagittal und transversal schneiden. Die Berechnung eines Gelenkwinkels erfolgt *zu* einer Ebene, jedoch wird bei der Angabe des Winkels von einer Bewegung *in* der Ebene gesprochen. In Abbildung 5.3) sind die drei Körperebenen und MoCap-Daten einer aufrecht stehenden Person dargestellt.

Bei einer aufrecht stehenden Person ist die Einzeichnung der Körperebenen noch intuitiv möglich. Wird eine Körperhaltung eingenommen, bei der z.B. der Rumpf oder die Gliedmaßen gekrümmt oder rotiert sind, ist es für die Berechnung von

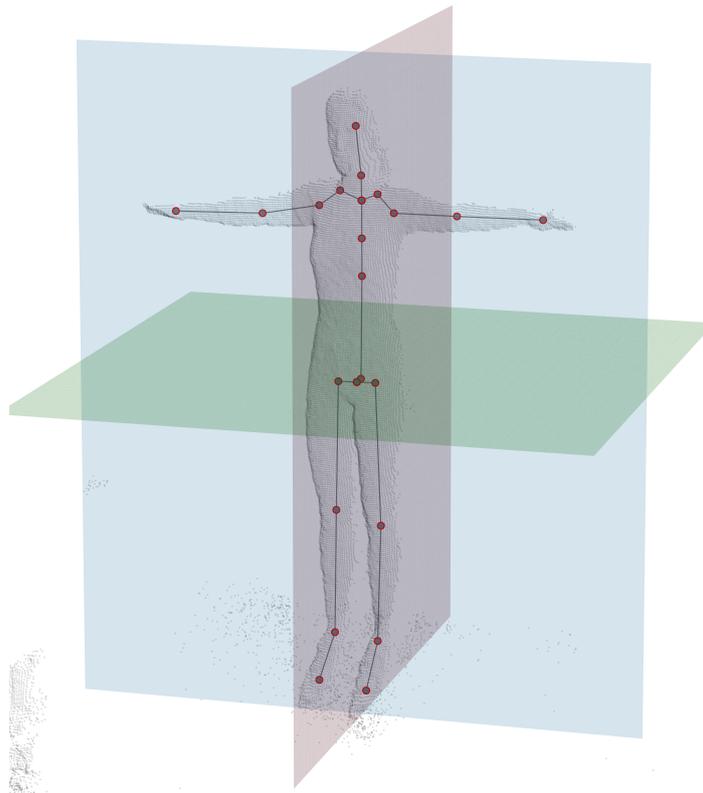


Abbildung 5.3.: Körperebenen und MoCap-Daten einer aufrecht stehenden Person; Sagittalebene (● rot), Frontalebene (● blau), Transversalebene (● grün)

Winkeln notwendig, die Körperebenen ebenfalls entsprechend zu kippen oder zu rotieren. Der Ansatz in dieser Arbeit ist es, für die jeweilige Gelenkwinkelberechnung die entsprechende Körperebene an die aktuelle Körperhaltung anzupassen. Dafür wird die Körperebene anhand der umliegenden Gelenkpunkte lokal ausgerichtet. Zum Beispiel wird für die Flexion der linken Schulter die Frontalebene an der Position (x, y, z) *shoulder_center* und *chest* orientiert und lokal in die Position *shoulder_left* gesetzt. Anschließend kann der Elevationswinkel zwischen der Position *elbow_left* und der Frontalebene berechnet werden. In diesem Beispiel wird der Winkel *zu* der Frontalebene berechnet, jedoch wird der Winkel als Bewegung *in* der Sagittalebene angegeben. In Abbildung 5.4a sind die lokal gesetzten Körperebenen für das genannte Beispiel dargestellt.

Zusammengefasst kann gesagt werden, dass für die Berechnung eines Gelenkwinkel-

kels zu einer Körperebene zwei Fixpunkte definiert werden müssen, um die jeweilige Körperebene zu setzen. Für die Berechnung wird anschließend eine Hilfsposition aus den zwei Fixpunkten gebildet. Diese kann durch eine lineare Verlängerung des Vektors zwischen den beiden Fixpunkten bestimmt werden. Anschließend können die Vektorlängen \vec{a} , \vec{b} und \vec{c} zwischen der gesuchten Gelenkposition a , der Hilfsposition b und der nächsten Gelenkposition in der Skelett-Hierarchie c auf Basis der Gleichung in 5.3 berechnet werden. Mit den berechneten Vektorlängen \vec{a} , \vec{b} und \vec{c} kann dann der geometrische Winkel α anhand von Gleichung 5.4 bestimmt werden.

$$v(P_1, P_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \quad (5.3)$$

$$\alpha = \arccos\left(\frac{\vec{a}^2 - \vec{b}^2 - \vec{c}^2}{-2 \cdot \vec{b} \cdot \vec{c}}\right) \quad (5.4)$$

In Tabelle 5.2 sind Gelenkpunkte und die Fixpunkte für die Orientierung zugeordneter Ebenen aufgelistet. Da bei den Händen und Füßen kein direkter Bezug zum Torso besteht und sich die Ebene entsprechend der Ausrichtung zum Körper ändert, wird in diesen Fällen keine Körperebene angegeben. Die Berechnung auf Basis der zugeordneten Fixpunkte ist jedoch vergleichbar.

Im nächsten Schritt wird die Bewegungsrichtung bestimmt. In dieser Arbeit wird die Bewegungsrichtung ausgehend von der Neutral-Null-Stellung entsprechend der Freiheitsgrade durch ein positives oder negatives Vorzeichen angegeben. Ein gerade nach vorne gerichteter Arm hätte z. B. einen Winkel von $+90^\circ$ (Flexion), ein nach hinten gerichteter Arm hätte -40° (Extension). Sind in einem Gelenk Bewegungen nach vorne, oben oder rechts möglich, werden diese mit einem positiven Vorzeichen angegeben. Bewegungen nach hinten, unten oder links werden mit einem negativen Vorzeichen markiert.

Mathematisch wird die Berechnung des Vorzeichens anhand des Kreuzprodukts aus Ebene und Gelenkpositionen gelöst. Das heißt, es wird betrachtet, ob sich die Gelenkposition c des nächsten Punktes in der Skelett-Hierarchie links oder rechts, vor oder hinter, über oder unter dem Vektor aus den beiden Fixpunkten a und b liegt. Mit Bezug auf das oben genannte Beispiel läge die Position des *elbow_left* bei einem gerade nach vorne gerichteten Arm ebenfalls *vor* der Frontalebene. Gleichung 5.5 zeigt die Berechnung des Kreuzprodukts. Dafür ist es notwendig, die zweidimensionalen Komponenten der Positionen einzusetzen, welche die Positionen

Tabelle 5.2.: Für die Berechnung von Gelenkwinkeln relevante Fixpunkte und Körperebenen

Gelenkwinkel	Berechnung zur Ebene	Winkel in Ebene	Fixpunkte für die Orientierung der Körperebene
neck_flex	frontal	sagittal	neck, shoulder_center
neck_latflex	sagittal	frontal	shoulder_left, shoulder_right
chest_flex	frontal	sagittal	chest, shoulder_center
chest_latflex	sagittal	frontal	shoulder_left, shoulder_right
lumbar_flex	frontal	sagittal	hip_center, spine_mid
lumbar_latflex	sagittal	frontal	hip_left, hip_right
shoulder_(left, right)_flex	frontal	sagittal	shoulder_center, chest
shoulder_(left, right)_elev	sagittal	frontal	shoulder_left, shoulder_right
shoulder_(left, right)_transrot	sagittal	transversal	shoulder_center, shoulder_(left, right)
shoulder_(left, right)_sagrot	frontal	sagittal	shoulder_center, chest
hand_(left, right)_flex			elbow_(left, right), wrist_(left, right)
hand_(left, right)_azim			elbow_(left, right), wrist_(left, right)
hip_(left, right)_flex	frontal	sagittal	hip_center, spine_mid
hip_(left, right)_elev	sagittal	frontal	hip_left, hip_right
foot_(left, right)_vers			knee_(left, right), ankle_(left, right)

des Winkels *in* der Ebene beschreiben. In Gleichung 5.6 ist die Zuordnung des Vorzeichens zu dem Gelenkwinkel α entsprechend des Ergebnisses aus der Berechnung des Kreuzprodukts d dargestellt.

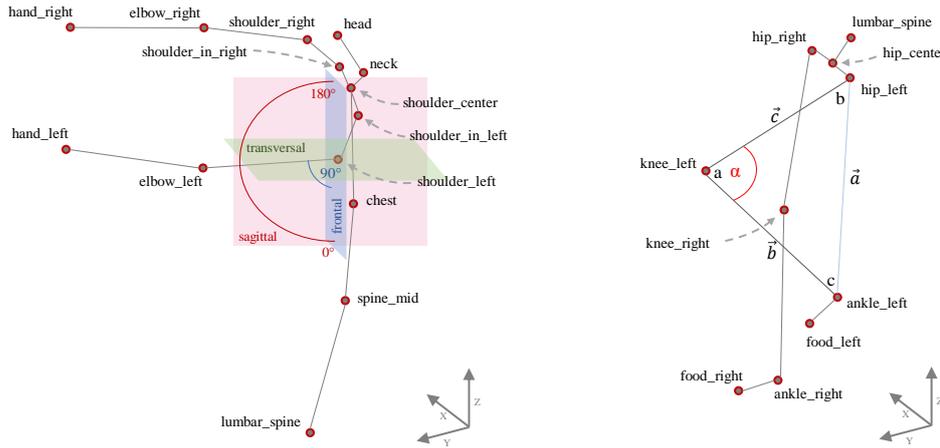
$$d = (b_1 - a_1) \times (c_2 - a_2) - (b_2 - a_2) \times (c_1 - a_1) \quad (5.5)$$

$$\alpha = \begin{cases} -\alpha, & \text{if } d < 0. \\ \alpha, & \text{otherwise.} \end{cases} \quad (5.6)$$

(2.) Winkel berechnet aus drei kartesischen Koordinaten

Die Gelenkwinkel *knee_(left, right)_flex* und *elbow_(left, right)_flex* können aus drei kartesischen Koordinaten, ohne Verwendung von Körperebenen, berechnet werden. In die Berechnung werden die Koordinaten des gesuchten Gelenks a und die Posi-

tionen der beiden benachbarten Gelenkpunkte b und c eingegeben. Für die Berechnung des Winkels $knee_left_flex$ müssten bspw. zunächst die Vektorlängen zwischen $hip_left - ankle_left$, $knee_left - ankle_left$ und $hip_left - knee_left$ oder \vec{a} , \vec{b} und \vec{c} mit Hilfe der bekannten Gleichung 5.3 berechnet werden. Anschließend kann ebenfalls anhand von Gleichung 5.4 der Winkel α in der Position a bestimmt werden. Eine Veranschaulichung dieses Beispiels ist in Abbildung 5.4b gezeigt.



(a) Bsp. für lokal gesetzte Körperebenen in $shoulder_left$; exemplarische Darstellung des Elevationswinkel zur Frontalebene und auf der Sagittalebene [Vox et al., 2021, Fig. 4b]

(b) Bsp. für die Berechnung der Flexion von $knee_left$ aus den kartesischen Koordinaten a , b , c von den Gelenkpunkten $knee_left$, hip_left , $ankle_left$ [Vox et al., 2021, Fig. 5]

Abbildung 5.4.: Berechnung von physiologischen Gelenkwinkeln aus MoCap-Daten mit Bezug zu Körperebenen und aus drei kartesischen Koordinaten

(3.) Rotationswinkel berechnet aus der relativen Drehung zwischen zwei Quaternionen

Rotationswinkel können durch die relative Drehung zwischen zwei verketteten Quaternionen berechnet werden [Duan et al., 2020]. Dafür ist es notwendig, das Quaternion q_1 des Start-Gelenkpunktes zu invertieren und mit dem nachfolgenden Quaternion q_2 zu multiplizieren (siehe Gleichung 5.7).

$$q_{12} = q_1^{-1} \otimes q_2 \quad (5.7)$$

Im nächsten Schritt muss das relative Quaternion q_{12} in Eulersche Winkel Yaw ,

Pitch, *Roll* transformiert werden. In Abbildung 5.5 ist die Verkettung der Gelenkpunkte *Elbow* - *Wrist* - *Hand* exemplarisch dargestellt. In diesem Fall beschreibt *Roll* den Winkel *elbow_rot* für die Rotation im Unterarm. Wie in Abschnitt 2.1 beschrieben, orientiert sich der Ursprung in einer Skelett-Hierarchie an der Lage des übergeordneten Quaternions. Die Ausgangslage des übergeordneten Quaternions ist in der Regel nicht gleich der Neutral-Null-Stellung der jeweiligen RoM. Aus diesem Grund kann es insbesondere an den äußersten Gelenkpunkten (Hände, Füße) zu unplausiblen Gelenkwinkeln außerhalb der RoM kommen. Im folgenden Abschnitt wird ein Verfahren für eine Offset-Korrektur bzw. das Setzen der Neutral-Null-Stellung beschrieben.

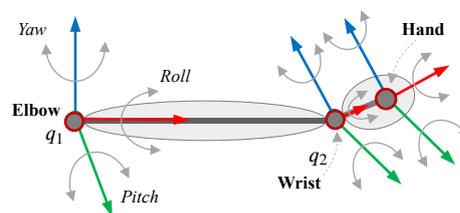


Abbildung 5.5.: Darstellung von verketteten Quaternionen und der Winkel *Yaw*, *Pitch*, *Roll* am Beispiel der Gelenkpunkte *Elbow* - *Wrist* - *Hand*

5.2.1. Kalibrierung und Setzen der Neutral-Null-Stellung

Ziel der Kalibrierung ist es, die Gelenkwinkelmessung ausgehend von der Neutral-Null-Stellung (siehe auch Abschnitt 2.5 *Klinischer Standard für die Gelenkwinkelmessung*) sicherzustellen. Prinzipiell wird bei der Berechnung von Gelenkwinkeln zu einer Körperebene bereits erreicht, dass der Wertebereich bei 0° startet, wenn die Gliedmaßen entsprechend parallel zu einer Körperebene liegen. Von diesem Fall kann jedoch in der Regel nicht ausgegangen werden. Beispielsweise bei einer krankhaften Kyphose [Rohlmann et al., 2001], einer verstärkten Krümmung der Brustwirbelsäule, muss bei einer Positionierung von einer Körperebene an Gelenkpunkten der Wirbelsäule mit einer Schiefelage der Ebene gerechnet werden. Ein ähnliches Paradigma zeigt sich, wie oben beschrieben, bei der Berechnung von Rotationswinkeln anhand von Quaternionen. Aus diesem Grund wird in dieser Arbeit ein Kalibrierungsschritt vor jeder Messung eingehalten.

Bei der Kalibrierung wird zu Beginn jeder Messung die jeweilige Person aufgefordert, für mindestens drei Sekunden die Neutral-Null-Stellung einzunehmen und sich in diesem Zeitraum nicht zu bewegen. Anschließend wird daraus ein Mittelwert (Offset) für jeden Gelenkwinkel berechnet. Der Offset wird dann von den nachkommenden Werten subtrahiert. Durch die Offset-Korrektur wird erreicht, dass die Gelenkwinkelmessung entsprechend des Wertebereichs in der RoM eingehalten wird.

Die Angabe von Wertebereichen in der RoM ist nicht für jeden Gelenkwinkel intuitiv interpretierbar und auch in der Literatur konnten verschiedene Auffassungen gefunden werden. Insbesondere bei den Gelenkwinkeln der Knie und Ellenbogen würde bei einer gestreckten Haltung, nach Neutral-Null-Methode, ein Gelenkwinkel von $\approx 0^\circ$ angegeben werden. Intuitiv wird dieser Winkel auch mit einem Wert von $\approx 180^\circ$ beschrieben. Diese Paradigmen sind in der Umsetzung aufgegriffen und können entsprechend angepasst werden. Für die Bewertung kritischer Gelenkwinkel anhand der genannten Bewertungsschemata wird jedoch einheitlich von einem Wertebereich entsprechend der RoM, beginnend in der Neutral-Null-Stellung, ausgegangen.

5.2.2. Offline- und Online-Filterung

Über die generelle Notwendigkeit der Anwendung eines Filters zur Reduktion von Rauschen oder Artefakten bei MoCap-Sensorik kann keine allgemeingültige Aussage getroffen werden. Bei der untersuchten Sensorik (RGB-Tiefenkamera, Lighthouse) konnten jedoch sowohl Positionssprünge zu unplausiblen Koordinaten als auch Rauschen festgestellt werden (siehe Abschnitte 5.3, 5.4). Für die Berechnung eines Gelenkwinkels werden mehrere Gelenkpositionen benötigt. Artefakte einer einzigen Position wirken sich unmittelbar auf die Berechnung des Winkels aus. Für die Behandlung von Positionssprüngen wird im folgenden Abschnitt eine Plausibilitätsprüfung vorgestellt, die auf Basis von Gelenkwinkeln unerreichbare Gelenkstellungen identifiziert.

Für die Bereinigung von Rauschen in Offline-Daten wurde in dieser Arbeit ein Tiefpassfilter mit einer Cutoff-Frequenz von 3 Hz angewendet. Die Filterung von Daten in einem Sensorstream in Echtzeit ist durch einen Moving Average Filter mit einer Fenstergröße von 3 Frames umgesetzt. Nachteil eines Moving Average Filters gegenüber einer Tiefpassfilterung ist, dass mit einer leicht reduzierten Maximalamplitude gerechnet werden muss ($\approx -1.5^\circ$ bei 30 Hz und einer Fenstergröße von

3 *Frames*). Vorteil ist jedoch eine hohe Robustheit bei Aussetzern und eine geringe Rechenlast.

In dieser Arbeit werden hauptsächlich Bewegungen mit niedrigen Bewegungsgeschwindigkeiten analysiert, bei denen ein Tiefpassfilter bei 3 *Hz* optimale Ergebnisse zeigt. Wird das System für die Analyse von sehr schnellen Bewegungen, wie z.B. für die Analyse eines Tremors [McAuley et al., 1997] angewendet, ist eine Filterung entsprechend abzuwägen.

5.2.3. Plausibilitätsprüfung anhand der RoM

Artefakte und Positionssprünge in MoCap-Daten können zu unplausiblen Gelenkwinkelangaben führen. Aus diesem Grund wird in dieser Arbeit eine Plausibilitätsprüfung auf Basis des physiologischen Bewegungsausmaßes angewendet. In Tabelle 2.3 wird die RoM für untersuchte Gelenkwinkel angegeben. Nach Berechnung des jeweiligen Winkels wird überprüft, ob ein physiologisch plausibler Winkel innerhalb der RoM vorliegt. Ebenfalls wird überprüft, ob sich der Wert in einen plausiblen Verlauf zu den drei davorliegenden Frames einreicht. Entsprechend unplausible Werte werden für die weitere Verarbeitung markiert und interpoliert oder von der Bewertung ausgeschlossen.

5.2.4. Überprüfung der Berechnungsmethode

Wie auch in [Vox et al., 2021] beschrieben, wurde die entwickelte Methode für die Berechnung von physiologischen Gelenkwinkeln mit WIDAAN verglichen. WIDAAN wurde für die Bewertung von Arbeitsplätzen nach ergonomischen Kriterien entwickelt. Um vergleichbare Ergebnisse unabhängig von individuellen Körpergrößen zu erhalten, werden die Daten auf ein Körpermodell, das sogenannte "Dortmunder" [Jäger et al., 2001], abgebildet. Bei der Methode in dieser Arbeit bleibt die Anthropometrie oder die individuelle Körpergröße der jeweiligen Person erhalten. Die Berechnungsgrundlage von WIDAAN ist nicht veröffentlicht. Trotz methodischer Unterschiede ist ein vergleichbarer Trend zwischen den Berechnungsverfahren erwartbar. Aus diesem Grund wurden Berechnungsergebnisse auf Plausibilität mit WIDAAN verglichen. Die Winkel [*neck_flex*, *neck_latflex*, *shoulder_(left, right)_elev*, *elbow_(left, right)_flex*, *elbow_(left, right)_azim*, *hip_(left, right)_flex*, *knee_(left, right)_flex*] zeigten eine plausible Übereinstimmung zwischen WIDAAN und der eigenen Berechnungsmethode [Vox et al., 2021].

Die Plausibilität von berechneten Gelenkwinkeln wurde ebenfalls in Vergleichen zu einer Messung mit dem Goniometer überprüft. Eine Evaluierung der Genauigkeit bei der Gelenkwinkelmessung, bei Anwendung der Berechnungsmethode mit Daten der Kinect v2 und einem HTC Vive System, ist in den folgenden Kapiteln untersucht.

5.3. Untersuchung des Sensorrauschens bei der Gelenkwinkelmessung mit der Kinect v2 in statischen Haltungen

In einer Vielzahl von Evaluierungsstudien von RGB-Tiefenkamera Systemen wurde die Genauigkeit von Gelenkpositionen und -winkeln in dynamischen Bewegungen untersucht. Ergänzend zu den Ergebnissen soll in diesem Experiment die statische Abweichung bzw. das Sensorrauschen bei der Gelenkwinkelmessung mit der Kinect v2 + Kinect SDK identifiziert werden. Ziel ist es, das System auf systematische Fehler zu untersuchen und den RMSE zu ermitteln. In diesem Kapitel wird eine Vergleichsmessung zwischen der Kinect v2 und dem optischen und markerbasierten System Qualisys Oqus bei statischen Haltungen vorgestellt. Teile der Ergebnisse sind bereits in [Vox and Wallhoff, 2016] publiziert.

5.3.1. Beschreibung der Sensorik

Kinect v2 + Kinect SDK

Die Kinect v2 ist, wie in Abschnitt 2.2 beschrieben, ein RGB-Tiefenkamera System. Für das Skeleton Tracking wurde in dieser Studie das Kinect SDK in der Version 2.0-1409 eingesetzt. Die Aufzeichnung der MoCap-Daten erfolgte mit einer Aufnahme rate von 30 *Hz*.

Qualisys Oqus

Als Referenzsystem wurde das optische und markerbasierte System Qualisys Oqus 300 Serie (Qualisys AB, Göteborg, Schweden) eingesetzt. Das System besteht aus neun Kameras mit einer Auflösung von 1280×1024 Pixeln bei einer Aufnahme rate von bis zu 1000 *Hz*. Das Messprinzip basiert auf emittierten Infrarot-Lichtimpulsen, welche von optischen Markern reflektiert und von den Kameras erfasst werden. Laut

Herstellerangaben liegt der Positionsfehler von einzelnen Markern bei $< 1 \text{ mm}$. Die resultierten Gelenkwinkel mit dem Qualisys System können als Grundwahrheit angenommen werden [Schmitz et al., 2014, van der Kruk and Reijnen, 2018]. In diesem Experiment wurden die MoCap-Daten mit dem Qualisys System synchron zu der Kinect v2 mit 30 Hz aufgezeichnet.

5.3.2. Experimenteller Aufbau

In dem Experiment wurde eine Plastikfigur mit menschlichen Proportionen ca. 2 m vor der Kinect v2 platziert und neun Kameras des Qualisys Oqus Systems kreisförmig um das Körpermodell herum gestellt. Die Aufnahme der MoCap-Daten ist mit beiden Systemen synchron erfolgt. Der experimentelle Aufbau ist in Abbildung 5.6 dargestellt.

Um die Gelenkmittelpunkte mit dem markerbasierten Qualisys System korrekt zu identifizieren, ist es notwendig, zwei Marker pro Gelenk anatomisch korrekt zu platzieren, sodass die Mittelachse zwischen den Markerpositionen das Zentrum des jeweiligen Gelenkes durchläuft. Aus diesem Grund wurde die Positionierung der Marker an dem Körpermodell durch eine Physiotherapeutin unterstützt.

Die Messungen wurden unter verschiedenen Bedingungen durchgeführt. Zum einen wurde der Einfluss von Tageslicht untersucht und eine Messung mit rein künstlichem und anteiligem Tageslicht durchgeführt. Zum anderen wurden Messungen mit und ohne Kleidung durchgeführt. Bei der Messung mit Kleidung wurden Löcher entsprechend der Positionen von optischen Markern in die Kleidung geschnitten.

5.3.3. Methodik

Verglichen wurden die Systeme auf Basis von berechneten Gelenkwinkeln anhand der in Abschnitt 5.2 beschriebenen Berechnungsmethode. Für den Systemvergleich wurden die Winkel [*shoulder_(left, right)_elev*, *elbow_(left, right)_flex*, *hip_(left, right)_flex*, *knee_(left, right)_flex*] ausgewählt und die Zeitreihen der Gelenkwinkelmessungen zwischen den beiden Systemen verglichen (siehe Abbildung 5.7). Für die Analyse wurde der Offset beider Signale subtrahiert, um die Differenzen ausgehend von 0° darzustellen. Ein Vergleich zwischen den Systemen wurde auf Basis des RMSE sowie Minimum (*min*), Maximum (*max*), Mittelwert (*mean*) und der Standardabweichung (*std*) durchgeführt.

