

VOL. 02

KARLSRUHE SERIES ON
HUMANOID ROBOTICS

CHRISTIAN MANDERY

Organisation, Repräsentation
und Analyse menschlicher
Ganzkörperbewegung für die
datengetriebene Bewegungsgene-
rierung bei humanoiden Robotern



Christian Mandery

Organisation, Repräsentation und Analyse menschlicher
Ganzkörperbewegung für die datengetriebene
Bewegungsgenerierung bei humanoiden Robotern

Karlsruhe Series on Humanoid Robotics

Edited by Prof. Dr.-Ing. Tamim Asfour

Vol. 02

Organisation, Repräsentation und Analyse menschlicher Ganzkörperbewegung für die datengetriebene Bewegungsgenerierung bei humanoiden Robotern

von
Christian Mandery

Dissertation, Karlsruher Institut für Technologie
KIT-Fakultät für Informatik

Tag der mündlichen Prüfung: 19. Juli 2017

Referenten: Prof. Dr.-Ing. Tamim Asfour, Prof. Dr. rer. nat. Helge Ritter

Impressum



Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.
Reprint using the book cover is not allowed.

www.ksp.kit.edu



*This document – excluding the cover, pictures and graphs – is licensed
under a Creative Commons Attribution-Share Alike 4.0 International License
(CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>*



*The cover page is licensed under a Creative Commons
Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0):
<https://creativecommons.org/licenses/by-nd/4.0/deed.en>*

Print on Demand 2017 – Gedruckt auf FSC-zertifiziertem Papier

ISSN 2512-0875

ISBN 978-3-7315-0729-1

DOI 10.5445/KSP/1000075715

Organisation, Repräsentation und Analyse menschlicher Ganzkörperbewegung für die datengetriebene Bewegungsgenerierung bei humanoiden Robotern

Zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Dipl.-Inform. Christian Mandery

aus Ludwigshafen am Rhein

Tag der mündlichen Prüfung: 19. Juli 2017
Referent: Prof. Dr.-Ing. Tamim Asfour
Korreferent: Prof. Dr. rer. nat. Helge Ritter

Danksagung

Die vorliegende Dissertation ist das Ergebnis meiner Arbeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Hochperformante Humanoide Technologien (H²T) des Instituts für Anthropomatik und Robotik, Karlsruher Institut für Technologie (KIT). An dieser Stelle möchte ich mich bei all denjenigen bedanken, die zum guten Gelingen dieser Arbeit beigetragen haben.

Allen voran möchte ich meinem Doktorvater und dem Leiter des H²T Prof. Dr. Tamim Asfour danken für das mir entgegengebrachte Vertrauen und die Chance, in diesem faszinierenden Forschungsfeld arbeiten zu dürfen. Seine Begeisterung und sein Einsatz für die humanoide Robotik und unsere zahlreichen konstruktiven Gespräche waren und sind für mich eine bedeutende Inspiration und ohne seine Unterstützung in den vergangenen Jahren hätte die vorliegende Arbeit mit Sicherheit nicht entstehen können. Zudem möchte ich Prof. Dr. Helge Ritter für das Interesse an meiner Arbeit und für seine Bereitschaft zur Übernahme des Korreferats danken.

Bei meinen Kollegen möchte ich mich für die Hilfsbereitschaft, die sehr angenehme Arbeitsatmosphäre und nicht zuletzt auch die vielen fachlichen Gespräche bedanken, die mich immer neue Ansatzpunkte entdecken ließen. Mein besonderer Dank geht hierbei an Dr. Júlia Borràs Sol, die in vielen Belangen wie ein Mentor für mich war, für ihre fortwährende Unterstützung und die zahlreichen fruchtbaren Diskussionen, in denen eine nicht geringe Anzahl der Ideen und Ansätze dieser Arbeit erst entstanden sind. Bei Jannik Steinbring möchte ich mich nicht nur für seinen Rat und die erfolgreiche, auch in dieser Arbeit dokumentierte wissenschaftliche Kooperation bedanken, sondern vor allem auch für die wertvolle langjährige Freundschaft. Ömer

Terlemez, Graziella Barbaro, Sebastian Bodenstedt und Darko Katic möchte ich für unsere angenehmen und entspannenden gemeinsamen Mittagessen danken. Darüber hinaus möchte ich allen weiteren Kollegen am H²T meinen Dank ausdrücken: Eren Erdal Aksoy, Diana Becker, Jonas Beil, Martin Do, Isabel Ehrenberger, Markus Grotz, Hans Haubert, Peter Kaiser, Lukas Kaul, Manfred Kröhnert, Michael Neaga, Simon Ottenhaus, Fabian Paus, Samuel Rader, David Schiebener, Nikolaus Vahrenkamp, Mirko Wächter, Isabelle Wappler, Pascal Weiner und You Zhou.

Weiterhin möchte ich mich bei allen Studenten bedanken, die mit ihren Abschlussarbeiten oder HiWi-Stellen einen Beitrag zum Gelingen dieser Arbeit geleistet haben. Mein besonderer Dank hier gilt Matthias Plappert für die zahlreichen interessanten Gespräche und die produktive und vielfältige Zusammenarbeit, aus der gleich mehrere Publikationen hervorgegangen sind. Außerdem danke ich Mirjam Jöchner, die unter anderem den Code der Bewegungsgenerierung von Python nach C++ portiert hat, und Andre Langenstein, der im Rahmen seiner Bachelorarbeit die Generierung von Klassifikationsvorschlägen in der KIT-Bewegungsdatenbank umgesetzt und evaluiert hat.

Abschließend gilt ein besonderer Dank meiner Familie und vor allem meinen Eltern Dorothea und Peter, die mich während meines gesamten Lebens bedingungslos bei all meinen Schritten unterstützt haben, immer ein offenes Ohr für meine Sorgen hatten und mir zur Seite standen und stehen. Ihnen widme ich diese Arbeit. Ebenso möchte ich mich bei meiner Freundin Katharina für den Rückhalt und ihre emotionale Unterstützung bedanken, sowie bei meinen Freunden, für die ich während meiner Promotion leider viel zu wenig Zeit hatte.

Karlsruhe, im Oktober 2017

Christian Mandery

Inhaltsverzeichnis

Abkürzungsverzeichnis	vii
1 Einleitung	1
1.1 Zielsetzung und Beiträge der Arbeit	5
1.2 Aufbau und Struktur der Arbeit	9
2 Erfassung und Repräsentation	
menschlicher Ganzkörperbewegung	11
2.1 Master Motor Map: Ein Ganzkörper-Referenzmodell des menschlichen Körpers	12
2.2 Erfassung menschlicher Ganzkörperbewegung	17
2.2.1 Markerbasierte Bewegungserfassung	17
2.2.2 Rekonstruktion von Gelenkwinkeln	23
2.3 Merkmale zur Beschreibung menschlicher Bewegung	30
2.3.1 Definition von Merkmalen	32
2.3.2 Verarbeitungsschritte	37
2.4 Übertragung menschlicher Bewegung auf humanoide Roboter	39
2.5 Zusammenfassung von Kapitel 2	43
3 Modellierung und Klassifikation von	
Ganzkörperbewegungen mit HMMs	45
3.1 Verwandte Arbeiten	46
3.2 HMMs für menschliche Ganzkörperbewegung	50
3.2.1 Grundlagen und Definition von HMMs	51

3.2.2	Training von HMMs für Bewegungsdaten	56
3.2.3	Anwendung zur Multi-Class-Bewegungsklassifikation	59
3.3	Merkmalsauswahl	62
3.4	Multi-Label-Bewegungsklassifikation	68
3.4.1	Problemtransformation mit Label-Potenzmengen-Methode	70
3.4.2	Problemtransformation mit Binäre-Relevanz-Methode	71
3.4.3	Anpassung von Algorithmen	74
3.5	Evaluation und Ergebnisse	75
3.5.1	Merkmalsauswahl	77
3.5.2	Multi-Label-Bewegungsklassifikation	84
3.6	Zusammenfassung von Kapitel 3	88
4	Die KIT-Bewegungsdatenbank	91
4.1	Verwandte Arbeiten und existierende Datenbanken	92
4.2	Aufbau der Bewegungsdatenbank	100
4.2.1	Datentypen	100
4.2.2	Softwarearchitektur	106
4.2.3	Schnittstellen	113
4.3	Organisation von Bewegungsaufnahmen mit dem Motion Description Tree	117
4.4	Datengetriebene Generierung von Klassifikationsvorschlägen	123
4.5	Inhalt der Bewegungsdatenbank	131
4.6	Zusammenfassung von Kapitel 4	138
5	Analyse menschlicher Bewegung anhand einer Ganzkörperposen-Taxonomie	141
5.1	Verwandte Arbeiten	142

5.2	Taxonomie menschlicher Ganzkörper-Support-Posen	152
5.2.1	Definition möglicher Support-Kontakte	153
5.2.2	Definition von Support-Posen aus Support-Kontakten	155
5.2.3	Taxonomie zur Beschreibung von Support-Posen und ihren Transitionen	156
5.3	Segmentierung von Bewegungsaufnahmen zur Extraktion von Support-Posen	161
5.4	Evaluation und Ergebnisse	170
5.4.1	Validierung des Segmentierungsverfahrens	173
5.4.2	Darstellung von Support-Posen-Transitionen als Graph	176
5.4.3	Häufigkeit spezifischer Support-Posen und ihrer Transitionen	178
5.4.4	Zeitdauer von Support-Posen-Transitionen	180
5.5	Zusammenfassung von Kapitel 5	185
6	Datengetriebene Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen	189
6.1	Verwandte Arbeiten	190
6.2	Definition der Problemstellung	202
6.3	Unterteilung von Support-Posen hinsichtlich kinematischer Konfigurationen	206
6.4	Statistische Modellierung von Posen-Transitionen	213
6.4.1	N-Gramm-Modell für Transitionswahrscheinlichkeiten	213
6.4.2	Modellierung der räumlichen Translation bei Posen-Transitionen	218
6.5	Planungsalgorithmus zur Generierung von Posen-Sequenzen	220
6.6	Evaluation und Ergebnisse	227

6.6.1	Parametrisierung des N-Gramm-Transitionsmodells .	230
6.6.2	Generierung von Posen-Sequenzen für verschiedene Beispielszenarien	232
6.7	Zusammenfassung von Kapitel 6	241
7	Zusammenfassung und Ausblick	245
7.1	Wissenschaftliche Beiträge der Arbeit	245
7.2	Diskussion und Ausblick	249
A	Bewegungsdatensätze der Evaluation in Abschnitt 3.5 . .	255
	Abbildungsverzeichnis	259
	Tabellenverzeichnis	265
	Algorithmenverzeichnis	269
	Literaturverzeichnis	271

Abkürzungsverzeichnis

- AAV** Algorithmenanpassungsverfahren (*algorithm adaption methods*), siehe Abschnitt 3.4
- BRM** Binäre-Relevanz-Methode (*binary relevance method*), siehe Abschnitt 3.4.2
- DMP** Dynamic Movement Primitive
- FHMM** Factorial Hidden Markov Model
- GKP-Taxonomie** Ganzkörperposen-Taxonomie, siehe Abschnitt 5.2
- H²T** Lehrstuhl für Hochperformante Humanoide Technologien
- HMM** Hidden Markov Model, siehe Abschnitt 3.2.1
- KKR** Korrektklassifikationsrate (*accuracy*), siehe Abschnitt 3.2.3
- KP** Konfigurationspose, siehe Abschnitt 6.3
- LPM** Label-Potenzmengen-Methode (*label powerset method*), siehe Abschnitt 3.4.1
- MMM** Master Motor Map, siehe Abschnitt 2.1
- PCA** Hauptkomponentenanalyse (*principal component analysis*)
- PHMM** Parametric Hidden Markov Model
- PTV** Problemtransformationsverfahren (*problem transformation methods*), siehe Abschnitt 3.4

SP	Support-Pose, siehe Abschnitt 5.2
SVM	Support Vector Machine
ZMP	Zero-Moment Point

1 Einleitung

Nach ihrem Einzug in die Produktion und dem dadurch ausgelösten industriellen Umbruch beginnen Roboter zunehmend, nun auch einen Platz im unmittelbaren Umfeld des Menschen zu erobern. Zum einen sollen Serviceroboter durch das Erfüllen lästiger Routineaufgaben im Haushalt den Alltag von Menschen erleichtern und beispielsweise älteren Menschen länger ein selbstbestimmtes Leben in ihrer vertrauten Wohnung ermöglichen. Zum anderen findet sich ein aussichtsreiches Einsatzgebiet für Roboter in menschengemachten Umgebungen bei ihrem Einsatz in Katastrophen- oder Bergungsszenarien, bei dem sie den Menschen in potentiell gefährlichen Umgebungen, bspw. nach Havarien in einem Kernkraftwerk, ersetzen sollen.

Die Anforderungen an Roboter in solchen menschenzentrierten Umgebungen weichen stark von denen in der industriellen Fertigung ab, wo die Umgebung durch die einzusetzenden Roboter bestimmt wird und ein hohes Maß an Strukturierung aufweist. Insbesondere liegt bei menschengemachten Umgebungen ein Umfeld vor, welches nicht nur *durch*, sondern vor allem auch *für* den Menschen und seine Fertigkeiten gestaltet wurde. Es verwundert daher nicht, dass besonders humanoide Roboter, deren Konstruktion dem menschlichen Vorbild nachempfunden wurde, für derartige Einsatzgebiete als vielversprechend angesehen werden. Die Ausstattung mit Armen und Beinen erlaubt diesen Robotern theoretisch eine optimale, am menschlichen Vorbild orientierte Nutzung der menschenzentrierten Umgebung und ihrer Elemente wie Geländern, Treppen, Leitern usw., und steigert nicht zuletzt auch die Akzeptanz des Roboters durch den Menschen.



Abbildung 1.1: Beispiele für menschliche Fortbewegung unter Nutzung des ganzen Körpers.

Die tatsächliche Fortbewegung humanoider Roboter in menschenzentrierten Umgebungen stellt allerdings heutzutage nach wie vor ein schwieriges, nicht zufriedenstellend gelöstes Problem dar. Während bei der Generierung von Laufbewegungen für humanoide Roboter unter Laborbedingungen – nämlich auf dem flachen Boden in Abwesenheit weiterer Umgebungsobjekte – bereits gewisse Fortschritte zu verzeichnen sind, bleiben die Leistungen dieser Roboter in realen Umgebungen immer noch weit hinter dem menschlichen Vorbild zurück, dessen Fähigkeiten zur Fortbewegung unter Nutzung des gesamten Körpers in Abbildung 1.1 beispielhaft illustriert werden. Als Demonstration der bestehenden Unzulänglichkeiten humanoider Roboter mag die *DARPA Robotics Challenge* (DRC) dienen, die in den Jahren 2011 bis 2015 vom US-amerikanischen Verteidigungsministerium als internationaler Wettbewerb für Rettungsroboter durchgeführt wurde. Im Rahmen der DRC sollten sich humanoide Roboter, neben anderen Aufgaben, über unwegsames Terrain und in komplexen menschengemachten Umgebungen wie beispielsweise auf Leitern fortbewegen. Bereits im Vorfeld des Finales in Pomona, Kalifornien, zeigte sich, dass die geplanten Anforderungen abgesenkt und die Aufgaben für die teilnehmenden Roboter vereinfacht werden müssen, wofür die vorgesehene Leiter beispielsweise durch eine Treppe ersetzt wurde. Dennoch konnten nur wenige der 24 am Finale teilnehmenden Roboter das Ziel auf eigenen

Beinen erreichen und eine Video-Zusammenstellung von Fehlschlägen wie dem Hinfallen teilnehmender Roboter¹ ist mit knapp 2 Millionen Abrufen auf YouTube zu einem der populärsten Resultate des Wettbewerbs avanciert.

Doch was ist die Ursache der Schwierigkeiten humanoider Roboter bei Fortbewegungsaufgaben, die aus der Sichtweise eines Menschen im wahrsten Sinne des Wortes kinderleicht erscheinen? Prof. Christopher G. Atkeson, welcher sich mit dem Team WPI-CMU als einziger sämtlichen Aufgaben der DRC stellte und 14 von 16 Punkten erreichen konnte, beschreibt rückblickend den Verzicht auf die Verwendung weiterer *Support-Kontakten* zwischen dem Roboter und seiner Umgebung zusätzlich zu den Füßen als einen der entscheidenden Gründe:

„It is startling to realize that we and all other teams failed to use the stair railings, put a hand on the wall to help cross the rough terrain, or grab the door frame to more safely get through the door in the DRC Finals. Even drunk people are smart enough to use nearby supports. [...] Why didn't any robot use railings, walls, door frames, or obstacles for support and stabilization? Humans use stair railings, and brace themselves by reaching out to a nearby wall when walking over difficult rough terrain. Why didn't any DRC robots do this? Full body locomotion (handholds, bracing, leaning against a wall or obstacles) should be easier than our current high performance minimum contact locomotion approaches.“ (Atkeson et al. 2015, ©2015 IEEE)

Tatsächlich wurde bei der Programmierung der teilnehmenden Roboter bei der DRC und bei vergleichbar herausfordernden Aufgaben zuvor meist darauf verzichtet, Kontakt mit den Händen oder anderen Teilen des Roboters zu in der Nähe erreichbaren Umgebungselementen wie Wänden, Geländern, Griffen usw. herzustellen, obwohl solche Umgebungskontakte vom menschlichen Vorbild oftmals zur Erhöhung seiner Stabilität durch Festhalten oder Abstützen eingesetzt werden, was beispielsweise bei (Klein-)Kindern, aber auch bei gebrechlichen oder sogar betrunkenen Menschen beobachtet werden kann. Der weitgehende Verzicht auf diese Strategie bei humanoiden Robotern führt dazu, dass während der Fortbewegung für einen Großteil der Zeit

¹ A Compilation of Robots Falling Down at the DARPA Robotics Challenge, IEEE Spectrum.

das gesamte Gewicht des Roboters über einen einzigen *Support-Kontakt*, nämlich einen seiner beiden Füße, balanciert werden muss. Insbesondere aber basiert die Nutzung der Füße als ausschließliche *Support-Kontakte* auf der Annahme, dass auftretende Kräfte oder Störungen des Roboters weitestgehend entlang der Richtung des Gravitationsvektors auftreten, wodurch eine Vielzahl von Fortbewegungsstrategien, wie etwa „Lehnen“, von vornherein ausgeschlossen werden.

Freilich scheiterte der Einsatz zusätzlicher *Support-Kontakte* bei der DRC nicht am fehlenden Erfindungsreichtum der beteiligten Wissenschaftler. Vielmehr bestehen neben Herausforderungen bei der Roboterkonstruktion (bspw. hinsichtlich der Konstruktion von Armen und Händen) Schwierigkeiten, derartige *Multi-Kontakt-Bewegungen* effizient zu beschreiben und zu planen: „More contacts make tasks mechanically easier, but algorithmically more complicated.“ (Atkeson et al. 2015). Ein wesentliches Problem besteht hierbei darin, dass traditionelle Ansätze für die zweibeinige Fortbewegung humanoider Roboter basierend auf dem Zero-Moment Point (ZMP) mit dem Vorliegen weiterer Kontakte bei *Multi-Kontakt-Bewegungen* nicht mehr eingesetzt werden können. Eine direkte Planung der auszuführenden Trajektorie für eine *Multi-Kontakt-Bewegung* hingegen stellt aufgrund der hohen Zahl an Bewegungsfreiheitsgraden humanoider Roboter und der damit verbundenen Komplexität ein sehr schwieriges Problem dar. Zwar können mit Techniken der mathematischen Optimierung, bspw. auf Basis quadratischer Programmierung, prinzipiell gültige Lösungen für dieses Problem gefunden werden, allerdings sind derartige Optimierungsansätze mit einem exorbitanten Rechenaufwand verbunden, der einen effizienten Einsatz für zuvor unbekannte Problemstellungen in der Praxis, und somit die Nutzung von *Multi-Kontakt-Bewegungen* überhaupt, verhindert.