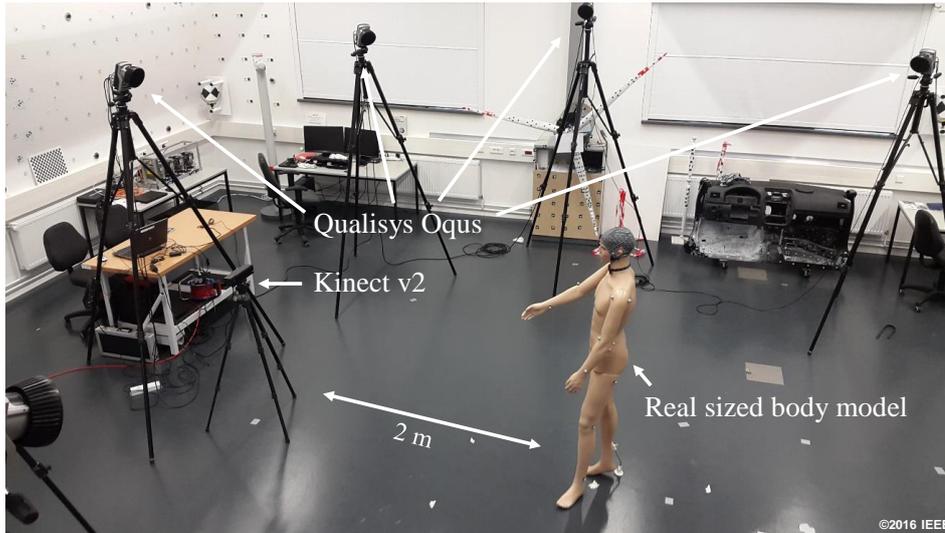
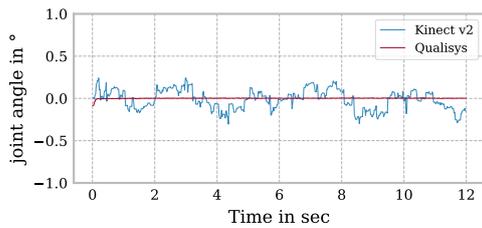
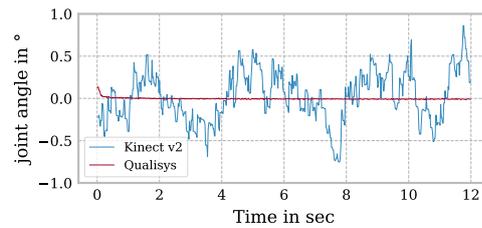


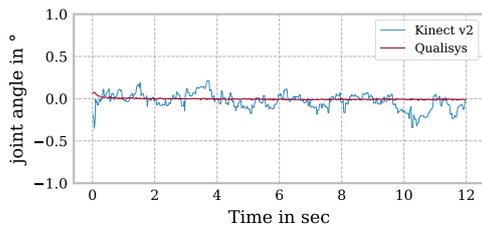
Abbildung 5.6.: Experimenteller Aufbau der Kinect v2 vor einer real großen menschlichen Plastikfigur und neun umringenden Kameras des Qualisys Oqus Systems [Vox and Wallhoff, 2016, Fig. 2]



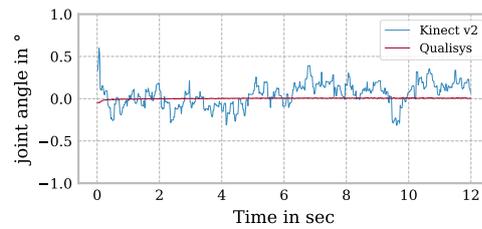
(a) shoulder_right_elev



(b) elbow_right_flex



(c) hip_right_flex



(d) knee_right_flex

Abbildung 5.7.: Vergleich zwischen der Kinect v2 und dem Qualisys als Referenzsystem auf Basis berechneter Gelenkwinkel in einer statischen Haltung [Vox and Wallhoff, 2016, Fig. 5]

5.3.4. Ergebnisse

Im Vergleich der Systeme konnte ein RMSE von [*shoulder_right_elev* 0.014°, *elbow_right_flex* 0.089°, *hip_right_flex* 0.01°, *knee_right_flex* 0.022°] für die einzelnen untersuchten Winkel und ein RMSE von 0.034° über alle Winkel und die gesamte Messung ermittelt werden. Für die Winkel der linken Körperhälfte können vergleichbare Werte angenommen werden. In Tabelle 5.3 sind die statistischen Parameter aus der Vergleichsmessung zwischen den Systemen aufgelistet. Bei der Kinect v2 zeigt sich ein maximales Rauschen mit einer Abweichung von 0.861°. Das ermittelte Rauschen bei dem Qualisys System kann bereits auf eine Fehlerfortpflanzung durch Rundung der Gelenkpositionen bei der Berechnung der Gelenkwinkel zurückgeführt werden.

Tabelle 5.3.: Statistische Auswertung der Vergleichsmessung zwischen Kinect v2 und dem Qualisys auf Basis berechneter Gelenkwinkel; *min*, *max*, *mean* und *std* in [°]

	shoulder_ right_elev	elbow_ right_flex	hip_ right_flex	knee_ right_flex
Qualisys				
<i>min</i>	-0.085	-0.017	-0.02	-0.048
<i>max</i>	0.007	0.136	0.076	0.013
<i>mean</i>	0.001	-0.004	0.00	0.003
<i>std</i>	0.007	0.013	0.011	0.008
Kinect v2				
<i>min</i>	-0.305	-0.752	-0.350	-0.316
<i>max</i>	0.242	0.861	0.217	0.598
<i>mean</i>	-0.017	0.028	-0.031	0.05
<i>std</i>	0.116	0.295	0.095	0.143

Des Weiteren konnten bei plötzlichen Veränderungen der Lichtverhältnisse sowie zu den Startzeitpunkten der Messungen kurzfristig höhere und unplausible Werte bei der Kinect v2 festgestellt werden. Eine Anpassung der Gelenkpositionen in einen plausiblen Bereich ist nach ca. 1–2 *Sekunden* nach dem Ereignis erfolgt. Unterschiede zwischen einer bekleideten oder unbekleideten menschlichen Plastikfigur konnten unter den statischen Bedingungen nicht festgestellt werden.

5.4. Evaluierung der Sensorgenauigkeit des HTC Vive Tracking Systems

In dieser Untersuchung wird das *HTC Vive* Lighthouse Tracking System in Kombination mit *Final IK* mit dem Qualisys System als Referenz verglichen. Eine Untersuchung der Sensorik unter statischen und dynamischen Bedingungen auf Basis der räumlichen Genauigkeit ist, wie in Abschnitt 3.1.2 beschrieben, bereits mehrfach erfolgt. Diese Studie konzentriert sich insbesondere auf eine Genauigkeit bei dynamischen menschlichen Bewegungen. Wie in Kapitel 1 *Einleitung* beschrieben, soll ein Lighthouse Tracking System primär für die ergonomische Bewertung von Arbeitsplätzen in VR eingesetzt und untersucht werden. Aus diesem Grund wird das System anhand von sechzehn Gelenkwinkeln evaluiert, welche hauptsächlich in den Bewertungsschemata RULA und REBA eingesetzt werden. In vergleichbaren Studien wurden, auf Basis der Positionen, für das HMD höchste Genauigkeiten erhoben [Ikbal et al., 2021]. Um insbesondere das HMD zu analysieren, werden in dieser Untersuchung, neben den bereits eingeführten Winkeln, auch die Winkel *head_inc* und *head_latinc* analysiert. Hiermit werden Bewegungen am Übergang zwischen Hals und Kopf dargestellt. Ziel ist es, die Genauigkeit bei der Gelenkwinkelmessung zu identifizieren, um eine Einschätzung treffen zu können, ob das System für eine Bewegungsbewertung eingesetzt werden kann.

In den Grundlagen zur Funktionsweise von Lighthouse Tracking Systemen (Abschnitt 2.3) wird erklärt, dass in die IK eine variable Anzahl von Tracking-Positionen eingegeben werden kann, um MoCap-Daten zu berechnen. Für die Minimalkonfiguration reichen die Tracking-Positionen der Hände (Controller) und des Kopfes (HMD). Maximal kann das HTC Vive System mit bis zu neun Trackern, mit zusätzlichen Tracking-Positionen an den äußeren Gliedmaßen und am Torso, betrieben werden. In dieser Studie wird auch der Einfluss der Anzahl verwendeter Tracker auf die Genauigkeit untersucht. Insgesamt werden zwei Forschungsfragen behandelt:

1. Wie genau ist das HTC Vive System in Kombination mit Final IK im Vergleich zu dem optischen und markerbasierten System von Qualisys für den Einsatz der Gelenkwinkelmessung?
2. Wie wirkt sich die Anzahl der Vive-Tracker auf die Genauigkeit aus?

Um diese Fragen zu beantworten, wurde eine technische Vergleichsstudie zwischen dem HTC Vive + Final IK (im Folgenden "Vive") und dem Qualisys Sys-

tem durchgeführt. Dabei wurden 20 dynamische Bewegungen mit jeweils drei Wiederholungen analysiert, welche die RoM der untersuchten Gelenkwinkel abdecken. Zur Untersuchung der Auswirkungen von verschiedenen Tracker-Konfigurationen (Anzahl) wurden die MoCap-Daten mit Final IK unter Berücksichtigung der unterschiedlichen Konfigurationen berechnet und abgespeichert. Die grundsätzlichen Verarbeitungsschritte für die Normalisierung der MoCap-Daten und Ableitung von Gelenkwinkeln sind entsprechend der Beschreibungen in den Abschnitten 5.1 und 5.2 durchgeführt. Im Kern werden Abweichungen zum Referenzsystem berechnet und mit entsprechenden statistischen Verfahren im Zeitbereich ausgewertet. Teile der Ergebnisse sind bereits in [Vox et al., 2021] veröffentlicht.

5.4.1. Beschreibung der Sensorik

HTC Vive + Final IK

In dieser Studie wird das Lighthouse Tracking System HTC Vive in Kombination mit Final IK in der Version 1.8 verwendet. Zur Visualisierung des Körpers der Teilnehmer*innen in einer virtuellen Umgebung wird die Game Engine *Unity* (Version 2018.3.14f1) eingesetzt und mit dem Vive System verbunden. Über das HMD werden die eigenen Bewegungen in Form eines menschlichen Avatars dargestellt. Um das Maß für die Größe des Avatars in der virtuellen Realität entsprechend zuzuordnen und die MoCap-Daten zu bestimmen, werden die anthropometrischen Daten nach [Jürgens et al., 1998] erhoben und als statische Variablen in die Berechnung in Final IK eingegeben. Das System ist in der Lage, MoCap-Daten mit einer Aufnahmezeit von 90 Hz zu erfassen und aufzuzeichnen.

Qualisys Oqus

Als Referenzsystem wird ebenfalls das Qualisys Oqus eingesetzt. Das System wird mit acht Kameras, mit einer Aufnahmezeit von 100 Hz unter Verwendung eines 48-Marker-Modells (Qualisys Animation Marker Set), betrieben. Für die Nachbearbeitung der Daten wird die Qualisys Software (QTM) eingesetzt, um verdeckte Markerpositionen (Lücken) zu interpolieren und die MoCap-Daten abzuleiten.

5.4.2. Experimenteller Aufbau und Datenerhebung

Teilnehmer*innen

An der Studie¹ nahmen 26 junge, gesunde Erwachsene (10 Frauen, 16 Männer) mit informierter Zustimmung teil. Alle Testpersonen hatten ein uneingeschränktes oder korrigiertes Sehvermögen und waren frei von muskuloskelettalen Beeinträchtigungen, die zu Bewegungseinschränkungen führen könnten. Die Altersspanne reichte von 21 bis 27 Jahren (Durchschnittsalter: 22.80 ± 1.54 ; Durchschnittsgröße: 1.75 ± 0.07 m; Durchschnittsgewicht: 74.92 ± 12.28 kg).

Bei vier der 26 gemessenen Testpersonen wurden nicht alle Datensätze für die untersuchten Tracker-Konfigurationen aufgezeichnet und gespeichert, sodass diese Personen ausgeschlossen wurden. Zwei weitere wurden ausgeschlossen, da einige Trackerdaten kontinuierlich unplausible Werte zeigten (z.B. Ellenbogen auf dem Rücken). Insgesamt wurden 20 Teilnehmer*innen in die Vergleichsanalyse einbezogen und MoCap-Daten mit einer Gesamtlänge von 144 Minuten analysiert.

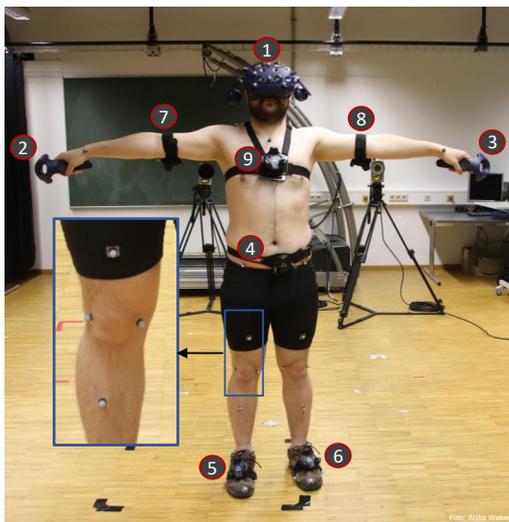
Messaufbau

In dem Messaufbau wurden vier Vive-Basisstationen in einem 5×5 m Viereck, in einer Höhe von 2.5 m, mit Blickrichtung in das Zentrum des Messbereiches aufgestellt. Um diesen Bereich wurden die Kameras des Qualisys Systems, ebenfalls mit Blickrichtung in den Messbereich, positioniert. Für die Messung mit dem Qualisys System wurden die retroreflektierenden Marker an anatomischen Orientierungspunkten direkt auf der Haut der Teilnehmer*innen befestigt. Für die Messung mit dem Vive System wurde das HMD am Kopf und sechs weitere Vive-Tracker mit Gurten an den Oberarmen, Brust, Hüfte und Füßen befestigt. Zusätzlich wurden zwei Vive-Controller in den Händen gehalten. Insgesamt wurden neun Positionen mit dem Vive System erfasst.

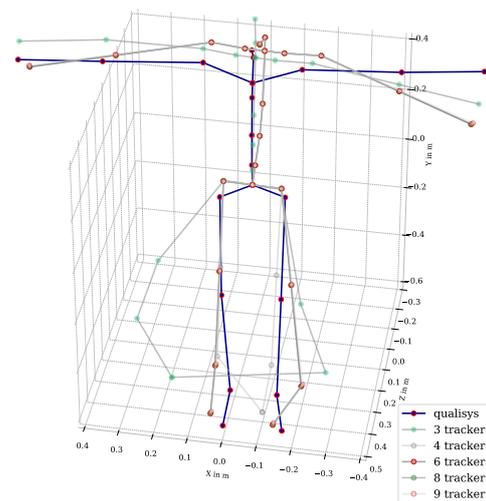
Die Berechnung der MoCap-Daten durch die IK wurde in fünf verschiedenen Tracker-Konfigurationen durchgeführt. In der Konfiguration **1–3** wurden nur die Positionen der Hände (Controller) und des Kopfes (HMD) eingegeben. In der Konfiguration **1–4** wurde die Eingabe um die Position der Hüfte erweitert. Bei der Konfiguration

¹Genehmigt durch die Ethikkommission der Hochschule Koblenz auf Grundlage der Positivliste (Stand 01/2018). Datenerhebung an der Hochschule Koblenz in der Abteilung Mathematik und Technik.

1–6 wurden die Positionen der FüÙe hinzugefügt und eine Erfassung der unteren Gliedmaßen ermöglicht. In der Konfiguration 1–8 fließen die Tracker an den Oberarmen mit in die Berechnung ein. Bei 1–9 wird die Position der Brust ergänzt. Somit wurden mit dem Vive System fünf Datensätze aus 3,4,6,8 und 9 Tracking-Positionen erzeugt, welche jeweils mit dem Qualisys verglichen wurden. In Abbildung 5.8a ist ein vollinstrumentierter Teilnehmer mit den auf der Haut platzierten retroreflektierenden Markern und den befestigten Vive-Trackern dargestellt. Die Positionen der Vive-Tracker sind entsprechend der beschriebenen Konfigurationen durchnummeriert. Ein Vergleich zwischen dem Qualisys System und den untersuchten Vive-Tracker-Konfigurationen, auf Basis eines einzelnen Frames, ist in Abbildung 5.8b gezeigt.



(a) Vollinstrumentierte Person in T-Position mit HMD, zwei Controllern und sechs Vive-Trackern sowie aufgeklebten retroreflektierenden Markern mit Zoom auf dem rechten Knie; Nummerierung entsprechend der untersuchten Konfigurationen [Vox et al., 2021, Fig. 3a]



(b) Vergleich der MoCap-Daten zwischen Qualisys und Vive in den Tracker-Konfigurationen 3,4,6,8 und 9; jeweils dargestellt als einzelner Frame [Vox et al., 2021, Fig. 3b]

Abbildung 5.8.: Experimenteller Aufbau für den technischen Vergleich zwischen der MoCap-Sensorik von Qualisys und Vive

Ablauf

Über das HMD konnten die Proband*innen sowohl die eigenen auf einen Avatar projizierten Bewegungen wahrnehmen als auch der Versuchsanleitung folgen. Dafür wurde ein zweiter Avatar eingeblendet, um die spezifischen Bewegungen für die Teilnehmer*innen zu demonstrieren. Zu Beginn wurde eine Synchronisationsbewegung durchgeführt: Die Versuchsperson stand in einer T-Haltung, beugte die Arme 90° und streckte sie dann wieder mit drei Wiederholungen. Dasselbe Prozedere wurde für die 20 untersuchten Bewegungen durchgeführt. Vor jeder neuen Bewegung und zwischen den drei Wiederholungen stand die Versuchsperson kurz in der Neutral-Null-Stellung. Jede Bewegung musste bei normaler Geschwindigkeit mit dem größtmöglichen Bewegungsumfang (RoM) ausgeführt und kurz in der Endposition gehalten werden. Jedem Gelenk sind bestimmte Bewegungen zugeordnet, um die Gelenkwinkel in der jeweiligen RoM zu untersuchen. In Tabelle 5.4 sind die untersuchten Gelenkwinkel und eine Beschreibung der Bewegung aufgelistet.

Tabelle 5.4.: Beschreibung der Bewegung und zugeordneter Gelenkwinkel

Gelenkwinkel	Beschreibung der Bewegung
head_inc, neck_flex	Neigung des Kopfes nach vorne und hinten
head_latinc, neck_latflex	seitliche Neigung des Kopfes
chest_flex, lumbar_flex	Beugung des Oberkörpers nach vorne (gerader Rücken)
chest_flex, lumbar_flex	Beugung des Oberkörpers nach vorne (gekrümmter Rücken)
shoulder_(left, right)_elev, shoulder_(left, right)_rot	Heben beider Arme seitlich und nach vorne, hinten
elbow_(left, right)_flex, shoulder_(left, right)_rot	Ellenbogenbeugung (Curls)
knee_left_flex	Kniebeugung: Stehen auf dem rechten Bein, linker Unterschenkel in der Luft und nach hinten gebeugt in einem 90° Winkel
knee_right_flex	Kniebeugung: Stehen auf dem linken Bein, rechter Unterschenkel in der Luft und nach hinten gebeugt in einem 90° Winkel
hip_left_flex	Ausfallschritt, linkes Bein vorne, Kniebeugung 90°, Ellenbogenbeugung 90°, Rückentorsion nach rechts, links
hip_right_flex	Ausfallschritt, rechtes Bein vorne, Kniebeugung 90°, Ellenbogenbeugung 90°, Rückentorsion nach links, rechts

5.4.3. Datenverarbeitung

Die MoCap-Daten beider Systeme wurden zunächst mit der Software WIDAAN zeitlich synchronisiert und manuell annotiert. Dabei wurde der Start- und Endzeitpunkt der untersuchten Bewegungen und zwischen den Wiederholungen manuell in den Zeitreihendaten markiert. Im nächsten Schritt wurde Resampling angewendet, um die Daten auf eine einheitliche Framerate zu bringen. Kurze zeitliche Unterbrechungen (Datenaussetzer während der Messung) in den Vive-Daten wurden interpoliert. Anschließend konnten Daten beider Systeme, wie in Abschnitt 5.1 beschrieben, in ein einheitliches Koordinatensystem transformiert werden. Nachfolgend wurden die Gelenkwinkel nach den in Abschnitt 5.2 beschriebenen Verfahren berechnet. In einem weiteren Schritt wurden systematische Abweichungen korrigiert [Vox et al., 2021, S. 12]. Insgesamt waren folgende Schritte notwendig, um die Daten für den Vergleich vorzubereiten:

- Manuelle Annotation
- Synchronisation, Resampling, Filterung und Interpolation
- Normalisierung und Transformation
- Systematische Korrektur
- Berechnung Gelenkwinkel

5.4.4. Methodik

Der Vergleich zwischen den Systemen konzentriert sich auf die berechneten Abweichungen zwischen dem untersuchten Vive System und dem Qualisys System als Referenz. Für jeden untersuchten Gelenkwinkel und jede untersuchte Vive Tracker-Konfiguration wurde die Differenz zum Qualisys ausgewertet. In Phasen, in denen ein untersuchtes Gelenk nicht bewegt wird und in der Neutral-Null-Stellung verbleibt, kann ein kleinstmöglicher Fehler angenommen werden, was zu einer Verbesserung der Fehlerstatistik führen würde. Aus diesem Grund wurden in der Auswertung nur die zeitlichen Segmente ausgewählt, in denen das jeweilige Gelenk bewegt wurde. Die statischen Sequenzen wurden ausgeschnitten. Die so ausgewählten Segmente von allen Proband*innen wurden konkateniert und insgesamt mit den folgenden Methoden ausgewertet.

Concordance Correlation Coefficient

Ein Vergleich zwischen allen Tracker-Konfigurationen und allen untersuchten Gelenkwinkeln wurde zunächst auf Basis des Concordance Correlation Coefficient (CCC) nach [Lin, 1989] durchgeführt. Der CCC kann als eine Erweiterung der Pearson Korrelation [Benesty et al., 2009] angesehen werden, bei dem zusätzlich zur Korrelation eine Orts- und Zeitverschiebung berücksichtigt wird. Die grundlegende Berechnung des CCC ist in Gleichung 5.8 dargestellt, wobei σ_{qv} die Kovarianz, μ_q, μ_v die Mittelwerte und σ_q, σ_v die Standardabweichungen von q Qualisys und v Vive beschreibt.

$$ccc = \frac{2\sigma_{qv}}{(\mu_q - \mu_v)^2 + \sigma_q^2 + \sigma_v^2} \quad (5.8)$$

Nach [McBride, 2005] werden Korrelationen < 0.90 als schlecht, > 0.90 bis < 0.95 als mäßig, > 0.95 bis < 0.99 als substantiell und > 0.99 als nahezu perfekt bewertet (siehe Abbildung 5.9).

Bland-Altman-Plot

Eine gängige Analysemethode zum Vergleich von zwei Signalen oder Messmethoden ist der Bland-Altman-Plot [Bland et al., 1986]. In einem Bland-Altman-Plot werden auf der X-Achse der Durchschnitt und auf der Y-Achse die Differenz aus Stichproben beider Systeme abgebildet. Des Weiteren werden die Limits of Agreement (LoA) ($\mu \pm 1,96 \times \sigma$) auf Basis des Mittelwerts (Mean) μ und der Standardabweichung (SD) σ eingezeichnet. Die LoA beschreiben einen Bereich von häufig aufgetretenen Abweichungen.

In dieser Auswertung wurde der Bland-Altman-Plot modifiziert und auf der X-Achse die Werte des Referenzsystems (Qualisys) dargestellt. Das birgt den Vorteil, dass die Abweichungen zwischen Qualisys und Vive entsprechend der RoM abgebildet werden. Somit zeigen die Werte auf der X-Achse den tatsächlichen Gelenkwinkel, gemessen mit dem Qualisys, und auf der Y-Achse wird die entsprechende Abweichung des Vive Systems abgebildet (Abbildungen 5.10).

Violin-Plot

In dieser Auswertung werden die Gelenkwinkel-Differenzen auch als Violin-Plot dargestellt. Vorteil ist, dass die Häufigkeit von entsprechenden Abweichungen in

der RoM identifiziert werden kann. Für den Vergleich wurden zwei Arten eines Violin-Plots angewendet. Erstens werden die beiden Systeme auf Basis der Häufigkeit erreichter Gelenkwinkel in der RoM, anhand von zwei gegenüber liegenden Violinen, dargestellt (Abbildung 5.11a). Im Idealfall ist die Form der Vive-Violine auf der linken Seite gleich der Form der Qualisys-Violine auf der rechten Seite.

In der zweiten Darstellung wird mithilfe der angezeigten Violin-Form die Häufigkeit von auftretenden Abweichungen zwischen den beiden Systemen demonstriert (Abbildung 5.11a). Diese Darstellung gibt Aufschluss darüber, in welchen Wertebereichen sich die Unterschiede häuften und welche Unterschiede seltener auftreten. Zur besseren Herausarbeitung des Wertebereichs, in dem die Unterschiede variierten, und zur Eliminierung von Ausreißern wurden Perzentile zwischen 5 % und 95 % berechnet. Die Genauigkeit kann somit durch den Inter-Percentile-Range (IPR) des 5 %- und 95 %-Perzentils beschrieben werden. Die Orientierung des Medianwertes zeigt an, ob ein Winkel eher über- oder unterschätzt wird. Ein Medianwert bei 0° würde bedeuten, dass die Differenz gleichermaßen überschätzt und unterschätzt wird. Oder er deutet darauf hin, dass die Signale eher übereinstimmen.

5.4.5. Ergebnisse

In diesem Abschnitt werden die Ergebnisse des Vergleichs zwischen Vive und dem Referenzsystem Qualisys vorgestellt. Eine Gesamtübersicht der sechzehn analysierten Gelenkwinkel mit 25 %-, 50 %- und 75 %-Quartilen, dem Mittelwert, der Standardabweichung, dem Minimum und dem Maximum ist in Tabelle A.1 im Anhang dargestellt. Im Folgenden werden die Ergebnisse auf Grundlage des CCC, der Bland-Altman-Plots und der Violin-Plots dargestellt.