1.1 Zielsetzung und Beiträge der Arbeit

Die Zielsetzung der vorliegenden Arbeit besteht darin, einen Beitrag zur Lösung des angesprochenen Problems der Multi-Kontakt-Bewegung bei humanoiden Robotern zu leisten. Hierfür wird ein neuartiger datengetriebener Ansatz zur Generierung von Multi-Kontakt-Bewegungen vorgeschlagen, welcher durch die Verknüpfung von menschlichen Bewegungsdaten mit einer am Lehrstuhl für Hochperformante Humanoide Technologien (H²T) entwickelten Ganzkörperposen-Taxonomie eine Dimensionalitäts- und Komplexitätsreduktion des Problems erreichen und gleichzeitig direkt vom menschlichen Vorbild lernen kann. Die Verwendung einer Taxonomie zur Beschreibung von Ganzkörperposen geht dabei auf einen im Bereich des Robotergreifens verbreiteten Ansatz zurück, wo Taxonomien seit längerer Zeit sehr erfolgreich zur Komplexitätsreduktion des ebenfalls hochdimensionalen Greifproblems zum Einsatz kommen. Analog hierzu lässt sich das Finden einer stabilen Ganzkörperpose als *duales Greifproblem* interpretieren, bei dem die Umgebung den Roboter über die bestehenden Support-Kontakte „greift“ (mehr hierzu in Kapitel 5). Auf der anderen Seite bildet die Verwendung von großen Mengen aufgezeichneter Bewegungsdaten die Grundlage, um mögliche Abfolgen dieser Ganzkörperposen direkt vom menschlichen Vorbild zu lernen. Die schlussendlich generierte Multi-Kontakt-Bewegung wird bei dem vorgestellten Ansatz folglich als Sequenz von Ganzkörperposen repräsentiert, die in zukünftigen Arbeiten, beispielsweise mit Hilfe von Bewegungsprimitiven², zu einer kontinuierlichen, stabilen Trajektorie verbunden werden können.

Die Beiträge der Arbeit bestehen demnach zum einen in einer Untersuchung, wie die erforderlichen großen Mengen menschlicher Bewegungsdaten erfasst, organisiert, repräsentiert und schlussendlich zur Anwendung verar-

² Diese Vorgehensweise greift eine seit langem in der Robotik und Neurowissenschaft vertretene Vorstellung auf, derzufolge komplexe Bewegungen als Abfolge von einer endlichen Anzahl möglicher „simpler Teilbewegungen“, sogenannter Bewegungsprimitive, dargestellt werden können.

beitet werden müssen. So wird eine im Rahmen dieser Arbeit entwickelte großskalige Datenbank menschlicher Ganzkörperbewegung beschrieben, die Bewegungsdaten mit einer semantischen Kennzeichnung in einer vereinheitlichenden Repräsentation bereitstellt. Zum anderen wird ein Verfahren zur Segmentierung menschlicher Bewegungsdaten hinsichtlich der erwähnten Ganzkörperposen-Taxonomie vorgestellt, welches zu einer (partiellen) Validierung dieser Taxonomie beiträgt. Abschließend wird das im Kontext dieser Arbeit entwickelte Verfahren zur datengetriebenen Generierung von Multi-Kontakt-Bewegungen als Sequenz von Ganzkörperposen, das auf den zuvor verarbeiteten und segmentierten Bewegungsdaten aufbaut, eingeführt und evaluiert. Im Folgenden sollen die angesprochenen Beiträge der Arbeit näher erläutert werden.

Repräsentation menschlicher Ganzkörperbewegung und deren Organisation in einer großskaligen Bewegungsdatenbank: Mit modernen Bewegungserfassungssystemen steht heutzutage eine Möglichkeit zur hochgenauen Erfassung von großen Mengen an Bewegungsdaten unterschiedlicher menschlicher Probanden zur Verfügung, welche die Grundlage für den in dieser Arbeit verfolgten Ansatz bildet. Der erste Teil der vorliegenden Arbeit beschäftigt sich daher mit der Fragestellung, wie menschliche Ganzkörperbewegungen erfasst, repräsentiert und in einer großskaligen Bewegungsdatenbank organisiert werden können, um diese für Anwendungen der humanoiden Robotik nutzbar zu machen. Hierfür werden im Rahmen dieser Arbeit für das am H²T entwickelte Master Motor Map (MMM) Framework zur Repräsentation menschlicher Ganzkörperbewegung zwei neue Verfahren basierend auf nicht-linearer Optimierung bzw. Kalman-Filterung vorgestellt, die eine Rekonstruktion der menschlichen Bewegung aus gelabelten oder ungelabelten Messungen eines markerbasierten Bewegungserfassungssystems und ihre Übertragung auf humanoide Roboter ermöglichen. Auf Basis der rekonstruierten MMM-Bewegung werden anschließend 29 unterschiedliche Merkmale zur Beschreibung menschlicher Ganzkörperbewegung definiert,

welche die Basis für eine Multi-Label-Bewegungsklassifikation basierend auf Hidden Markov Models (HMMs) bilden und hierfür zu hochdeskriptiven niedrigdimensionalen Teilmengen dieser Merkmale kombiniert werden.

Die Organisation erfasster Bewegungsdaten erfolgt anschließend in der im Rahmen dieser Arbeit entwickelten *KIT Whole-Body Human Motion Database*. Im Unterschied zu etablierten Bewegungsdatenbanken werden hierbei zusätzlich zu den Bewegungen der menschlichen Probanden auch die Posen von in der Bewegung involvierten Objekten und Umgebungselementen (bspw. ein Geländer) berücksichtigt, was für die nachfolgende Analyse der Interaktion zwischen Mensch und Umgebung von entscheidender Bedeutung ist. Die Konzeption der Bewegungsdatenbank berücksichtigt insbesondere die Fragestellung, wie die Bewegungsdaten derart organisiert werden können, dass eine systematische Abfrage des Datenbestands zur Suche nach spezifischen Bewegungstypen möglich wird, und es wird hierfür mit dem *Motion Description Tree* eine neue Vorgehensweise zur Organisation von Bewegungsdaten auf Basis hierarchisch strukturierter Tags vorgeschlagen. Diese Tags werden zum einen manuell durch die Benutzer zugewiesen, zum anderen wird jedoch auch das in dieser Arbeit vorgestellte HMM-basierte Multi-Label-Bewegungsklassifikationssystem zur Generierung von Tagging-Vorschlägen eingesetzt.

Segmentierung menschlicher Bewegung zur Extraktion von Support-

Posen: Wie eingangs erwähnt, basiert der im Rahmen dieser Arbeit verfolgte Ansatz auf der am H²T entwickelten, aus 46 Posen bestehenden Ganzkörperposen-Taxonomie. In dieser Arbeit wird ein Verfahren zur automatischen Segmentierung menschlicher Bewegungsdaten hinsichtlich dieser Taxonomie vorgestellt und evaluiert, dessen Ausgangsbasis die durch die Bewegungsdatenbank bereitgestellte MMM-Repräsentation menschlicher Bewegungen einschließlich beteiligter Umgebungsobjekte ist. Mit Hilfe einer hierfür entwickelten Heuristik, welche auf den kartesischen Abständen der Körpersegmente zur Umgebung und den jeweiligen Segment-

Geschwindigkeiten beruht, werden für jeden Zeitpunkt der Bewegung die Support-Kontakte des menschlichen Körpers mit seiner Umgebung extrahiert. Auf Basis der erkannten Support-Kontakte kann anschließend jedem Zeitpunkt einer Bewegung die korrespondierende Support-Pose der Taxonomie zugeordnet und die Bewegung somit als Abfolge von Ganzkörperposen dargestellt werden.

Mit einer Untersuchung zur Häufigkeit des Auftretens einzelner Support-Posen und der Transitionen zwischen ihnen, sowie den mit diesen Transitionen assoziierten Zeitdauern liefert das Segmentierungsverfahren die Grundlage für eine partielle Validierung der vorgeschlagenen Ganzkörperposen-Taxonomie und im nächsten Schritt zudem die Trainingsdaten für das im Kontext dieser Arbeit entwickelte Verfahren zur datengetriebenen Generierung von Multi-Kontakt-Bewegungen.

Datengetriebene Generierung von Multi-Kontakt-Bewegungen für humanoide Roboter: Grundlage für den vorgeschlagenen Ansatz zur datengetriebenen Generierung von Multi-Kontakt-Bewegungen ist zunächst eine weitergehende, auf einer Analyse menschlicher Bewegungsdaten basierende Unterteilung der durch die Ganzkörperposen-Taxonomie vorgegebenen Support-Posen hinsichtlich ihrer kinematischen Konfiguration. Diese Unterteilung ist notwendig, da die teilweise erheblichen Konfigurationsunterschiede innerhalb einer Support-Pose (bspw. Arme nach vorne vs. nach hinten ausgestreckt) bei der Planung von real ausführbaren Posen-Sequenzen nicht ignoriert werden dürfen. Die resultierenden, mit konkreten Gelenkwinkeln des menschlichen Körpers assoziierten Konfigurationsposen können im Anschluss auf die Kinematik humanoider Roboter übertragen werden.

Aufbauend auf den segmentierten Bewegungsdaten und der beschriebenen Unterteilung der Support-Posen wird in dieser Arbeit ein neuartiger statistischer Ansatz zur datengetriebenen Bewegungsgenerierung eingeführt, welcher auf einer linguistischen Sichtweise auf das Problem der Multi-Kontakt-Bewegungsgenerierung beruht, bei der eine Pose als *Wort* und eine Bewegung

als *Satz*, als Abfolge von *Wörtern*, angesehen wird. Von dieser Sichtweise ausgehend wird mit segmentierten menschlichen Bewegungsdaten ein N-Gramm-Modell trainiert, welches die bedingte Auftretenswahrscheinlichkeit möglicher Ganzkörperposen basierend auf den vorhergehenden Posen modelliert und sich somit als ein statistisches Modell für die Abfolge dieser Posen bzw. ihrer Transitionen darstellt. Analog zu Problemstellungen der natürlichen Sprachverarbeitung stellt sich die Suche nach der wahrscheinlichsten Sequenz von Ganzkörperposen unter dem statistischen Modell für die gegebenen Nebenbedingungen einer Multi-Kontakt-Bewegungsaufgabe nun dar als Wegfindungsproblem in dem Baum, welcher durch Transitionen zwischen den vorgegebenen Ganzkörperposen im statistischen Modell aufgespannt wird. Zur Lösung dieses Wegfindungsproblems wird in dieser Arbeit ein heuristischer Suchalgorithmus mit *Pruning* vorgeschlagen und evaluiert.

1.2 Aufbau und Struktur der Arbeit

Die vorliegende Arbeit ist in sieben Kapitel unterteilt, die sich inhaltlich mit den zuvor angesprochenen wissenschaftlichen Beiträgen der Arbeit und den dazu erforderlichen Grundlagen auseinandersetzen. Im Einzelnen gliedert sich die Arbeit wie folgt:

Kapitel 2 erläutert die Grundlagen zur Erfassung, Repräsentation und Verarbeitung menschlicher Ganzkörperbewegung, welche in den weiteren Teilen der Arbeit Verwendung finden. Genauer werden hierfür das MMM-Framework mit dem MMM-Referenzmodell des menschlichen Körpers, die Rekonstruktion menschlicher Bewegung auf Basis markerbasierter Bewegungserfassung, die Definition von Merkmalen zur Beschreibung menschlicher Ganzkörperbewegung und ein Verfahren zur kinematischen Übertragung menschlicher Bewegungen auf humanoide Roboter beschrieben.

Kapitel 3 befasst sich mit der Modellierung und Klassifikation menschlicher Ganzkörperbewegungen mit Hilfe von HMMs. Hierbei werden insbesondere die Vorgehensweise zur systematischen Merkmalsauswahl und

die Anwendung von HMMs für die Multi-Label-Bewegungsklassifikation erläutert und evaluiert.

Kapitel 4 führt mit der *KIT Whole-Body Human Motion Database* die im Rahmen dieser Arbeit entstandene großskalige Datenbank menschlicher Ganzkörperbewegung ein. Neben einer Beschreibung des Aufbaus, der Struktur und des gegenwärtigen Inhalts der Bewegungsdatenbank wird untersucht, wie menschliche Bewegungsdaten mit einem *Motion Description Tree* in einer solchen Datenbank systematisch organisiert und wie Benutzer mit Hilfe maschineller Bewegungsklassifikation hierbei unterstützt werden können.

Kapitel 5 behandelt die Segmentierung menschlicher Bewegungsdaten auf Basis der bereits angesprochenen Ganzkörperposen-Taxonomie. Hierzu wird zunächst diese Taxonomie eingehend beschrieben und ein Verfahren zur Segmentierung menschlicher Bewegungsaufnahmen bezüglich der verwendeten Support-Posen vorgestellt. Mit Hilfe dieses Verfahrens wird anschließend eine statistische Analyse der Auftretenshäufigkeit einzelner Posen und ihrer Transitionen, sowie den mit diesen Transitionen verknüpften Zeitdauern durchgeführt, die zu einer partiellen Validierung der Taxonomie beiträgt.

Kapitel 6 befasst sich mit dem im Rahmen dieser Arbeit entwickelten datengetriebenen Ansatz zur Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen humanoider Roboter. Nach einer Definition der vorliegenden Problemstellung und einer Unterteilung der durch die Ganzkörperposen-Taxonomie vorgegebenen Support-Posen hinsichtlich ihrer kinematischen Konfiguration wird beschrieben, wie segmentierte menschliche Bewegungsdaten zum Lernen eines statistisches Modells für Posen-Transitionen genutzt werden können. Im Anschluss wird das entwickelte Verfahren zur Generierung von Posen-Sequenzen für eine gegebene Multi-Kontakt-Fortbewegungsaufgabe auf Basis des gelernten statistischen Modells vorgestellt und evaluiert.

Kapitel 7 beinhaltet abschließend eine kurze Zusammenfassung der wissenschaftlichen Beiträge der Arbeit und einen Ausblick auf zukünftige Forschungsarbeiten.

2 Erfassung und Repräsentation menschlicher Ganzkörperbewegung

Menschliche Ganzkörperbewegung nimmt in der vorliegenden Arbeit eine zentrale Rolle ein. Dieses Kapitel befasst sich deshalb zunächst mit den Grundlagen zu ihrer Erfassung, Repräsentation und Verarbeitung, die in den nachfolgenden Kapiteln Verwendung finden. Die Basis zur Repräsentation menschlicher Ganzkörperbewegung stellt hierbei das am H²T entwickelte MMM-Framework dar, welches in Abschnitt 2.1 näher beschrieben wird und als Kernkomponente ein Referenzmodell des menschlichen Körpers beinhaltet. In Abschnitt 2.2 werden anschließend die Vorgehensweise zur Erfassung der menschlichen Bewegung mit Hilfe eines markerbasierten passiv-optischen Bewegungserfassungssystems sowie zwei im Rahmen dieser Arbeit entwickelte Verfahren zur sich daran anschließenden Rekonstruktion von Gelenkwinkeltrajektorien erläutert. Abschließend werden in Abschnitt 2.3 Merkmale zur Beschreibung menschlicher Ganzkörperbewegung definiert, welche insbesondere in Kapitel 3 zur Klassifikation von Ganzkörperbewegungen Verwendung finden, und in Abschnitt 2.4 ein Verfahren zur kinematischen Übertragung menschlicher Bewegung auf humanoide Roboter erläutert.

2.1 Master Motor Map: Ein Ganzkörper-Referenzmodell des menschlichen Körpers

Bereits vor der Jahrtausendwende wurde in der Forschungsgemeinschaft die Notwendigkeit einer einheitlichen Darstellungsform und Repräsentation für menschliche Bewegung erkannt und es wurden entsprechende Anstrengungen unternommen (Gavrila et al. 1996; Ju et al. 1996; Sidenbladh et al. 2000), die in der Entwicklung verschiedener Frameworks für die Modellierung und Repräsentation menschlicher Bewegungen mündeten. Delp et al. (2007) beschreiben das OpenSim-Framework für die Erstellung von muskuloskelettalen Modellen des Menschen und ihrer dynamischen Simulation und präsentieren eine Fallstudie für ein Modell mit 21 Bewegungsfreiheitsgraden und insgesamt 92 Muskeln. Weitere Frameworks finden sich beispielsweise bei der HuMAnS Toolbox (Wieber et al. 2006) für die Analyse und Simulation menschlicher Bewegung mit einem Modell mit 36 Bewegungsfreiheitsgraden und dem HUMANOID Environment (Boulic et al. 1995) für die Modellierung und Animation des Menschen im Bereich der Computergrafik. Am H²T wurde mit einer ähnlichen Motivation das MMM-Framework ins Leben gerufen, welches zum ersten Mal in (Azad et al. 2007) vorgestellt wurde und im Anschluss Verwendung in zahlreichen wissenschaftlichen Arbeiten fand (Asfour et al. 2011; Do et al. 2011; Gärtner et al. 2010; Terlemez et al. 2014). Aufgeteilt auf die beiden Komponenten *MMMCORE* und *MMMTOLS* wird das MMM-Framework inzwischen als Open-Source-Projekt einem weltweiten Publikum frei zur Verfügung gestellt¹ und von verschiedenen Forschungsgruppen genutzt. Auch die vorliegende Arbeit baut zur Repräsentation und Verarbeitung menschlicher Ganzkörperbewegung auf dem MMM-Framework auf und erweitert es, wie beispielsweise in Abschnitt 2.2.2 gezeigt, um neue Ansätze.

¹ <https://gitlab.com/mastermotormap/>

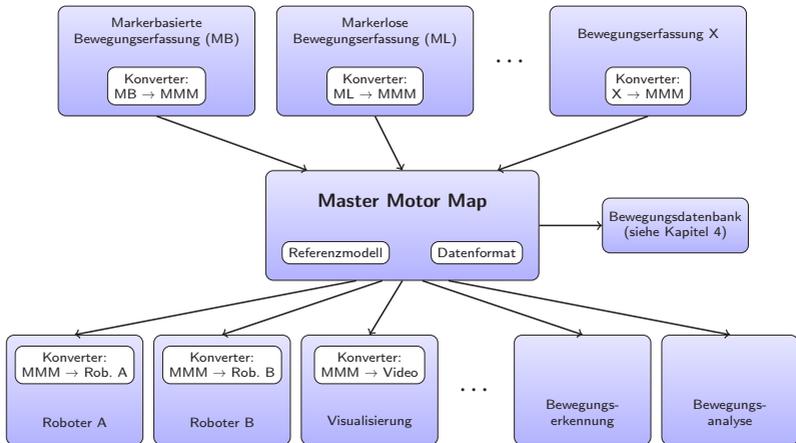


Abbildung 2.1: Durch das MMM-Framework realisierte Entkopplung der Erfassung menschlicher Bewegung von ihren verschiedenen Anwendungsfeldern, wie der Übertragung der Bewegung auf humanoide Roboter, Bewegungserkennung oder -analyse (modifiziert nach Mandery et al. 2016c, ©2016 IEEE). Als zentrale Komponenten des MMM-Frameworks fungieren hierbei das MMM-Referenzmodell des menschlichen Körpers und das MMM-Datenformat zur Speicherung und Verarbeitung menschlicher Bewegungsaufnahmen.

Der grundsätzliche Aufbau des MMM-Frameworks ist in Abbildung 2.1 dargestellt. Als zentralen Bestandteil beinhaltet das MMM-Framework ein Referenzmodell des menschlichen Körpers zusammen mit einem XML-basierten Dateiformat zur Repräsentation von Bewegungen basierend auf diesem Modell und den notwendigen Softwareroutinen, um Bewegungen in diesem Dateiformat zu lesen und zu speichern. Zur Überführung von aufgezeichneten Daten verschiedener Bewegungserfassungssysteme in die vereinheitlichende MMM-Darstellung (oberer Teil von Abbildung 2.1) stellt das MMM-Framework verschiedene sogenannte Konverter bereit, welche mit einer einheitlichen Schnittstelle angesteuert und auch mittels der durch die *MMMConverterGUI* angebotenen Benutzerschnittstelle genutzt werden können. Durch die offene Architektur des MMM-Frameworks können vom Benutzer jederzeit weitere Konverter für neuartige Bewegungserfassungs-

techniken hinzugefügt werden. Ausgehend von der vereinheitlichenden Darstellung der menschlichen Bewegung kann diese dann weiterverarbeitet und beispielsweise mit den Werkzeugen des MMM-Frameworks visualisiert, mit Konvertern auf humanoide Roboter übertragen oder zur Bewegungserkennung oder -analyse, beispielsweise für Dynamikberechnungen, eingesetzt werden (unterer Teil von Abbildung 2.1).



Abbildung 2.2: Multimodale Bewegungserfassung im Labor des H²T. Zusätzlich zur markerbasierten passiv-optischen Bewegungserfassung werden Farb- und Tiefenbilder aus verschiedenen Blickwinkeln und Ton mit einem 8-Kanal-Mikrofonarray aufgenommen.

Ein Hauptmerkmal des MMM-Konzepts ist der bereits in Abbildung 2.1 dargestellte multimodale Ansatz, welcher die Verwendung unterschiedlicher Sensormodalitäten zur Erfassung menschlicher Bewegung und die Repräsentation von Bewegungen auch als parallele (zeitsynchrone) Aufzeichnung verschiedener Modalitäten erlaubt. Ein Beispiel für eine derartige multimodale Bewegungsaufnahme ist in Abbildung 2.2 dargestellt. Insbesondere für die in Kapitel 4 beschriebene Bewegungsdatenbank kommen verschiedene Sensormodalitäten zum Einsatz, während sich die verbleibenden Teile der vorliegenden Arbeit auf die Verwendung von Daten aus markerbasierter Bewegungserfassung konzentrieren. Weitere Informationen über verwendete

Sensormodalitäten und allgemeine zusätzliche Informationen zum MMM-Framework finden sich in (Mandery et al. 2016c).