Vergleich auf Grundlage des CCC

Ein Vergleich zwischen den fünf Vive-Tracker-Konfigurationen und Qualisys auf Basis des CCC ist für alle Gelenkwinkel in Abbildung 5.9 dargestellt. Korrelationen größer als 0.85 sind farblich herausgestellt.

Nach [McBride, 2005] erreichen *head_inc*, *head_latinc* und *neck_flex* eine substanzielle Korrelation in der 3-Tracker-Konfiguration und eine schlechte Korrelation mit vier oder mehr Trackern. Die *shoulder_(left, right)_elev*-Winkel erreichen eine substanzielle Korrelation für alle Tracker-Konfigurationen. Die *elbow_(left, right)_flex*-Winkel erreichen eine mäßige bis schlechte Korrelation für alle Konfigurationen.

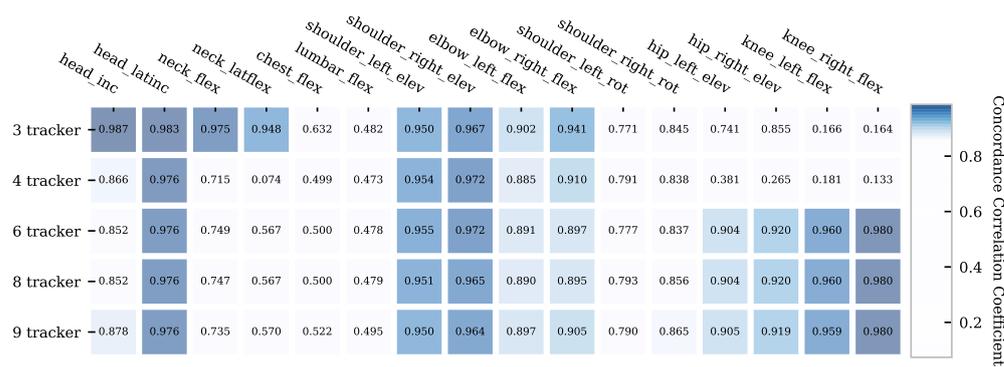


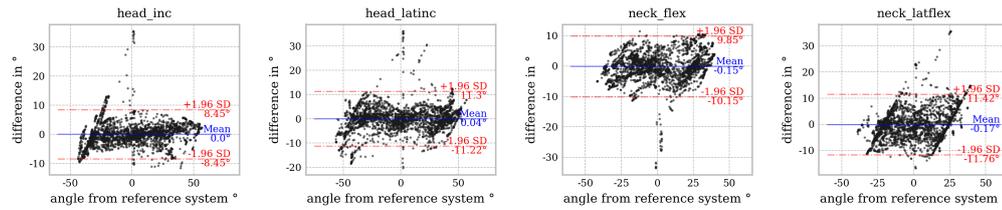
Abbildung 5.9.: Concordance Correlation Coefficient für alle untersuchten Gelenkwinkel und Tracker-Konfigurationen [Vox et al., 2021, Fig. 11]

Darüber hinaus ist bei den *hip*- und *knee*-Winkeln für die Tracker-Konfigurationen sechs und höher eine mäßige bis substanzielle Korrelation festzustellen. Bei allen Tracker-Konfigurationen zeigen die *chest_flex*, *lumbar_flex*, *shoulder_rot* und *elbow_azim*-Winkel schlechte Korrelationen. Der *hip_elev*-Winkel erreichte eine mäßige Korrelation für die 6-Tracker-Konfigurationen und höher. Für *knee_flex* ist eine substanzielle Korrelation ab der 6-Tracker-Konfiguration und höher zu erkennen.

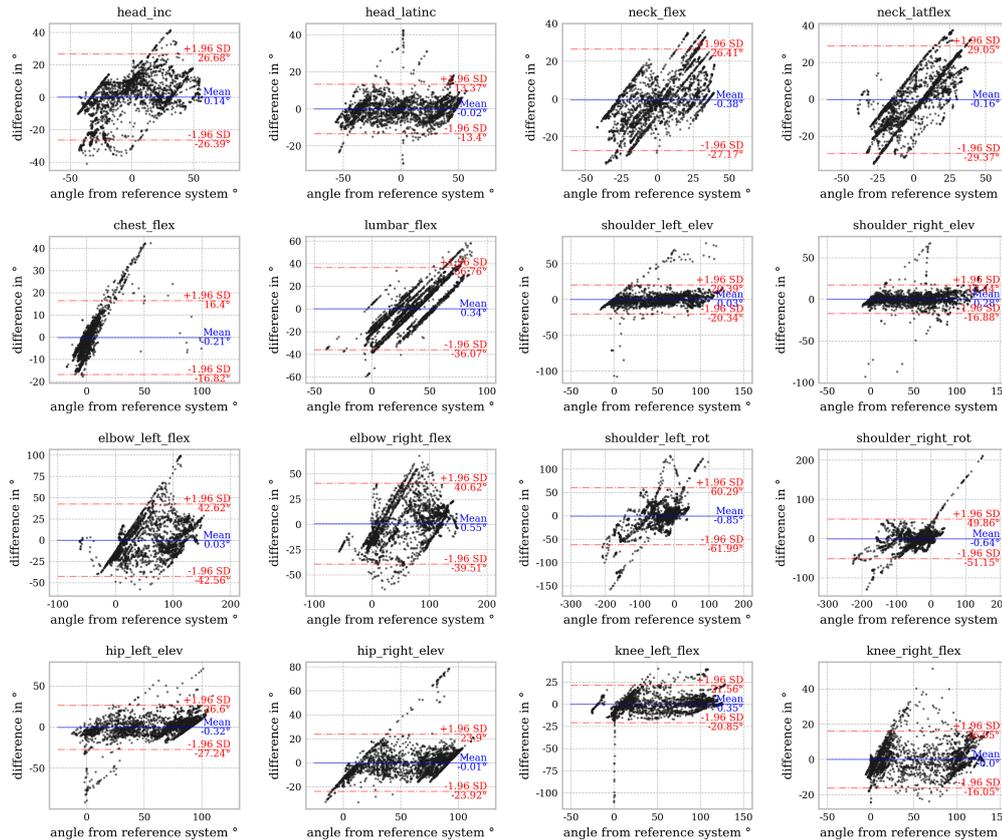
Insgesamt zeigen sich erhebliche Unterschiede zwischen der 3-Tracker-Konfiguration und den Konfigurationen mit sechs oder mehr Trackern. Bei drei Trackern gibt es bessere Korrelationen bei den *head*- und *neck*-Winkeln. Erst ab der 6-Tracker-Konfiguration können die unteren Gliedmaßen (*hip*- und *knee*-Winkel) erfasst werden. Wesentliche Unterschiede zwischen der 6-, 8- und 9-Tracker-Konfiguration sind nicht zu erkennen. Aus diesem Grund werden im Folgenden die Gelenkwinkeldifferenzen von *head*- und *neck*-Winkeln zwischen der 3- und der 6-Tracker-Konfiguration gegenübergestellt und näher untersucht. Eine vollständige Präsentation aller Differenzen und ein Vergleich zwischen allen Konfigurationen ist in [Vox et al., 2021, S. 14ff] dargestellt.

Darstellung der Genauigkeit auf Basis von modifizierten Bland-Altman-Plots

Die Bland-Altman-Plots sind mit 2500 gleichverteilten randomisierten Indexwerten aus den Zeitreihen der Differenzen berechnet. Die Bland-Altman-Plots für die *head*- und *neck*-Winkel der 3-Tracker-Konfiguration sind in Abbildung 5.10a dargestellt. Bland-Altman-Plots für die 6-Tracker-Konfiguration von allen untersuchten Gelenkwinkeln sind in Abbildung 5.10b präsentiert.



(a) Bland-Altman-Plots für die *head-inc*- und *neck*-Winkel von der 3-Tracker-Konfiguration [Vox et al., 2021, Zsfg. Fig. 12 und A1]



(b) Bland-Altman-Plots für alle untersuchten Gelenkwinkel von der 6-Tracker-Konfiguration [Vox et al., 2021, Zsfg. Fig. 14 und A3]

Abbildung 5.10.: Untersuchung der Abweichung zwischen Vive und Qualisys als Referenz auf Basis von modifizierten Bland-Altman-Plots

Im Vergleich zwischen den *head-inc*-, *neck_flex*- und *neck_latflex*-Winkeln der 3- und 6-Tracker-Konfiguration zeigen die LoA (rot gepunktete Linien) einen um ca. $\pm 10^\circ$ kleineren Bereich bei der 3-Tracker-

Konfiguration liegen die LoA von den Winkeln [*head_inc*, *neck_flex*, *neck_latflex*] bei $< \pm 12^\circ$.

Bei den Winkeln *chest_flex* und *lumbar_flex* der 6-Tracker-Konfiguration können über die gesamte RoM lineare Abweichungen wahrgenommen werden. Des Weiteren können sehr hohe LoA ($> \pm 30^\circ$) bei den Winkeln [*lumbar_flex*, *elbow_flex*, *shoulder_rot*] identifiziert werden. Mäßige LoA ($< \pm 30^\circ$ and $> \pm 15^\circ$) zeigen sich bei [*head_inc*, *head_latinc*, *neck_flex*, *neck_latflex*, *chest_flex*, *shoulder_(left, right)_elev*, *hip_(left, right)_elev*, *knee_(left, right)_flex*]. Der Mittelwert liegt bei allen Winkeln und Konfigurationen bei ca. 0° .

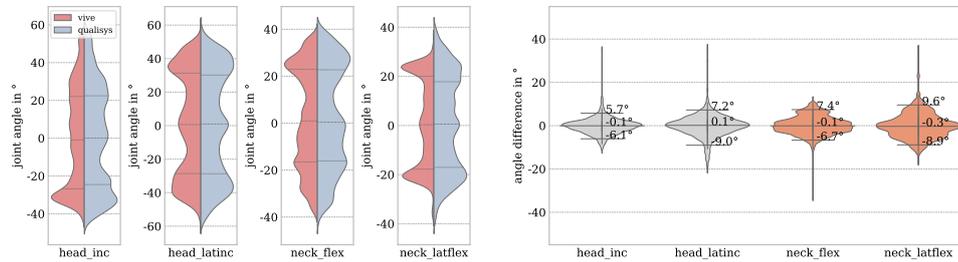
Darstellung der Genauigkeit auf Basis von Violin-Plots

Wie im vorherigen Abschnitt erläutert, konnten gute Übereinstimmungen bei den *head*- und *neck*-Winkeln für die 3-Tracker-Konfiguration festgestellt werden. Diese Beobachtung zeigt sich auch bei Betrachtung der Violinen (Häufigkeit von erreichten Winkeln in der RoM) von Vive und Qualisys in Abbildung 5.11a. Zu sehen ist eine gute Übereinstimmung der Violin-Formen bei den Winkeln [*head_inc*, *head_latinc*, *neck_flex*]. Bei *neck_latflex* ist zu erkennen, dass die Minima und Maxima der RoM bei Vive gegenüber Qualisys nicht ganz erreicht werden. In Abbildung 5.11b sind Häufigkeitsverteilungen der Abweichungen zwischen Vive und Qualisys dargestellt. Alle *head*- und *neck*-Winkel zeigen gute Genauigkeiten mit einem IPR $< \pm 10^\circ$.

Bei Betrachtung der Violinen in der 6-Tracker-Konfiguration zeigt sich die Ausprägung einer verminderten RoM des Vive Systems bei *neck_latflex* noch gravierender. In Abbildung 5.12a kann ein Abflachen bei $\pm 20^\circ$ erkannt werden. Im Vergleich zeigt der gemessene Winkel mit dem Qualisys System ein Bewegungsausmaß von ca. $\pm 40^\circ$. Diese schlechte Übereinstimmung von *neck_latflex* ist ebenfalls anhand des IPR in Abbildung 5.12b nachzuvollziehen. Hier konnte ein IPR von -26° bis 24.7° gemessen werden. Für die weiteren Gelenkwinkel zeigen sich ähnlich hohe Abweichungen bzw. ein hoher IPR. Eine Ausnahme sind die Winkel *shoulder_(left, right)_elev*. Bei diesen zeigt sich sowohl eine gute Übereinstimmung der Violinen als auch ein geringer IPR von ca. $\pm 10^\circ$.

Eine lineare Abweichung bei dem Winkel *lumbar_flex*, welche bereits in den Bland-Altman-Plots (Abbildung 5.10b) zu sehen ist, zeigt sich auch bei Betrachtung der Violin-Form von *lumbar_flex* in Abbildung 5.12. Eine lineare Abweichung deutet

darauf hin, dass sich der Winkel zwar messbar in den Daten von Qualisys bewegt, aber mit dem Vive System kaum erfasst wird. Resultat ist eine ausgeprägte Violine über die gesamte RoM mit hohen Werten im IPR.



(a) Violin-Plot für das Aufkommen von Gelenkwinkeln zwischen Vive (linke Violine, rot) und Qualisys (rechte Violine, blau); Quartile (25%, 50% und 75%) als gestrichelte Linien [Vox et al., 2021, Zsfg. Fig. 9]

(b) Häufigkeit von auftretenden Winkeldifferenzen; Angabe der 5 %- und 95 %- Perzentile und des Medians [Vox et al., 2021, Zsfg. Fig. 10]

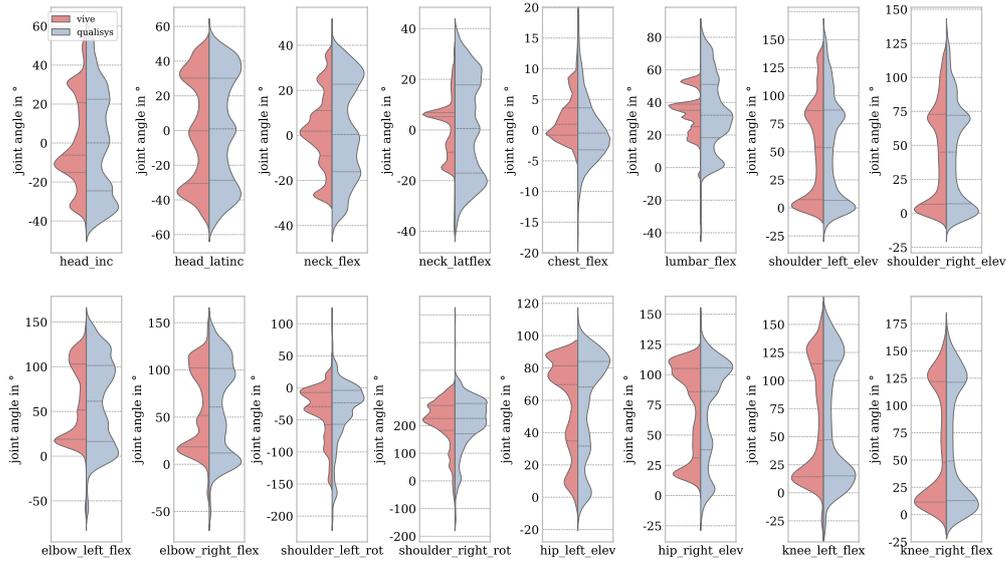
Abbildung 5.11.: Violin-Plots für die *head*- und *neck*-Winkel bei der 3-Tracker-Konfiguration

Zusammenfassung

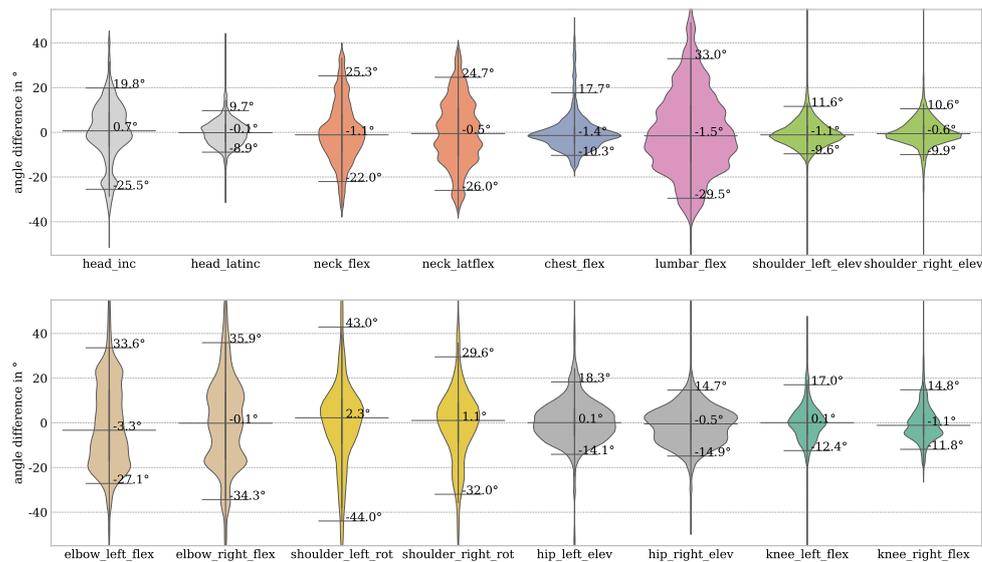
Im Allgemeinen können die Gelenkwinkeldifferenzen und somit die Genauigkeit des Vive Systems per Winkel anhand des IPR in *gut* (0° bis $\pm 10^\circ$), *mittel* ($\pm 10^\circ$ bis $\pm 15^\circ$) und *schlecht* ($> \pm 15^\circ$) unterteilt werden.

- Gut: *head_inc*, *head_latinc*, *neck_flex* und *neck_latflex* in der 3-Tracker-Konfiguration.
- Mittel: *shoulder_(left, right)_elev* und *chest_flex* ab der 6-Tracker-Konfiguration.
- Schlecht: *lumbar_flex*, *elbow_(left, right)_flex*, *shoulder_(left, right)_rot*, *hip_(left, right)_elev* und *knee_(left, right)_flex* ab der 6-Tracker-Konfiguration.

Insgesamt (alle Gelenkwinkel und Konfigurationen) reicht der IPR von $\pm 6^\circ$ bis $\pm 44^\circ$. Bei *shoulder_(left, right)_rot* sowie den *head*-, *neck*-, *chest*-, *hip*- und *knee*-Winkeln (3-, 6-Tracker-Konfigurationen) liegt der IPR zwischen $\pm 6^\circ$ und $\pm 18^\circ$. Der IPR von *elbow_(left, right)_flex* und *lumbar_flex* liegt zwischen $\pm 30^\circ$ und $\pm 44^\circ$. Ein Bias in den Ergebnissen für spezifische Typen von Proband*innen konnte nicht



(a) Violin-Plots für das Aufkommen von Gelenkwinkeln zwischen Vive (linke Violine, rot) und Qualisys (rechte Violine, blau); Quartile (25 %, 50 % und 75 %) als gestrichelte Linien [Vox et al., 2021, Zsfg. Fig. 9]



(b) Häufigkeit von auftretenden Winkeldifferenzen; Angabe der 5 %- und 95 %- Perzentile und des Medians [Vox et al., 2021, Zsfg. Fig. 10]

Abbildung 5.12.: Violin-Plots für alle untersuchten Gelenkwinkel bei der 6-Tracker-Konfiguration

festgestellt werden. Die Ergebnisse zeigen, dass bei der Gelenkwinkelmessung von dynamischen Bewegungen mit dem HTC Vive + Final IK hohe Abweichungen,

insbesondere bei den Gelenken der äußeren Gliedmaßen, zu erwarten sind.

5.5. Entwicklung eines Modells für die Bewegungserkennung

Die Entwicklung von Modellen für die Bewegungserkennung in MoCap-Daten ist eine anspruchsvolle Aufgabe. Eine effektive Erkennung sollte robust gegenüber Rauschen und invariant gegenüber zeitlichen Abhängigkeiten sein, sowie eine hohe Genauigkeit bei geringem Rechenaufwand liefern. Zwei wesentliche Herausforderungen bei dieser Aufgabe sind die effektive Darstellung von Features und wie daraus diskriminierende Merkmale für eine automatische Klassifikation erlernt werden können. In dieser Arbeit werden die Verfahren SVM und ANN evaluiert. Schwerpunkt bei der Modellentwicklung ist der Umgang mit einer geringen Trainings-Datenmenge und der Einsatz auf mobilen Endgeräten.

Dabei wird eine Erkennung von einzelnen Frames in den MoCap-Daten verfolgt. Korrekterweise kann in einem einzelnen Frame keine vollständige Bewegung abgebildet werden, da zeitliche Informationen fehlen. Ein einzelner Frame zeigt eine einzelne Körperhaltung. Eine Bewegung besteht aus einer Abfolge von zeitlich aufeinanderfolgenden Körperhaltungen. Annahme in dieser Arbeit ist, wenn aufeinanderfolgende Körperhaltungen erkannt werden, dann kann auch eine Aussage getroffen werden, um welche Bewegung es sich handelt. Vorteil dieser Vorgehensweise ist, dass zeitliche Eigenschaften in den Daten, wie die Bewegungsgeschwindigkeit, für die Erkennung keine tragende Rolle spielen. Grundlage für die Modellentwicklung sind MoCap-Daten bzw. Zeitreihendaten mit zeitlich annotierten Bewegungssequenzen.

Die Entwicklungsschritte sind bereits in der folgenden Publikationshistorie veröffentlicht: In [Vox and Wallhoff, 2017b] wurde die grundlegende Machbarkeit für eine Klassifikation auf Basis von SVM in einer *one vs. one*-Strategie und mit Gelenkwinkeln als Features untersucht. In der darauffolgenden Veröffentlichung [Vox and Wallhoff, 2018c] wurde an einer Verbesserung des Modells durch eine Erweiterung der Featurebasis gearbeitet. Dabei wurden auch Möglichkeiten evaluiert, normalisierte Positionswerte als Features zu verwenden. In [Vox and Wallhoff, 2018b] wurde eine Modellentwicklung auf Basis von ANN untersucht. Ein Vergleich zwischen SVM und ANN ist in [Vox and Wallhoff, 2018a] publiziert. Diese Arbeit hat sich

auch insbesondere mit Verfahren der künstlichen Erweiterung von Zeitreihendaten beschäftigt, um das Training der Modelle mit Daten von nur einer Person zu ermöglichen.

Für das Training und Testen der entwickelten Modelle wurden MoCap-Daten mit der Kinect v2 in einer Bewegungsstudie erhoben. Die Durchführung der Studie wird im nächsten Abschnitt beschrieben. Die Ergebnisse aus den Publikationen wurden für diese Dissertation mit aktuellen Versionen der genutzten Frameworks (*Python* 3.7.4, *numpy* 1.19.3, *sklearn* 1.0.1, *tensorflow* 2.7.0) reproduziert.

5.5.1. Datenerhebung

In dieser Studie² wurden MoCap-Daten von 21 Proband*innen im Alter zwischen 19 und 81 Jahren (Durchschnittsalter: 34.80 ± 10.22) mit der Kinect v2 mit einer Aufnahmerate von 30 Hz aufgezeichnet. Die Kinect v2 wurde 2.5 – 3.5 m mit frontalem Blickwinkel vor den Teilnehmer*innen positioniert. Die Teilnehmer*innen trugen Alltagskleidung und der Messbereich wurde mit Tages- und künstlichem Licht ausgeleuchtet. Die Proband*innen wurden angewiesen, 19 Bewegungsübungen, vorgemacht durch einen Übungsleiter, fünfmal zu wiederholen. Dieser Ablauf wurde zweimal durchgeführt, sodass insgesamt zehn Wiederholungen pro Bewegung aufgezeichnet wurden. Neben den 19 Bewegungsübungen wurden auch Bewegungen in Ruhephasen und Pausen aufgezeichnet. Insgesamt wurden somit MoCap-Daten von 20 Klassen (19 Bewegungsübungen und eine *noExercise*-Klasse) generiert (siehe Tabelle 5.5). Anschließend wurden die Daten mit einer eigens entwickelten Software (siehe Abbildung A.2 im Anhang) manuell annotiert. Dabei wurde das Label und der Start- und Endzeitpunkt von jeder Bewegungswiederholung in den Zeitreihendaten gesetzt.

5.5.2. Methodik

Ziel der Erkennung ist, in den MoCap-Daten automatisch eine jeweils bekannte Klasse zu identifizieren. Dafür wird in dieser Arbeit *überwachtes* Training angewendet (siehe auch Abschnitt 2.4 *Algorithmen für die Erkennung von Körperhaltungen*). Die Datenverarbeitung besteht im Wesentlichen aus drei Prozessschritten. Bei dem Preprocessing werden die MoCap-Daten wie in Abschnitt 5.1 *Transformation und*

²Genehmigt durch die Kommission für Folgenabschätzung und Ethik der Carl von Ossietzky Universität Oldenburg (Drs. 61/2016). Durchgeführt an der Jade Hochschule Oldenburg in der Abteilung Technik und Gesundheit für Menschen (TGM).

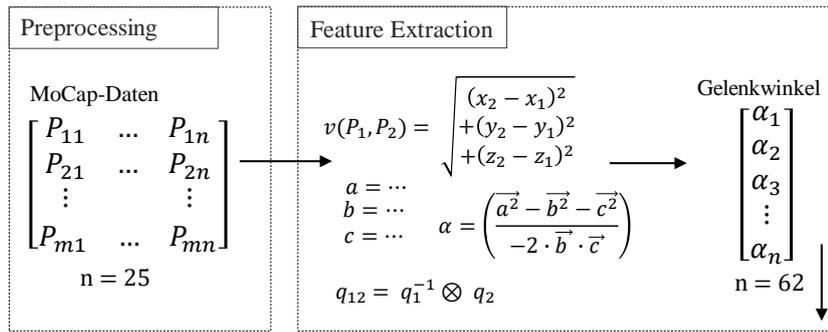
Tabelle 5.5.: Beschreibung der Bewegungsübungen (Klassen) in den MoCap-Daten

No.	Abkürzung	Beschreibung
1	abdArmElbow90L	90° Abduktion linker Ellenbogen nach oben
2	abdArmElbow90R	90° Abduktion rechter Ellenbogen nach oben
3	abdLegL	Abduktion mit dem linken Bein
4	armsFrontLBackR	linker Arm nach vorne und rechter Arm nach hinten
5	diagoArmLLegR	linker Arm nach oben, rechtes Bein abduziert
6	flexElbowL	Beugung des linken Ellenbogens
7	flexElbowR	Beugung des rechten Ellenbogens
8	fullSquat	Kniebeuge mit erhobenen Armen
9	jumpInStraddle	Sprung in Grätschstellung
10	jumpingJack	Hampelmann
11	kneeUpL	Beugung linke Hüfte, Knie hoch
12	kneeUpR	Beugung rechte Hüfte, Knie hoch
13	lungeLegL	Ausfallschritt mit dem linken Bein nach vorne
14	noExercise	Regeneration oder ähnliches
15	sideLungeLegL	Ausfallschritt zur linken Seite
16	sideLungeLegR	Ausfallschritt zur rechten Seite
17	sideSquatL	Kniebeuge in der Grätsche nach links
18	sideSquatR	Kniebeuge in der Grätsche nach rechts
19	swingArmBackward	Arme rückwärts kreisen
20	swingArmForward	Arme vorwärts kreisen

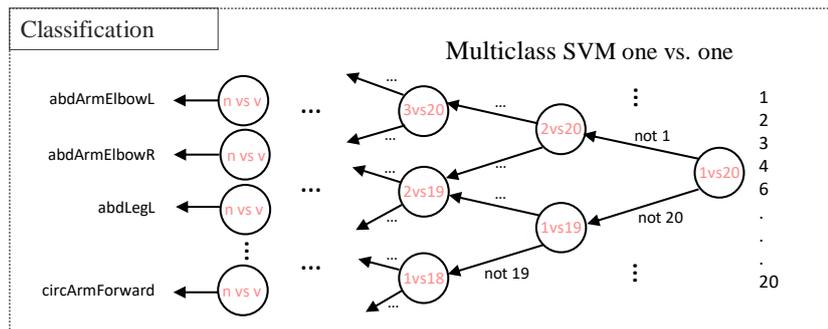
Normalisierung von MoCap-Daten vereinheitlicht, um eine Erkennung unabhängig von der Sensorik zu ermöglichen. Das Feature Extraction hat das Ziel, möglichst aussagekräftige Features mit einer guten Diskriminationsfähigkeit aus den MoCap-Daten zu generieren. In dieser Arbeit werden Winkel zwischen Gelenkpunkten als Features eingesetzt. Dafür werden neben den in Abschnitt 5.2 vorgestellten Gelenkwinkeln auch physiologisch nicht interpretierbare Winkel berechnet. Insgesamt wird eine Featurebasis mit 62 normalisierten Winkeln eingesetzt. Für die Klassifikation wird sowohl eine Multiclass SVM in einer *one vs. one*-Strategie mit einem Polynomial-Kernel untersucht als auch ein ANN mit drei Hidden Layern in einer 20-100-20 Architektur. Eine vereinfachte Darstellung der Verarbeitungsschritte und der Modellentwicklung ist in Abbildung 5.13 dargestellt. In den folgenden Abschnitten werden das Feature Extraction, die Modellentwicklung und die Ergebnisse dargestellt.

Feature Extraction und künstliche Erweiterung der Anzahl an Samples

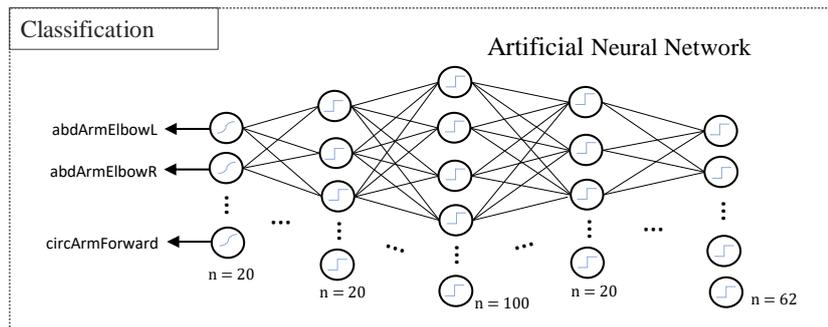
Ansatz in dieser Arbeit ist es, Winkel als Features einzusetzen. Die Auswahl von geeigneten Winkeln ist anhand einer Varianzanalyse erfolgt. Dabei wird angenom-



(a) Preprocessing von MoCap-Daten zu Gelenkwinkeln als Features



(b) Modell-Darstellung einer SVM in einer one vs. one Strategie



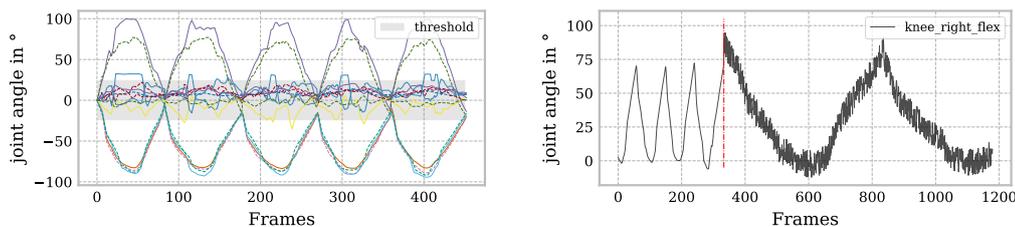
(c) Modell-Darstellung eines ANN in einer 62-20-100-20-20 Architektur

Abbildung 5.13.: Vereinfachte Darstellung der Verarbeitungsschritte und Modellentwicklung für die framewise Erkennung von Körperhaltungen

men, dass Winkel, in denen sich eine messbar hohe Varianz im zeitlichen Verlauf zeigt, auch gute Diskriminationsfähigkeiten besitzen. Aus diesem Grund wurde eine Vielzahl von möglichen Winkel-Kombinationen als Features analysiert und anhand eines Schwellwert-Verfahrens selektiert. Insgesamt wurden 62 Features mit einer Varianz > 0.02 ausgewählt.

Des Weiteren wurde ein Grenzwert-Filter (threshold) eingesetzt, um schwer zuzuordnende Bewegungen mit wenig Bewegungsausmaß zu filtern. Diese Bewegungen zeigen sich insbesondere im Übergang aus der Neutral-Null-Stellung in eine Bewegungsübung. Bei der Filterung wurden Winkel $< \pm 25^\circ$ als nicht zuordenbar deklariert und der *noExercise*-Klasse zugewiesen. In Abbildung 5.14a ist der Grenzwert-Filter anhand von exemplarischen Winkeln illustriert.

Für das Training der Modelle wurde die Anzahl der Samples in den Daten künstlich erweitert. Anforderung an die Erkennung ist, dass die Modelle mit Daten von jeweils einer Person trainiert werden können. Dadurch wird erreicht, dass das System flexibel und nachträglich auf neue Bewegungsmuster angepasst werden kann. Wenig Daten für das Training von Erkennungs-Modellen einzusetzen, ist eine besondere Herausforderung. Aus diesem Grund wurde ein Verfahren entwickelt, bei dem die Trainingsdaten durch gleichverteilte, randomisierte Werte in einem Grenzbereich von $\pm 10^\circ$, um den Originalwert hochgesamlet (verrauscht) werden. Ziel ist es, die einer Klasse zugeordneten Frames in den MoCap-Daten zu erhöhen, um mehr Trainingsdaten zu gewinnen. Annahme ist, dass innerhalb des Grenzbereichs die Varianz zu anderen Nutzer*innen abgebildet und somit weitere Testpersonen simuliert werden können. Dieses Verfahren ist in Abbildung 5.14b am Beispiel des Winkels *knee_right_flex* dargestellt. Auf der linken Seite der roten Linie ist der normale Winkelverlauf von drei Bewegungsmustern zu sehen. Auf der rechten Seite ist der künstlich verrauschte Verlauf dargestellt. Dieses Verfahren wird auf allen genutzten Features angewendet.



(a) Exemplarische Darstellung ausgewählter Features (Winkel) von fünf Bewegungswiederholungen; grenzwertbasierte Filterung von schwer zuzuordnenden Bewegungsbereichen

(b) Künstliche Erweiterung der Trainingsdaten durch hochsamplen der Werte in einem Grenzbereich von $\pm 10^\circ$ um den originalen Wert; originaler Winkelverlauf (linke Seite), künstlich verrauschter Winkelverlauf (rechte Seite)

Abbildung 5.14.: Vorverarbeitung der MoCap-Daten für die Modellentwicklung

Modellentwicklung

Für die Modellentwicklung wurden die Daten, auf Basis der berechneten Winkel, in Trainings- und Testdaten unterteilt. Der Trainingsdatensatz besteht aus einer Aufnahme einer Testperson und umfasst 9.79 MB, was Zeitreihendaten von 11:23 min entsprechen. Dieser Datensatz wurde durch das beschriebene Verfahren um das Sechsfache, auf 58.787 MB, 67:42 min erweitert. Getestet wurde mit 20 davon unabhängigen Proband*innen, bzw. Datensätzen mit einer Gesamtgröße von 136.30 MB, 165:56 min.

SVM

Für die frameweise Erkennung wurden zunächst SVMs untersucht. Wie in den Grundlagen (Abschnitt 2.4) beschrieben, können bei SVMs verschiedene Kernel eingesetzt werden. Die Auswahl des effizientesten Kernels wurde empirisch ermittelt. Mit dem RBF- und Polynomial-Kernel wurden die höchsten Genauigkeiten erzielt. Der Rechenaufwand beim Einsatz eines RBF-Kernels liegt im Vergleich zum Polynomial-Kernel ca. um das Zehnfache höher. Aus diesem Grund wurde der Polynomial-Kernel favorisiert. Um SVMs für ein Multiklassen-Problem zu verwenden, wurde eine *one vs. one*-Strategie angewendet. Bei dieser Strategie wird eine SVM für jedes Klassenpaar gebildet. Entsprechend wird die Klasse mit durchgehender *Gewinnquote* als prognostiziertes Label herausgegeben. SVMs können im Vergleich zum ANN nicht re-trainiert werden. Für eine Anpassung, sei es auch nur das Hinzufügen einer weiteren Klasse, muss das Modell vollständig neu trainiert werden.

ANN

Ebenfalls wurde eine Erkennung auf Basis eines ANN untersucht. Dafür wurde, zusätzlich zu den bereits erwähnten Verarbeitungsschritten, eine Normalisierung der Features in einen Wertebereich zwischen $[0,1]$ vorgenommen. Die Architektur sowie die Auswahl geeigneter Aktivierungsfunktionen und Parameter wie Lernrate, Batchsize und Epochen wurden empirisch ermittelt. Jeweils die Parameter mit den höchsten Genauigkeiten wurden ausgewählt. Dabei konnte ein ANN mit 62 Neuronen im Input-Layer (entsprechend der Anzahl der Features) und 20 Neuronen im Output-Layer (entsprechend der Anzahl von gesuchten Klassen) und drei Hidden Layer mit einer 20-100-20 Architektur identifiziert werden. Als Aktivierungsfunktionen wurden die ReLU bei den Neuronen im Input- und Hidden Layer und die Sigmoidfunktion in den Neuronen im Output-Layer verwendet. Des Weiteren wurde

eine Lernrate von 0.001 sowie eine Batch-Size von 100 als geeignet identifiziert und ein Training mit 150 Epochen durchgeführt.

5.5.3. Ergebnisse

Die Modelle (SVM, ANN) wurden auf drei Endgeräten trainiert und getestet. Im Folgenden werden Rechenzeiten und Erkennungsleistung vorgestellt. Die Rechenzeit ist insbesondere relevant, wenn das System mobil eingesetzt wird und die Erkennung auf unbekannte Bewegungsmuster angepasst werden soll. In Tabelle 5.6 sind die eingesetzten Endgeräte und Berechnungszeiten aufgelistet.

Tabelle 5.6.: Berechnungszeiten für die Erkennen-Modelle (SVM, ANN) auf verschiedenen Endgeräten

Spezifikationen Endgerät	SVM	ANN
i9-11900k, 3.50 GHz, 32.0 GB RAM, GPU RTX3700	2:37 min	11:07 min
i7-6600U, 2.60 GHz, 16.0 GB RAM	4:17 min	19:47 min
Raspi 4b + aktive Kühlung, 1.50 GHz, 8.0 GB RAM	39:02 min	72:34 min

Das erste System (*i9-11900k*) ist ein Desktop-PC mit zusätzlicher GPU. Das zweite untersuchte Endgerät (*i7-660U*) ist ein mobiler Laptop (Abbildung 5.19a). Das dritte Endgerät (*Raspberry Pi 4b*) stellt einen Einplatinencomputer dar (Abbildung 5.19b). Bei allen Systemen konnten Test- und Trainingsdaten vollständig in den Random Access Memory (RAM) geladen werden, was die Performance gegenüber einem sequentiellen Laden von einer Festplatte deutlich steigert. Insgesamt zeigen sich kürzere Berechnungszeiten für eine Modellentwicklung auf Basis von SVM.

Bei der Erkennungsleistung bzw. der Genauigkeit für die frameweise Klassifikation können keine wesentlichen Unterschiede zwischen SVM und ANN festgestellt werden. Insgesamt konnten Genauigkeiten von [SVM 87 %, ANN 89 %] erreicht werden. Anhand der Konfusionsmatrizen für die Erkennung mit SVM (Abbildung 5.15) und ANN (Abbildung 5.16) kann gezeigt werden, dass beide Modelle Bewegungen in der *noExercise*-Klasse (Bewegungen mit geringem Bewegungsausmaß) nahezu perfekt erkennen. Eine schwache Erkennung [SVM 31 %, ANN 41 %] ist für die Klasse *jumpInStraddle* wahrzunehmen, da vergleichbare Bewegungsmuster auch in den Klassen *noExercise* und *sideLungeLegR* auftauchen. Eine ähnliche Auffälligkeit ist bei den Klassen *sideSquatL* und *sideSquatR*, ebenfalls in beiden Konfusionsma-

trizen, zu sehen. Bei den weiteren Klassen liegt die Erkennungsleistung zwischen $\approx 70\%$ bis $\approx 100\%$. Eine technische Evaluierung der Erkennungsleistung bei kritischen Körperhaltungen ist in Abschnitt 6.2 beschrieben.

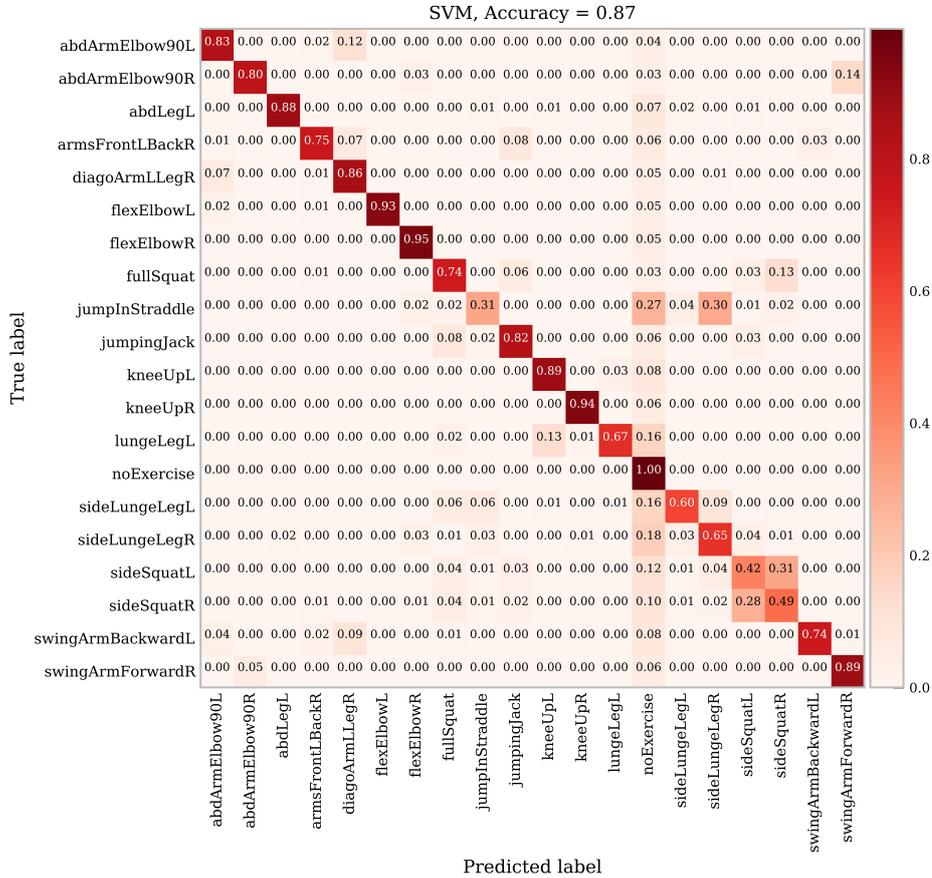


Abbildung 5.15.: Konfusionsmatrix mit normalisierten Genauigkeiten [0, 1] für die frameweise Erkennung auf Basis von SVMs

5.6. Algorithmus für die zeitliche Segmentierung von Bewegungssequenzen

Für die zeitliche Segmentierung von Bewegungssequenzen wurde ein wissens- und regelbasierter Algorithmus entwickelt. Der Algorithmus profitiert einerseits von dem Wissen über die jeweils mögliche RoM der physiologischen Gelenkwinkel. Andererseits werden Regeln verwendet, um den Start- und Endzeitpunkt von Bewegungs-

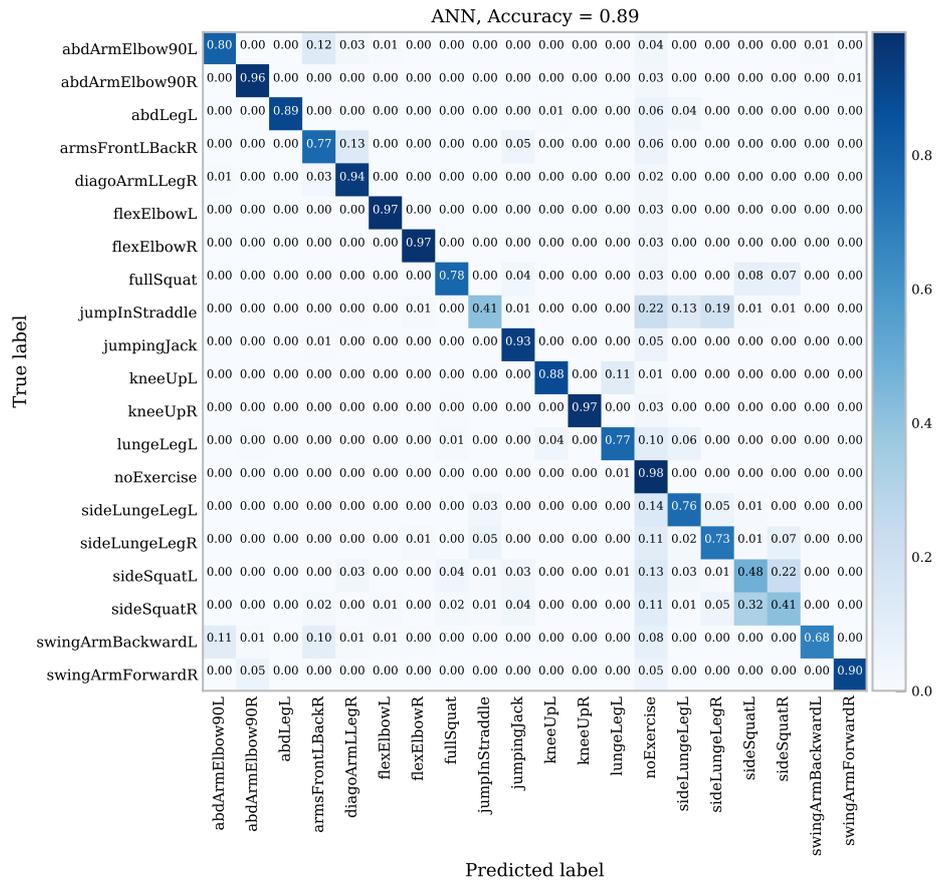


Abbildung 5.16.: Konfusionsmatrix mit normalisierten Genauigkeiten [0, 1] für die frameweise Erkennung auf Basis eines ANN

sequenzen automatisch zu identifizieren.

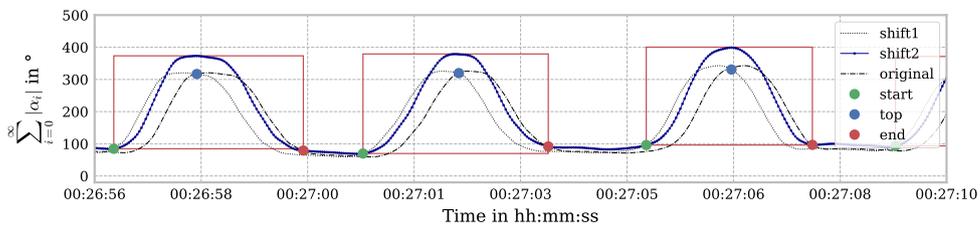
Eingabe in den Algorithmus sind die berechneten und physiologischen Gelenkwinkel (Abschnitt 5.2), bei denen das mögliche Bewegungsausmaß bekannt ist. Im ersten Schritt wird der Betrag von allen Gelenkwinkeln α_i summiert (Gleichung 5.9).

$$\sum_{i=0}^{\infty} |\alpha_i| \quad (5.9)$$

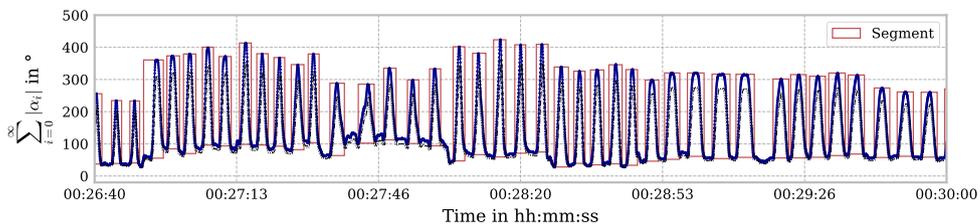
Dadurch wird erreicht, dass aus den multidimensionalen Gelenkwinkeldaten ein einzelnes Signal erzeugt wird. Dieses Signal kann als Zusammenfassung von Bewegungen über alle Gelenkwinkel verstanden werden. Dabei ist es irrelevant, welches

Gelenk bewegt wird. Jegliche Bewegungen werden in der Summe aus allen Gelenkwinkeln sichtbar gemacht. Aufgrund des Wissens über das mögliche Bewegungsausmaß einzelner Gelenke ist die Dimension bzw. der größtmögliche Amplitudenausschlag bekannt.

Die Idee des Algorithmus ist, das Summensignal (*original*) zweimal zeitlich zu verschieben und anhand von Schnittpunkten zwischen den versetzten Signalen (*shift1*, *shift2*) den Start- und Endzeitpunkt (*start*, *end*) sowie den Peak (*top*) zu identifizieren. Der Segmentierungs-Algorithmus ist anhand von drei exemplarischen Bewegungswiederholungen in Abbildung 5.17a dargestellt. Für eine bessere Visualisierung der Schnittpunkte (*top*) zwischen *original* und *shift2* wurden die Amplituden verringert. Eine Darstellung der erkannten Segmente, über einen Stream von Bewegungssequenzen in einem Zeitraum von 3:20 *min*, ist in Abbildung 5.17b dargestellt.



(a) Start- und Endzeitpunkte (*start*, *end*) sowie Peaks (*top*) von drei exemplarischen Bewegungssequenzen; erkannt anhand von Schnittpunkten zwischen dem Summensignal *original* und zeitlich versetzten Signalen (*shift1*, *shift2*)



(b) Erkannte Segmente über einen exemplarischen Messzeitraum von 3:20 *min*

Abbildung 5.17.: Algorithmus für die zeitliche Segmentierung von Bewegungssequenzen

Eine eindeutige Identifikation der Schnittpunkte *start*, *end* oder *top* wird anhand der zeitlichen Orientierung des originalen und der versetzten Signale ausgehend vom Schnittpunkt erreicht. Ein Schnittpunkt zwischen *shift1* und *shift2* stellt den Startzeitpunkt dar, Schnittpunkte zwischen *shift1* und *original* den Peak usw.

Durch Schwankungen in der Amplitude kann es zu ungeplanten Schnittpunkten kommen. Diese werden anhand von Regeln mit festgelegten Grenzwerten herausgefiltert. Dadurch wird erreicht, dass nur Segmente erkannt werden, bei denen Start- und Endzeitpunkt eine gewisse zeitliche Differenz aufweisen und der Amplitudenausschlag eine gewisse Mindesthöhe erreicht. Insgesamt müssen für den Algorithmus folgende Parameter festgelegt werden:

- *shift*: Zeitlicher Versatz in Sekunden.
- *width*: Mindestdauer eines Segments bzw. Bewegung in Sekunden.
- *height*: Mindesthöhe zwischen (*start*, *end*) und *top* in $[\circ]$.
- *bufferize*: Maximal möglicher Suchbereich in Sekunden.