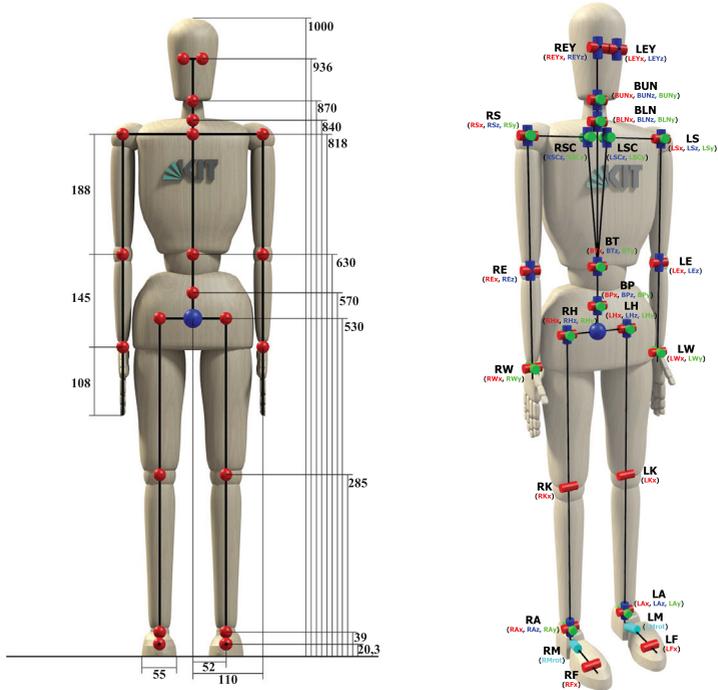


Abbildung 2.3: Kinematik des MMM-Referenzmodells für den menschlichen Körper (entnommen aus Mandery et al. 2016c, ©2016 IEEE). Links: Segmentlängen des Referenzmodells in Millimetern bezogen auf eine Körpergröße von einem Meter. Rechts: Rotatorische Bewegungsfreiheitsgrade des Modells mit dazugehörigen Gelenknamen und farblich gekennzeichneten Rotationsachsen (Kinematik der Hand nicht dargestellt).

Wie bereits erwähnt, beinhaltet das MMM-Framework als zentrale Komponente ein Ganzkörper-Referenzmodell des Menschen für die Repräsentation menschlicher Bewegungen. Dieses Referenzmodell umfasst sowohl kinematische als auch dynamische Spezifikationen für den menschlichen Körper, welche der biomechanischen Literatur entnommen wurden, und liegt

im Simox-Modellformat (Vahrenkamp et al. 2012) vor, sodass notwendige Berechnungen beispielsweise zur Vorwärtskinematik auf die vorhandenen Implementierungen in der Simox-Bibliothek zurückgreifen können.

Die kinematischen Spezifikationen des Modells umfassen die Definition von Gelenken zusammen mit den sie verbindenden Segmentlängen und biomechanisch fundierten Gelenkwinkelgrenzen. Insgesamt werden 104 Bewegungsfreiheitsgrade definiert, die mit Ausnahme der Hände in Abbildung 2.3 dargestellt sind. 6 dieser Bewegungsfreiheitsgrade repräsentieren die Root-Pose für den in der Hüfte des Modells gelegenen Wurzelknoten und insgesamt 52 Bewegungsfreiheitsgrade wurden für rotatorische Gelenke in Körpertorso, Extremitäten, Kopf und Augen festgelegt. Die Modellierung dieser 52 Bewegungsfreiheitsgrade orientiert sich weitestgehend an den Arbeiten von Winter (2009), und im Weiteren Simonidis et al. (2007) für die angenäherte Modellierung der Wirbelsäule mit drei Segmenten, Myers et al. (2004) für die Modellierung des Fußes mit vier Segmenten, sowie Dempster et al. (1959) zur Beschreibung der Kinematik des menschlichen Arms. Die verbleibenden 46 Bewegungsfreiheitsgrade verteilen sich auf die beiden Hände mit jeweils 23 Bewegungsfreiheitsgraden und basieren auf den Angaben von León et al. (2014) und Buchholz et al. (1992). Obwohl mit den 104 Bewegungsfreiheitsgraden ein sehr detailliertes Modell zur Beschreibung menschlicher Bewegung zur Verfügung steht, werden bei verschiedenen Anwendungen oftmals nicht alle Gelenke des Modells tatsächlich benötigt und das MMM-Framework erlaubt aus diesem Grund in allen Verarbeitungsschritten die Beschränkung der Kinematik auf eine Teilmenge der verfügbaren Gelenke. Zusätzlich zur Kinematik beinhaltet das MMM-Modell ebenfalls dynamische Spezifikationen für den menschlichen Körper durch die Angabe der Masse, des Massenschwerpunkts und des Trägheitstensors für jedes der Modellsegmente. Da die Erfassung der dynamischen Parameter bei einem lebenden Probanden nur mit einem hohen Aufwand und indirekt möglich

ist², beruhen diese Spezifikationen ebenfalls auf statistischen Analysen in der biomechanischen Forschung (Leva 1996; Contini 1972).

Alle Spezifikationen des MMM-Modells beziehen sich hierbei auf eine normalisierte Darstellung für eine Körpergröße von einem Meter und einer Gesamtmasse von einem Kilogramm, wodurch zur Verwendung des Modells zunächst noch ein Adaptionsschritt mit Hilfe der im MMM-Framework vorgesehenen Modell-Prozessoren (*model processors*) erforderlich ist. Im einfachsten Fall, welcher im Rahmen dieser Arbeit angenommen wird, werden die kinematischen Spezifikationen in Form von Segmentlängen, sowie die dynamischen Spezifikationen von Segmentmassen und -massenschwerpunkten linear auf die gemessene und somit bekannte Körpergröße und das Körpergewicht des menschlichen Probanden skaliert. Die Trägheitstensoren der Modellsegmente können anschließend durch Anwendung des Parallelachsen-Theorems adaptiert werden (Winter 2009). Für weitere Informationen zum MMM-Referenzmodell sei auf (Mandery et al. 2016c) und (Terlemez 2017) verwiesen, welche umfassendere Spezifikationen und eine ausführliche Diskussion getroffener Entwurfsentscheidungen beinhalten.

2.2 Erfassung menschlicher Ganzkörperbewegung

2.2.1 Markerbasierte Bewegungserfassung

Die Erfassung der menschlichen Bewegung nimmt eine wichtige Rolle in einer Vielzahl von Anwendungen ein, neben der Robotik beispielsweise auch in der biomechanischen Forschung, in Rehabilitation und Sportwissenschaften, aber auch in der Unterhaltungsindustrie zur Produktion von Filmen und Videospielen. Bedingt durch dieses weite Spektrum an Anwendungsfeldern

² Eine Möglichkeit stellt das in (Venture et al. 2008) vorgestellte Verfahren zur Bestimmung von Segmentmassen, -massenschwerpunkten und -trägheitsparametern aus Daten eines Bewegungserfassungssystems und von Kraftmessplatten dar, welches in (Venture et al. 2009) auf eine Echtzeitanwendung erweitert wurde.

wurde eine große Zahl unterschiedlicher Verfahren zur Bewegungserfassung und eine mannigfaltige Auswahl an einsatzbereiten Produkten entwickelt. Field et al. (2011, 2009) liefern eine Übersicht über gängige in Anwendungen der Robotik zum Einsatz kommende Verfahren und Sensoren. Sie schlagen eine Kategorisierung der Verfahren in optische, beschleunigungsbasierte, mechanische, magnetische und akustische Ansätze vor, die in Abbildung 2.4 schematisch dargestellt ist und als Grundlage für die nachfolgenden Ausführungen dienen soll.

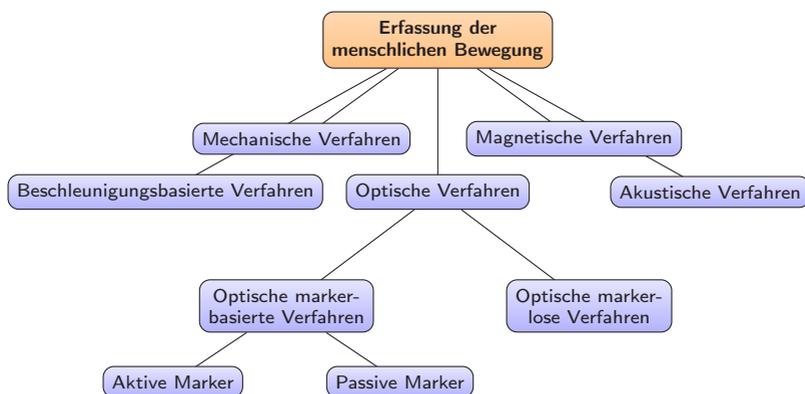


Abbildung 2.4: Kategorisierung von Verfahren zur Erfassung der menschlichen Bewegung nach Field et al. (2011, 2009).

Verfahren zur optischen Bewegungserfassung lassen sich weiter unterteilen in markerbasierte und markerlose Ansätze. Bei markerbasierten Ansätzen kommen am menschlichen Körper angebrachte Marker zum Einsatz, welche manuell an zuvor definierten anatomischen Orientierungspunkten (*anatomical landmarks*) platziert und vom System lokalisiert werden. Die Lokalisierung dieser Marker erfolgt mit Hilfe eines Mehrkamerasystems, welches eine Triangulation der 3D-Koordinaten der Marker im Raum anhand der 2D-Kamerabilder mit Hilfe der bekannten Techniken der Epipolargeometrie erlaubt und zur Erfassung von Ganzkörperbewegungen sinnvollerweise

aus einer allseitigen Anordnung von Kameras aufgebaut wird³. Zum Einsatz kommen können sowohl aktive Marker, welche über eine eigenständige Stromversorgung verfügen und beispielsweise mit Hilfe einer LED eine gepulste Codefolge emittieren, als auch passive Marker, die im infraroten Spektrum des Lichts reflektive Eigenschaften aufweisen und von einer an den Kameras angebrachten infraroten Lichtquelle beleuchtet werden. Als Vorteile von optischen markerbasierten Verfahren ist ihre potentiell sehr hohe Genauigkeit der Markerlokalisierung im Submillimeter-Bereich zu nennen, sowie die (in Innenräumen) hohe Robustheit gegenüber Störeinflüssen und die hohen erzielbaren Bildfrequenzen von bis zu 2 kHz (Field et al. 2011). Als Nachteile fallen allerdings ihr hoher Preis und der apparative Aufwand für Kameras und deren Kalibrierung ins Gewicht, zusammen mit der mangelnden Mobilität dieser Systeme. Markerlose Ansätze zur optischen Bewegungserfassung versuchen eine direkte Erfassung der menschlichen Pose aus Farb- und gegebenenfalls Tiefenbildern ohne die Notwendigkeit von Markern zu realisieren. Für den Einsatz markerloser Verfahren im Bereich des maschinellen Sehens findet sich in (Moeslund et al. 2006, 2001) eine ausführliche Übersicht über 424 Arbeiten, welche allerdings neuere Arbeiten seit 2006 nicht berücksichtigt, und eine Einordnung verschiedener Verfahren, die bei markerloser Bewegungserfassung zum Beispiel für die Initialisierung und Schätzung der menschlichen Pose zum Einsatz kommen können. Insgesamt handelt es sich bei der markerlosen optischen Bewegungserfassung um ein aktives Forschungsgebiet, wobei die Verfahren hinsichtlich Genauigkeit und Fehlerrate bislang nicht mit markerbasierten Ansätzen konkurrieren können und zudem oft eine hohe Anfälligkeit gegenüber Beleuchtungsbedingungen und eine durch den Berechnungsaufwand der Verfahren niedrigere Bildfrequenz aufweisen. Exemplarisch findet sich in (Puthenveetil et al. 2013) eine

³ Obwohl bereits zwei Kameras als Stereokamerasystem eine eindeutige Rekonstruktion der Markerposition erlauben würden, ist eine höhere Zahl an Kameras erforderlich, um die auf allen Seiten des Menschen angebrachten Marker erfassen und das Auftreten von Verdeckungen reduzieren zu können.

Gegenüberstellung von markerbasierter und markerloser Bewegungserfassung für einen konkreten Anwendungsfall und ein Vergleich der Genauigkeit beider Verfahren anhand der gemessenen Gelenkwinkel und Segmentlängen des menschlichen Körpers, bei welcher die aus der Anwendung markerbasierter Bewegungserfassung resultierende durchschnittliche Standardabweichung um ca. den Faktor 8 (Gelenkwinkel) bzw. 4 (Segmentlängen) unter der des markerlosen Ansatzes lag.

Beschleunigungsbasierte Verfahren zur Bewegungserfassung basieren auf Messungen von Beschleunigungssensoren bzw. inertialen Messeinheiten, welche analog zu Markern an zuvor definierten anatomischen Orientierungspunkten des menschlichen Körpers platziert werden. Als wesentlicher Nachteil dieser Verfahren gilt die durch Integration von Beschleunigungswerten entstehende Drift, die zu im Zeitverlauf anwachsenden Positionsungenauigkeiten führt und somit die Bestimmung der absoluten Position des Menschen bei längeren Bewegungen unmöglich macht.

Mechanische Ansätze nehmen eine direkte Messung der menschlichen Gelenkwinkel durch die Verwendung eines mit entsprechenden Sensoren versehenen Exoskeletts vor. Da nur Gelenkwinkel und nicht die Position gemessen werden, können zahlreiche Bewegungen wie beispielsweise Dreh- oder Sprungbewegungen nicht erfasst werden. Hinzu kommen die bei Exoskeletten bekannten Probleme der Misalignments, welche durch die Schwierigkeit einer mechanischen Abbildung der hochkomplexen menschlichen Kinematik entstehen.

Magnetische Verfahren basieren wie markerbasierte optische Ansätze auf am menschlichen Körper angebrachten Markern, die hier jedoch als Transmitter durch Messung ihres elektromagnetischen Feldes mit mehreren Empfangsantennen lokalisiert werden. Die Verwendung von magnetischen Messungen umgeht die Probleme optischer Systeme bei visuellen Verdeckungen, beschränkt allerdings den Arbeitsraum stark und erhöht die Ungenauigkeit und Störanfälligkeit gegenüber äußeren Einflüssen.

Abschließend können auch akustische Verfahren zum Einsatz kommen, bei denen am menschlichen Körper angebrachte Mikrofone eine Lokalisierung anhand von periodisch ausgesendeten Ultraschall-Impulsen erlauben. Wie bei magnetischen Verfahren ist hier eine Störanfälligkeit gegenüber Umgebungseinflüssen gegeben, zu denen wie bei optischen Verfahren Probleme mit Verdeckungen hinzukommen können.

Aufgrund der genannten Vorteile von markerbasierten passiv-optischen Verfahren zur Bewegungserfassung wie der hohen Robustheit und Präzision wurde diese Erfassungstechnik für die im Rahmen dieser Arbeit verwendeten Bewegungsaufnahmen gewählt. Hierzu steht am H²T ein MX10-System des Herstellers Vicon (Vicon o.D.) zur Verfügung, welches als kommerziell erhältliches System einen derartigen Ansatz der Bewegungserfassung umsetzt. Das System arbeitet mit zehn Vicon T10-Kameras, die über eine leuchtstarke, auf das Reflektionsverhalten der passiven Marker abgestimmte Lichtquelle im Infrarotspektrum verfügen. Die Kameras verfügen jeweils über einen Vicon Vegas-S CMOS-Sensor und werden mit einer Auflösung von 1120x896 Pixeln (1 Megapixel) und einer Bildfrequenz von 100 Hz betrieben, wobei acht der Kameras an stationären Positionen entlang der oberen Wandbereiche montiert sind und die verbleibenden zwei Kameras auf Stativen aufgabenspezifisch positioniert werden können. Als Vorschrift zur Positionierung der Marker auf dem menschlichen Körper wird das am H²T entwickelte *KIT Reference Marker Set* verwendet, welches die Positionierung der in Abbildung 2.5 gezeigten 56 Marker festlegt. Weitere Informationen über dieses Marker Set sind in (Mandery et al. 2015b, 2016c) verfügbar und online abrufbar⁴. Zusätzlich zur markerbasierten Bewegungserfassung können je nach Fragestellung zeitsynchron weitere Sensormodalitäten aufgezeichnet werden, beispielsweise Videoaufnahmen, Daten inertialer Messeinheiten oder Messungen von Kraftsensoren. Diese zusätzlichen Daten können ebenfalls über die in Kapitel 4 beschriebene Bewegungsdatenbank bereitgestellt

⁴ https://motion-database.humanoids.kit.edu/marker_set/

werden, wurden allerdings im Rahmen dieser Arbeit mit Ausnahme der zu Evaluations- und Illustrationszwecken fast immer durchgeführten Videoaufnahmen nicht weiter verwendet.



Abbildung 2.5: Positionierung von passiv-reflektierenden Markern am menschlichen Körper gemäß dem *KIT Reference Marker Set* (entnommen aus Mandery et al. 2016c, ©2016 IEEE).

Nach Abschluss der markerbasierten Bewegungserfassung liegt die aufgezeichnete Bewegung zunächst in einem proprietären Rohformat vor, in welchem die lokalisierten 3D-Positionen der Marker zunächst nur teilweise und fehlerbehaftet den korrespondierenden Markerbezeichnungen (*marker labels*) zugeordnet wurden und Trajektorien für Marker während visueller Verdeckungen gänzlich fehlen. Mittels manueller Nachbearbeitung durch den Menschen werden fehlerhafte Markerbezeichnungen korrigiert, fehlende Markerbezeichnungen ergänzt, Artefakte wie fehlerhafte Störmarker entfernt und die Lücken (*gaps*) in Markertrajektorien bei Verdeckungen durch

verschiedene Verfahren wie Spline-Interpolation geschlossen (siehe auch Aristidou et al. 2013). Im Anschluss stehen danach vollständig definierte Trajektorien für jeden der 56 Marker zur Verfügung, welche im C3D-Format (C3D o.D.) als etabliertem Standard-Format markerbasierter Bewegungsaufnahmen exportiert werden und dadurch für vielfältige Anwendungen und Frameworks wie das MMM-Framework, aber auch beispielsweise das in (Barre et al. 2014) vorgestellte Biomechanical ToolKit (BTK), nutzbar sind.

2.2.2 Rekonstruktion von Gelenkwinkeln

Der erste Schritt zur weiteren Verarbeitung markerbasierter Bewegungsdaten und zur Bestimmung der in Abschnitt 2.3 beschriebenen Merkmale besteht in der Rekonstruktion von Gelenkwinkeltrajektorien für das MMM-Referenzmodell des menschlichen Körpers auf Basis der vorliegenden Markertrajektorien im kartesischen Raum. Wie in Abschnitt 2.1 beschrieben, kommen dazu in der Terminologie des MMM-Frameworks sogenannte Konverter zum Einsatz. Als Verbesserung der im MMM-Framework existierenden Ansätze (Azad et al. 2007; Terlemez et al. 2014), wurde hierfür in (Mandery et al. 2015a) ein Ansatz auf Basis nicht-linearer Optimierung vorgestellt, der nach seiner Integration als neuer Konverter in das MMM-Framework auch in (Mandery et al. 2016c) beschrieben wurde. Die Bewegungsrekonstruktion wird hierbei in Übereinstimmung mit etablierten Arbeiten aus der Biomechanik (Delp et al. 2007; Lu et al. 1999) als Optimierungsproblem mit einem *Least-Squares-Ansatz* formuliert. Zur Durchführung wird zunächst das MMM-Referenzmodell durch das Hinzufügen von *virtuellen* Markern erweitert, welche mit den im Marker Set definierten Markern auf dem menschlichen Körper korrespondieren und analog zu den Vorschriften für den menschlichen Probanden auf den Modellsegmenten positioniert werden. Als nächster Schritt wird das Referenzmodell dann auf die bekannte Körpergröße des menschlichen Probanden angepasst, wobei alle Modellsegmente linear skaliert werden.