Bei einer MoCap-Sensorik mit einer Aufnahmezeit von 30 Hz wurde der Algorithmus mit dem Setup [*shift* 0.2 s, *width* 1 s, *height* 10°, *bufferize* 120 s] betrieben. Dabei wird angenommen, dass eine Bewegung mindestens 1 s andauert und das Summensignal eine Höhe von mindestens 10° erreicht. Insbesondere bei Anwendung des Algorithmus mit Daten von einem Sensorstream in Echtzeit ist es notwendig, die Größe des Buffers (*bufferize*) festzulegen. Anhand der Größe des Buffers wird der maximal zurückliegende Suchbereich definiert. Die Schnittpunkte von Start- und Endzeitpunkten einer Bewegung können nur dann erkannt werden, wenn diese innerhalb des Suchbereichs liegen. Das bedeutet auch, dass durch die *bufferize* die maximal mögliche Bewegungsdauer eingegrenzt wird. Ausgaben des Algorithmus sind eindeutige Indices von erkannten Start- und Endzeitpunkten sowie Peaks einer Bewegung.

Eine Eigenschaft des Algorithmus ist, dass ein Segment nicht in Echtzeit, sondern erst nach dem festgelegten zeitlichen Versatz (*shift*) erkannt wird. Des Weiteren kann eine robuste Segmentierung nur dann garantiert werden, wenn zwischen jeder Bewegung die Neutral-Null-Stellung eingenommen wird oder ein eindeutiger Richtungswechsel erfolgt. Um die Anforderungen an den Algorithmus für Anwender*innen verständlich zu machen, kann ein akustisches oder visuelles Feedback für die Signalisierung von erkannten Start- und Endzeitpunkten verwendet werden. Der Ansatz des Segmentierungs-Algorithmus ist bereits in [Vox et al., 2021] publiziert.

5.7. Automatische Bewegungsbewertung mit Einsatz von MoCap-Sensorik

In diesem Kapitel wird die Umsetzung einer automatischen Bewegungsbewertung in einem prototypischen System beschrieben. Ziel ist es, die in Abschnitt 2.6 *Bewegungsbewertung auf Basis von Gelenkwinkeln* beschriebenen Schemata zu digitalisieren und mit MoCap-Daten bzw. berechneten Gelenkwinkeln zu speisen. Wie in Kapitel 4 *Eigener Ansatz* beschrieben, wird dabei der Einfluss von Gewicht und somit biomechanischen Belastungen vernachlässigt. Neben der Bewertung von kritischen Gelenkwinkeln ist das System in der Lage, die Bewegung zu benennen (Klassifikation) und die Dauer (Segmentierung) in einer Körperhaltung oder für eine Bewegung anzugeben. Diese Angaben sind insbesondere für die Interpretation einer Risikoabschätzung von Relevanz. Beispielsweise sind sporadisch, kurzfristig auftretende, hohe Belastungen mit einem geringeren Risiko zu bewerten als lang andauernde Haltungen in kritischen Gelenkwinkeln.

5.7.1. Digitalisierung von Bewertungsschemata

Eingabe für die Bewertungsschemata sind physiologische Gelenkwinkel. In Abschnitt 2.6 sind Tabellen mit Regeln zu kritischen Gelenkwinkelbereichen für die Schemata DGOU, RULA, REBA und dem IFA Ampelschema beschrieben. Aufgrund der bereits vereinheitlichten Beschreibung der Gelenkwinkel und Freiheitsgrade, mit konkreter Zuordnung zu den in Abschnitt 5.2 berechneten Gelenkwinkeln, konnten die angegebenen Regeln durch Wenn-Dann Anweisungen implementiert werden. In dieser Arbeit wurde insbesondere das IFA Ampelschema und eine Kombination aus RULA und REBA (*RULA+*) eingesetzt.

IFA Ampelschema

Das IFA Ampelschema wurde in eine Regelbasis auf Grundlage von Tabelle 2.7 und 2.8 implementiert. Dieses Schema eignet sich insbesondere für die Bewertung von einzelnen Gelenkwinkeln. Für die Visualisierung von kritischen Grenzwinkeln wurde eine Gauge-Darstellung in einer prototypischen Software umgesetzt (siehe Abbildung A.3).

RULA+

Die Schemata RULA und REBA wurden zu einem gemeinsamen Schema *RULA+* zusammengefasst, um Bewegungen sowohl von den oberen als auch den unteren Gliedmaßen zu bewerten. Dafür wurden die Regeln in den Tabellen 2.5 und 2.6 implementiert und die tabellarische Zuordnung für die Berechnung des Gesamtscores nach [McAtamney and Nigel Corlett, 1993, Hignett and McAtamney, 2000] umgesetzt. *RULA+* bietet die Möglichkeit, eine Körperhaltung anhand einer einzigen Zahl zu bewerten. Der Score ist anhand der Skala [1 *kein Risiko*, 2-3 *leichtes Risiko*, 4-7 *mittleres Risiko*, 8-10 *erhöhtes Risiko*, > 11 *sehr hohes Risiko*] einzuschätzen. Ein Beispiel für eine Bewertung nach *RULA+*, bei Ausführung einer Kniebeuge (*fullSquat*), ist in Abbildung 5.18b zu sehen.

Anpassung an Sensorik

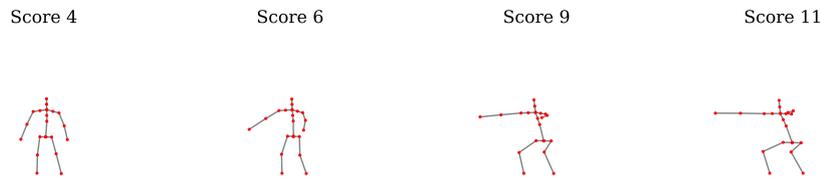
Des Weiteren werden sensorspezifische Eigenschaften bei der Bewegungsbewertung berücksichtigt. Zum Beispiel muss aufgrund der Controller bei einem Lighthouse Tracking System, welche in den Händen gehalten werden, mit einer veränderten Stellung der Handgelenke gerechnet werden. Aus diesem Grund wurde die Möglichkeit geschaffen, bestimmte Gelenkwinkel von der Bewertung auszuschließen. Gleiches Verfahren kann für MoCap-Sensorik angewendet werden, bei der bspw. nur die obere oder die untere Körperhälfte erfasst werden kann. In diesen Fällen ist die Skala entsprechend der betrachteten Gelenkwinkel anzupassen.

5.7.2. Framework für die automatische Bewegungsbewertung

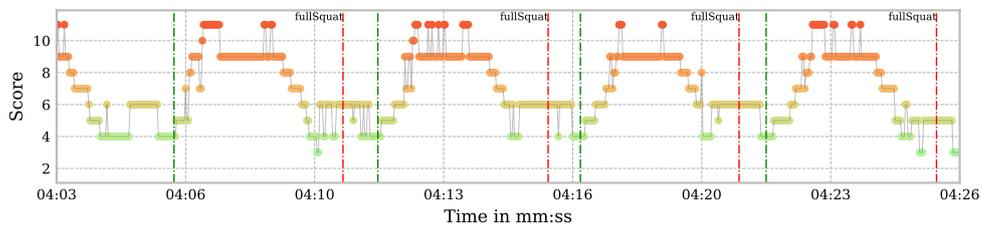
Im Zuge dieser Arbeit ist ein Framework entstanden, welches die Möglichkeit bietet, Bewegungen automatisch zu erkennen, zeitlich zu segmentieren und zu bewerten. Dafür kann unterschiedliche MoCap-Sensorik eingesetzt werden. Das Framework bietet die Möglichkeit, MoCap-Daten aus einem Sensorstream in einer Server-Client Struktur zu verarbeiten und Ergebnisse bereitzustellen. Des Weiteren kann MoCap-Sensorik auch direkt angebunden und die Bewegungsbewertung in prototypischen Anwendungen visualisiert werden. Erste prototypische Applikationen dieses Frameworks sind in den Abbildungen A.1, A.2 und A.3 im Anhang gezeigt. Im Folgenden wird zusammenfassend erläutert, welche Informationen für die Bewegungsbewertung aus den MoCap-Daten gewonnen werden.

In Abbildung 5.18 sind die Ergebnisse der automatischen Bewegungsbewertung

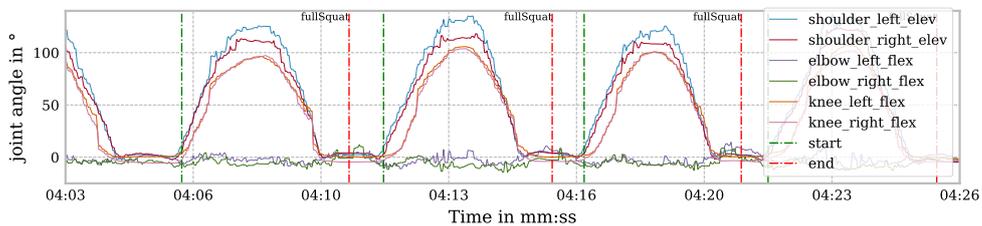
am Beispiel von Kniebeugen (*fullSquat*) dargestellt. In (a) sind einzelne Frames mit zugehörigem Score gezeigt. In (b) ist der Score über die Zeit am Beispiel von Kniebeugen gezeigt. In (c) sind exemplarische Gelenkwinkel als Zeitreihe dargestellt. Ebenfalls sind in (b) und (c) Start- und Endzeitpunkte (Segmentierung) der Bewegungswiederholungen eingezeichnet.



(a) Darstellung einzelner Frames mit zugehörigem Score (*RULA+*)



(b) Darstellung des Scores (*RULA+*) als Zeitreihe und Einzeichnung der Start- und Endzeitpunkte von Bewegungswiederholungen



(c) Darstellung exemplarischer physiologischer Gelenkwinkel als Zeitreihe und Einzeichnung der Start- und Endzeitpunkte von Bewegungswiederholungen

Abbildung 5.18.: Automatische Bewegungsbewertung, Erkennung und Segmentierung am Beispiel von Kniebeugen (*fullSquat*)

Der zeitliche Verlauf des Scores in (b) zeigt ein kurzfristiges Auftreten eines sehr hohen Risikos. Der Unterschied zwischen einem mittleren Risiko (*Score 9*) und einem sehr hohen Risiko (*Score 11*) ist anhand der dargestellten Körperhaltungen in (a) zu sehen. Ein mittleres Risiko ist bereits bei einer Körperhaltung in einer Hockstellung durch die Gelenkwinkel der Knie zu erwarten. Das sporadische Aufkommen

eines sehr hohen Risikos tritt auf, wenn sich die Arme zusätzlich zur Hockstellung oberhalb der Schultergelenke befinden (Flexion oder Abduktion $> 90^\circ$) und damit einen kritischen Winkelbereich erreichen. Des Weiteren können Informationen über die Bewegung (Label), die Häufigkeit von Bewegungswiederholungen sowie Entspannungsphasen erfasst und ausgegeben werden.

In einer abschließenden Beurteilung der Informationen und Bewertungsergebnisse empfiehlt es sich, Expert*innen zu konsultieren, um mögliche Konsequenzen zu diskutieren. Individuelle Faktoren wie Krankheiten oder Bewegungseinschränkungen und zusätzliche Belastungen in der Umgebung (Hitze, Kälte, etc.) werden durch die beschriebenen Verfahren nicht miteinbezogen. Ebenfalls werden in der aktuellen Version Vitaldaten wie Herzfrequenz, Körpertemperatur u.w., die auf Erschöpfung der Nutzenden hinweisen könnten, nicht berücksichtigt.

5.7.3. Anwendung der automatischen Bewegungsbewertung auf mobilen Endgeräten

Alle in dieser Arbeit beschriebenen Algorithmen wurden auf mobilen Endgeräten getestet. Die Systemperformance hängt dabei im Wesentlichen von der verwendeten MoCap-Sensorik ab. Wird die Kinect v2 eingesetzt, muss mit einer Latenz von $\approx 70\text{ ms}$ gerechnet werden [Guffanti et al., 2020]. Das Skeleton Tracking in dem Kinect SDK basiert auf einem Random Forrest Algorithmus. Bei diesem liegt die Entscheidungsfindung für die Identifikation von Gelenkpunkten in einem variablen Zeitraum. Des Weiteren muss bei kamerabasierter Sensorik in Situationen der Umgebungsveränderung (z.B. der Lichtverhältnisse) mit Latenzen durch die automatische Anpassung von Filtereinstellungen und bildstabilisierenden Verfahren gerechnet werden. Beim Einsatz des HTC Vive Tracking Systems liegt die Latenz bei $\approx 30\text{ ms}$ [Viitanen et al., 2018]. VR-Tracking Systeme haben in der Regel hohe Systemanforderungen und erfordern eine zusätzliche Grafikerunterstützung.

Die in dieser Arbeit entwickelten Verfahren für die frameweise Verarbeitung zeigen keine nennenswerte Erhöhung des Rechenaufwandes. Wie in Abschnitt 2.1 beschrieben, liegt das Datenaufkommen eines Samples in den MoCap-Daten bei ca. 2.4 kB . Erfüllt ein genutztes Endgerät die Systemanforderungen für die eingesetzte MoCap-Sensorik, kann davon ausgegangen werden, dass die in dieser Arbeit entwickelten Verarbeitungsprozesse ohne wesentliche Latenz berechnet werden können. Eine Ausnahme bildet die zeitliche Segmentierung. Bedingt durch den Algorithmus

ist eine Erkennung von Start- und Endzeitpunkt nur mit einer Latenz um den zeitlichen Versatz (*shift*) möglich.

Bei einer Anpassung des Modells für die Bewegungserkennung auf einem mobilen Endgerät ist jedoch mit den beschriebenen Trainingszeiten in Tabelle 5.6 zu rechnen. Bei einer Reduzierung der Klassen und somit einer geringeren Komplexität für die Modell-Berechnung, ist auch eine entsprechende Reduzierung der Trainingszeit zu erwarten. In Abbildung 5.19 ist die Anwendung der automatischen Bewegungsbewertung auf mobilen Endgeräten mit Einsatz der Kinect v2 zu sehen. In Abbildung (a) wird die prototypische Software auf einem Laptop (i7-6600U, 2.60 GHz, 16.0 GB RAM) ausgeführt. Abbildung (b) zeigt die Kinect v2 verbunden mit einem Einplatinencomputer (Raspberry Pi 4b + aktive Kühlung, 1.50 GHz, 8.0 GB RAM).



(a) Anwendung der automatischen Bewegungsbewertung auf einem Laptop (i7-6600U, 2.60 GHz, 16.0 GB RAM) mit der Kinect v2



(b) Kinect v2 verbunden mit einem Einplatinencomputer (Raspberry Pi 4b + aktive Kühlung, 1.50 GHz, 8.0 GB RAM)

Abbildung 5.19.: Testung der automatischen Bewegungsbewertung auf mobilen Endgeräten

Die Anwendung auf einem mobilen Laptop ist mit der beschriebenen Latenz und den genannten Trainingszeiten möglich. Der Einplatinencomputer eignet sich nur bedingt für die Anwendung. Hiermit ist zwar eine Aufzeichnung von RGB- und Tiefenbildern möglich, jedoch kann das Skeleton Tracking nicht parallel oder nur mit erheblichen Latenzen ausgeführt werden. Eine Bereitstellung von MoCap-Daten in Echtzeit konnte auf diesem System nicht erreicht werden. Somit ist eine automatische Bewegungsbewertung ebenfalls nicht einsetzbar.

6. Technische Evaluation der Ergebnisse

In diesem Kapitel werden die Ergebnisse der Studie zur Sensorgenauigkeit des HTC Vive Tracking Systems (Abschnitt 5.4) und der Bewegungserkennung auf Basis von ANN und SVMs (Abschnitt 5.5) mittels Informationen aus der Bewegungsbewertung (Abschnitt 5.7) technisch evaluiert. Insgesamt ist die Systemperformance in kritischen Körperhaltungen von hoher Relevanz, da nach den Bewertungsschemata (Abschnitt 2.6) entsprechende Maßnahmen zur Risikoreduktion zu treffen sind. Eine Fehleinschätzung würde entweder zu unnötigen Maßnahmen führen oder Risiken unentdeckt lassen. Im Kern werden folgende Fragen untersucht:

- Wie ist die Sensorgenauigkeit des HTC Vive Systems in kritischen Gelenkwinkelbereichen?
- Wie hoch sind die Genauigkeiten der Erkennung von kritischen Körperhaltungen mit den Modellen auf Basis von ANN und SVMs?

Zur Beantwortung werden zunächst kritische Gelenkwinkel bzw. Körperhaltungen anhand des IFA Ampelschemas und RULA+ identifiziert und Sub-Datensätze aus den Daten der beschriebenen Studien generiert. Anschließend wird die Sensorgenauigkeit und Erkennungsleistung auf Basis der Sub-Datensätze evaluiert.

6.1. Sensorgenauigkeit in kritischen Körperhaltungen

Wie in Abschnitt 5.4 beschrieben, wurde die Genauigkeit des HTC Vive in Kombination mit Final IK auf Basis von Bewegungen einzelner Gelenkwinkel unter Ausnutzung der gesamten RoM evaluiert. Mit dem IFA Ampelschema ist eine Kategorisierung des Risikos ebenfalls auf Basis von einzelnen Gelenkwinkeln möglich. Das IFA Ampelschema wurde daher genutzt, um den Sub-Datensatz zu bilden, welcher zur Bestimmung der Erkennungsgenauigkeit von kritischen Gelenkwinkeln benötigt wird.

In Tabelle 6.1 sind die untersuchten Gelenkwinkel mit den zugehörigen Regeln

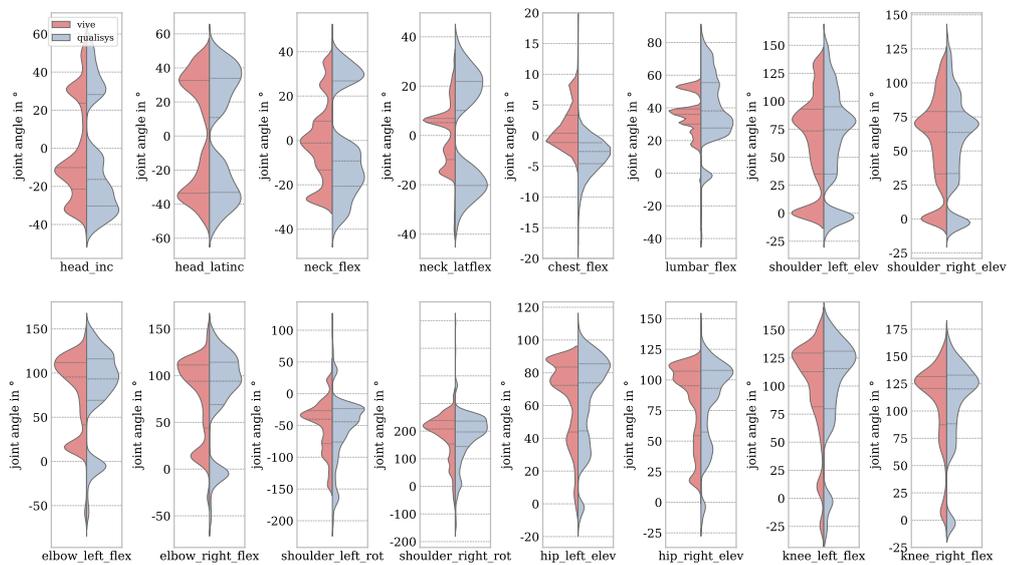
für die Untermenge der kritischen Winkelbereiche (● gelb) und (● rot) nach IFA Ampelschema aufgelistet (siehe auch Tabellen 2.7 und 2.8). Anhand dieser Regeln wurde ausgehend vom Referenzsystem (Qualisys) ein Sub-Datensatz gebildet und das HTC Vive in der 6-Tracker-Konfiguration erneut bewertet. Da die Winkel $hip_{(left, right)}_{elev}$ in dem IFA Ampelschema nicht berücksichtigt werden, wurden für diese Winkel die Regeln von $shoulder_{(left, right)}_{elev}$ übernommen. Dabei wurde die Annahme getroffen, dass aufgrund des ähnlichen Bewegungsausmaßes auch vergleichbare kritische Grenzwinkel existieren. Diese Annahme wurde nur für die Behandlung dieses Sonderfalls in der vorliegenden Sensorevaluierung getroffen und ist nicht auf die allgemeine Bewertung von weiteren Gelenkwinkeln übertragbar.

Tabelle 6.1.: Untersuchte Gelenkwinkel mit zugehörigen Regeln nach IFA Ampelschema für die Auswahl von kritischen Winkeln (\geq ●)

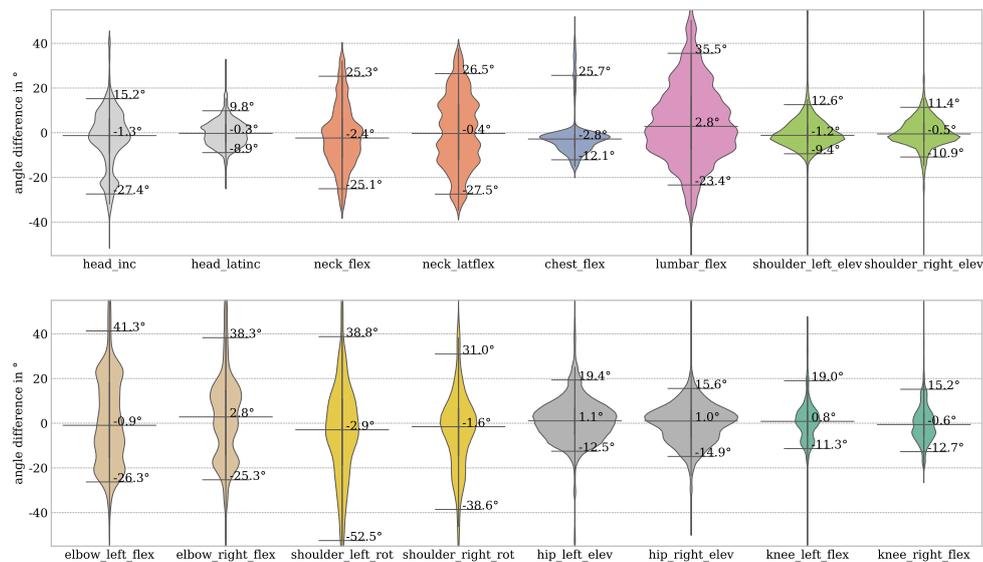
Gelenkwinkel	Erste Regel	Zweite Regel
head_inc	● $\alpha < 0^\circ$	● $\alpha > 25^\circ$
head_latinc	● $\alpha < -10^\circ$	● $\alpha > 10^\circ$
neck_flex	● $\alpha < 0^\circ$	● $\alpha > 25^\circ$
neck_latflex	● $\alpha < -10^\circ$	● $\alpha > 10^\circ$
chest_flex	● $\alpha < 0^\circ$	● $\alpha > 20^\circ$
lumbar_flex	● $\alpha < 0^\circ$	● $\alpha > 20^\circ$
shoulder_(left, right)_elev	● $\alpha < 0^\circ$	● $\alpha > 20^\circ$
elbow_(left, right)_flex	● $\alpha < 0^\circ$	● $\alpha > 60^\circ$
shoulder_(left, right)_rot	● $\alpha < -30^\circ$	● $\alpha > 60^\circ$
$hip_{(left, right)}_{elev}$	● $\alpha < 0^\circ$	● $\alpha > 20^\circ$
knee_(left, right)_flex	● $\alpha < 0^\circ$	● $\alpha > 45^\circ$

In Abbildung 6.1 sind die Genauigkeiten in kritischen Gelenkwinkelbereichen nach IFA Ampelschema anhand von Violin-Plots für die 6-Tracker-Konfiguration dargestellt. Anhand der Violin-Formen in (a) können zunächst die Grenzbereiche wahrgenommen werden. Wie am Beispiel von $head_{inc}$ zu erkennen, treten keine Gelenkwinkel innerhalb des Regelbereichs von 0° bis 25° auf. Das heißt, in diesem Bereich zeigen die Violinen beider Systeme eine flache Form. Der IPR von $head_{inc}$ liegt zwischen -27.4° und 15.2° und ist somit $\approx 2.5^\circ$ geringer als in der Untersuchung über die gesamte RoM (siehe Abbildung 5.12). Für die weiteren Gelenkwinkel liegt der Unterschied im Vergleich zwischen dem IPR in kritischen Gelenkwinkeln des Sub-Datensatzes und über die gesamte RoM in ähnlichen Größenordnungen. Diese Beobachtung zeigt, dass bei kritischen Grenzwinkeln die gleichen Genauigkeiten angenommen werden können, wie über die gesamte RoM. Das HTC Vive System weist für die untersuchten Gelenkwinkel, auch in höheren Winkelbereichen,

keine Präferenzen von zusätzlicher Über- oder Unterschätzung auf.



(a) Violin-Plots für das Aufkommen von Gelenkwinkeln zwischen Vive (linke Violine, rot) und Qualisys (rechte Violine, blau); Quartile (25 %, 50 % und 75 %) als gestrichelte Linien



(b) Häufigkeit von auftretenden Winkeldifferenzen; Angabe der 5 %- und 95 %-Perzentile und des Medians

Abbildung 6.1.: Violin-Plots für alle untersuchten Gelenkwinkel bei der 6-Tracker-Konfiguration in kritischen Gelenkwinkelbereichen

6.2. Genauigkeit der Erkennung von kritischen Körperhaltungen

Zur Ermittlung der Erkennungsgenauigkeit von kritischen Körperhaltungen wurde ebenfalls ein Sub-Datensatz aus den bereits erhobenen Daten (Abschnitt 5.5.1) gebildet. In diesem Fall wurde RULA+ verwendet, um kritische Haltungen zu identifizieren. Dies führt dazu, dass die Evaluierung für potenziell höhere Winkelamplituden durchgeführt wird und ein erhöhtes Aufkommen von Winkeldifferenzen um ca. 0° von der Analyse ausgeschlossen werden. Dafür wurden fünf Ganzkörperbewegungen (*diagoArmLLegR*, *fullSquat*, *jumpingJack*, *sideLungeLegL*, *sideLungeLegR*) aus den ursprünglich 20 Bewegungen ausgewählt, bei denen Haltungen (Frames) mit einem Score von ≥ 6 (*leichtes Risiko*) regelmäßig auftreten, und die ANN und SVM Modelle mit diesen getestet.

In Abbildung 6.2 ist der Score-Verlauf nach RULA+, berechnet aus den Gelenkwinkeln des Sub-Datensatzes, mit Kennzeichnung der Haltungen mit einem Score ≥ 6 (*threshold*) dargestellt.

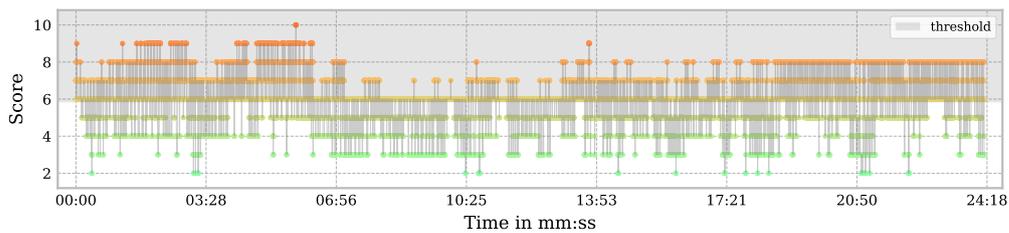


Abbildung 6.2.: Darstellung des Sub-Datensatzes aus den Bewegungen (*diagoArmLLegR*, *fullSquat*, *jumpingJack*, *sideLungeLegL*, *sideLungeLegR*) mit Einzeichnung eines *threshold* für Haltungen mit einem Score ≥ 6

In Abbildung 6.3 sind die Konfusionsmatrizen für die Erkennung der fünf Ganzkörperbewegungen mit einem Score ≥ 6 dargestellt. Die Erkennung auf Basis von ANN in (a) zeigt eine höhere Genauigkeit von 94% im Vergleich zu den 89% über dem Datensatz mit allen 20 Bewegungen (Abbildung 5.16). Die Konfusionsmatrix in (b) zeigt für die Erkennung auf Basis von SVMs eine vergleichbare Genauigkeit von 86% zu der Genauigkeit von 87% des 20-Klassen-Problems. Insgesamt zeigen die Ergebnisse, dass bei kritischen Körperhaltungen die überwiegende Anzahl an Frames richtig erkannt werden können.

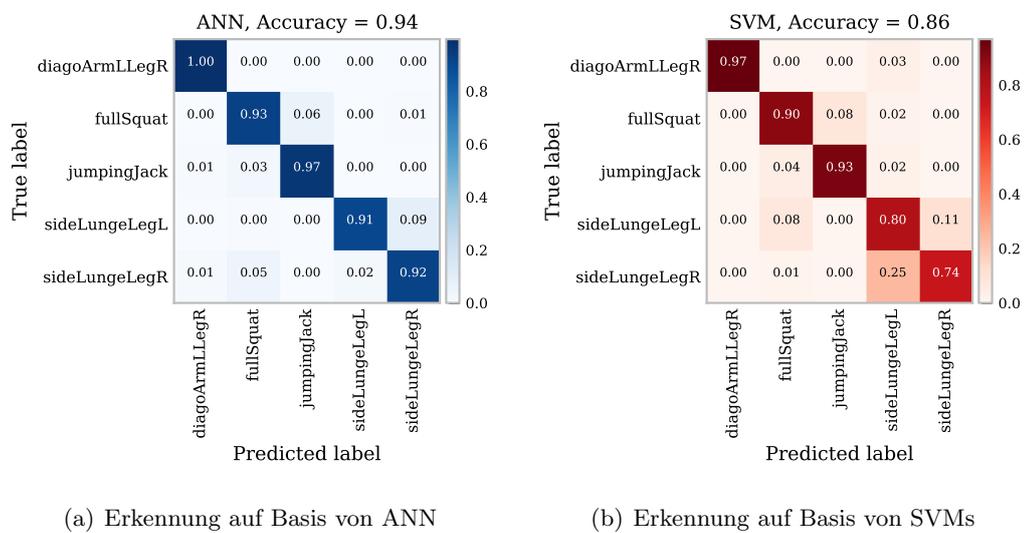


Abbildung 6.3.: Konfusionsmatrizen mit normalisierten Genauigkeiten $[0, 1]$ für die frameweise Erkennung von fünf Ganzkörperbewegungen mit einem Score ≥ 6

7. Diskussion

Im Folgenden werden die präsentierten Ergebnisse aus Kapitel 5 anhand der aufgestellten Forschungsfragen diskutiert und untersucht, ob die zentrale Hypothese belegt oder widerlegt werden kann.

„Motion-Capture-Sensorik im Consumer-Bereich bietet eine ausreichende Genauigkeit für die Erkennung und Bewertung von Bewegungen und Körperhaltungen anhand von Gelenkwinkeln.“

7.1. Fragestellung I - Sensorgenauigkeit

Bietet Motion-Capture-Sensorik (RGB-Tiefenkamera, Lighthouse-Verfahren) eine ausreichende Gelenkwinkelgenauigkeit für eine Bewegungsbewertung?

Um die erhobenen Genauigkeitsangaben bei der Gelenkwinkelmessung mit MoCap-Sensorik interpretieren zu können, ist es notwendig, einen tolerierbaren Bereich zu ermitteln. Für die bessere Einschätzung der im Folgenden genannten Dimensionen ist in Abbildung 7.1 die Flexion des Ellenbogengelenks in 10°-Schritten von (a) 60° nach (d) 90°, gemessen mit einem Standard-Goniometer, dargestellt.

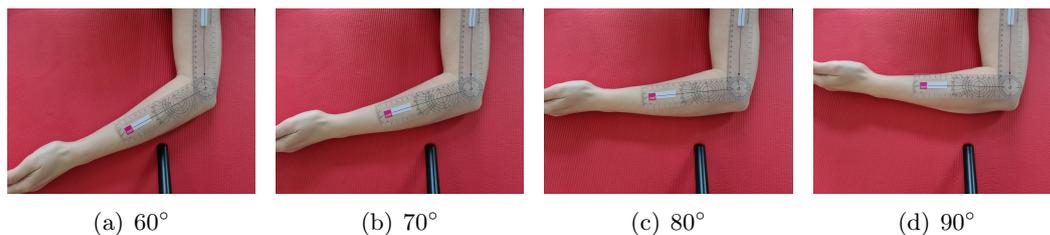


Abbildung 7.1.: Ellenbogenflexion von 60° bis 90° in 10° Schritten gemessen mit einem Standard-Goniometer [Vox et al., 2021, Fig. 21]

Wie in Abschnitt 2.5 beschrieben, ist das Goniometer der klinische Standard für die Gelenkwinkelmessung [Gajdosik and Bohannon, 1987]. Bei der Positionierung

des Goniometers auf definierten Gelenkpunkten können Fehler nicht ausgeschlossen werden. In [Gogia et al., 1987, Boone et al., 1978] wurde die Reliabilität von Therapeut*innen und medizinischem Personal bei der Gelenkwinkelmessung mit dem Goniometer untersucht. Dabei konnten Abweichungen zwischen 5° und 10° festgestellt werden, welche als tolerierbar bewertet wurden.

7.1.1. Optische und markerbasierte Systeme als technische Referenz

In dieser Arbeit, wie auch in einigen referenzierten Arbeiten in Abschnitt 3.1, wurden in technischen Vergleichsstudien und für die Evaluierung von MoCap-Sensorik optische und markerbasierte Systeme wie das Qualisys eingesetzt. Dabei ist zu beachten, dass es trotz der hohen Positionsgenauigkeit der einzelnen Marker zu Fehlern bei der Positionierung der Marker auf der Haut, bzw. den anatomischen Strukturen kommen kann. Ebenfalls kann es bei der Ausübung von Bewegungen, aufgrund der Elastizität von Haut, Fettgewebe und Muskulatur zu Positionsabweichungen kommen. Da eine starre Positionierung von Markern an Knochenstrukturen nicht möglich ist, kann diese Fehlerquelle nicht ausgeschlossen werden.

Eine perfekte Genauigkeit für die Gelenkwinkelmessung, ohne Einfluss der genannten Fehlerquellen, bietet eine Untersuchung auf Basis von Röntgenbildern [Gogia et al., 1987]. Dieses Verfahren ist jedoch nur für statische Momentaufnahmen geeignet und nicht für eine Bewegungsmessung bei zeitlichen Verläufen.

Demzufolge ist eine perfekte Referenz für die Gelenkwinkelmessung nicht vorhanden. Dennoch kann bei einer Messung mit optischen und markerbasierten Systemen, trotz der bekannten Fehlerquellen, von einer sehr hohen Genauigkeit ausgegangen werden [Liu et al., 2007, Windolf et al., 2008]. Für die Bewertung der Genauigkeiten wird im Folgenden eine tolerierbare Abweichung von 5° bis 10° , entsprechend der größtmöglichen Bewegungsamplitude in der RoM, angenommen.

7.1.2. RGB-Tiefenkamera

Wie in Abschnitt 3.1.1 *Sensorgenauigkeit von RGB-Tiefenkameras* beschrieben, liegt die Genauigkeit (RMSE) für die Systeme Azure und Kinect v2 mit entsprechenden Verfahren für das Skeleton Tracking bei $\approx \pm 10^\circ$. Unter ungünstigen Bedingungen (verdeckte Gliedmaße) kann der RMSE auf bis zu $\approx \pm 30^\circ$ ansteigen oder in einen unplausiblen Messbereich außerhalb der RoM abweichen. Die Ergebnisse

aus der eigenen Untersuchung in Abschnitt 5.3 zeigen, dass unter statischen Bedingungen eine maximale Abweichung von 0.861° eintreten kann. Technologiebedingt eignen sich RGB-Tiefenkameras nur in Umgebungen mit wenig Tageslicht. Plötzliche Veränderungen von Lichtverhältnissen im Messbereich können zu unplausiblen Messwerten führen. Die Neu-Fokussierung der Kameralinse sowie die automatische Anpassung an die Lichtverhältnisse kann einen Zeitraum von bis zu 2 *Sekunden* benötigen. Des Weiteren ist der Messraum innerhalb des Blickwinkels eingeschränkt. Die angegebenen Genauigkeiten können nur für Bewegungen in aufrechten Körperhaltungen mit einem frontalen Blickwinkel zur Kamera garantiert werden.

Insgesamt kann gesagt werden, dass **RGB-Tiefenkamera Systeme unter optimalen Bedingungen eine Genauigkeit in einem tolerierbaren Rahmen bieten**. Ebenfalls eignen sich die Systeme für eine Anwendung auf mobilen Endgeräten (Abschnitt 5.7.3) und zeichnen sich durch eine schnelle Einsetzbarkeit aus.

7.1.3. Lighthouse Tracking System

Wie in Abschnitt 3.1.2 im Stand der Technik beschrieben, können mit Lighthouse Tracking Systemen hohe Genauigkeiten für die Erfassung räumlicher Positionen und Drehungen einzelner Tracker (HMD, Controller, Tracker) erzielt werden. Für die Erfassung von MoCap-Daten ist jedoch eine IK notwendig, um weitere Gelenkpunkte zu berechnen und menschliche Bewegungen vollständig abbilden zu können. Die jeweiligen Tracker werden in der Regel mit Gurten an Gliedmaßen befestigt. Bei Bewegungen kann es zu einem Verrutschen der Tracker kommen, was zu einer Fehlerfortpflanzung bei der Berechnung von Gelenkpositionen durch die IK führen kann. In dieser Arbeit wurde das Lighthouse Tracking System HTC Vive in Kombination mit Final IK für die Gelenkwinkelmessung untersucht. Die Berechnung eines Gelenkwinkels ist nach der entwickelten Methode in Abschnitt 5.2 von mehreren Gelenkpunkten abhängig. Bereits die Ungenauigkeit von nur einem Gelenkpunkt in der Berechnungsgrundlage führt zu einer Abweichung der Berechnung des jeweiligen Gelenkwinkels. Demzufolge kann das Verrutschen eines Trackers an den Gliedmaßen zu einer Fehlerfortpflanzung auf zwei Arten führen. Zum einen kann es zu einer ungenauen Berechnung von umliegenden Gelenkpunkten durch die IK kommen. Zum anderen wirkt sich diese Ungenauigkeit unmittelbar auf die Berechnung von Gelenkwinkeln aus, welche mit den jeweiligen Gelenkpunkten zusammenhängen.

In der durchgeführten Vergleichsstudie zwischen dem HTC Vive und dem Qualisys

als Referenz wurde die Genauigkeit in fünf verschiedenen Vive-Tracker-Konfigurationen evaluiert. In der kleinsten Tracker-Konfiguration mit zwei Controllern in den Händen und dem HMD am Kopf wird die IK mit drei Positionen gespeist. Die Ergebnisse in 5.4 zeigen, dass bei der 3-Tracker-Konfiguration lediglich die oberen Gliedmaßen erfasst werden können. Dabei liegen Genauigkeiten für die *head*- und *neck*-Winkel in einem IPR $\leq \pm 10^\circ$. Durch die geringere RoM bei den *head*- und *neck*-Winkeln ist jedoch auch eine verhältnismäßig bessere Genauigkeit gegenüber den Winkeln der Gliedmaßen zu erwarten.

Erst ab der 6-Tracker-Konfiguration ist eine Erfassung aller Gliedmaßen möglich. Die Genauigkeit mit der 6-Tracker-Konfiguration liegt über alle Gelenkwinkel bei einem IPR zwischen $\pm 10^\circ$ und $\pm 44^\circ$. Ab der 6-Tracker-Konfiguration ist auch eine Verringerung der Genauigkeit von den *head*- und *neck*-Winkeln gegenüber der 3-Tracker-Konfiguration wahrzunehmen. Das lässt darauf schließen, dass bei Zunahme von Positionen als Eingabe in die IK die Komplexität der Berechnung und somit das Fehler-Potenzial steigt. Wesentliche Unterschiede zwischen der 6-Tracker-Konfiguration und den 7-, 8- und 9-Tracker-Konfigurationen konnten nicht festgestellt werden.

Demzufolge ist der Einsatz eines Lighthouse Tracking Systems + IK für die Gelenkwinkelmessung nur für ausgewählte Winkel [3-Tracker: *head_inc*, *head_latinc*, *neck_flex*, *neck_latflex*; 6-Tracker: *shoulder_elev*] geeignet. Hohe Abweichungen ($> \pm 15^\circ$ in einem nicht tolerierbaren Ausmaß) konnten bei den Gelenkwinkeln *chest_flex*, *lumbar_flex*, *shoulder_rot* und *elbow_flex* festgestellt werden. Als grenzwertig können *hip_elev* und *knee_flex* mit einer Abweichungen von $\approx \pm 15^\circ$ betrachtet werden. In der Evaluierung der Genauigkeit in Abschnitt 6.1 konnte nachgewiesen werden, dass **das System auch in den nach IFA Ampelschema kritischen Gelenkwinkelbereichen keine ausreichende Genauigkeit bietet.**

7.2. Fragestellung II - Bewegungserkennung

Mit welcher Genauigkeit können Körperhaltungen und Bewegungen auf Basis von Motion-Capture-Daten erkannt werden?

In Abschnitt 5.5 wurde eine framewise Erkennung von Körperhaltungen auf Basis von SVM und ANN untersucht. Für das Training der Modelle wurden die Daten von

einer Person eingesetzt und künstlich erweitert. Getestet wurden die Modelle mit Daten von 20 davon unabhängigen Personen. Dabei konnten **Erkennungsgenauigkeiten und Trainingszeiten auf einem mobilen Endgerät (Laptop) von [SVM 87 %, 4:17 min; ANN 89 %, 19:47 min] bei einem 20-Klassen-Problem** erreicht werden. Mit dem Modell auf Basis von ANN können somit etwas höhere Genauigkeiten erzielt werden, jedoch ist für das Training ein höherer Rechenaufwand erforderlich.

Um herauszufinden, wie sich die Erkennungsleistung insbesondere bei kritischen Körperhaltungen verhält, wurden die Modelle erneut mit einem Sub-Datensatz von ausgewählten Klassen (Ganzkörperbewegungen) mit einem balancierten Aufkommen von Frames mit einem kritischen Score (RULA+) evaluiert. Für ein 5-Klassen-Problem und Körperhaltungen mit einem Score ≥ 6 (*leichtes Risiko*) konnten Genauigkeiten von [SVM 86 %; ANN 94 %] erzielt werden.

Kann eine Bewegungssequenz mit dem in Abschnitt 5.6 beschriebenen Algorithmus zeitlich segmentiert werden, ist es anhand der entwickelten Modelle möglich, die überwiegende Anzahl der Frames zugehörig zu einer Bewegungssequenz richtig zu klassifizieren und somit eindeutig zu identifizieren. Die evaluierten Bewegungssequenzen, welche Frames (Körperhaltungen) mit einem Score ≥ 6 enthalten, konnten auf Basis der überwiegend richtig klassifizierten einzelnen Frames zu 100 % richtig erkannt werden.

Anforderung an den Algorithmus für die Segmentierung ist, dass Start- und Endzeitpunkt einer Bewegung entweder in der Neutral-Null-Stellung liegen oder ein klarer Richtungswechsel in einer Gelenkwinkelamplitude eintritt. Aufgrund dieser Anforderungen ist ein aktives Eingreifen in die Bewegungsabläufe, insbesondere bei der Analyse von Bewegungen am Arbeitsplatz, von den Nutzer*innen erforderlich. In einem Interview mit Therapeut*innen und Expert*innen für Arbeitsplatzergonomie wurde dieses Verhalten als unkritisch eingestuft und gilt in einer technischen Beobachtungssituation als tolerierbar. Ein wesentlicher Einfluss des angepassten Bewegungsablaufs auf die Ergebnisse in der Bewegungsbewertung ist nicht zu erwarten.

7.3. Fragestellung III - Bewegungsbewertung

Ist eine Bewegungsbewertung auf Basis physiologischer Gelenkwinkel und erkannter Körperhaltungen mit Einsatz von Motion-Capture-Sensorik in einem mobilen System möglich?

Durch das im Zuge dieser Arbeit entwickelte Framework ist es möglich, Bewegungen zeitlich zu segmentieren, zu erkennen und den zeitlichen Verlauf anhand von Schemata für die Bewegungsbewertung auf Basis von Gelenkwinkeln zu bewerten. Dadurch wird erreicht, dass Arbeitsprozesse oder Bewegungstraining statistisch erfasst, der Typ der Bewegung, die Anzahl von Wiederholungen sowie die Dauer in kritischen Körperhaltungen automatisch dokumentiert werden können. Die entwickelten Verarbeitungsprozesse bieten sowohl die Möglichkeit, MoCap-Daten offline zu analysieren, als auch einen Sensorstream in Echtzeit auszuwerten. Dadurch bietet der Einsatz der automatischen Bewegungsbewertung das Potenzial zur Optimierung von Bewegungsabläufen und zur Vermeidung von Fehlbelastungen bzw. Schädigungen des MSS.

Grundlage für die Anwendung ist eine mobil einsetzbare MoCap-Sensorik. In dieser Arbeit wurde das Framework mit einer RGB-Tiefenkamera und mit MoCap-Daten eines Lighthouse Tracking Systems für VR-Anwendungen getestet. Aus technischer Sicht ist die Nutzbarkeit mit einem Lighthouse Tracking System möglich, jedoch kann aufgrund der ermittelten Ungenauigkeiten bei der Gelenkwinkelmessung (Abschnitt 7.1.3) keine korrekte Bewegungsbewertung erreicht werden. Die Systemanforderung der MoCap-Sensorik ist dabei im Wesentlichen ausschlaggebend für die Nutzbarkeit der entwickelten Datenverarbeitung auf einem mobilen Endgerät. Eine nennenswerte Erhöhung der Rechenauslastung und eine Zunahme der Latenz, neben der von der Sensorik bekannten Latenz, konnte nicht festgestellt werden. Wie in Abschnitt 5.7.3 beschrieben, ist der **Einsatz der automatischen Bewegungsbewertung auf einem mobilen Endgerät (Laptop) mit einer RGB-Tiefenkamera (Kinect v2) möglich.**

Anzumerken ist, dass ein technisches System eine Therapie, ein Training oder die Ergonomiebewertung nur unterstützen kann. Individuelle Faktoren wie krankheitsbedingte Einschränkungen, Belastungsfaktoren der Umgebung, Konzentrationsvermögen und die körperliche Konstitution können nicht durch das System,

sondern nur durch Therapeut*innen oder Ergonomieexpert*innen berücksichtigt werden. Die Ergebnisse der vorgestellten automatischen Bewegungsbewertung sind ausdrücklich als Empfehlungen zu betrachten.

7.4. Personenbezogene Informationen in MoCap-Daten

Wie in [Kwon and Lee, 2020] gezeigt, ist es möglich, eine Person anhand der anthropometrischen Informationen (Gliedmaßenlängen) eindeutig auf Basis von MoCap-Daten zu identifizieren. Aus diesem Grund ist es notwendig, datenschutzrechtliche Aspekte bei der Verarbeitung von MoCap-Daten in einem System jederzeit zu berücksichtigen. Obwohl in dieser Arbeit nicht explizit untersucht, wird im Folgenden der Personenbezug in Gelenkwinkeldaten diskutiert.

Die Erkennung, Segmentierung und Bewertung basiert in dieser Arbeit auf berechneten Gelenkwinkeln. Für die Informationserhaltung aus einer Bewegungsbewertung wäre es demnach möglich, nur die berechneten Winkel zu speichern. Eine Rekonstruktion der Gelenkpunkte aus den Gelenkwinkeln ist nur bedingt möglich. Nach aktuellem Stand der Recherche konnte keine Veröffentlichung gefunden werden, welche eine Identifizierung von Personen auf Basis von Gelenkwinkeldaten untersucht hat. Es kann die Annahme getroffen werden, dass charakteristische Bewegungsauffälligkeiten, hervorgerufen z.B. durch Bewegungsbeeinträchtigungen, einen gewissen Wiedererkennungswert in den Gelenkwinkel-Verläufen bieten. Dieser ist jedoch mit geringerem Anteil von personenbezogenen Informationen gegenüber MoCap-Daten zu bewerten. Ein System, welches menschliche Bewegungen analysieren und Ergebnisse speichern kann und dabei sicherstellt, keine personenbezogenen Informationen von Nutzer*innen zu enthalten, ist zum aktuellen Zeitpunkt nur als Ausblick zu betrachten. Jedoch bietet die Speicherung der Informationen auf Basis von Gelenkwinkeldaten einen potenziellen Ansatz zur Sicherstellung des (privatheitsbewahrenden) Datenschutzes.

8. Fazit und Ausblick

Im Folgenden werden die erreichten Ergebnisse dieser Arbeit zusammengefasst und jeweils zu den untersuchten Zielsetzungen ein Fazit und Ausblick gegeben.

Insgesamt wurden in dieser Dissertation die Sensorgenauigkeit und Anwendung eines RGB-Tiefenkamera Systems und eines Lighthouse Tracking Systems im Kontext der Bewegungsbewertung untersucht. Es wurde eine Berechnungsgrundlage für die Bestimmung von physiologischen Gelenkwinkeln entwickelt. Zudem wurden Methoden (SVM, ANN) für die automatische Bewegungserkennung evaluiert und ein Algorithmus für die zeitliche Segmentierung von Bewegungssequenzen entworfen. Des Weiteren wurden Schemata für die Bewegungsbewertung digitalisiert und insgesamt ein Framework für die Echtzeit- und Offline-Analyse von Bewegungen implementiert. Das entwickelte Framework wurde auf einem mobilen Endgerät mit einer prototypischen Benutzeroberfläche integriert und getestet.

Die **Sensorgenauigkeit** der verwendeten MoCap-Sensorik hat insbesondere für die Gelenkwinkelmessung und Bewegungsbewertung einen hohen Stellenwert. MoCap-Sensorik basierend auf einer RGB-Tiefenkamera und entsprechender Software für das Skeleton Tracking bieten eine ausreichende Genauigkeit. Der technologisch begrenzte Messbereich (Blickwinkel) führt jedoch zur Einschränkung des Einsatzes. Bei einem Lighthouse Tracking System in Kombination mit einer IK konnte eine Ungenauigkeit in einem nicht mehr tolerierbaren Maß festgestellt werden. Aufgrund hoher Genauigkeiten von einzelnen Trackern (ohne IK) bei räumlichen Positionen und Drehungen bieten Lighthouse Tracking Systeme jedoch das Potenzial, den Fehler durch eine Anpassung der Berechnungsmethode zu reduzieren. Wie in [Sitole et al., 2020] untersucht, ist die Berechnung von Gelenkwinkeln auch auf Basis der absoluten Rohdaten von Trackern mit hohen Genauigkeiten möglich. Denkbar wäre ein hybrider Ansatz für die Berechnung von Gelenkwinkeln auf Basis der MoCap-Daten (berechnet durch die IK) und den Rohdaten der Tracker für eine Optimierung der Genauigkeit.

Für die **Bewegungserkennung** wurden Modelle auf Basis von ANN und SVMs untersucht und hohe Genauigkeiten für die Erkennung von kritischen Körperhaltungen (nach RULA+) identifiziert. Durch ein Verfahren für die künstliche Erweiterung der Daten konnte erreicht werden, dass die Modelle mit Daten von nur einer Person trainiert werden können. Dadurch wird eine hohe Flexibilität für die Anpassbarkeit der Erkennungsaufgabe geboten. Bei dem Modell auf Basis von ANN wurde eine geringfügig bessere Erkennungsgenauigkeit bei höherem Rechenaufwand erreicht. Des Weiteren bieten Modelle auf Basis von ANN die Möglichkeit von *Re-Training* und *Transfer Learning*, wodurch eine Verbesserung der Modelle durch weitere Daten oder eine Anpassung der Bewegungserkennung erreicht werden kann. Aufgrund der genannten Vorteile wird in dieser Arbeit die Verwendung von ANN für die frameweise Erkennung von Körperhaltungen bevorzugt.