Seien nun $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_n)$ die gemessenen kartesischen Koordinaten der n Marker in einem Zeitschritt (im Falle des MMM-Modells gilt $n = 56$) und $\mathbf{x} = (p_x, p_y, p_z, \alpha, \beta, \gamma, \theta_1, \dots, \theta_m)$ der Vektor, welcher die resultierende Pose des MMM-Modells als Kombination aus der Root-Pose des Modells (Root-Position und -Rotation) und seinen m Gelenkwinkeln beschreibt. Zusätzlich seien $\mathbf{V}(\mathbf{x}) = (\mathbf{v}_1(\mathbf{x}), \dots, \mathbf{v}_n(\mathbf{x}))$ die Positionen der zuvor angesprochenen virtuellen Marker, welche für bekanntes \mathbf{x} über die Vorwärtskinematik des Modells bestimmt werden können. Die Bestimmung der Pose des MMM-Modells zu den beobachteten Markerkoordinaten in einem Zeitschritt wird dann formuliert als Minimierung von

$$f(\mathbf{x}) = \sum_{i=1}^n \|\mathbf{u}_i - \mathbf{v}_i(\mathbf{x})\|^2, \quad (2.1)$$

wobei als Nebenbedingung die Einhaltung der durch das Modell bzw. letztendlich durch die Anatomie des Menschen vorgegebenen Gelenkwinkelgrenzen für $\theta_1, \dots, \theta_m$ sichergestellt werden muss. Zur Rekonstruktion einer Bewegung wird Gleichung 2.1 mit Hilfe der Neuimplementierung des von Rowan (1990) vorgestellten Subplex-Algorithmus in der NLOpt-Bibliothek (Johnson o.D.) für nicht-lineare Optimierung gelöst. Als Startwert für \mathbf{x} wird hierbei in jedem Zeitschritt das Resultat der Optimierung des vorhergehenden Zeitschritts bzw. zur Initialisierung der Nullvektor verwendet.

Wie bereits erwähnt, setzt das vorgestellte Verfahren zur Rekonstruktion von Gelenkwinkeln korrekt gekennzeichnete (gelabelte) Markertrajektorien voraus, damit in jedem Schritt die kartesischen Abstände zwischen den korrespondierenden Paaren von beobachteten und virtuellen Markern berechnet werden können. Bei gewöhnlich im C3D-Format vorliegenden Bewegungsaufnahmen muss diese Information in der Regel zunächst in einem aufwendigen Nachbearbeitungsprozess mit menschlicher Unterstützung erstellt werden. Sehr interessant sind daher Verfahren, welche diese Information nicht benötigen und direkt auf ungelabelten Marker-Punktwolken arbeiten

können, wie sie von Maycock et al. (2015), Meyer et al. (2014) und Schubert et al. (2015) in der jüngeren Vergangenheit im Bereich der Robotik vorgestellt wurden. Vom Autor selbst wurde in (Steinbring et al. 2016a) im Rahmen einer Kooperation mit dem Lehrstuhl für Intelligente Sensor-Aktor-Systeme (ISAS) des KIT ein Verfahren präsentiert, um das Problem der Rekonstruktion von Ganzkörperbewegungen auf Basis markerbasierter ungelabelter Bewegungsdaten zu lösen⁵.

Für dieses Verfahren wird das Rekonstruktionsproblem als Problem der Zustandsschätzung für ein zeitdiskretes nicht-lineares dynamisches System interpretiert, wobei der zu schätzende Systemzustand der Pose des Menschen entspricht, wie zuvor repräsentiert durch die Gelenkwinkel und die 6D-Root-Pose. Ähnlich wie in (Bonnet et al. 2016) wird eine Variante eines Kalman-Filters zur Schätzung dieses Systemzustands eingesetzt, hier der am ISAS entwickelte Smart Sampling Kalman Filter (Steinbring et al. 2016b), bei welchem es sich um eine Variante eines *Linear Regression Kalman Filters* (LRKF) handelt. Da die Formalisierung als Kalman-Filter zunächst keine Integration der durch das Modell vorgegebenen Gelenkwinkelgrenzen ermöglicht, werden durch eine Parametertransformation auf Basis einer periodischen Funktion Werte aus dem Intervall $[-\infty, \infty]$ auf den vom Modell erlaubten Wertebereich für jedes Gelenk abgebildet und so die Einhaltung von Gelenkwinkelgrenzen durch das Kalman-Filter bei Verwendung von unbegrenzten Ersatzparametern garantiert. Zur Berechnung der bei ungelabelten Daten selbst zu bestimmenden Marker-Labels bzw. Korrespondenzen zwischen beobachteten und virtuellen Markern wird ein *Global-Nearest-Neighbor*-Ansatz auf Basis des LAPJV-Algorithmus (Jonker et al. 1987)

⁵ In (Steinbring et al. 2015) wurde zunächst eine Variante des Verfahrens vorgestellt, welche ein Labeling der Eingabedaten voraussetzt und somit in Anforderungen und Ergebnissen dem vorgestellten Ansatz auf Basis nicht-linearer Optimierung entspricht. Durch Formulierung als MHT-Problem wurde dieser Ansatz dann in (Steinbring et al. 2016a) für eine Anwendung mit ungelabelten Eingabedaten erweitert.

angewendet, der eine effiziente Variante des bekannten Ungarischen Algorithmus (Kuhn 1955) darstellt.

Ein wesentliches Problem hierbei besteht in der korrekten Etablierung der initialen Korrespondenzen im ersten Zeitschritt und der im Fehlerfall möglichen Konvergenz in falsche lokale Minima, die auch in darauffolgenden Zeitschritten nicht mehr verlassen werden können. Will man die mögliche Startpose des Menschen nicht künstlich einschränken (bspw. Armorientierungen vorgeben) oder im Extremfall gar eine fixe Startpose für den Menschen festlegen, kann das Filter-Problem als *Multiple-Hypotheses-Tracking-Problem* (MHT-Problem) (Blackman et al. 1999) aufgefasst werden, für das initial verschiedene angenommene Startkonfigurationen verfolgt und mit jeweils einer eigenen Instanz des Kalman-Filters geschätzt werden. Durch systematisches Sampling der Kinematik des MMM-Modells wurden hierzu 100 mögliche initiale Ganzkörperkonfigurationen des Menschen festgelegt, von denen im Zeitablauf unwahrscheinliche Filter, welche die beobachteten Markermessungen nicht hinreichend erklären können, oder solche, die gegen einen identischen Systemzustand konvergiert haben, eliminiert werden, bis schlussendlich lediglich eine Filter-Instanz weiter verfolgt werden muss.

Zur Evaluation des beschriebenen ungelabelten Ansatzes wurden exemplarisch 20 nicht nachbearbeitete und somit ungelabelte Aufnahmen komplexer Ganzkörper-Bewegungsabläufe mit verschiedenen initialen Posen betrachtet, für die nur in zwei Fällen keine Konvergenz des beschriebenen Verfahrens gegen das korrekte globale Optimum und somit keine korrekte Rekonstruktion der menschlichen Bewegung beobachtet werden konnte. Ein Beispiel für die Rekonstruktion eines derartigen Bewegungsablaufs ist in Abbildung 2.6 zu sehen, wobei hier der menschliche Proband erschwerend sogar außerhalb des Arbeitsbereichs des Bewegungserfassungssystems startet, womit initial keine Marker erfasst werden können und diese erst im Zeitablauf sukzessive sichtbar werden. Für die evaluierten Bewegungen wurde gezeigt, dass der Ansatz nach der Initialisierungsphase, und somit der Elimination aller Filter-Instanzen bis auf eine, bei einer Bildfrequenz von 100 Hz echtzeitfähig ist.

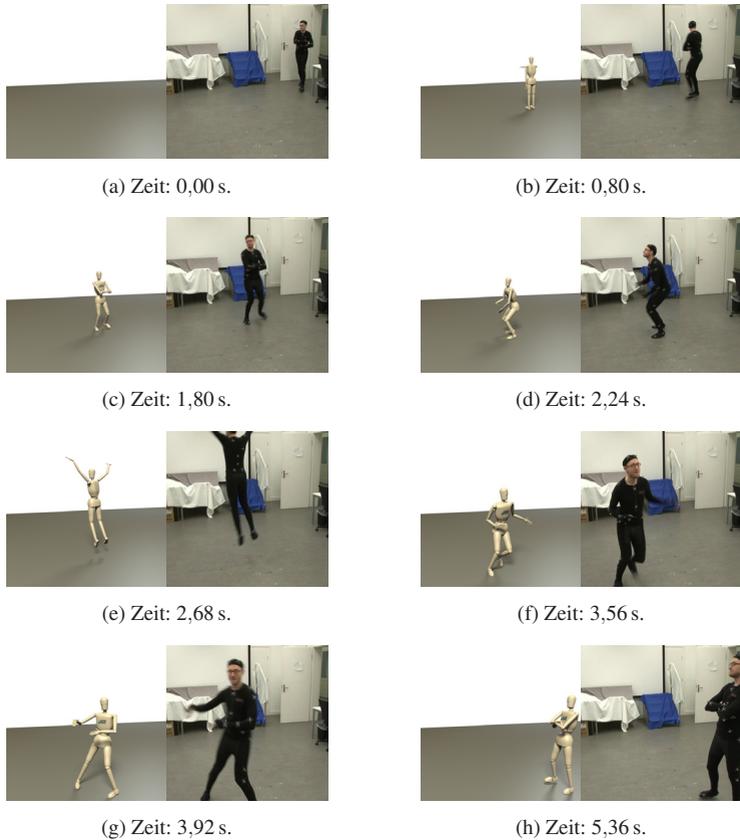


Abbildung 2.6: Rekonstruktion eines komplexen menschlichen Ganzkörper-Bewegungsablaufs basierend auf ungelabelten Markermessungen (entnommen aus Steinbring et al. 2016a, ©2016 IEEE).

Für eine ausführlichere Diskussion des Verfahrens und seiner Evaluation sei der Leser auf (Steinbring et al. 2016a) verwiesen. Obwohl der hier besprochene ungelabelte Ansatz mit dem MMM integriert wurde und seine Evaluation vielversprechende Ergebnisse gezeigt hat, wurde auf eine Anwendung des Verfahrens im Rahmen dieser Arbeit letztendlich verzichtet, da die Bereitstellung aufgezeichneter Bewegungsdaten im etablierten C3D-Format als

Grundanforderung an die in Kapitel 4 vorgestellte Bewegungsdatenbank ohnehin die vollständige manuelle Nachbearbeitung der Bewegungsaufnahmen zwingend erforderlich macht. Somit könnte der größte Vorteil des vorgestellten Verfahrens – das nicht erforderliche manuelle Labeling der Markerdaten – nicht mehr genutzt werden, da alle Eingabedaten bereits in einem gelabelten Format vorliegen.

Abschließend soll noch erwähnt werden, dass das eingangs beschriebene Verfahren zur Gelenkwinkelrekonstruktion basierend auf nicht-linearer Optimierung auch zur Rekonstruktion der Pose von Umgebungsobjekten als gelenklose Starrkörper zum Einsatz kommt. Wie bei der Diskussion der in Kapitel 4 beschriebenen Bewegungsdatenbank ausführlicher begründet, ist die Kenntnis der Lage und Orientierung von Umgebungselementen und Objekten von essenzieller Bedeutung für die Analyse der Interaktion zwischen Mensch und Umwelt, beispielsweise für die Mensch-Objekt-Interaktion bei Manipulationsaufgaben sowie auch für die Erkennung von Support-Kontakten bei dem in Kapitel 5 beschriebenen Verfahren. Zur Bewegungserfassung werden relevante Umgebungsobjekte daher ebenfalls mit Markern versehen und zu den Objekten korrespondierende 3D-Modelle erstellt, in welchen analog zum MMM-Referenzmodell virtuelle Marker definiert werden. Als Mindestzahl zur eindeutigen Rekonstruktion der Objektpose werden drei auf eine nicht-symmetrische Weise angebrachte Marker benötigt, wobei sich aus Gründen der Robustheit in den meisten Fällen eine größere Anzahl an Markern als nützlich erwiesen hat, sofern das Objekt deren Anbringung ermöglicht. Im Anschluss können diese am Objekt angebrachten Marker zeitsynchron zum menschlichen Probanden vom Bewegungserfassungssystem lokalisiert und das beschriebene Verfahren zur Gelenkwinkelrekonstruktion im Spezialfall $m = 0$ (nur Root-Pose und keine Gelenkwinkel) angewendet werden.

Abbildung 2.7 zeigt verschiedene Umgebungselemente und Objekte, die nach der beschriebenen Vorgehensweise mit Markern versehen wurden und am H²T für Bewegungsaufnahmen zur Verfügung stehen. Abbildung 2.8 zeigt verschiedene Szenen einer solchen Bewegungsaufnahme, in welcher

der Proband mit mehreren Umgebungsobjekten interagiert, zusammen mit einer Visualisierung der rekonstruierten Bewegung. Wie wir in Kapitel 4 und Kapitel 5 sehen werden, erlaubt die Kenntnis der exakten Posen für die als bekannt angenommenen Objektmodelle eine Untersuchung der menschlichen Interaktion mit seiner Umgebung, die zum Verständnis menschlicher Ganzkörperbewegung unerlässlich ist.



Abbildung 2.7: Mit Markern versehene Umgebungselemente und Objekte zur zeitsynchronen Erfassung der Bewegungen von Mensch und Objekten (entnommen aus Mandery et al. 2016c, ©2016 IEEE).

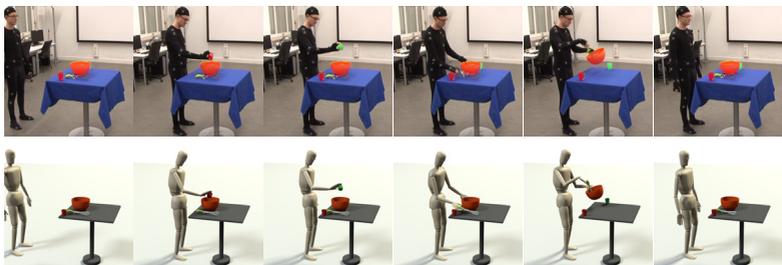


Abbildung 2.8: Rekonstruktion von Gelenkwinkeln für eine Bewegung (entnommen aus Mandery et al. 2015b, ©2015 IEEE). Zusätzlich zur menschlichen Ganzkörperbewegung werden auch die Posen der ebenfalls mit Markern versehenen Objekte rekonstruiert.

2.3 Merkmale zur Beschreibung menschlicher Bewegung

Wie zuvor bereits erläutert, kann eine menschliche Ganzkörperbewegung sowohl durch die Markertrajektorien eines markerbasierten Bewegungserfassungssystems als auch durch rekonstruierte Gelenkwinkeltrajektorien eines Modell des menschlichen Körpers dargestellt werden. Allerdings bleibt zunächst festzuhalten, dass zum einen noch unklar ist, welche Repräsentation der Bewegung für einen spezifischen Anwendungsfall besser geeignet ist, und zum anderen beide Repräsentationsformen hochdimensionale und stark redundante Merkmale darstellen⁶. Insbesondere die Hochdimensionalität der Repräsentation kann sich bei Verfahren wie der Bewegungsklassifikation auf Basis von Hidden Markov Models, welche in Kapitel 3 vorgestellt wird, negativ sowohl auf die benötigte Rechenzeit, als auch auf die numerische Stabilität der Verfahren auswirken. Im Folgenden sollen in Abschnitt 2.3.1 daher zusätzliche Merkmale zur Beschreibung menschlicher Ganzkörperbewegung mit einem Fokus auf niedrigdimensionale Merkmale definiert werden. In Abschnitt 2.3.2 werden abschließend zentrale Verarbeitungsschritte näher erläutert, welche zur Berechnung der eingeführten Merkmale erforderlich sind. Die vorliegende Definition von Merkmalen wurde zusammen mit der in Abschnitt 3.3 und Abschnitt 3.5.1 beschriebenen Merkmalsauswahl in (Mandery et al. 2016b) publiziert und im Folgenden werden zur Erhöhung der Übersichtlichkeit die dort eingeführten, vom Englischen abgeleiteten Merkmal-Bezeichnungen beibehalten.

⁶ Als Beispiele für Redundanz sei angeführt, dass beobachtete Positionen unterschiedlicher Marker bei Kenntnis der menschlichen Anatomie offensichtlich nicht vollständig unabhängig voneinander sein können und die Wahl eines Gelenkwinkels zur Vermeidung von Selbstkollisionen auch einen Einfluss auf den erlaubten Wertebereich anderer Gelenke haben muss.

Tabelle 2.1: Übersicht über die definierten Merkmale zur Beschreibung menschlicher Ganzkörperbewegung, zusammen mit der jeweiligen Dimensionalität der Merkmale (modifiziert nach Mandery et al. 2016b, ©2016 ISIF).

Merkmal-Gruppe	Merkmal-Bezeichnung	Dim.
Marker-Merkmale	marker_pos	$56 \cdot 3 = 168$
	marker_vel	$56 \cdot 3 = 168$
	marker_vel_norm	1
	marker_acc	$56 \cdot 3 = 168$
	marker_acc_norm	1
Gelenkwinkel-Merkmale	joint_pos	40
	joint_vel	40
	joint_vel_norm	1
	joint_acc	40
	joint_acc_norm	1
Root-Pose-Merkmale	root_pos	3
	root_vel	3
	root_vel_norm	1
	root_acc	3
	root_acc_norm	1
	root_rot	3
	root_rot_norm	1
Massenschwerpunkt-Merkmale	CoM_pos	3
	CoM_vel	3
	CoM_vel_norm	1
	CoM_acc	3
	CoM_acc_norm	1
Endeffektor-Merkmale	end_effectors_pos	$4 \cdot 3 = 12$
	end_effectors_vel	$4 \cdot 3 = 12$
	end_effectors_vel_norm	4
	end_effectors_acc	$4 \cdot 3 = 12$
	end_effectors_acc_norm	4
Drehimpuls-Merkmale	angular_momentum	3
	angular_momentum_norm	1

2.3.1 Definition von Merkmalen

Insgesamt wurde eine Gesamtzahl von 29 unterschiedlichen Merkmalen zur Beschreibung menschlicher Ganzkörperbewegung definiert, welche sich in sechs unterschiedliche Gruppen unterteilen lassen. Tabelle 2.1 bietet einen Überblick über diese Merkmale, die im Folgenden in der durch die Tabelle vorgegebenen Reihenfolge erläutert werden sollen.

Zunächst können direkt die vom Bewegungserfassungssystem triangulierten und manuell nachbearbeiteten Markertrajektorien als Merkmal zur Beschreibung menschlicher Ganzkörperbewegung verwendet werden. Das Merkmal *marker_pos* beschreibt die kartesischen Koordinaten der am menschlichen Körper angebrachten Marker zu einem bestimmten Zeitpunkt. Da insgesamt 56 dieser Marker verwendet werden (vgl. Abschnitt 2.2.1), handelt es sich somit mit 168 Dimensionen ($56 \cdot 3$) um ein ausgesprochen hochdimensionales Merkmal. Ausgehend von der im Merkmal *marker_pos* enthaltenen Position \mathbf{p}_t^i des Markers i zum Zeitpunkt t kann anschließend mit Hilfe des zentralen Differenzenquotienten die lokale Ableitung und somit der Geschwindigkeitsvektor \mathbf{v}_t^i des Markers i angenähert werden:

$$\mathbf{v}_t^i = \frac{\mathbf{p}_{t+1}^i - \mathbf{p}_{t-1}^i}{2\Delta t}. \quad (2.2)$$

Δt entspricht hierbei der Länge eines Zeitschritts, also dem Kehrwert der Bildfrequenz des zugrunde liegenden Bewegungserfassungssystems. Das Merkmal *marker_vel* beinhaltet die Geschwindigkeitsvektoren aller 56 Marker, womit es sich ebenfalls um ein 168-dimensionales Merkmal handelt. Für das eindimensionale Merkmal *marker_vel_norm* wird die euklidische Norm des kombinierten Geschwindigkeitsvektors berechnet, die einen niedrigdimensionalen Anhaltspunkt für das gesamte Ausmaß der Markerbewegungen liefert. Analog zur Berechnung der Geschwindigkeitsmerkmale aus den Markerpositionen werden ausgehend von *marker_vel* die beiden Merkmale

marker_acc und *marker_acc_norm* berechnet, welche die Markerbeschleunigungen beschreiben.

Ebenfalls können die mit dem in Abschnitt 2.2.2 beschriebenen Verfahren bestimmten Gelenkwinkeltrajektorien als Merkmal zur Beschreibung einer Bewegung verwendet werden, was durch das Merkmal *joint_pos* dargestellt wird. Bei der Verwendung von markerbasierten Verfahren zur Bewegungserfassung können allerdings nur für eine Teilmenge der 98 in Abschnitt 2.1 vorgestellten rotatorischen Gelenke des Modells die Gelenkwinkel rekonstruiert werden, da beispielsweise mit dem verwendeten Marker Set keine Rekonstruktion der Hand-Pose und aus naheliegenden Gründen auch keine Erkennung der Augen-Pose in Form der Blickrichtung möglich ist. Tabelle 2.2 zeigt eine Übersicht über die 40 Gelenkwinkel, die bei der Rekonstruktion einer Bewegung berücksichtigt werden und in das Merkmal *joint_pos* einfließen, sowie ihre Zuordnung zu verschiedenen Körperteilen des MMM-Modells. Das Merkmal *joint_vel* repräsentiert den Geschwindigkeitsvektor im Gelenkwinkelraum, der wie zuvor als zentraler Differenzenquotient mit Gleichung 2.2 berechnet wird. Analog zu *marker_vel_norm* wird *joint_vel_norm* als euklidische Norm des Geschwindigkeitsvektors und somit Anhaltspunkt für die Gesamtbewegung im Gelenkwinkelraum bestimmt.

Tabelle 2.2: Verwendete Gelenke des MMM-Referenzmodells (vgl. Abbildung 2.3) zur Repräsentation menschlicher Ganzkörperbewegung in den Merkmalen *joint_pos*, *joint_vel* und *joint_acc*.

Hals	Torso	Bein (l.)	Bein (r.)	Arm (l.)	Arm (r.)
BLNx	BPx	LHx	RHx	LSx	RSx
BLNy	BPy	LHy	RHy	LSy	RSy
BLNz	BPz	LHz	RHz	LSz	RSz
BUNx	BTx	LKx	RKx	LEx	REx
BUNy	BTy	LAx	RAx	LEz	REz
BUNz	BTz	LAy	RAy	LWx	RWx
		LAz	RAz	LWy	RWy

Ausgehend von den Gelenkwinkelgeschwindigkeiten werden ebenfalls die Beschleunigungen im Gelenkwinkelraum bestimmt und durch die Merkmale *joint_acc* und *joint_acc_norm* repräsentiert.

Zusammen mit den Gelenkwinkeltrajektorien wird bei der Bewegungsrekonstruktion ebenfalls die 6-dimensionale Root-Pose des Modells bestimmt, welche sich aus den kartesischen Koordinaten für den Wurzelknoten des Modells (vgl. Abbildung 2.3) und seiner Rotation zusammensetzt. Das Merkmal *root_pos* beschreibt den Verlauf der Root-Position, wovon mit Hilfe von Gleichung 2.2 wiederum Geschwindigkeitsvektoren als *root_vel* und Beschleunigungsvektoren als *root_acc* abgeleitet und davon die euklidischen Normen als *root_vel_norm* und *root_acc_norm* bestimmt werden. Die Root-Rotation wird in Form von Roll-Nick-Gier-Winkeln (*roll-pitch-yaw angles*) durch das Merkmal *root_rot* repräsentiert und ebenfalls als eindimensionales Merkmal in *root_rot_norm* bereitgestellt.

Ausgehend von den bislang beschriebenen Merkmalen zur Beschreibung von Root-Pose und Gelenkwinkeln, die durch die Bewegungsrekonstruktion unmittelbar zur Verfügung stehen, können unter Verwendung des MMM-Modells mit seinen assoziierten kinematischen und dynamischen Spezifikationen weitere Merkmale abgeleitet werden. Die dynamischen Parameter des Modells erlauben für eine bekannte Ganzkörperkonfiguration die Berechnung der Position des Ganzkörper-Massenschwerpunkts (*Center of Mass*, CoM), welche durch das Merkmal *CoM_pos* beschrieben wird und in der linken Hälfte von Abbildung 2.9 beispielhaft für eine Laufbewegung dargestellt wird. Im Unterschied zum Wurzelknoten des Modells ist der Ganzkörper-Massenschwerpunkt nicht fest mit einem Segment des Modells assoziiert, wodurch es bei verschiedenen Ganzkörperbewegungen zu einem stark abweichenden Verlauf dieser beiden Merkmale kommt, die für die Nullpose des Modells noch fast zur Deckung kommen. Als Beispiele hierfür seien eine Winkbewegung genannt, bei welcher sich die Root-Position im Gegensatz zum Ganzkörper-Massenschwerpunkt nicht bewegt, oder eine Verbeugungsbewegung, bei welcher sich die Root-Position in posteriore Richtung, der

Massenschwerpunkt sich aber stattdessen in anterior-distale Richtung bewegt. Wie zuvor bereits erläutert, werden unter Verwendung von Gleichung 2.2 und der euklidischen Norm für den Ganzkörper-Massenschwerpunkt die zusätzlichen Merkmale CoM_{vel} , CoM_{acc} , CoM_{vel_norm} und CoM_{acc_norm} abgeleitet.

Die Vorwärtskinematik des MMM-Modells erlaubt die Berechnung der kartesischen Positionen für beliebige Segmente des menschlichen Körpers, wovon im Kontext menschlicher Ganzkörperbewegung die Positionen der Hände und Füße als Endeffektoren der menschlichen Kinematik besonders interessant sind, da diese vorrangig zur Interaktion mit der Umgebung verwendet werden. Das Merkmal $end_effectors_pos$ trägt dem Rechnung und repräsentiert die jeweils 3-dimensionalen Positionen der Handinnenflächen und der Hinterfüße. Analog zu den vorhergehenden Merkmalen werden die ersten und zweiten Ableitungen in den Merkmalen $end_effectors_vel$, $end_effectors_acc$, $end_effectors_vel_norm$ und $end_effectors_acc_norm$ bestimmt.

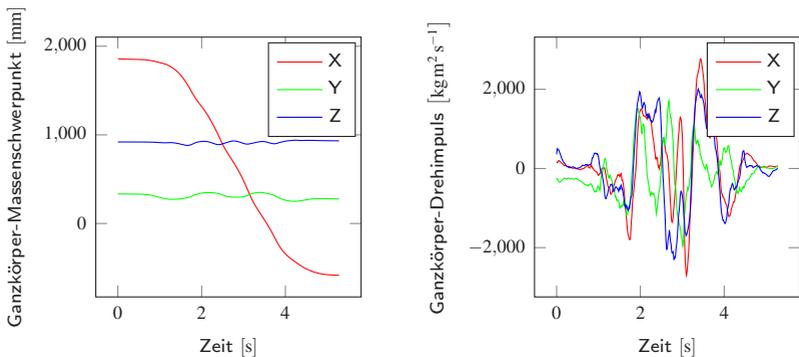


Abbildung 2.9: Verlauf des (nicht normalisierten) Ganzkörper-Massenschwerpunkts (Merkmal CoM_{pos}) und des Ganzkörper-Drehimpulses (Merkmal $angular_momentum$) für eine exemplarische Laufbewegung mit vier Schritten.

Abschließend erlauben die dynamischen Spezifikation des MMM-Modells auch die Berechnung des Ganzkörper-Drehimpulses zu jedem Zeitpunkt einer Bewegung. Der Drehimpuls weist für zahlreiche Bewegungstypen wie Laufbewegungen einen charakteristischen Verlauf auf, welcher bereits in verschiedenen Arbeiten der Biomechanik untersucht wurde (Herr et al. 2008) und auch Anwendung in der Robotik zur Regelung zweibeiniger Systeme gefunden hat (Hofmann et al. 2009; Popovic et al. 2004). Dies macht den Drehimpuls als Größe potentiell sehr interessant zur Beschreibung und Klassifikation derartiger Bewegungen. Ausgehend von den $N = 55$ Segmenten des MMM-Modells berechnet sich der Ganzkörper-Drehimpuls \mathbf{L} als

$$\mathbf{L} = \sum_{i=1}^N \left[m_i (\mathbf{r}_i^c \times \mathbf{v}_i^c) + \mathbf{I}_i^c \boldsymbol{\omega}_i \right], \quad \mathbf{L} \in \mathbb{R}^3 \quad (2.3)$$

mit

$$\begin{aligned} \mathbf{r}_i^c &= \mathbf{r}_{CoM_i} - \mathbf{r}_{CoM}, \\ \mathbf{v}_i^c &= \dot{\mathbf{r}}_{CoM_i} - \dot{\mathbf{r}}_{CoM}. \end{aligned}$$

Der erste Summand in Gleichung 2.3 beschreibt den Drehimpuls des Modellsegments i um den Massenschwerpunkt, wobei m_i für die Masse des Segments i , \mathbf{r}_i^c für seine Position und \mathbf{v}_i^c für seinen Geschwindigkeitsvektor in Bezug auf den Massenschwerpunkt stehen, welche mit dem Kreuzprodukt verknüpft werden. Der zweite Summand beschreibt den Eigendrehimpuls des Modellsegments i , der sich als Produkt seines Trägheitstensors \mathbf{I}_i^c und seiner Winkelgeschwindigkeit $\boldsymbol{\omega}_i$ berechnen lässt. Der Verlauf des Ganzkörper-Drehimpulses ist für eine exemplarische Laufbewegung in der rechten Hälfte von Abbildung 2.9 gezeigt und wird durch das Merkmal *angular_momentum* repräsentiert, wobei zur Berechnung von \mathbf{L} hier das lokale Koordinatensystem des Modells zugrunde gelegt wird. Das Merkmal *angular_momentum_norm* repräsentiert wiederum die Norm des Drehimpulses

und liefert somit einen Anhaltspunkt zur Erkennung von Drehungen bei dynamischen Bewegungsverläufen.

2.3.2 Verarbeitungsschritte

Im vorangegangenen Abschnitt 2.3.1 wurde eine Definition von 29 Merkmalen vorgenommen, die zur Beschreibung menschlicher Ganzkörperbewegung zum Einsatz kommen können. In diesem Abschnitt sollen nun drei wesentliche (Vor-)Verarbeitungsschritte erläutert werden, welche eine wichtige Rolle bei der praktischen Berechnung dieser Merkmale spielen.

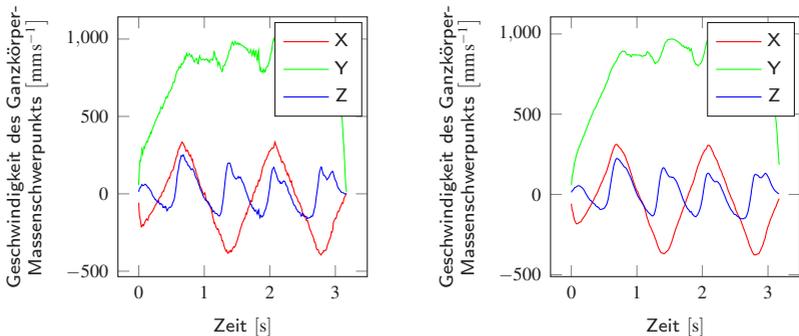


Abbildung 2.10: Vergleich der Verläufe des ungeglätteten (links) und geglätteten (rechts) Geschwindigkeitsvektors des Ganzkörper-Massenschwerpunkts (Merkmal CoM_vel) für eine exemplarische Laufbewegung mit vier Schritten.

1. Als erster Verarbeitungsschritt werden die vorliegenden Rohdaten des Bewegungserfassungssystems und die Ergebnisse der Bewegungsrekonstruktion einer Glättung unterzogen, um Rauschen oder einzelne Fehlmessungen zu beseitigen. Insbesondere für die Bestimmung von abgeleiteten Merkmalen (Geschwindigkeiten und Beschleunigungen) ist dies von großer Bedeutung, da durch die numerische Annäherung der Ableitung in Gleichung 2.2 kleine derartige Fehler verstärkt werden. Zur Glättung kommt ein Tiefpassfilter in Form des gleiten-

den Mittelwerts mit einer Fenstergröße von drei Frames zum Einsatz. Abbildung 2.10 zeigt beispielhaft die Anwendung der Glättung auf den Geschwindigkeitsvektors des Ganzkörper-Massenschwerpunkts (Merkmal *CoM_vel*).

2. Bei Merkmalen, welche sich auf Positionen in einem kartesischen Koordinatensystem beziehen, muss dieses Koordinatensystem festgelegt werden. So werden beispielsweise die von einem Bewegungserfassungssystem gelieferten Markertrajektorien in einem vom Aufnahmeszenario abhängigen globalen Koordinatensystem⁷ dargestellt, welches allerdings aufgrund der Abhängigkeit der Trajektorien von der Startposition und -orientierung des Menschen zu einer einheitlichen Repräsentation von Bewegungen ungeeignet ist. Aus diesem Grund werden die Positionsmerkmale *marker_pos*, *root_pos* und *CoM_pos* bezogen auf die initiale Root-Pose des Menschen zu Beginn der Bewegung dargestellt. Dabei wird das in 4-dimensionalen homogenen Koordinaten gegebene kartesische Merkmal \boldsymbol{x} als

$$\hat{\boldsymbol{x}}_t = \boldsymbol{T}_0^{-1} \boldsymbol{x}_t$$

normalisiert bestimmt, wobei \boldsymbol{T}_0 die 4×4 -Transformationsmatrix für die Root-Pose des MMM-Modells zu Beginn der Bewegung beschreibt. Analog hierzu werden die vier in *end_effectors_pos* enthaltenen Positionen der Endeffektoren durch Bezug auf die jeweils aktuelle Root-Pose des Modells normalisiert.

3. Abschließend wird für die berechneten Werte eines Merkmals eine merkmalspezifische Skalierung vorgenommen. Für verschiedene

⁷ In allen Fällen wird das Bewegungserfassungssystem jedoch derart kalibriert, dass sich der ebenerdige Boden bei $y = 0$ befindet. Hierdurch können Kontakte des Menschen mit dem Boden beispielsweise in Abschnitt 5.3 einfacher ermittelt werden, ohne dass zusätzliche Marker auf dem Boden angebracht werden müssten.

Anwendungen wie auch die in Kapitel 3 vorgestellte Klassifikation menschlicher Ganzkörperbewegung mit HMMs ist es erforderlich, dass sich die Wertebereiche der unterschiedlichen Merkmale ähneln, da Abstände zwischen Merkmalsvektoren mit Hilfe der euklidischen Norm berechnet werden und somit ähnliche Abmessungen des Merkmalsraums entlang jeder Dimension angenommen werden. Betrachtet man verschiedene der definierten Merkmale, beispielsweise Gelenkwinkel in Bogenmaß bzw. Positionen in Millimetern, so ist offensichtlich, dass die definierten Merkmale diese Anforderung nicht erfüllen können. Daher wird für den skalaren Wert x der Dimension eines Merkmals ein skaliertes Wert \hat{x} bestimmt als

$$\hat{x}_t = 2 \cdot \frac{x_t - x_{\min}}{x_{\max} - x_{\min}} - 1.$$

x_{\min} und x_{\max} sind für jede Dimension jedes Merkmals unterschiedlich, aber konstant und werden einmalig auf einer hinreichend großen und diversen Menge von Trainingsdaten bestimmt, womit \hat{x} später auch für ungesehene Bewegungsdaten näherungsweise nur Werte aus $[-1, 1]$ annimmt.

2.4 Übertragung menschlicher Bewegung auf humanoide Roboter

Die Übertragung menschlicher Bewegung auf andere kinematische Umgebungen wie beispielsweise humanoide Roboter ist ein aktives Forschungsfeld, zu dem in der Robotik und weiteren anderen Disziplinen bereits eine Vielzahl an Beiträgen veröffentlicht wurden. Oft kommen zur Lösung dieses Problems Optimierungsansätze zum Einsatz, wie sie beispielsweise von Do et al. (2008), Minato et al. (2007), Safonova et al. (2003) und Kim et al. (2006) beschrieben sind. In (Ude et al. 2004) wird ein solcher Optimierungsansatz vorgestellt, der zur Modellierung der Gelenkwinkeltrajektorien B-Spline-

Wavelets verwendet. Dahingegen wird in (Shon et al. 2005) die Übertragung der Bewegung mit Hilfe von Gauß-Prozessen als Regressionsproblem formuliert und in (Matsui et al. 2005) ein Ansatz präsentiert, welcher mit Hilfe eines neuronalen Netzes eine Bewegung des Roboters aus der des Menschen erzeugt. Dynamische Aspekte wie Stabilität können bei der Übertragung einer Bewegung entweder mit Hilfe einer Regelung innerhalb des Verfahrens berücksichtigt werden (Zheng et al. 2013; Yamane et al. 2009, 2003a; Naksuk et al. 2005) oder als nachträglicher Filter-Schritt zur dynamischen Stabilisierung einer vorgegebenen Referenztrajektorie realisiert werden (Boutin et al. 2010; Yamane et al. 2003b). Zusätzlich zur Robotik finden sich in der Computergrafik weitere Ansätze zur Übertragung der menschlichen Bewegung auf animierte Charaktere (dort als *Motion Retargeting* bezeichnet). Von Monzani et al. (2000) wird nicht ganz unähnlich zum MMM-Ansatz die Verwendung eines Intermediärmodells zur Repräsentation der menschlichen Bewegung vorgeschlagen. In weiteren Arbeiten werden Verfahren vorgestellt, welche die Form der Umgebung berücksichtigen. Shin et al. (2001) nimmt beispielsweise eine Priorisierung einzelner Merkmale vor, womit das Beibehalten von Kontakten sichergestellt werden kann, und von Al-Asqhar et al. (2013) wird ein Verfahren zur On-the-Fly-Anpassung von Bewegungen an die Umgebungsgeometrie vorgestellt, das auch in stark begrenzenden Umgebungen wie beispielsweise dem Fahrersitz eines Kraftfahrzeugs zum Einsatz kommen kann.

Im Folgenden soll nun der im Rahmen dieser Arbeit verfolgte Ansatz zur kinematischen Übertragung menschlicher Bewegungen auf humanoide Roboter erläutert werden, welcher analog zu dem in Abschnitt 2.2.2 beschriebenen Verfahren zur Bewegungsrekonstruktion arbeitet und somit ebenfalls als Problem der nicht-linearen Optimierung formuliert ist. In der Terminologie des MMM-Frameworks handelt es sich hierbei wie bei dem Ansatz zur Bewegungsrekonstruktion um einen Konverter, der hier jedoch als Ausgangspunkt keine Markermessungen eines Bewegungserfassungssystem, sondern Gelenkwinkel des MMM-Modells und als Zielkinematik einen humanoiden

Roboter verwendet (derartige Konverter sind in Abbildung 2.1 unten links dargestellt). Der hier vorgestellte Ansatz wird anschließend in Abschnitt 6.3 zur Übertragung menschlicher Referenzposen auf die Kinematik eines humanoiden Roboters verwendet, um dort durch die Definition sogenannter Konfigurationsposen zur Generierung von Multi-Kontakt-Bewegungen für humanoide Roboter beizutragen. Im Vergleich zu aktuellen Ansätzen in der Literatur wird hier daher ein einfacher Ansatz ohne Berücksichtigung der Bewegungsdynamik gewählt, da für die Übertragung der Konfigurationsposen anstelle einer kontinuierlichen Trajektorie lediglich diskrete Posen verarbeitet werden und somit an dieser Stelle keine dynamischen Merkmale berechnet werden können.

Für die Übertragung der Bewegung wird angenommen, dass der Roboter eine grob ähnliche Morphologie zum Menschen und somit eine humanoide Kinematik aufweist und dass diese Kinematik des Roboters als Modell vorliegt. Im ersten Schritt werden dann vergleichbar mit (Yamane et al. 2003a) virtuelle Marker auf dem Modell des Roboters analog zu den Vorschriften für menschliche Probanden positioniert. Ausgangspunkt für die Übertragung einer Bewegung ist immer die normalisierte Darstellung dieser im MMM-Format, womit für jeden Zeitschritt die Root-Pose und m Gelenkwinkel des MMM-Referenzmodells bekannt sind. Diese Annahme erlaubt eine Abstraktion der Bewegungsübertragung von den individuellen Eigenschaften einzelner menschlicher Probanden und vom Bewegungserfassungsprozess, womit beispielsweise ein zukünftiger Wechsel auf ein anderes Verfahren zur Bewegungserfassung leichter zu realisieren wäre. Für jede zu übertragende Pose werden mit Hilfe der bekannten Konfiguration des MMM-Modells und der bekannten Vorwärtskinematik virtuelle Markermessungen erzeugt, d.h. es werden die Positionen der in Abschnitt 2.2.2 erwähnten virtuellen Marker auf dem MMM-Modell für diese Pose berechnet. Weicht die Größe des Roboters von der des menschlichen Probanden ab, wird nachfolgend eine Skalierung der erzeugten Markerpositionen mit dem Größenverhältnis der beiden Kinematiken vorgenommen. Im Anschluss bildet Gleichung 2.1 die Grund-

lage zur Übertragung der Bewegung, wobei U nun die erzeugten virtuellen Markermessungen bezeichnet und $\mathbf{x} = (p_x, p_y, p_z, \alpha, \beta, \gamma, \theta_1, \dots, \theta_o)$ die Roboterpose, welche sich aus Root-Pose und den o Gelenkwinkeln des Roboters zusammensetzt⁸. $V(\mathbf{x})$ sind hier die durch die Vorwärtskinematik des Roboters berechneten Positionen der virtuellen Marker auf dem Robotermodell. Wie in Abschnitt 2.2.2 wird die in NLOpt bereitgestellte Neuimplementierung des Subplex-Algorithmus zur Lösung des Optimierungsproblems eingesetzt.