Der Algorithmus für die automatische **Segmentierung** von Bewegungssequenzen bietet eine Möglichkeit, Start- und Endzeitpunkte zu identifizieren. Voraussetzung ist jedoch, dass der Bewegungsablauf einen eindeutigen Richtungswechsel beinhaltet oder mit der Neutral-Null-Stellung startet und endet. Durch ein akustisches oder visuelles Feedback bei der Identifizierung von Start- und Endzeitpunkten kann die Nutzbarkeit für Anwender*innen verbessert werden. Für abwechslungsreiche und sporadische Bewegungskombinationen ist die Segmentierung nicht geeignet. Bewegungswiederholungen bei Bewegungstraining oder in Arbeitsprozessen können unter den genannten Anforderungen robust segmentiert werden.

Die entwickelte **Bewegungsbewertung** bietet die Möglichkeit, Bewegungen auf Basis von MoCap-Daten und digitalisierten Schemata mit Regeln für kritische Gelenkwinkelbereiche zu beurteilen. Für weiterführende Entwicklungsschritte empfiehlt sich die Untersuchung der Simulation von Gewichten und biomechanischen Belastungen. Des Weiteren könnten Biosignale wie Herzfrequenz und Körpertemperatur, welche sich mittlerweile auch durch Geräte im Consumer-Bereich messen lassen, einbezogen werden, um Erschöpfungen und Überbelastung zu erkennen und rechtzeitig zu vermeiden.

Insgesamt ist festzuhalten, dass die zentrale Hypothese belegt werden kann und somit die Erkennung und Bewertung von Bewegungen und Körperhaltungen mit Einsatz von MoCap-Sensorik aus dem Consumer-Segment mit den genannten Ein-

schränkungen möglich ist. Für die Anwendung mit einem Lighthouse Tracking System sind jedoch Optimierungen notwendig, um das Fehlerpotenzial bei der Gelenkwinkelmessung zu reduzieren und eine Genauigkeit in einem tolerierbaren Bereich zu erreichen.

In einer abschließenden Konsultierung von Expert*innen aus dem medizinischen Bereich zur Vorstellung der entwickelten Funktionalität und Systemleistung konnte eine äußerst positive Resonanz eingeholt werden. Abschließend kann festgestellt werden, dass die dargestellten Ergebnisse und das entwickelte System für die automatische Bewegungsbewertung mit mobil einsetzbarer MoCap-Sensorik einen hervorragenden Ausgangspunkt bilden, um potentiell schädliche Bewegungen und Körperhaltungen automatisch zu identifizieren. Insbesondere durch das hohe Maß an Flexibilität (Anpassung der Bewertungsschemata, Erweiterung der Bewegungserkennung, Austausch der MoCap-Sensorik) birgt das System hohes Potenzial, um Erkrankungen des Muskel-Skelett-Systems (MSS) vorzubeugen oder gar vermeiden zu können. Zudem ist es denkbar, dass mithilfe des Systems ein schnellerer Rehabilitationserfolg mit konsekutiver Wiedereingliederung in den Beruf erreicht und die Zufriedenheit der Betroffenen erhöht werden könnte.

Zukünftig ist für eine weiterführende Forschung eine Untersuchung der Wirksamkeit und des Nutzungsverhaltens des in dieser Arbeit entwickelten Assistenzsystems zu empfehlen.

A. Anhang

A.1. Benutzeroberflächen der eigens entwickelten Applikationen

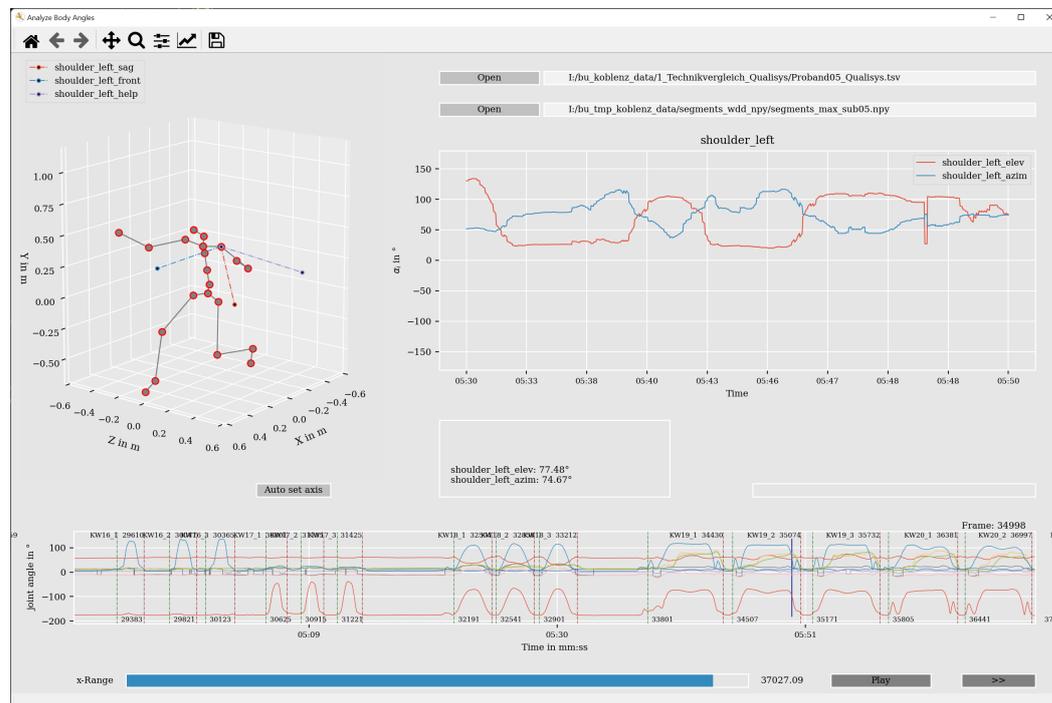


Abbildung A.1.: Eigens entwickelte Software für die Berechnung und Analyse von Gelenkwinkeln in MoCap-Daten verschiedener Sensortypen; Skelettdarstellung anhand von Gelenkpunkten mit Einzeichnung der Körperebenen (oben, links); Ausschnitt der Zeitreihendarstellung der jeweils analysierten Gelenkwinkel (oben, rechts); vollständige Zeitreihendarstellung und Einzeichnung zeitlicher Segmente (unten)

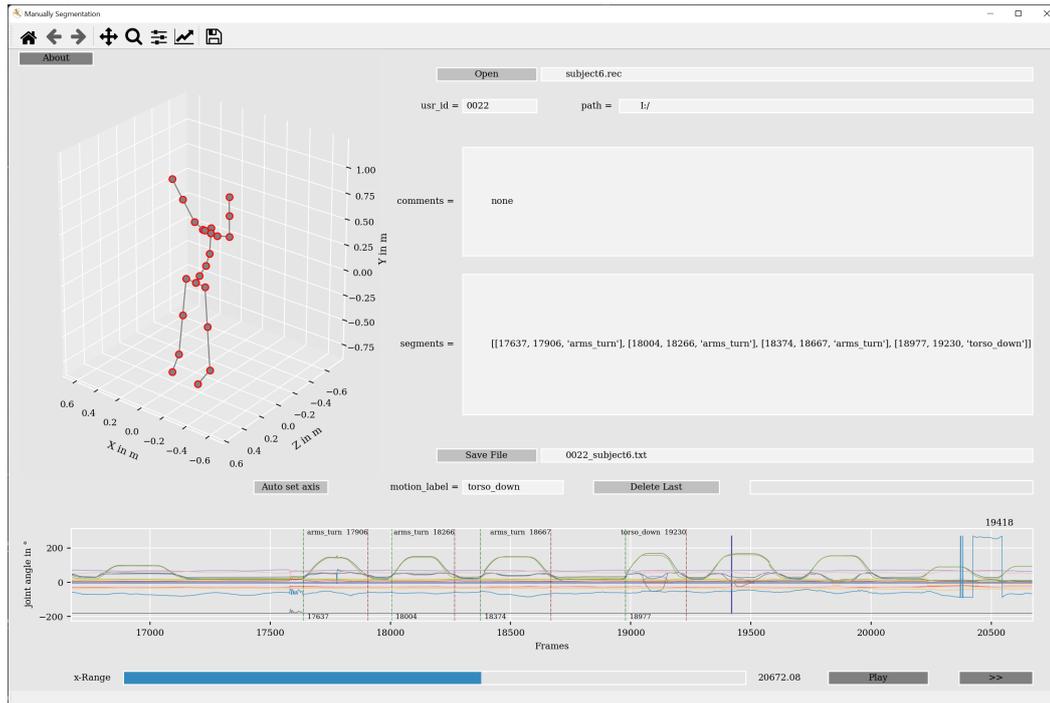


Abbildung A.2.: Eigens entwickelte Software für die manuelle Annotation von MoCap-Daten; Benennung der Bewegung (Label) und Setzen von Start- und Endzeitpunkt (Segmentierung) in den Zeitreihendaten

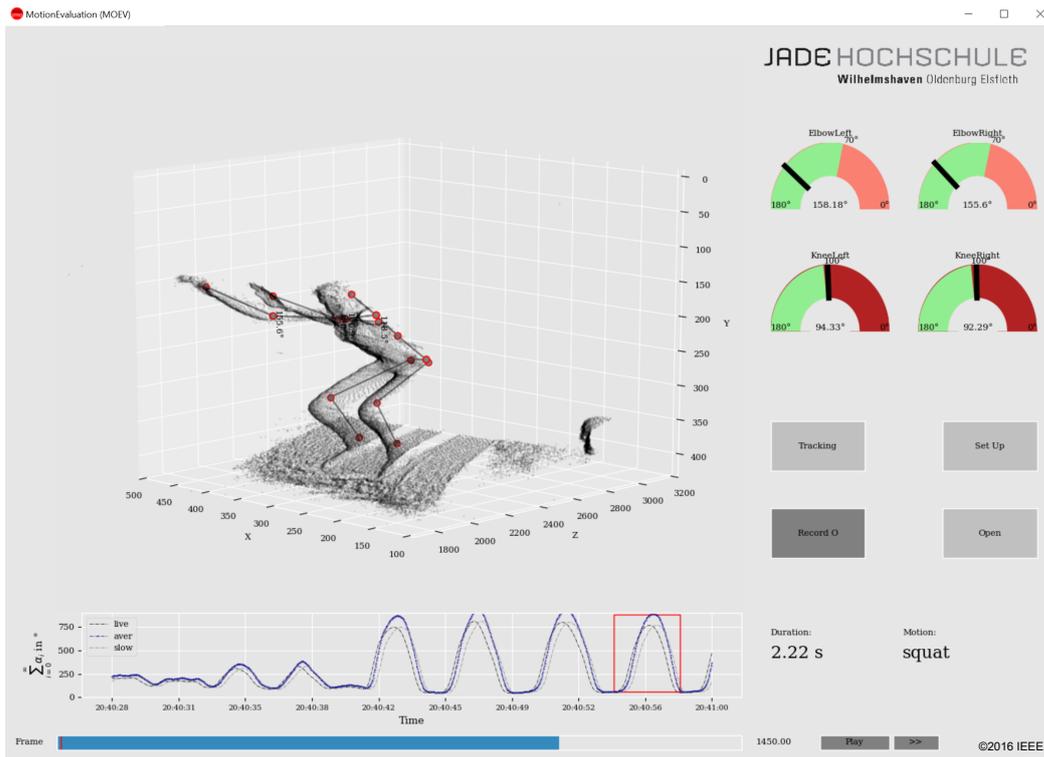


Abbildung A.3.: Eigens entwickelte Software-Benutzeroberfläche für die Anwendung der Bewegungsbewertung mit der Kinect v2 in einem mobilen Endgerät; zeigt eine 3D-Skelett-Ansicht mit einem Probanden in hockender Haltung mit auf Kopfhöhe ausgestreckten Armen; summierte Gelenkwinkel in Zeitreihen mit dem Segment der erkannten Bewegung (roter Kasten), Gauge-Darstellung mit kritischen Gelenkwinkelbereichen. Die Dauer und das Label werden in einem Textfeld (rechts unten) angezeigt. In der aktuellen Version sind vier Funktionen (Tracking, Aufzeichnung, System einrichten, Datei öffnen) implementiert [Vox and Wallhoff, 2019, Fig. 4]

A.2. Statistik der analysierten Gelenkwinkel von Vive und Qualisys

A.3. Weitere Publikationen

- Vox, J. P. and Wallhoff, F. (2017a). Autonomous Learning of Motion Sequences within a Human-Robot Interaction. In *6. Interdisziplinärer Workshop Kognitive Systeme (KogSys); Menschen, Teams, Systeme und Automaten, Neubiberg, Germany*, pages 1–6, (Paper in Tagungsband, Präsentation)
- Theuerkauff, T., Wagner, Y., Vox, J. P., Wolf, K. I., and Wallhoff, F. (2017). Robust adaptable human-robot dialogues for production process. In *6. Interdisziplinärer Workshop Kognitive Systeme (KogSys); Menschen, Teams, Systeme und Automaten, Neubiberg, Germany*, pages 1–5, (Paper in Tagungsband, Präsentation)
- Ohlenbusch, M., Bartner, N. F. H., Vöge, S., Vox, J. P., Fleßner, J., Frenken, M., and Wallhoff, F. (2018). Installation and Control of Building Automation Systems Using Human-Robot-Interaction. In *2018 23rd IEEE International Conference on Methods Models in Automation Robotics (MMAR), Miedzydroje, Poland*, pages 493–497, <http://dx.doi.org/10.1109/MMAR.2018.8485854>, (Konferenzpaper, Poster)
- Tessmer, J., Vox, J. P., and Wallhoff, F. (2019). Human robot interaction with the purpose to motivate for performing movement exercises. In *8. Interdisziplinärer Workshop Kognitive Systeme (KogSys); Verstehen, Beschreiben und Gestalten von komplexen Mensch-Maschine-Szenarien, Duisburg-Essen, Germany*, page 18f <https://doi.org/10.17185/dupublico/48470>, (Abstractband)
- Wallhoff, F., Vox, J. P., and Theuerkauff, T. (2019). *Assistenz- und Servicerobotik – die Gestaltung der Mensch-Maschine-Schnittstelle als Grundlage des Anwendungserfolgs*, pages 99–122. Springer Berlin Heidelberg, Berlin, Heidelberg https://doi.org/10.1007/978-3-662-57611-3_7, (Buchkapitel)

Literaturverzeichnis

- [Albert et al., 2020] Albert, J. A., Owolabi, V., Gebel, A., Brahms, C. M., Granacher, U., and Arnrich, B. (2020). Evaluation of the pose tracking performance of the azure kinect and kinect v2 for gait analysis in comparison with a gold standard: A pilot study. *Sensors*, 20(18).
- [Ameler et al., 2019] Ameler, T., Blohme, K., Brandt, L., Brüngel, R., Hensel, A., Huber, L., Kuper, F., Swoboda, J., Warnecke, M., Warzecha, M., Heß, D., Frömke, J., Schmitz-Stolbrink, A., and Friedrich, C. M. (2019). A comparative evaluation of steamvr tracking and the optitrack system for medical device tracking. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1465–1470.
- [Amell and Kumar, 2001] Amell, T. and Kumar, S. (2001). Work-related musculoskeletal disorders: design as a prevention strategy. a review. *Journal of occupational rehabilitation*, 11(4):255–265.
- [Amstutz et al., 2018] Amstutz, N., Geisen, T., Hassler, B., Diezi, J., Widmer, L., Steiner, L., Kraus, K., and Wenger, N. (2018). „Arbeiten, solange der Körper mitmacht“: Betriebliche Herausforderungen im Zusammenhang mit Einfacharbeit und alternden Belegschaften. *Arbeit*, 27(1):5–25.
- [Bajpai and Joshi, 2021] Bajpai, R. and Joshi, D. (2021). Movenet: A deep neural network for joint profile prediction across variable walking speeds and slopes. *IEEE Transactions on Instrumentation and Measurement*, 70:1–11.
- [Bauer et al., 2021a] Bauer, P., Lienhart, W., and Jost, S. (2021a). Accuracy investigation of the pose determination of a vr system. *Sensors*, 21(5).
- [Bauer et al., 2021b] Bauer, P., Lienhart, W., and Jost, S. (2021b). Genauigkeitsuntersuchung eines VR-Systems zur 3D-Koordinatenbestimmung. *AVN. Allgemeine Vermessungsnachrichten*, 128(1):3–12.

- [Bazarevsky et al., 2020] Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., and Grundmann, M. (2020). Blazepose: On-device real-time body pose tracking.
- [Benesty et al., 2009] Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). *Pearson Correlation Coefficient*, pages 1–4. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Bland et al., 1986] Bland, J. M., Altman, D., Martin Bland, J., and Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The lancet*, 327(8476):307–310. Originally published as Volume 1, Issue 8476.
- [BMAS, 2020] BMAS (2020). Bundesministerium für Arbeit und Soziales. Sicherheit und Gesundheit bei der Arbeit–Berichtsjahr 2019. Unfallverhütungsbericht Arbeit.
- [Boone et al., 1978] Boone, D. C., Azen, S. P., Lin, C.-M., Spence, C., Baron, C., and Lee, L. (1978). Reliability of Goniometric Measurements. *Physical Therapy*, 58(11):1355–1360.
- [Borges et al., 2018] Borges, M., Symington, A., Coltin, B., Smith, T., and Ventura, R. (2018). Htc vive: Analysis and accuracy improvement. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2610–2615.
- [Brenscheidt et al., 2021] Brenscheidt, S., Siefer, A., Hinnenkamp, H., Hünefeld, L., Meyer, S.-C., and N., B. (2021). Arbeitswelt im Wandel: Zahlen - Daten - Fakten (2021). Ausgabe 2021.
- [Büker et al., 2021] Büker, L. C., Zuber, F., Hein, A., and Fudickar, S. (2021). HRDepthNet: Depth Image-Based Marker-Less Tracking of Body Joints. *Sensors*, 21(4).
- [Cao et al., 2021] Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., and Sheikh, Y. (2021). Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):172–186.

- [Capecci et al., 2016] Capecci, M., Ceravolo, M., Ferracuti, F., Iarlori, S., Kyrki, V., Longhi, S., Romeo, L., and Verdini, F. (2016). Physical rehabilitation exercises assessment based on hidden semi-markov model by kinect v2. pages 256–259.
- [Caserman et al., 2019] Caserman, P., Achenbach, P., and Göbel, S. (2019). Analysis of inverse kinematics solutions for full-body reconstruction in virtual reality. pages 1–8.
- [Chen et al., 2015] Chen, C., Jafari, R., and Kehtarnavaz, N. (2015). Utd-mhad: A multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. pages 168–172.
- [Chen et al., 2018] Chen, D., Liu, H., and Ren, Z. (2018). Application of wearable device htc vive in upper limb rehabilitation training. pages 1460–1464.
- [Cho and Chen, 2014] Cho, K. and Chen, X. (2014). Classifying and visualizing motion capture sequences using deep neural networks. *CoRR*, abs/1306.3874.
- [Cippitelli et al., 2016] Cippitelli, E., Gasparri, S., Gambi, E., and Spinsante, S. (2016). A human activity recognition system using skeleton data from rgb-d sensors. *Computational Intelligence and Neuroscience*, 2016:1–14.
- [Clark et al., 2018] Clark, R., Mentiplay, B., Hough, E., and Pua, Y.-H. (2018). Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and kinect alternatives. *Gait and Posture*, 68.
- [Delp et al., 2007] Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E., and Thelen, D. G. (2007). Opensim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering*, 54(11):1940–1950.
- [Destatis, 2021] Destatis (2021). Statistisches Bundesamt. 14. koordinierte Bevölkerungsvorausberechnung für Deutschland. <https://service.destatis.de/bevoelkerungspyramide/>. aufgerufen am 20. Juli 2021.
- [DGOU, 2021] DGOU (2021). Deutsche Gesellschaft für Orthopädie und Unfallchirurgie. Nachbehandlungsempfehlung 2021. *Arbeitskreis Traumarehabilitation Sektion Physikalische Therapie und Rehabilitation der DGOU*, 7.

- [DGUV, 2018] DGUV (2018). Deutsche Gesetzliche Unfallversicherung. Muskel-Skelett-System. <https://www.dguv.de/de/praevention/themen-a-z/muskel-skelett-system/index.jsp>. aufgerufen am 20. Juli 2021.
- [DIN EN ISO 7250-1, 2017] DIN EN ISO 7250-1 (2017). Wesentliche Maße des menschlichen Körpers für die technische Gestaltung - Teil 1: Körpermaßdefinitionen und -messpunkte. *Deutsches Institut für Normung e.V., Berlin*.
- [DIN ISO 8855, 2013] DIN ISO 8855 (2013). Straßenfahrzeuge - Fahrzeugdynamik und Fahrverhalten. *Deutsches Institut für Normung e.V., Berlin*.
- [Doorenbosch et al., 2003] Doorenbosch, C., Harlaar, J., and Veeger, D. (2003). The globe system: An unambiguous description of shoulder positions in daily life movements. *Journal of rehabilitation research and development*, 40:147–55.
- [Du et al., 2015] Du, Y., Wang, W., and Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. pages 1110–1118.
- [Duan et al., 2020] Duan, Y., Zhang, X., and Li, Z. (2020). A new quaternion-based kalman filter for human body motion tracking using the second estimator of the optimal quaternion algorithm and the joint angle constraint method with inertial and magnetic sensors. *Sensors*, 20(21).
- [D’Antonio et al., 2020] D’Antonio, E., Taborri, J., Palermo, E., Rossi, S., and Patanè, F. (2020). A markerless system for gait analysis based on openpose library. pages 1–6.
- [Ellegast et al., 2009] Ellegast, R., Hermanns, I., and Schiefer, C. (2009). Workload assessment in field using the ambulatory cuela system. In Duffy, V. G., editor, *Digital Human Modeling*, pages 221–226, Berlin, Heidelberg. Springer, Springer Berlin Heidelberg.
- [Ellegast et al., 2010] Ellegast, R., Hermanns, I., and Schiefer, C. (2010). *Feldmesssystem CUELA zur Langzeiterfassung und-analyse von Bewegungen an Arbeitsplätzen*, volume 375. GRIN Verlag.
- [Fehr and Slesina, 2004] Fehr, M. and Slesina, W. (2004). Berufsbezogene Informationen zur Steigerung des Reha-Erfolgs bei Muskel-Skelett-Erkrankungen. *Das Gesundheitswesen*, 66(08/09):14.

- [Feichtenhofer et al., 2016] Feichtenhofer, C., Pinz, A., and Wildes, R. P. (2016). Spatiotemporal residual networks for video action recognition. *CoRR*, abs/1611.02155.
- [Gaglio et al., 2015] Gaglio, S., Re, G. L., and Morana, M. (2015). Human activity recognition process using 3-d posture data. *IEEE Transactions on Human-Machine Systems*, 45(5):586–597.
- [Gajdosik and Bohannon, 1987] Gajdosik, R. L. and Bohannon, R. W. (1987). Clinical Measurement of Range of Motion: Review of Goniometry Emphasizing Reliability and Validity. *Physical Therapy*, 67(12):1867–1872.
- [Gardner and Dorling, 1998] Gardner, M. and Dorling, S. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14):2627–2636.
- [Gellert et al., 2018] Gellert, F. J., Kesselmann, M., and Wilke, C. B. (2018). Arbeitswelt im wandel. *Prävention und Gesundheitsförderung*, 13(1):12–17.
- [Gesnouin et al., 2020] Gesnouin, J., Pechberti, S., Bresson, G., Stanciulescu, B., and Moutarde, F. (2020). Predicting intentions of pedestrians from 2d skeletal pose sequences with a representation-focused multi-branch deep learning network. *Algorithms*, 13(12).
- [Gogia et al., 1987] Gogia, P. P., Braatz, J. H., Rose, S. J., and Norton, B. J. (1987). Reliability and Validity of Goniometric Measurements at the Knee. *Physical Therapy*, 67(2):192–195.
- [Göpfert et al., 2009] Göpfert, B., Frère, J., Huber, C., Nüesch, C., and Romkes, J. (2009). Bewegungsanalyse: wichtige Informationen für Sport und Alltag. 9(48):875–8.
- [Guffanti et al., 2020] Guffanti, D., Brunete, A., Hernando, M., Rueda, J., and Navarro Cabello, E. (2020). The accuracy of the microsoft kinect v2 sensor for human gait analysis. a different approach for comparison with the ground truth. *Sensors*, 20(16).
- [Haller, 2007] Haller, N. (2007). *Bewegungsgeschwindigkeit im Krafttraining : Konzentrische und exzentrische Muskelarbeitsweise und deren Metabolismus*. PhD thesis, Universität Konstanz, Konstanz.