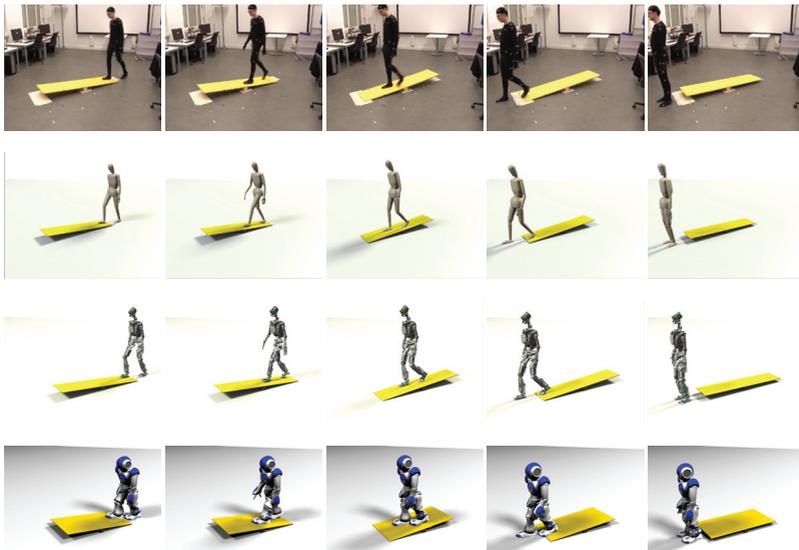


Abbildung 2.11: Rekonstruktion einer menschlichen Laufbewegung (erste Reihe) als MMM-Repräsentation (zweite Reihe) und Übertragung der Bewegung auf die humanoiden Roboter ARMAR-4 (dritte Reihe) und NAO (vierte Reihe) (entnommen aus Mandery et al. 2016c, ©2016 IEEE).

Eine Anwendung des Verfahrens ist in Abbildung 2.11 dargestellt, in der eine menschliche Laufbewegung zunächst mit dem in Abschnitt 2.2.2 eingeführten Verfahren rekonstruiert und anschließend mit dem in diesem Abschnitt

⁸ In der Regel verfügt der Roboter über weniger Bewegungsfreiheitsgrade als das MMM-Referenzmodell, d.h. $o < m$.

vorgestellten Verfahren auf die beiden humanoiden Roboter ARMAR-4 (Asfour et al. 2013) in der dritten und NAO (Nao o.D.) in der vierten Reihe übertragen wird.

2.5 Zusammenfassung von Kapitel 2

In diesem Kapitel wurden notwendige Grundlagen zur Erfassung, Repräsentation und Verarbeitung menschlicher Ganzkörperbewegungen eingeführt, welche eine wesentliche Voraussetzung für die nachfolgenden Teile dieser Arbeit darstellen. Hierzu wurde zunächst das am H²T entwickelte MMM-Framework diskutiert, das eine Abstraktion erfasster Bewegungen von individuellen Unterschieden menschlicher Probanden und somit eine vereinheitlichende Repräsentation menschlicher Ganzkörperbewegungen ermöglicht. Die Basis dieser Repräsentation stellt hierbei das MMM-Referenzmodell dar, welches kinematische und dynamische Spezifikationen für den menschlichen Körper mit einer Gesamtzahl von 104 Bewegungsfreiheitsgraden festlegt. Im Anschluss wurde ein Überblick über mögliche Ansätze für die Erfassung menschlicher Bewegungen gegeben, wobei der Fokus auf den im Kontext dieser Arbeit eingesetzten markerbasierten passiv-optischen Ansätzen lag. Zur Rekonstruktion von Gelenkwinkeltrajektorien der MMM-Repräsentation ausgehend von Messungen eines solchen markerbasierten Ansatzes wurden zwei unterschiedliche Verfahren vorgestellt, die auf einer Formulierung des Problems als nicht-lineares Optimierungsproblem bzw. auf dem Einsatz eines Kalman-Filters basieren, und es wurde demonstriert, wie mit Hilfe von nicht-linearer Optimierung darüber hinaus auch eine kinematische Übertragung menschlicher Bewegung auf humanoide Roboter realisiert werden kann. Schließlich wurden ausgehend von der markerbasierten Erfassung einer Bewegung und der rekonstruierten MMM-Bewegung 29 unterschiedliche, vorwiegend niedrigdimensionale Merkmale zur Beschreibung menschlicher Ganzkörperbewegung eingeführt, welche sich in Marker-, Gelenkwinkel-, Root-Pose-, Massenschwerpunkt-, Endeffektor- und Drehimpuls-Merkmale

unterteilen lassen, und es wurden die zur ihrer Berechnung notwendigen Verarbeitungsschritte erläutert. Die definierten Merkmale bilden hierbei die Grundlage der in Kapitel 3 behandelten Modellierung menschlicher Ganzkörperbewegung mit Hidden Markov Models und werden im Rahmen der dort vorgestellten Merkmalsauswahl weitergehend untersucht.

3 Modellierung und Klassifikation von Ganzkörperbewegungen mit HMMs

Die Sammlung größerer Mengen an Bewegungsdaten mit der in Kapitel 2 beschriebenen Methodik ist eng verbunden mit der Fragestellung, wie diese Bewegungsdaten modelliert und im nächsten Schritt automatisch klassifiziert werden können. Hierfür soll in diesem Kapitel ein Ansatz zur Modellierung und Klassifikation menschlicher Ganzkörperbewegung mit Hilfe von Hidden Markov Models (HMMs) vorgestellt werden, welcher im Rahmen einer Merkmalsauswahl für eine Dimensionalitätsreduktion der Bewegungsrepräsentation zum Einsatz kommt und in einem System zur Multi-Label-Bewegungsklassifikation mündet, das eine vorgegebene Ganzkörperbewegung automatisch einer beliebigen Teilmenge von verfügbaren Bewegungslabels zuordnen kann. Dieses Bewegungsklassifikationssystem wiederum legt eine wichtige Grundlage für die Organisation von Bewegungsdaten in der in Kapitel 4 vorgestellten Bewegungsdatenbank.

Das Kapitel ist wie folgt gegliedert. Abschnitt 3.1 liefert einen Überblick über in der Literatur existierende Arbeiten zur Modellierung und Klassifikation menschlicher Bewegungen mit einem Fokus auf Ganzkörperbewegungen. Im Anschluss wird in Abschnitt 3.2 der im Rahmen dieser Arbeit eingesetzte Ansatz zur Modellierung von Ganzkörperbewegung mit Hilfe von HMMs vorgestellt und formalisiert. Abschnitt 3.3 befasst sich mit einer Merkmalsauswahl basierend auf den in Abschnitt 2.3 definierten Merkmalen, welche auf die Definition von niedrigdimensionalen Merkmalsvektoren für

eine effiziente Bewegungsklassifikation abzielt, und schlägt hierzu einen Wrapper-Ansatz basierend auf einer im Rahmen dieser Arbeit entwickelten Metaheuristik vor. Abschnitt 3.4 beschreibt anschließend die Anwendung des vorgestellten HMM-basierten Ansatzes auf das angesprochene Problem der Multi-Label-Bewegungsklassifikation. Abschließend werden in Abschnitt 3.5 die Ergebnisse der Evaluation für die Merkmalsauswahl und die Multi-Label-Bewegungsklassifikation vorgestellt und diskutiert.

3.1 Verwandte Arbeiten

In der Literatur finden sich zahlreiche Arbeiten, welche sich mit dem Problem der Modellierung menschlicher Bewegung und ihrer maschinellen Klassifikation befassen. Der folgende Abschnitt bietet einen Überblick über diese bereits vorhandenen Ansätze, die in der Robotik von Mensch-Roboter-Interaktion und Imitationslernen bis hin zu Rehabilitationsrobotik und der menschlichen Bewegungsanalyse reichen. Anwendungen im Bereich des maschinellen Sehens werden hierbei im Anschluss separat behandelt, da aufgrund der Verfügbarkeit von 2D-Bildern als einzigen Eingabedaten die Anforderungen und Voraussetzungen in diesem Bereich deutlich von Anwendungen in der Robotik abweichen.

Es existieren zahlreiche Ansätze, bei welchen die auch in dieser Arbeit verwendeten HMMs zur Modellierung der menschlichen Ganzkörperbewegung zum Einsatz kommen. HMMs bilden die Grundlage für den in (Kulić et al. 2012, 2008, 2007a) und weiteren Veröffentlichungen derselben Autoren vorgestellten Ansatz, bei dem aus der Segmentierung menschlicher Ganzkörperbewegungen gelernte Bewegungsprimitive durch HMMs modelliert und im nächsten Schritt zu neuen Ganzkörperbewegungen verkettet werden. In (Inamura et al. 2004) werden HMMs für die Modellierung von Bewegungs-Protosymbolen eingesetzt, die zur Bewegungsgenerierung für einen humanoiden Roboter dienen. Takano et al. (2015, 2006) verwenden HMMs auf ähnliche Weise zur Modellierung von Bewegungs-Protosymbolen,

die dann im nächsten Schritt statistisch mit Wörtern der natürlichen Sprache verknüpft werden, wodurch eine statistische wechselseitige Inferenz zwischen Bewegungen und dazugehörigen textuellen Beschreibungen ermöglicht wird. In (Asfour et al. 2008) wird für Imitationslernen aus mehreren Beobachtungen einer demonstrierten Armbewegung ein HMM gelernt, mit dessen Hilfe dann wiederum eine generalisierte Trajektorie dieser Bewegung erzeugt werden kann.

Zusätzlich zu regulären HMMs haben auch mehrere Varianten bzw. Erweiterungen jener Einsatz zur Bewegungsmodellierung gefunden. Kulić et al. (2007b) untersuchen die Modellierung menschlicher Ganzkörperbewegungen mit Hilfe von Factorial Hidden Markov Models (FHMMs), bei denen im Unterschied zu HMMs die beobachteten Emissionen von mehreren unabhängigen statt nur von einer Markov-Kette beeinflusst werden (Ghahramani et al. 1997). In (Kulić et al. 2008) wird ein effizienterer Trainingsalgorithmus für diese FHMMs vorgestellt und mit Bewegungsdaten evaluiert. Eine weitere Modifikation von HMMs, die ebenfalls zur Bewegungsmodellierung eingesetzt wird, stellen Parametric Hidden Markov Models (PHMMs) dar, bei denen ein zusätzlicher Parametervektor das Bewegungsmodell erweitert und so eine Parametrisierung des modellierten Bewegungsablaufs, bspw. die Veränderung der Schrittlänge bei einer Laufbewegung, ermöglicht. Krüger et al. (2010) zeigen eine Anwendung von PHMMs zur Repräsentation und Erkennung von Aktionsprimitiven, die zu komplexen Aktionen verkettet werden können. Wilson et al. (1999) untersuchen eine Gestenerkennung auf Basis von PHMMs und vergleichen sie mit einem HMM-basierten Ansatz, und in (Herzog et al. 2008) werden PHMMs zur Erkennung und Imitation von Zeige- und Greifbewegungen eingesetzt.

Neben HMMs und ihren Varianten wie FHMMs und PHMMs werden zahlreiche weitere Techniken des maschinellen Lernens zur Modellierung und Klassifikation menschlicher Bewegungen eingesetzt. Fest etabliert ist die Modellierung mittels Dynamic Movement Primitives (DMPs) (Ijspeert et al. 2002, 2013; Jenkins et al. 2007; Nakanishi et al. 2004; Ernesti et al. 2012),

bei denen die Charakteristik der Bewegung durch ein nicht-lineares System, im Falle periodischer Bewegungen einen Oszillator, beschrieben wird. In weiteren Arbeiten kommen Baumstrukturen zum Einsatz, beispielsweise in einem durch hierarchisches Clustering der Trainingsdaten erzeugten Binärbaum (Yamane et al. 2011) oder in einem Präfixbaum (Kapadia et al. 2013), oder eine Bewegungsrepräsentation basierend auf einem Graphen (Kulić et al. 2012; Breazeal et al. 2005). In (Falco et al. 2017) wird ein Merkmal basierend auf paarweisen Korrelationen zwischen den Gelenken, die den höchsten Informationswert aufweisen, mit einem darauf definierten Ähnlichkeitsmaß eingesetzt. Baca (2012) beschreibt die Erkennung von Bewegungen mit Hilfe künstlicher neuronaler Netze und Ogata et al. (2005) trainieren rekurrente neuronale Netze für Bewegungen einer Aufgabenstellung der Mensch-Roboter-Interaktion. Auch in (Y. Du et al. 2015) werden rekurrente neuronale Netze eingesetzt, indem mit einer hierarchischen Anordnung solcher Netze die entsprechend den Körperteilen aufgeteilten Gelenkpositionen schrittweise verarbeitet und fusioniert werden. Taylor et al. (2009, 2006) stellen eine Modellierung menschlicher Bewegungen mit Hilfe von *Conditional Restricted Boltzmann Machines* vor, welche sich zur Erstellung von *Deep Belief Networks* und zur Synthese neuer Bewegungen eignen. Abschließend wird in (Calinon et al. 2007) ein Framework für *Programmieren durch Vormachen* vorgestellt, das auf Gauß-/Bernoulli-Mischmodellen basiert.

Wie bereits erwähnt, existieren insbesondere im Bereich des maschinellen Sehens zahlreiche Arbeiten zur Erkennung menschlicher Bewegungen bzw. Aktivitäten, etwa für Videoüberwachung, Mensch-Maschine-Interaktion, Rehabilitationsanwendungen oder für die Organisation und Abfrage von Videoaufzeichnungen. Aggarwal et al. (2011) und Poppe (2010) liefern einen umfassenden Überblick über Arbeiten in diesem Anwendungsgebiet. In vielen Fällen kommen hier bereits beschriebene Techniken wie etwa HMMs (Yamato et al. 1992) zum Einsatz, allerdings unterscheiden sich diese Ansätze durch die gänzlich anderen Ausgangsdaten in Form von 2D-Bildern,

der daraus folgenden abweichenden Verarbeitung und der Definition anderer Merkmale dennoch stark von dem in dieser Arbeit vorgestellten Ansatz.

Allen Ansätzen zur Bewegungsmodellierung und -klassifikation ist gemeinsam, dass eine Repräsentation für die Eingabedaten in Form der erfassten menschlichen Bewegung gewählt werden muss. Je nach verwendetem Ansatz können hierfür hochdimensionale Merkmalsvektoren, beispielsweise basierend auf Markertrajektorien, zum Einsatz kommen, während für andere Verfahren wie für die weit verbreiteten und auch im Rahmen dieser Arbeit zum Einsatz kommenden HMMs auf eine stärkere Beschränkung der Dimensionalität des Merkmalsraums geachtet werden muss. Zur Durchführung einer derartigen Dimensionalitätsreduktion können Techniken der Merkmalsextraktion und Merkmalsauswahl zum Einsatz kommen. Bei der Merkmalsextraktion, etwa auf Basis einer Hauptkomponentenanalyse (*principal component analysis*, PCA), werden synthetische Merkmale aus den Eingabedaten abgeleitet, die beispielsweise die Varianz in einem niedrigdimensionalen Raum maximieren. Dieser Ansatz kommt in einigen Arbeiten zur Verarbeitung menschlicher Bewegung zum Einsatz (Yamane et al. 2011; Calinon et al. 2007; Lin et al. 2014), erlaubt jedoch keine Identifikation nützlicher Größen in den hochdimensionalen Eingabedaten und erfordert demzufolge auch nach der Merkmalsextraktion stets die Erfassung des ursprünglichen hochdimensionalen Merkmalsvektors. Interessant sind daher Ansätze zur Merkmalsauswahl, bei denen in einer vorgegebenen Merkmalsmenge möglichst deskriptive Untermengen gesucht werden. Freeman (2014) liefert einen Überblick und eine Bewertung unterschiedlicher Filtermaße (*filter measures*) zur Merkmalsauswahl und beschreibt zwei Anwendungen für menschliche Bewegungsdaten, wobei neben einer Merkmalsauswahl für die Erkennung von Handgesten basierend auf elektromyografischen Messungen auch die Anwendung auf Daten menschlicher Ganzkörperbewegungen unter Verwendung des in (Kulić et al. 2008) beschriebenen Datensatzes vorgestellt wird. Zhou et al. (2017) beschreiben eine Anwendung eines modifizierten Filter-Ansatzes zur Merkmalsauswahl auf Basis eines *Markov Blanket* für

ein System zur Bewegungserkennung, welches HMMs bzw. dynamische Bayessche Netze einsetzt. Hierbei werden insgesamt 213 Merkmalsdimensionen berücksichtigt, die aus mit einem Kinect-Sensor aufgenommenen Bewegungsdaten bestimmt werden. In (Ren et al. 2005) wird ein Verfahren aus der Computergrafik vorgestellt, das auf mit drei Kameras aufgenommenen Silhouetten arbeitet und zur Merkmalsauswahl für die verwendete Ensemble-Klassifikation eine Variante des AdaBoost-Algorithmus einsetzt. Lösch et al. (2008) beschreiben eine Merkmalsauswahl für die Erkennung menschlicher Bewegungen, bei der die Merkmale in einer Taxonomie organisiert werden und der Prozess der Merkmalsauswahl durch menschliche Unterstützung verbessert wird.

3.2 HMMs für menschliche Ganzkörperbewegung

HMMs bilden die Basis für die Modellierung menschlicher Ganzkörperbewegung, auf der im Rahmen dieser Arbeit die anschließend beschriebenen Ansätze zur Merkmalsauswahl und zur Multi-Label-Bewegungsklassifikation aufbauen. In diesem Abschnitt sollen die Grundlagen zum Einsatz von HMMs für ihre Anwendung auf Bewegungsdaten erläutert werden. Hierzu werden in Abschnitt 3.2.1 zunächst HMMs eingeführt und ein Überblick über die grundlegenden im Zusammenhang mit ihnen betrachteten Problemstellungen gegeben. In Abschnitt 3.2.2 wird dann das Training von HMMs für Ganzkörperbewegungen mit Hilfe des Baum-Welch-Algorithmus genauer beschrieben und es werden Fragestellungen der Parameterinitialisierung und Topologiewahl erörtert. Abschließend wird in Abschnitt 3.2.3 eine erste einfache Anwendung von HMMs für das Problem der Multi-Class-Bewegungsklassifikation gezeigt, welches sich mit der Zuordnung einer Bewegung zu genau einer von M Bewegungsklassen beschäftigt und zu dem später behandelten komplexeren Problem der Multi-Label-Bewegungsklassifikation hinleitet.

3.2.1 Grundlagen und Definition von HMMs

Bei HMMs handelt es sich um eine Klasse sehr populärer stochastischer Modelle, die in verschiedensten Einsatzgebieten von Problemen der automatischen Spracherkennung und der Computerlinguistik über Anwendungen im Bereich des maschinellen Sehens wie etwa Gestenerkennung bis hin zu Fragestellungen der Bioinformatik wie der Erkennung von Proteinsequenzen zum Einsatz kommen. Grundsätzlich eignen sich HMMs gut für die Modellierung sequentieller Daten, wie sie bei der Repräsentation einer menschlichen Bewegung als zeitliche Abfolge eines Merkmalsvektors vorliegen, der aus einer (noch festzulegenden) Teilmenge der in Abschnitt 2.3 beschriebenen Merkmale besteht. HMMs können dabei eine den sequentiellen Daten zugrunde liegende Struktur anhand von Trainingsdaten erlernen und im Anschluss durch die Berechnung ihrer Übereinstimmung mit zu erkennenden Daten für die Bewegungsklassifikation eingesetzt werden¹. Die Erläuterungen in diesem Abschnitt basieren im Wesentlichen auf in der Literatur bekannten Beschreibungen von HMMs in (Bishop 2006; Elliott et al. 1995; Rabiner 1989), auf die auch für weiterführende Informationen verwiesen sei.

Grundlage eines HMM ist eine diskrete Markow-Kette erster Ordnung, die aus einem Zustandsraum mit K unterschiedlichen Zuständen s_1, \dots, s_K besteht. Der Zustand q_t der Markow-Kette zum Zeitpunkt t geht zu diskreten Zeitpunkten in einen Folgezustand über. Bei einer Markow-Kette erster Ordnung wird hierbei angenommen, dass die bedingten Übergangs-

¹ Eine ebenfalls sehr populäre Anwendung besteht im Einsatz von HMMs als generative Modelle, mit denen basierend auf dem gelernten stochastischen Modell neue synthetische Daten erzeugt werden können. Aufgrund der zahlreichen Nebenbedingungen wie Gelenkwinkelgrenzen und Kollisionsfreiheit ist dieser Ansatz für Ganzkörperbewegungen im Bereich der Robotik aber nicht ohne Weiteres anwendbar und soll im Rahmen dieser Arbeit auch nicht weiter untersucht werden.

wahrscheinlichkeiten zum Zeitpunkt t nur vom aktuellen Zustand q_t der Markow-Kette, nicht aber von vorhergehenden Zuständen abhängen:

$$P(q_{t+1} = s_i \mid q_t = s_j, q_{t-1} = s_k, \dots) = P(q_{t+1} = s_i \mid q_t = s_j).$$

Dies wird auch als Gedächtnislosigkeit oder als sogenannte Markow-Eigenschaft bezeichnet.

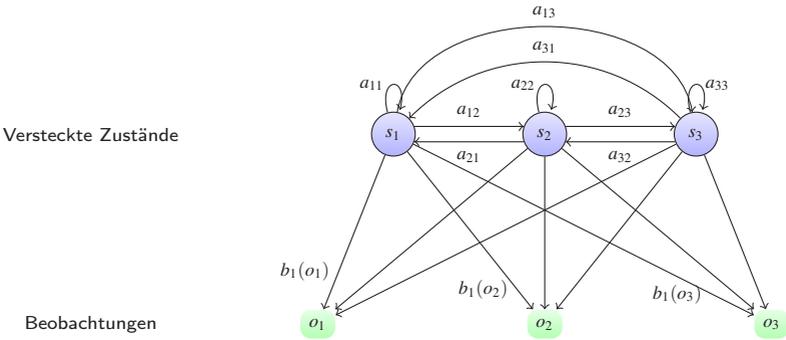


Abbildung 3.1: Darstellung eines beispielhaften HMM mit drei versteckten Zuständen s_1, s_2 und s_3 und durch drei Ausgabesymbole o_1, o_2 und o_3 repräsentierte diskrete Emissionen. a_{ij} bezeichnet die Wahrscheinlichkeit des Übergangs vom Zustand s_i zu s_j und b_i die Emissionsverteilung im Zustand i .

In einem HMM wird die beschriebene diskrete Markow-Kette erster Ordnung zur Modellierung des namensgebenden versteckten (*hidden*) Zustands verwendet. Zusätzlich zu den versteckten Zuständen werden Emissionen, oft auch Ausgaben genannt, festgelegt, die entweder diskret in Form eines vorgegebenen Ausgabealphabets oder kontinuierlich modelliert sein können. Jeder versteckte Zustand definiert eine Wahrscheinlichkeitsverteilung über die Emissionen, wodurch versucht werden kann, aus den beobachtbaren Emissionen indirekt Rückschlüsse auf den nicht beobachtbaren versteckten Zustand des HMM zu ziehen. Im Rahmen dieser Arbeit werden – wie in vergleichbaren Anwendungsbereichen – lediglich stationäre HMMs betrachtet, bei denen die Zustandsübergangs- und Emissionswahrscheinlichkeiten des

Modells zeitunabhängig sind. Eine Veranschaulichung eines HMM mit drei versteckten Zuständen und diskreten Emissionen in Form dreier Ausgabesymbole ist in Abbildung 3.1 dargestellt.

Bei der Anwendung von HMMs für die Modellierung der menschlichen Ganzkörperbewegung bestehen die Emissionen aus einer (noch genauer festzulegenden) Kombination der in Abschnitt 2.3 definierten Merkmale als D -dimensionalem Merkmalsvektor zur Beschreibung eines Zeitschritts der Bewegung. Bei einem Blick auf die Definition dieser Merkmale wird deutlich, dass für die Modellierung des Merkmalsvektors von kontinuierlichen Emissionen auszugehen ist, sofern man keine – mit einem Informationsverlust behaftete – Quantisierung der Merkmale auf diskrete Ausgabesymbole vornehmen möchte. In der vorliegenden Arbeit wird daher eine Modellierung der Emissionswahrscheinlichkeiten auf Basis einer multivariaten Normalverteilung vorgenommen. Eine solche multivariate Normalverteilung ist definiert über einen Mittelwertsvektor $\boldsymbol{\mu} \in \mathbb{R}^D$ und eine Kovarianzmatrix $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$. Die Wahrscheinlichkeitsdichtefunktion der Verteilung ergibt sich nach (Gut 2009) dann als

$$f(\boldsymbol{x}) = \frac{\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)}{\sqrt{(2\pi)^D \det(\boldsymbol{\Sigma})}}.$$

Wie zuvor erwähnt, werden die Emissionswahrscheinlichkeiten in Form von $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$ abhängig vom Zustand des HMM definiert.

Formal lassen sich die im Rahmen dieser Arbeit verwendeten HMMs als 3-Tupel $\boldsymbol{\lambda} = (\boldsymbol{A}, \boldsymbol{\pi}, \phi)$ beschreiben mit:

- $\boldsymbol{A} = (a_{ij}) \in [0, 1]^{K \times K}$ beschreibt die Übergangswahrscheinlichkeiten zwischen den versteckten Zuständen des HMM². Der Eintrag a_{ij} repräsentiert hierbei die Übergangswahrscheinlichkeit in den Folgezustand s_j ausgehend vom momentanen Zustand s_i . Da jede Spalte in \boldsymbol{A} eine

² Die Anzahl K der versteckten Zustände ist implizit bereits über die Größe der quadratischen Matrix \boldsymbol{A} festgelegt.

Wahrscheinlichkeitsverteilung repräsentiert, ist \mathbf{A} zeilenstochastisch, d.h. es gilt

$$\forall i : \sum_{j=1}^K a_{ij} = 1. \quad (3.1)$$

- $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K) \in [0, 1]^K$ legt die Anfangsverteilung fest. π_i beschreibt hierbei die Wahrscheinlichkeit, dass s_i der Startzustand ist. Da $\boldsymbol{\pi}$ eine Wahrscheinlichkeitsverteilung repräsentiert, muss gelten:

$$\sum_{i=1}^K \pi_i = 1. \quad (3.2)$$

- $\boldsymbol{\phi} = (\phi_1, \dots, \phi_K)$ modelliert die Emissionsverteilungen für die kontinuierlichen Emissionen des HMM. $\phi_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ repräsentiert hierbei den Mittelwertsvektor und die Kovarianzmatrix für die angenommene multivariate Normalverteilung der Emissionen im Zustand i .

Im Zusammenhang mit HMMs werden drei klassische Problemstellungen beschrieben, welche in unterschiedlichen Anwendungsbereichen vielfältig auftreten: Das Evaluationsproblem, das Dekodierungsproblem und das Trainingsproblem.

Das Evaluationsproblem fragt nach der Wahrscheinlichkeit $P(\mathbf{V} \mid \lambda)$ einer gegebenen Beobachtungssequenz $\mathbf{V} = (v_1, \dots, v_T)$ unter dem ebenfalls gegebenen HMM λ . Anschaulich beschreibt $P(\mathbf{V} \mid \lambda)$, mit welcher Wahrscheinlichkeit das HMM die vorgegebene Beobachtungssequenz erzeugt, und kann interpretiert werden als Maß dafür, wie gut das Modell auf die Beobachtungssequenz passt. Dies bildet die Grundlage für die im Rahmen dieser Arbeit beschriebene Bewegungsklassifikation. Das Evaluationsproblem kann mit Hilfe des Forward-Algorithmus effizient³ gelöst werden. Hierzu werden

³ Die Laufzeit wächst mit $\mathcal{O}(K^2 \cdot T)$ linear zur Länge der Beobachtungssequenz.

iterativ für jeden Zustand $k = 1, \dots, K$ und Zeitschritt t die Vorwärtsvariablen $\alpha_t(k)$ berechnet:

$$\alpha_1(k) = \pi_k \cdot P(v_1 | s_k),$$

$$\alpha_{t+1}(k) = \left(\sum_{i=1}^K \alpha_t(i) \cdot a_{ik} \right) \cdot P(v_{t+1} | s_k) \quad \text{für } t = 1, \dots, T-1.$$

Die bedingte Emissionswahrscheinlichkeit $P(v_t | s_k)$ kann hierbei direkt aus der bekannten Emissionsverteilung ϕ_k im Zustand k berechnet werden. Schlussendlich ergibt sich die Wahrscheinlichkeit der Beobachtungssequenz unter dem HMM zu

$$P(\mathbf{V} | \lambda) = \sum_{i=1}^K \alpha_T(i). \quad (3.3)$$

Beim Dekodierungsproblem wird für eine gegebene Beobachtungssequenz $\mathbf{V} = (v_1, \dots, v_T)$ und ein gegebenes HMM λ nach der wahrscheinlichsten Sequenz von versteckten Zuständen $\mathbf{Z} = (z_1, \dots, z_T)$ mit $z_t \in \{s_1, \dots, s_K\}$ gesucht. Dieses Problem kann mit dem Viterbi-Algorithmus als Abwandlung des Forward-Algorithmus ebenfalls effizient gelöst werden, ist jedoch für den vorgestellten Ansatz zur Bewegungsklassifikation ohne weitere Relevanz.

Als drittes Problem wird das Trainings- oder auch Optimierungsproblem beschrieben, bei welchem für eine gegebene Beobachtungssequenz $\mathbf{V} = (v_1, \dots, v_T)$ die Parameter eines HMM λ , also die Übergangswahrscheinlichkeiten \mathbf{A} , die Anfangsverteilung π und die bedingten Emissionsverteilungen ϕ , derart bestimmt werden sollen, dass die Wahrscheinlichkeit $P(\mathbf{V} | \lambda)$ der Beobachtungssequenz unter dem HMM maximiert wird. Die Anzahl an versteckten Zuständen K wird hierbei durch die Größe von \mathbf{A} vorgegeben. Die Lösung dieses Trainingsproblems ist von entscheidender Bedeutung für das Lernen von Modellen für menschliche Ganzkörperbewegungen ausgehend von Bewegungsaufnahmen und wird deshalb im folgenden Abschnitt ausführlicher diskutiert.

3.2.2 Training von HMMs für Bewegungsdaten

Beim Training eines HMM sollen für Trainingsdaten, die sich aus mehreren gegebenen Beobachtungssequenzen zusammensetzen, die Parameter des HMM $\lambda = (\mathbf{A}, \boldsymbol{\pi}, \phi)$ derart bestimmt werden, dass die Wahrscheinlichkeit dieser Trainingsdaten unter dem HMM maximiert wird. Eine analytische Lösung dieses Problems ist abgesehen von Spezialfällen nicht möglich, allerdings kann mit Hilfe des Baum-Welch-Algorithmus ein lokales Maximum approximiert werden. Hierfür wechseln sich im Baum-Welch-Algorithmus als einer Instanz des Expectation-Maximization-Algorithmus (EM-Algorithmus) ausgehend von gegebenen Initialwerten für \mathbf{A} , $\boldsymbol{\pi}$ und ϕ iterativ zwei Schritte ab: Im Expectation-Schritt wird für alle Zeitschritte und Zustände der Wert der Funktion $\xi_t(i, j)$ als Wahrscheinlichkeit, zum Zeitpunkt t im Zustand s_i und zum Zeitpunkt $t + 1$ in s_j zu sein, mit Hilfe des Forward-Backward-Algorithmus bestimmt. Im Maximization-Schritt werden anschließend neue Parameter \mathbf{A}' , $\boldsymbol{\pi}'$ und ϕ' für λ geschätzt. Beide Schritte werden abwechselnd für eine vorgegebene Zahl an Iterationen oder bis zum Erreichen von Konvergenz wiederholt. Für eine ausführlichere Darstellung des Baum-Welch-Algorithmus einschließlich notwendiger Anpassungen für die Handhabung multipler Beobachtungssequenzen beim Training sei auf (Rabiner 1989) verwiesen. Numerische Instabilitäten treten beim Baum-Welch-Algorithmus insbesondere dann auf, wenn wie bei der Verarbeitung von Bewegungsdaten sehr lange Beobachtungssequenzen vorgegeben werden, bei denen die dementsprechend hohe Zahl an Faktoren bei der Bestimmung von bedingten Wahrscheinlichkeiten zum Auftreten extrem kleiner Zahlenwerte führt. Im Rahmen dieser Arbeit wird deshalb als einfache und weit verbreitete Möglichkeit zur Abschwächung dieses Problems intern eine logarithmierte Darstellung der bedingten Wahrscheinlichkeiten gewählt.

Da der Baum-Welch-Algorithmus nur gegen ein lokales Maximum konvergiert und das Finden des globalen Maximums nicht garantiert, kommt der Wahl der initialen Werte für \mathbf{A} , $\boldsymbol{\pi}$ und ϕ eine entscheidende Bedeutung

zu, um eine hinreichend gute Lösung in der Nähe des globalen Maximums finden zu können. \mathbf{A} und $\boldsymbol{\pi}$ können hierbei entweder gleichverteilt oder zufällig (unter Berücksichtigung von Gleichung 3.1 bzw. 3.2) initialisiert werden. Ebenso können die in ϕ für jeden Zustand i enthaltenen Mittelwertsvektoren $\boldsymbol{\mu}_i$ und Kovarianzmatrizen $\boldsymbol{\Sigma}_i$ der multivariaten Normalverteilungen zufällig oder durch unüberwachtes Clustering der Trainingsdaten initialisiert werden. Entscheidet man sich für Clustering, können hierfür mit dem Lloyd-Algorithmus, einer Instanz des K-Means-Algorithmus, K Cluster für die in den Trainingsdaten enthaltenen Beobachtungen extrahiert und im Anschluss $\boldsymbol{\mu}_i$ und $\boldsymbol{\Sigma}_i$ mit dem Mittelwertsvektor und der Kovarianzmatrix des jeweiligen i -ten Clusters initialisiert werden. Ergänzend zu einer passenden Wahl der Initialisierungsstrategie kann eine deutliche Verbesserung der numerischen Eigenschaften des Baum-Welch-Algorithmus durch die Beschränkung der Kovarianzmatrizen auf Diagonalmatrizen und die damit einhergehende drastische Reduktion der zu optimierenden Parameter erreicht werden, obwohl die dahinterstehende Annahme, dass die einzelnen Dimensionen einer Beobachtung unkorreliert sind, nach der Definition der Merkmale in Abschnitt 2.3 rein theoretisch in vielen Fällen wenig gerechtfertigt erscheint. Basierend auf vorhergehenden Untersuchungen am H²T (Mandery et al. 2016b; Plapert 2015) wurde im Rahmen dieser Arbeit als Initialisierungsstrategie die gleichverteilte Initialisierung von \mathbf{A} und $\boldsymbol{\pi}$, die Initialisierung von ϕ mittels Clustering und zusätzlich die angesprochene Beschränkung der in ϕ enthaltenen Kovarianzmatrizen auf Diagonalmatrizen gewählt.

Ein weiterer wesentlicher Aspekt beim Training von HMMs stellt die sogenannte Topologie dar, welche die möglichen Übergänge zwischen den K versteckten Zuständen des HMM beschreibt. Der Baum-Welch-Algorithmus garantiert, dass Einträge in der Übergangswahrscheinlichkeitsmatrix \mathbf{A} und der Anfangsverteilung $\boldsymbol{\pi}$ mit dem Wert Null diesen Wert auch während des Trainings beibehalten. Diese Eigenschaft kann genutzt werden, um einen Übergang zwischen zwei Zuständen bzw. das Starten in einem bestimmten Zustand bereits bei der Initialisierung auszuschließen und so durch die Nut-

zung von Domänenwissen das Training mit dem Baum-Welch-Algorithmus zu vereinfachen, indem weniger Parameter gelernt werden müssen. Im einfachsten Fall wird keine der Übergangswahrscheinlichkeiten mit Null initialisiert, wodurch eine in Abbildung 3.2a beispielhaft gezeigte vollständig verbundene (*fully connected*) Topologie resultiert. Diese Topologie gehört zur Klasse der ergodischen Topologien, bei denen ausgehend von jedem Zustand jeder andere Zustand in einer endlichen Anzahl an Schritten erreichbar ist. Eine zweite wichtige Topologie stellt die Links-Rechts- oder Bakis-Topologie dar, bei der eine implizite „Reihenfolge“ der versteckten Zustände angenommen wird und die bei der Verarbeitung von Bewegungs-, aber auch Sprachdaten weit verbreitet eingesetzt wird. Bei dieser Topologie wird s_1 als Startzustand festgelegt und es werden grundsätzlich nur reflexive Zustandsübergänge oder Übergänge zu einem nachfolgenden Zustand erlaubt. Die Übergangswahrscheinlichkeitsmatrix \mathbf{A} nimmt daher mit

$$\forall j < i : a_{i,j} = 0$$

die Form einer oberen Dreiecksmatrix an:

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,K} \\ 0 & a_{2,2} & \dots & a_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{K,K} \end{pmatrix}.$$

Als zusätzliche Einschränkung kann über einen Parameter $\Delta \in \mathbb{N}$ die Anzahl der maximal zu überspringenden Zustände bei einem Übergang festgelegt werden:

$$\forall j > i + \Delta : a_{i,j} = 0.$$

Abbildung 3.2b und Abbildung 3.2c zeigen schematisch eine solche Links-Rechts-Topologie für $\Delta = 1$ bzw. $\Delta = 2$.

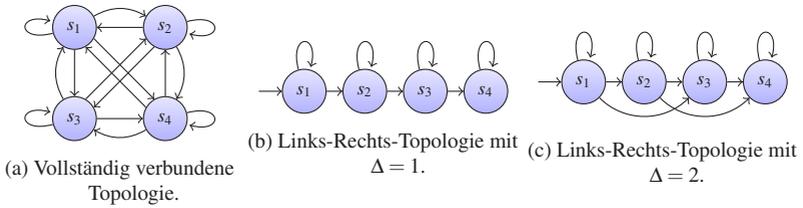


Abbildung 3.2: Schematische Darstellung verschiedener Topologien für ein HMM mit vier versteckten Zuständen.

3.2.3 Anwendung zur Multi-Class-Bewegungsklassifikation

Der folgende Abschnitt befasst sich mit der Anwendung der gelernten HMMs auf das Problem der Multi-Class-Bewegungsklassifikation, bei dem eine Bewegung genau einer von endlich vielen disjunkten Bewegungsklassen C_1, \dots, C_M zugeordnet werden soll. Diese Multi-Class-Bewegungsklassifikation stellt die Grundlage für den in Abschnitt 3.3 vorgestellten Ansatz zur Merkmalsauswahl dar und bildet anschließend zusammen mit dem bestimmten Merkmalsvektor die Basis zur Lösung des in Abschnitt 3.4 beschriebenen Problems der Multi-Label-Bewegungsklassifikation, bei dem eine Bewegung einer beliebigen Anzahl an Bewegungsklassen zugeordnet wird.

Unter der Annahme, dass für alle M Bewegungsklassen eine ausreichende⁴ Anzahl an Bewegungsaufnahmen als Trainingsdaten vorhanden sind, kann für jede dieser Bewegungsklassen unabhängig voneinander ein HMM λ_i gelernt werden. Für eine zu klassifizierende Bewegung \mathbf{V} werden dann die bedingten Wahrscheinlichkeiten $P(\mathbf{V} \mid \lambda_i)$ unter allen HMMs mit dem

⁴ Die Menge der erforderlichen Trainingsdaten hängt insbesondere von der Komplexität der Bewegung und der Varianz innerhalb der Bewegungsklasse ab. So werden für simple Bewegungen, die bereits anhand der beteiligten Körperteile oder Gelenke zuverlässig erkannt werden können, erfahrungsgemäß nur wenige Trainingsbeispiele benötigt. Für komplexe heterogene Bewegungsklassen, die durch unterschiedliche Parametrisierungen zu stark abweichenden Bewegungsabläufen führen können (z.B. Wischbewegungen), werden dementsprechend deutlich mehr Trainingsbeispiele benötigt.

Forward-Algorithmus (siehe Gleichung 3.3) wie zuvor erwähnt als Log-likelihoods (logarithmierte Likelihoods) berechnet, die im Folgenden der Kürze halber nur noch als Likelihoods bezeichnet werden.

Eine hohe Likelihood für λ_i bedeutet hierbei, dass λ_i gut auf die gegebene Bewegung \mathbf{V} passt und legt dementsprechend eine Zuordnung dieser Bewegung zur Bewegungsklasse C_i nahe. Daher kann für das Problem der Multi-Class-Bewegungsklassifikation die Bewegung \mathbf{V} wie in Abbildung 3.3 veranschaulicht der Klasse mit der höchsten Likelihood, also $C_{z(\mathbf{V})}$ mit dem Klassifikationsergebnis

$$z(\mathbf{V}) = \operatorname{argmax}_{i=1,\dots,M} P(\mathbf{V} | \lambda_i), \tag{3.4}$$

zugeordnet werden.