- [Hearst et al., 1998] Hearst, M., Dumais, S., Osuna, E., Platt, J., and Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28.
- [Hignett and McAtamney, 2000] Hignett, S. and McAtamney, L. (2000). Rapid entire body assessment (reba). *Applied Ergonomics*, 31(2):201–205.
- [Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.
- [Huang et al., 2016] Huang, G., Liu, Z., and Weinberger, K. Q. (2016). Densely connected convolutional networks. *CoRR*, abs/1608.06993.
- [Huang et al., 2006] Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501. Neural Networks.
- [Hämel and Schaeffer, 2013] Hämel, K. and Schaeffer, D. (2013). Who cares? fachkräftemangel in der pflege. *Zeitschrift für Sozialreform*, 59(4):413–432.
- [icura, 2021] icura (2021). Digitally supported training for everyone. User-friendly and motivating training technology. <https://www.icura.dk/>. aufgerufen am 20. Juli 2021.
- [IFA, 2015] IFA (2015). Institut für Arbeitsschutz der Deutschen Gesetzlichen Unfallversicherung. Bewertung physischer Belastungen gemäß DGUV-Information 208-033 (bisher: BGI/GUV-I 7011) (Anhang3). *Deutsche Gesetzliche Unfallversicherung (Hrsg)*.
- [IFM, 2021] IFM (2021). Institut für Mechatronik e.V. Chemnitz. Mensch-Technik-Interaktion, Dynamicus-tools. <https://www.ifm-chemnitz.de/produkte/mensch-technik-interaktion/dynamicus-tools/>. aufgerufen am 20. Juli 2021.
- [Ikbal et al., 2021] Ikbal, M. S., Ramadoss, V., and Zoppi, M. (2021). Dynamic pose tracking performance evaluation of htc vive virtual reality system. *IEEE Access*, 9:3798–3815.
- [Ionescu et al., 2014] Ionescu, C., Papava, D., Olaru, V., and Sminchisescu, C. (2014). Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339.

- [Ivorra et al., 2021] Ivorra, E., Pérez, M., and Alcañiz Raya, M. (2021). Azure kinect body tracking under review for the specific case of upper limb exercises. *MM Science Journal*, 2021:4333–4341.
- [Jäger et al., 2001] Jäger, M., Luttmann, A., Göllner, R., and Laurig, W. (2001). "the dortmunder biomechanical model for quantification and assessment of the load on the lumbar spine. *SAE Transactions*, 110:2163–2171.
- [Jürgens et al., 1998] Jürgens, H., Matzdorff, I., Windberg, J., Albert, J. A., Owolabi, V., Gebel, A., Brahms, C. M., Granacher, U., and Arnrich, B. (1998). Internationale anthropometrische daten als voraussetzung für die gestaltung von arbeitsplätzen und maschinen. *Arbeitswissenschaftliche Erkenntnisse, Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, Dortmund*, 108.
- [Kale and Patil, 2016] Kale, G. V. and Patil, V. H. (2016). A study of vision based human motion recognition and analysis. *CoRR*, abs/1608.06761.
- [Karhu et al., 1977] Karhu, O., Kansu, P., and Kuorinka, I. (1977). Correcting working postures in industry: a practical method for analysis. *Applied ergonomics*, 8(4):199–201.
- [Kay et al., 2017] Kay, W., Carreira, J., Simonyan, K., Zhang, B., Hillier, C., Vijayanarasimhan, S., Viola, F., Green, T., Back, T., Natsev, P., Suleyman, M., and Zisserman, A. (2017). The kinetics human action video dataset. *CoRR*, abs/1705.06950.
- [Kinetisense, 2021] Kinetisense (2021). Kinetisense is Your Portable Biomechanics Lab. Kinetisense Turns the Movement Theorist into a Movement Scientist. University Validated for Accuracy and Objectivity. <https://www.kinetisense.com/>. aufgerufen am 20. Juli 2021.
- [Kjellberg et al., 2016] Kjellberg, K., Lundin, A., Falkstedt, D., Allebeck, P., and Hemmingsson, T. (2016). Long-term physical workload in middle age and disability pension in men and women: a follow-up study of swedish cohorts. *International archives of occupational and environmental health*, 89(8):1239–1250.
- [KKH, 2008] KKH (2008). *Kaufmännische Krankenkasse Hannover. Beweglich? : Muskel-Skelett-Erkrankungen - Ursachen, Risikofaktoren und präventive Ansätze. Weißbuch Prävention ; 2007/08. 1st ed. 2008. edition.*

- [Kong et al., 2018] Kong, Y.-K., yong Lee, S., Lee, K.-S., and Kim, D.-M. (2018). Comparisons of ergonomic evaluation tools (alla, rula, reba and owas) for farm work. *International Journal of Occupational Safety and Ergonomics*, 24(2):218–223. PMID: 28301984.
- [Kraft, 2019] Kraft, S. (2019). Welche Auswirkungen hat der Fachkräftemangel? *Pflegezeitschrift*, 72(6):58–59.
- [Kwon and Lee, 2020] Kwon, B. and Lee, S. (2020). Human skeleton data augmentation for person identification over deep neural network. *Applied Sciences*, 10(14).
- [Landis and Koch, 1977] Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
- [Laraba et al., 2015] Laraba, S., Tilmanne, J., and Dutoit, T. (2015). Adaptation procedure for hmm-based sensor-dependent gesture recognition. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, MIG '15, page 17–22, New York, NY, USA. Association for Computing Machinery.
- [Leung and Haykin, 1991] Leung, H. and Haykin, S. (1991). The complex backpropagation algorithm. *IEEE Transactions on Signal Processing*, 39(9):2101–2104.
- [Li et al., 2020] Li, L., Martin, T., and Xu, X. (2020). A novel vision-based real-time method for evaluating postural risk factors associated with musculoskeletal disorders. *Applied Ergonomics*, 87:103138.
- [Li et al., 2019] Li, R., Si, W., Weinmann, M., and Klein, R. (2019). Constraint-based optimized human skeleton extraction from single-depth camera. *Sensors*, 19(11).
- [Li et al., 2010] Li, W., Zhang, Z., and Liu, Z. (2010). Action recognition based on a bag of 3d points. pages 9–14.
- [Liebers et al., 2013] Liebers, F., Brendler, C., and Latza, U. (2013). Alters-und berufsgruppenabhängige Unterschiede in der Arbeitsunfähigkeit durch häufige Muskel-Skelett-Erkrankungen. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 56(3):367–380.
- [Lin et al., 2021a] Lin, K., Wang, L., Jin, Y., Liu, Z., and Sun, M.-T. (2021a). Learning nonparametric human mesh reconstruction from a single image without ground truth meshes. pages 964–968.

- [Lin et al., 2021b] Lin, K., Wang, L., and Liu, Z. (2021b). End-to-end human pose and mesh reconstruction with transformers. *CoRR*, abs/2012.09760.
- [Lin, 1989] Lin, L. I.-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255–268.
- [Lins, 2021] Lins, C. (2021). *Evolutionär optimierte Haltungs- und Bewegungsmodelle auf Basis von Motion-Capture-Daten als Teil gesundheitsbezogener Assistenzsysteme*. PhD thesis.
- [Liu et al., 2007] Liu, H., Holt, C., and Evans, S. (2007). Accuracy and repeatability of an optical motion analysis system for measuring small deformations of biological tissues. *Journal of Biomechanics*, 40(1):210–214.
- [Liu, 2019] Liu, Z. (2019). 3D Skeletal Tracking on Azure Kinect – Azure Kinect Body Tracking SDK. *CVPR2019 Workshop on 3D Computer Vision in Medical Environments and ICIP2019 Microsoft Industry Workshop – Machine Learning and Computer Vision Applications (presentation)*.
- [Liu et al., 2016] Liu, Z., Zhou, L., Leung, H., and Shum, H. P. (2016). Kinect posture reconstruction based on a local mixture of gaussian process models. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2437–2450.
- [Lowe et al., 2019] Lowe, B. D., Dempsey, P. G., and Jones, E. M. (2019). Ergonomics assessment methods used by ergonomics professionals. *Applied Ergonomics*, 81:102882.
- [Lubetzky et al., 2019] Lubetzky, A. V., Wang, Z., and Krasovsky, T. (2019). Head mounted displays for capturing head kinematics in postural tasks. *Journal of Biomechanics*, 86:175–182.
- [Lockett, 2018] Lockett, E. (2018). *A quantitative evaluation of the htc vive for virtual reality research*. PhD thesis, The University of Mississippi.
- [Ma et al., 2019] Ma, Y., Mithraratne, K., Wilson, N. C., Wang, X., Ma, Y., and Zhang, Y. (2019). The validity and reliability of a kinect v2-based gait analysis system for children with cerebral palsy. *Sensors*, 19(7).
- [Maczewska et al., 2020] Maczewska, A., Polak-Sopinska, A., and Wisniewski, Z. (2020). Computer-aided occupational risk assessment of physical workload in the logistics 4.0. pages 378–390.

- [Manghisi et al., 2017] Manghisi, V. M., Uva, A. E., Fiorentino, M., Bevilacqua, V., Trotta, G. F., and Monno, G. (2017). Real time rula assessment using kinect v2 sensor. *Applied Ergonomics*, 65:481–491.
- [Mannerfelt et al., 2005] Mannerfelt, L., Biedermann, L., Latta, L. L., and Stoltze, K.-D. (2005). *Orthopädische Technik*, volume 9., überarbeitete und neu gestaltete Auflage. Georg Thieme Verlag, Stuttgart.
- [McAtamney and Nigel Corlett, 1993] McAtamney, L. and Nigel Corlett, E. (1993). Rula: a survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics*, 24(2):91–99.
- [McAuley et al., 1997] McAuley, J., Rothwell, J., and Marsden, C. (1997). Frequency peaks of tremor, muscle vibration and electromyographic activity at 10 hz, 20 hz and 40 hz during human finger muscle contraction may reflect rhythmicities of central neural firing. *Experimental brain research*, 114(3):525–541.
- [McBride, 2005] McBride, G. (2005). A proposal for strength-of-agreement criteria for lin’s concordance correlation coefficient. *NIWA client report: HAM2005-062*, 62.
- [Mehta et al., 2017] Mehta, D., Sridhar, S., Sotnychenko, O., Rhodin, H., Shafiei, M., Seidel, H.-P., Xu, W., Casas, D., and Theobalt, C. (2017). Vnect: Real-time 3d human pose estimation with a single rgb camera. *ACM Trans. Graph.*, 36(4).
- [Müller et al., 2007] Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B., and Weber, A. (2007). Documentation Mocap Database HDM05. Technical Report CG-2007-2, Universität Bonn.
- [Munea et al., 2020] Munea, T. L., Jembre, Y. Z., Weldegebriel, H. T., Chen, L., Huang, C., and Yang, C. (2020). The progress of human pose estimation: A survey and taxonomy of models applied in 2d human pose estimation. *IEEE Access*, 8:133330–133348.
- [Nanni et al., 2017] Nanni, L., Ghidoni, S., and Brahmam, S. (2017). Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172.
- [Niehorster et al., 2017] Niehorster, D. C., Li, L., and Lappe, M. (2017). The accuracy and precision of position and orientation tracking in the htc vive virtual rea-

- lity system for scientific research. *i-Perception*, 8(3):2041669517708205. PMID: 28567271.
- [Noble, 2006] Noble, W. S. (2006). What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567.
- [Nusser et al., 2009] Nusser, S., Otte, C., Hauptmann, W., Leirich, O., Krätschmer, M., and Kruse, R. (2009). Maschinelles lernen von validierbaren klassifikatoren zur autonomen steuerung sicherheitsrelevanter systeme (machine learning of verifiable classifiers for autonomous control of safety-related systems). *Autom.*, 57(3):138–145.
- [Offi et al., 2013] Offi, F., Chaudhry, R., Kurillo, G., Vidal, R., and Bajcsy, R. (2013). Berkeley mhad: A comprehensive multimodal human action database. pages 53–60.
- [Ohlenbusch et al., 2018] Ohlenbusch, M., Bartner, N. F. H., Vöge, S., Vox, J. P., Fleßner, J., Frenken, M., and Wallhoff, F. (2018). Installation and Control of Building Automation Systems Using Human-Robot-Interaction. In *2018 23rd IEEE International Conference on Methods Models in Automation Robotics (MMAR)*, Miedzyzdroje, Poland, pages 493–497.
- [Ota et al., 2020] Ota, M., Tateuchi, H., Hashiguchi, T., Kato, T., Ogino, Y., Yamagata, M., and Ichihashi, N. (2020). Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait Posture*, 80:62–67.
- [Papandreou et al., 2018] Papandreou, G., Zhu, T., Chen, L.-C., Gidaris, S., Tompson, J., and Murphy, K. (2018). Personlab: Person pose estimation and instance segmentation with a bottom-up, part-based, geometric embedding model.
- [Patsadu et al., 2012] Patsadu, O., Nukoolkit, C., and Watanapa, B. (2012). Human gesture recognition using kinect camera. In *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, pages 28–32.
- [Paulich et al., 2018] Paulich, M., Schepers, M., Rudigkeit, N., and Bellusci, G. (2018). Xsens mtw awinda: Miniature wireless inertial-magnetic motion tracker for highly accurate 3d kinematic applications.

- [Perumal, 2011] Perumal, L. (2011). Quaternion and its application in rotation using sets of regions. *International Journal of Engineering and Technology Innovation*, 1(1):35.
- [Pham et al., 2014] Pham, C. H., Le, Q. K., and Le, T. H. (2014). Human Action Recognition Using Dynamic Time Warping and Voting Algorithm.
- [Pham et al., 2019] Pham, H. H., Salmane, H., Khoudour, L., Crouzil, A., Zegers, P., and Velastin, S. A. (2019). Spatio-temporal image representation of 3d skeletal movements for view-invariant action recognition with deep convolutional neural networks. *Sensors*, 19(8).
- [Pjontek et al., 2013] Pjontek, R., Scheibe, F., and Tabataba, J. (2013). *Heidelberger Standarduntersuchung*, volume 2. Projekt Heidelberger Klinische Standards HeiCuMed, Heidelberg.
- [Plantard et al., 2017] Plantard, P., Shum, H. P., Le Pierres, A.-S., and Multon, F. (2017). Validation of an ergonomic assessment method using kinect data in real workplace conditions. *Applied Ergonomics*, 65:562–569.
- [Punnett and Wegman, 2004] Punnett, L. and Wegman, D. H. (2004). Work-related musculoskeletal disorders: the epidemiologic evidence and the debate. *Journal of electromyography and kinesiology*, 14(1):13–23.
- [Qualisys, 2022] Qualisys (2022). Qualisys AB. Qualisys Track Manager (QTM). <https://www.qualisys.com/software/qualisys-track-manager/>. aufgerufen am 05. Januar 2022.
- [Rashid, 2016] Rashid, T. (2016). *Make Your Own Neural Network*. CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 1st edition.
- [Richards, 1999] Richards, J. G. (1999). The measurement of human motion: A comparison of commercially available systems. *Human Movement Science*, 18(5):589–602.
- [Roetenberg et al., 2009] Roetenberg, D., Luinge, H., and Slycke, P. (2009). Xsens mvn: Full 6dof human motion tracking using miniature inertial sensors. *Xsens Motion Technologies BV, Tech. Rep*, 1.
- [Rohlmann et al., 2001] Rohlmann, A., Klöckner, C., and Bergmann, G. (2001). Biomechanik der kyphose. *Der Orthopäde*, 30(12):915–918.

- [Ryf and Weymann, 1995] Ryf, C. and Weymann, A. (1995). The neutral zero method — a principle of measuring joint function. *Injury*, 26:1–11.
- [Sabo et al., 2016] Sabo, A., Kafka, P., Scherer, M., Forjan, M., and David, V. (2016). Rehabilitation-konzepte für rehabilitation im häuslichen umfeld. *Forschungsforum der österreichischen Fachhochschule (FFH)*.
- [Sandler et al., 2018] Sandler, M., Howard, A. G., Zhu, M., Zhmoginov, A., and Chen, L. (2018). Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381.
- [Schaub et al., 2013] Schaub, K., Caragnano, G., Britzke, B., and Bruder, R. (2013). The european assembly worksheet. *Theoretical Issues in Ergonomics Science*, 14(6):616–639.
- [Schmitz et al., 2014] Schmitz, A., Ye, M., Shapiro, R., Yang, R., and Noehren, B. (2014). Accuracy and repeatability of joint angles measured using a single camera markerless motion capture system. *Journal of Biomechanics*, 47(2):587–591.
- [Seidenari et al., 2013] Seidenari, L., Varano, V., Berretti, S., Del Bimbo, A., and Pala, P. (2013). Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses. pages 479–485.
- [Shahroudy et al., 2016] Shahroudy, A., Liu, J., Ng, T., and Wang, G. (2016). NTU RGB+D: A large scale dataset for 3d human activity analysis. *CoRR*, abs/1604.02808.
- [Sharma et al., 2020] Sharma, S., Sharma, S., and Athaiya, A. (2020). Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, 04:310–316.
- [Shotton et al., 2013] Shotton, J., Sharp, T., Kipman, A., Fitzgibbon, A., Finocchio, M., Blake, A., Cook, M., and Moore, R. (2013). Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124.
- [Shuster, 2008] Shuster, M. D. (2008). The nature of the quaternion. *The Journal of the Astronautical Sciences*, 56(3):359–373.
- [Sitole et al., 2020] Sitole, S. P., LaPre, A. K., and Sup, F. C. (2020). Application and evaluation of lighthouse technology for precision motion capture. *IEEE Sensors Journal*, 20(15):8576–8585.

- [Solà, 2017] Solà, J. (2017). Quaternion kinematics for the error-state kalman filter.
- [Spitzley and Karduna, 2019] Spitzley, K. A. and Karduna, A. R. (2019). Feasibility of using a fully immersive virtual reality system for kinematic data collection. *Journal of Biomechanics*, 87:172–176.
- [Sung et al., 2012] Sung, J., Ponce, C., Selman, B., and Saxena, A. (2012). Unstructured human activity detection from rgb-d images. pages 842–849.
- [Tasnim et al., 2021] Tasnim, N., Islam, M. K., and Baek, J.-H. (2021). Deep learning based human activity recognition using spatio-temporal image formation of skeleton joints. *Applied Sciences*, 11(6).
- [Tessmer et al., 2019] Tessmer, J., Vox, J. P., and Wallhoff, F. (2019). Human robot interaction with the purpose to motivate for performing movement exercises. In *8. Interdisziplinärer Workshop Kognitive Systeme (KogSys); Verstehen, Beschreiben und Gestalten von komplexen Mensch-Maschine-Szenarien, Duisburg-Essen, Germany*, page 18f.
- [Theuerkauff et al., 2017] Theuerkauff, T., Wagner, Y., Vox, J. P., Wolf, K. I., and Wallhoff, F. (2017). Robust adaptable human-robot dialogues for production process. In *6. Interdisziplinärer Workshop Kognitive Systeme (KogSys); Menschen, Teams, Systeme und Automaten, Neubiberg, Germany*, pages 1–5.
- [Toshev and Szegedy, 2014] Toshev, A. and Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. pages 1653–1660.
- [Trawny and Roumeliotis, 2005] Trawny, N. and Roumeliotis, S. I. (2005). Indirect kalman filter for 3d attitude estimation - a tutorial for quaternion algebra. *University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep.*, 2.
- [Tölgyessy et al., 2021a] Tölgyessy, M., Dekan, M., Chovanec, L., and Hubinsky, P. (2021a). Evaluation of the azure kinect and its comparison to kinect v1 and kinect v2. *Sensors*, 21(2).
- [Tölgyessy et al., 2021b] Tölgyessy, M., Dekan, M., and Chovanec, (2021b). Skeleton tracking accuracy and precision evaluation of kinect v1, kinect v2, and the azure kinect. *Applied Sciences*, 11(12).
- [van der Kruk and Reijne, 2018] van der Kruk, E. and Reijne, M. M. (2018). Accuracy of human motion capture systems for sport applications; state-of-the-art review. *European Journal of Sport Science*, 18(6):806–819. PMID: 29741985.

- [van der Veen et al., 2019] van der Veen, S. M., Bordeleau, M., Pidcoe, P. E., France, C. R., and Thomas, J. S. (2019). Agreement analysis between vive and vicon systems to monitor lumbar postural changes. *Sensors*, 19(17).
- [Viitanen et al., 2018] Viitanen, M., Vanne, J., Hämäläinen, T. D., and Kulmala, A. (2018). Low latency edge rendering scheme for interactive 360 degree virtual reality gaming. pages 1557–1560.
- [von Salis-Soglio, 2015] von Salis-Soglio, G. F. (2015). *Die Neutral-0-Methode : Mit Langen- und Umfangsmessung*. Springer, Berlin, Heidelberg.
- [Vox et al., 2016] Vox, J. P., Franz, S., and Wallhoff, F. (2016). Adaptive Bewegungsanalyse von physiotherapeutischen Übungen für eine optimierte Mensch-Roboter-Trainingsinteraktion. *Kognitive Systeme*, 2016(1).
- [Vox and Wallhoff, 2016] Vox, J. P. and Wallhoff, F. (2016). Evaluation of motion tracking methods for therapeutic assistance in everyday living environments. In *2016 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI), Baden-Baden, Germany*, pages 96–101.
- [Vox and Wallhoff, 2017a] Vox, J. P. and Wallhoff, F. (2017a). Autonomous Learning of Motion Sequences within a Human-Robot Interaction. In *6. Interdisziplinärer Workshop Kognitive Systeme (KogSys); Menschen, Teams, Systeme und Automaten, Neubiberg, Germany*, pages 1–6.
- [Vox and Wallhoff, 2017b] Vox, J. P. and Wallhoff, F. (2017b). Recognition of human motion exercises using skeleton data and SVM for rehabilitative purposes. In *2017 IEEE Life Sciences Conference (LSC), Sydney, NSW, Australia*, pages 266–269.
- [Vox and Wallhoff, 2018a] Vox, J. P. and Wallhoff, F. (2018a). Artificial Increase of 3D-Skeleton-Data for Human Motion Recognition using supervised SVM and NN. In *2018 IEEE International Conference on Intelligent Systems (IS), Funchal, Portugal*, pages 602–607.
- [Vox and Wallhoff, 2018b] Vox, J. P. and Wallhoff, F. (2018b). Human Motion Recognition Using 3D-Skeleton-Data and Neural Networks. In Bramer, M. and Petridis, M., editors, *Artificial Intelligence XXXV, SGAI International Conference on Artificial Intelligence, Cambridge, England*, pages 204–209, Cham. Springer International Publishing.

- [Vox and Wallhoff, 2018c] Vox, J. P. and Wallhoff, F. (2018c). Preprocessing and Normalization of 3D-Skeleton-Data for Human Motion Recognition. In *2018 IEEE Life Sciences Conference (LSC), Montreal, QC, Canada*, pages 279–282.
- [Vox and Wallhoff, 2019] Vox, J. P. and Wallhoff, F. (2019). A Framework for the Analysis of Biomechanical Loading Using Human Motion Tracking. In *2019 IEEE 20th International Conference on Information Reuse and Integration for Data Science (IRI), Los Angeles, CA, USA*, pages 39–46.
- [Vox et al., 2021] Vox, J. P., Weber, A., Wolf, K. I., Izdebski, K., Schüler, T., König, P., Wallhoff, F., and Friemert, D. (2021). An Evaluation of Motion Trackers with Virtual Reality Sensor Technology in Comparison to a Marker-Based Motion Capture System Based on Joint Angles for Ergonomic Risk Assessment. *Sensors*, 21(9).
- [Wallhoff et al., 2019] Wallhoff, F., Vox, J. P., and Theuerkauff, T. (2019). *Assistenz- und Servicerobotik – die Gestaltung der Mensch-Maschine-Schnittstelle als Grundlage des Anwendungserfolgs*, pages 99–122. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Weber, 2011] Weber, B. (2011). Entwicklung und evaluation eines bewegungsmesssystems zur analyse der physischen aktivität. *Deutsche Gesetzliche Unfallversicherung e. V. (DGUV)*.
- [Wessinghage and Morsch, 2013] Wessinghage, T. and Morsch, A. (2013). Muskel-Skelett-Erkrankungen: Bedeutung von Bewegungsmangel und sportlicher Aktivität. *Public Health Forum*, 21(2):21–22.
- [Wiedemann et al., 2015] Wiedemann, L., Planinc, R., Nemeč, I., and Kampel, M. (2015). Performance evaluation of joint angles obtained by the kinect v2. pages 1–6.
- [Windolf et al., 2008] Windolf, M., Götzen, N., and Morlock, M. (2008). Systematic accuracy and precision analysis of video motion capturing systems—exemplified on the vicon-460 system. *Journal of Biomechanics*, 41(12):2776–2780.
- [Xia et al., 2012] Xia, L., Chen, C.-C., and Aggarwal, J. K. (2012). View invariant human action recognition using histograms of 3d joints. pages 20–27.

- [Xu et al., 2011] Xu, X., chi Chang, C., Faber, G. S., Kingma, I., and Dennerlein, J. T. (2011). The validity and interrater reliability of video-based posture observation during asymmetric lifting tasks. *Human Factors*, 53(4):371–382. PMID: 21901934.
- [Xu et al., 2015] Xu, X., McGorry, R. W., Chou, L.-S., hua Lin, J., and chi Chang, C. (2015). Accuracy of the microsoft kinect™ for measuring gait parameters during treadmill walking. *Gait Posture*, 42(2):145–151.
- [Yadav et al., 2021] Yadav, S. K., Tiwari, K., Pandey, H. M., and Akbar, S. A. (2021). Skeleton-based human activity recognition using convlstm and guided feature learning. *Soft Computing*, pages 1–14.
- [Yan et al., 2018] Yan, S., Xiong, Y., and Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. *CoRR*, abs/1801.07455.
- [Yeung et al., 2021] Yeung, L.-F., Yang, Z., Cheng, K. C.-C., Du, D., and Tong, R. K.-Y. (2021). Effects of camera viewing angles on tracking kinematic gait patterns using azure kinect, kinect v2 and orbbec astra pro v2. *Gait & Posture*, 87:19–26.
- [Yun et al., 2012] Yun, K., Honorio, J., Chattopadhyay, D., Berg, T. L., and Samaras, D. (2012). Two-person interaction detection using body-pose features and multiple instance learning. pages 28–35.
- [Zhang et al., 2017] Zhang, S., Liu, X., and Xiao, J. (2017). On geometric features for skeleton-based action recognition using multilayer lstm networks. pages 148–157.