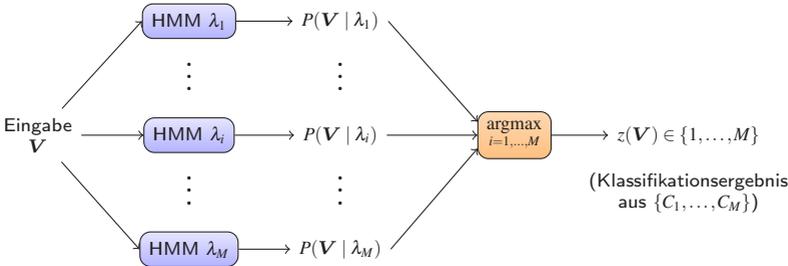


Abbildung 3.3: Multi-Class-Bewegungsklassifikation: Für eine zu klassifizierende Bewegung \mathbf{V} werden die Likelihoods unter den HMMs der M Bewegungsklassen bestimmt und die Klasse mit der höchsten Likelihood als Klassifikationsergebnis $z(\mathbf{V})$ ausgegeben.

Zur Beurteilung der Klassifikationsleistung bei der Multi-Class-Bewegungsklassifikation kommen im Rahmen dieser Arbeit mit der Korrektklassifikationsrate (*accuracy*, KKR) und dem F_1 -Maß (F_1 Score) zwei sehr gebräuchliche Gütemaße zum Einsatz. Für die Bestimmung dieser Gütemaße wird zunächst jede Bewegungsklasse isoliert betrachtet und für einen binären Klassifikator, der für eine beobachtete Bewegung die Zugehörigkeit dieser

Bewegung zu jener Bewegungsklasse vorhersagt, die Anzahl der richtig-positiv (r_p) und richtig-negativ (r_n) klassifizierten Testdaten, bei denen das Klassifikationsergebnis mit einer vorgegebenen Referenzklassifikation übereinstimmt, sowie die Anzahl der falsch-positiv (f_p) und falsch-negativ (f_n) klassifizierten Daten ermittelt. Darauf aufbauend wird die KKR für eine Bewegungsklasse definiert als

$$\text{KKR} = \frac{r_p + r_n}{r_p + r_n + f_p + f_n}. \quad (3.5)$$

Obwohl es sich bei der KKR um ein sehr einfaches und einleuchtendes Gütemaß handelt, bestehen Probleme bei einer großen Diskrepanz in der Häufigkeit zwischen positiven und negativen Klassenzugehörigkeiten. Da eine Bewegung im Zuge der Multi-Class-Bewegungsklassifikation stets nur einer Klasse zugeordnet wird und somit die Klassifikatoren aller Bewegungsklassen, bis auf diese eine, ein negatives Ergebnis liefern sollen, tritt dieses Problem mit einer zunehmenden Anzahl an Bewegungsklassen immer stärker zutage. Das F_1 -Maß (Rijsbergen 1979), welches in derartigen Fällen oftmals herangezogen wird, weist diesen Nachteil nicht auf und wird basierend auf Genauigkeit (*precision*) und Trefferquote (*recall*) wie folgt definiert:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (3.6)$$

mit

$$\text{precision} = \frac{r_p}{r_p + f_p}, \quad \text{recall} = \frac{r_p}{r_p + f_n}.$$

Die KKR bzw. F_1 -Maße für die einzelnen Bewegungsklassen können anschließend für eine Beurteilung des Gesamtsystems zur Multi-Class-Bewegungsklassifikation in einem gewichteten Mittelwert kombiniert werden, wobei die Bewegungsklasse C_i hierbei proportional zum *Support* dieser Bewegungsklasse, d.h. zur Anzahl der im Datensatz vertretenen Bewegungen der Klasse C_i , gewichtet wird (siehe Abschnitt 3.3 für eine Anwendung). Da

sowohl die KKR als auch das F_1 -Maß (und auch Genauigkeit und Trefferquote) stets im Intervall $[0, 1]$ liegen, gilt dies auch für diesen gewichteten Mittelwert.

3.3 Merkmalsauswahl

Die Struktur des Merkmalsvektors in Form von v_1, \dots, v_T einer Beobachtungssequenz V wurde im Rahmen der zuvor beschriebenen Modellierung menschlicher Ganzkörperbewegung mit HMMs zunächst nicht diskutiert, obwohl diese Struktur mutmaßlich einen entscheidenden Faktor für die Leistungsfähigkeit der Modellierung und der darauf aufbauenden Bewegungsklassifikation darstellt. Im Folgenden soll dies nun nachgeholt werden, indem ein Verfahren zur systematischen Merkmalsauswahl als Wrapper-Ansatz mit einer im Rahmen dieser Arbeit entwickelten Metaheuristik vorgestellt wird, welches eine Identifikation von niedrigdimensionalen, hochdeskriptiven Merkmalsvektoren als Teilmengen der in Abschnitt 2.3 definierten Merkmale erlaubt. Die Inhalte dieses Abschnitts wurden, zusammen mit der Evaluation in Abschnitt 3.5.1, zuvor in (Mandery et al. 2016b) publiziert.

Bei einem Blick auf die in Abschnitt 2.3 definierten Merkmale ist unmittelbar ersichtlich, dass eine Kombination all dieser Merkmale keinen für die HMM-basierte Modellierung menschlicher Ganzkörperbewegung brauchbaren Merkmalsvektor hervorbringen kann. Nicht nur liegt diese Kombination mit einer Gesamtzahl von 702 Dimensionen weit über der Anzahl an Dimensionen, für die ein HMM mit dem vorgestellten Ansatz überhaupt numerisch stabil trainiert werden kann, sondern zusätzlich würden weitere Probleme bei dermaßen hochdimensionalen Merkmalsvektoren ihre praktische Anwendung ausschließen: So ist eine steigende Dimensionalität des Merkmalsvektors mit wachsenden Anforderungen an die Menge der Trainingsdaten bzw. einem erhöhten Risiko für Overfitting bei gleichbleibender Menge an Trainingsdaten assoziiert (Guyon et al. 2003) und resultiert nicht zuletzt auch in einer höheren für das HMM-Training benötigten Rechenzeit.

Es ist somit erforderlich, die Dimensionalität des Merkmalsvektors stärker zu begrenzen.

In existierenden Arbeiten (siehe Abschnitt 3.1) wird die Fragestellung der Merkmalsauswahl meist ausgeklammert und zum Training von HMMs lediglich ein einzelnes hinreichend niedrigdimensionales Merkmal, oftmals (einige) Gelenkwinkel oder die kartesischen Positionen von ausgewählten Markern oder Körpersegmenten, manchmal in Verbindung mit Techniken der Merkmalsextraktion wie PCA herangezogen. Weder findet hierbei eine Untersuchung statt, ob sich die verwendeten Merkmale tatsächlich gut für die Modellierung menschlicher Bewegungen eignen, noch werden alternative Merkmale wie beispielsweise der auf dynamischen Bewegungseigenschaften basierende Ganzkörper-Massenschwerpunkt oder -Drehimpuls berücksichtigt. Die hier vorgestellte Merkmalsauswahl ist für eine Untersuchung dieser beiden Aspekte geeignet und stellt ein Verfahren zur systematischen Konstruktion von niedrigdimensionalen, dennoch hochdeskriptiven Merkmalsvektoren für die HMM-basierte Modellierung menschlicher Ganzkörperbewegung dar. Da es sich bei der Modellierung menschlicher Bewegungsdaten um eine universelle Fragestellung handelt, welche weit über die Anwendungen zur Bewegungsklassifikation hinausgeht, sind die Ergebnisse dieser Merkmalsauswahl ebenfalls relevant für solche andere Anwendungen menschlicher Ganzkörperbewegung mit HMMs, beispielsweise für die Synthese von Bewegungen (Takano et al. 2015), und potentiell für die Modellierung menschlicher Ganzkörperbewegung bei anderen Lernverfahren.

Verfahren zur Merkmalsauswahl lassen sich nach Guyon et al. (2003) in drei Kategorien unterteilen: Eingebettete (*embedded*) Ansätze, Filter-Ansätze und Wrapper-Ansätze. Bei eingebetteten Ansätzen wird die Merkmalsauswahl als integraler Teil des Trainings eines Lernverfahrens vorgenommen, wodurch solche Ansätze eng mit diesem Lernverfahren zusammenhängen und die Wahl des einzusetzenden Lernverfahrens einschränken. Während für Lernverfahren wie Support Vector Machines (SVMs) oder Entscheidungsbäume derartige eingebettete Ansätze bekannt und im Einsatz sind, existieren

keine Ansätze, die für die Modellierung von Zeitreihendaten mit HMMs, wie sie in der vorliegenden Arbeit betrachtet wird, Verbreitung gefunden hätten.

Filter-Ansätze hingegen sind von einem konkreten Lernverfahren unabhängig, indem sie eine Bewertung der verfügbaren Merkmale ohne Notwendigkeit eines Trainingsprozesses mit Hilfe eines Filtermaßes durchführen. In der Literatur finden sich eine Vielzahl solcher Filtermaße, beispielsweise basierend auf Merkmalskorrelationen. Durch die Unabhängigkeit von einem konkreten Lernverfahren und die vergleichsweise einfache Berechenbarkeit der meisten Filtermaße sind Filter-Ansätze potentiell sehr performant, auch wenn größere Mengen an hochdimensionalen Trainingsdaten verarbeiten werden sollen. Die für den erfolgreichen Einsatz dieser Ansätze entscheidende Wahl eines geeigneten Filtermaßes stellt in der Praxis allerdings oft eine erhebliche Herausforderung dar, weil die Eignung eines Filtermaßes von dem später einzusetzenden Lernverfahren, aber auch stark von den verfügbaren Trainingsdaten abhängt (Freeman et al. 2015). Eine weitere Schwierigkeit beim Einsatz von Filter-Ansätzen besteht darin, dass durch die individuelle Bewertung einzelner Merkmale mit Hilfe eines Filtermaßes vorhandene Beziehungen und Korrelationen zwischen den Merkmalen nicht berücksichtigt werden. Insbesondere wenn Merkmale – wie die in Abschnitt 2.3 beschriebenen – teilweise mathematisch voneinander abhängen, kann dies zu einer Auswahl von redundanten Merkmalen und damit zu suboptimalen Ergebnissen führen, wodurch sich Filter-Ansätze für solche Einsatzzwecke eher für eine Vorverarbeitung zum Ausschluss definitiv uninteressanter bzw. irrelevanter Größen eignen.

Bei Wrapper-Ansätzen (Kohavi et al. 1997) als dritter Kategorie von Verfahren zur Merkmalsauswahl wird das Lernverfahren in die Merkmalsauswahl miteinbezogen, ähnlich wie bei eingebetteten Ansätzen. Im Unterschied zu diesen wird das Lernverfahren hierbei allerdings lediglich zur Evaluation der Klassifikationsleistung für gegebene Teilmengen der verfügbaren Merkmale eingesetzt – das Lernverfahren wird dadurch ohne Modifikationen *gewrappt* und als Black Box behandelt. Durch diese Vorgehensweise kann

somit grundsätzlich jedes Lernverfahren in Verbindung mit einem Wrapper-Ansatz zum Einsatz kommen, sofern die benötigten Ressourcen für die wiederholte, potentiell rechenintensive Ausführung des Lernverfahrens für jede zu evaluierende Merkmalsmenge gegeben sind. Ein weiterer Vorteil von Wrapper-Ansätzen besteht darin, dass durch die Evaluation von Merkmalsmengen anstelle einzelner Merkmale Abhängigkeiten zwischen Merkmalen berücksichtigt werden können und dadurch anders als bei Filter-Ansätzen eine Auswahl stark redundanter Merkmale verhindert werden kann.

Als Folge der vorangestellten Überlegungen wurde als Vorgehensweise für die Merkmalsauswahl im Rahmen dieser Arbeit ein Wrapper-Ansatz gewählt, welcher die in Abschnitt 3.2.3 beschriebene Anwendung der Multi-Class-Bewegungsklassifikation für die Evaluation der Klassifikationsleistung heranzieht. Der durch Betrachtung beliebiger Merkmalsmengen sehr große Suchraum sorgt dafür, dass der zur Exploration dieses Suchraums eingesetzten Metaheuristik eine entscheidende Bedeutung für das Gelingen und die Performanz der Merkmalsauswahl zukommt. Da die Anzahl möglicher Kombinationen von Merkmalen exponentiell mit der Zahl dieser Merkmale wächst, stellt eine erschöpfende Suche als *Brute-Force-Methode*, bei der alle denkbaren Merkmalsmengen evaluiert werden, grundsätzlich schon bei einer geringen Anzahl an Merkmalen keine Option dar. Existierende Arbeiten zur Merkmalsauswahl basieren daher oft auf einer Vorwärtsselektion (Guyon et al. 2003), bei der ausgehend von einer leeren Menge iterativ jeweils dasjenige Merkmal hinzugefügt wird, welches die Klassifikationsleistung des unterliegenden Lernverfahrens am stärksten erhöht⁵. Diese einfache Metaheuristik birgt jedoch zwei erhebliche Nachteile in sich: Zum einen wird die Dimensionalität der verfügbaren Merkmale nicht berücksichtigt, wodurch einem sehr hochdimensionalen Merkmal selbst bei marginalen Unterschieden in der Verbesserung der Klassifikationsleistung gegenüber einem niedrigdimensionalen Merkmal der Vorzug gegeben werden kann. Dadurch

⁵ Analog hierzu wird oft auch eine Rückwärtselimination eingesetzt, bei der ausgehend von der Menge aller Merkmale iterativ Merkmale entfernt werden.

wiederum können sehr hochdimensionale Merkmalsmengen entstehen, die sich, wie zuvor erwähnt, schlecht für eine Anwendung mit HMMs eignen. Zum anderen repräsentiert eine Vorwärtsselektion einen *Greedy-Algorithmus*, bei dem in jeder Iteration stets das lokal beste Merkmal ausgewählt und somit die optimale, global beste Merkmalsmenge mit hoher Wahrscheinlichkeit nicht gefunden wird. Im Rahmen dieser Arbeit wird daher eine neue Metaheuristik, *N-Best Feature Subset Exploration*, vorgeschlagen und zum Einsatz gebracht, die den iterativen Ansatz der Vorwärtsselektion aufgreift, dabei aber die Dimensionalität der Merkmale berücksichtigt und durch paralleles Verfolgen alternativer Merkmalsmengen die Nachteile einer strikten Greedy-Vorgehensweise abmildert.

Die zur Exploration verwendete Metaheuristik ist in Algorithmus 1 gezeigt und verfolgt wie erwähnt eine iterative Vorgehensweise. Hierbei werden im k -ten Durchlauf ausschließlich Merkmalsmengen der Dimensionalität k betrachtet und der Abbruch erfolgt nach dem Erreichen einer vorgegebenen maximalen Dimensionalität (als kumulative Dimensionalität aller in Abschnitt 3.3 beschriebenen Merkmale höchstens 702). Zu Beginn jeder Iteration werden mögliche Merkmalsmengen der jeweiligen Dimensionalität k erstellt, indem die aus vorhergehenden Iterationen gespeicherten niedrigdimensionaleren Merkmalsmengen durch Hinzunahme eines einzelnen, jeweils noch nicht enthaltenen Merkmals auf die Dimensionalität k erweitert werden (Zeilen 4-15). Im Anschluss wird die Klassifikationsleistung der gerade erzeugten Merkmalsmengen für das in Abschnitt 3.2.3 beschriebene Problem der Multi-Class-Bewegungsklassifikation ermittelt (Zeilen 17-25). Hierfür kommt für jede zu evaluierende Merkmalsmenge eine 3-fache stratifizierte Kreuzvalidierung (*3-fold stratified cross-validation*) zum Einsatz, bei der die verfügbaren Bewegungsdaten in drei gleich große Teilmengen, die sogenannten *Folds*, aufgeteilt werden (Zeile 20). In jeder von drei aufeinanderfolgenden Durchläufen fungieren jeweils zwei dieser Folds als Trainingssatz und der verbleibende Fold als Testsatz (Zeilen 21-22), sodass nach dem Abschluss der Kreuzvalidierung jeder Fold genau einmal als Testsatz verwendet wurde.

Algorithmus 1 N-Best Feature Subset Exploration (modifiziert nach Mandery et al. 2016b, ©2016 ISIF)

```

1: possibleFeatures ← {...} (siehe Abschnitt 2.3)
2: fssForDim[0] ← {∅}
3: for curDim ← 1 to maxDim do
4:   // Erstelle mögliche Merkmalsmengen der Dimensionalität curDim
5:   fssToEvaluate ← {}
6:   for pastDim ← 0 to curDim − 1 do
7:     for all oldFss ∈ fssForDim[pastDim] do
8:       for all feature ∈ possibleFeatures do
9:         if pastDim + dim(feature) = curDim and
           feature ∉ oldFss then
10:            newFss ← oldFss + {feature}
11:            Füge newFss zu fssToEvaluate hinzu
12:          end if
13:        end for
14:      end for
15:    end for
16:
17:    // Evaluiere die in fssToEvaluate enthaltenen Merkmalsmengen
18:    for all fss ∈ fssToEvaluate do
19:      for round ← 1 to 3 do
20:        Teile Daten stratifiziert in Trainings- und Test-Folds auf
21:        Lerne HMMs mit Bewegungsdaten der Trainings-Folds
           (Abschnitt 3.2.2)
22:        Klassifiziere Bewegungen des Test-Folds mit den HMMs
           (Abschnitt 3.2.3)
23:      end for
24:      Berechne das gewichtete Mittel der  $F_1$ -Maße aller HMMs aus
           den drei Durchläufen
25:    end for
26:
27:    // Behalte die n besten Merkmalsmengen der Dim. curDim
28:    fssForDim[curDim] ← n-best(fssToEvaluate)
29: end for

```

Die Stratifikation der Kreuzvalidierung (Kohavi 1995) stellt hierbei sicher, dass die Verteilung der Bewegungsklassen in jedem Fold repräsentativ für den gesamten Datensatz ist⁶. Nach Abschluss der Kreuzvalidierung wird wie in Abschnitt 3.2.3 beschrieben der gewichtete Mittelwert der F_1 -Maße (siehe Gleichung 3.6) berechnet (Zeile 24), welcher die Klassifikationsleistung der evaluierten Merkmalsmenge in einem skalaren Maß zusammenfasst. Dieses Maß wird am Ende einer Iteration verwendet, um die besten n der evaluierten Merkmalsmengen dieser Iteration zu bestimmen und zu speichern (Zeile 28), damit diese in nachfolgenden Iterationen wiederum zur Konstruktion höherdimensionaler Merkmalsmengen eingesetzt werden können. Gleichermaßen repräsentieren die gespeicherten Merkmalsmengen auch das Ergebnis der Exploration, womit für den in dieser Arbeit gewählten Wert von $n = 10$ folglich die jeweils zehn besten gefundenen Merkmalsmengen jeder Dimensionalität bis zur vorgegebenen maximalen Dimensionalität resultieren.

3.4 Multi-Label-Bewegungsklassifikation

Während für einzelne Teilbereiche menschlicher Bewegung die Festlegung disjunkter Bewegungsklassen zur Kategorisierung vorgegebener Bewegungsaufnahmen mit der in Abschnitt 3.2.3 beschriebenen Multi-Class-Bewegungsklassifikation durchaus realisierbar scheint, ist ein derartiges Vorhaben aufgrund dessen Vielfalt und Mannigfaltigkeit kaum auf das gesamte Repertoire menschlicher Ganzkörperbewegungen übertragbar. Um die Klassifikation eines möglichst breiten Spektrums solcher Ganzkörperbewegungen zu ermöglichen, sind daher Methoden der Multi-Label-Bewegungsklassifikation interessant, bei denen eine Bewegung nicht genau einer Bewegungsklasse, sondern einer beliebigen Zahl von gegebenen Labels L_1, \dots, L_M zugeordnet wird. Bei einer Wahl von geeigneten Labels, welche

⁶ Eine Missachtung dieser Anforderung könnte im ungünstigsten Fall ansonsten dazu führen, dass eine Bewegungsklasse überhaupt nicht mehr in den Trainingsfolds enthalten ist und somit im Anschluss nicht erkannt werden kann.

verschiedene Bewegungstypen oder Ausführungsmodalitäten beschreiben können, ermöglicht dieser Ansatz im Vergleich zu disjunkten Bewegungsklassen eine wesentlich flexiblere und mächtigere Form der Bewegungsbeschreibung und kommt beispielsweise für die in Kapitel 4 vorgestellten KIT-Bewegungsdatenbank mit einer hierarchischen Organisation von Labels in einem *Motion Description Tree* zum Einsatz.

Die zuvor in Abschnitt 3.2.3 vorgestellte, vergleichsweise einfache Vorgehensweise zur Multi-Class-Bewegungsklassifikation kann allerdings nicht direkt auf die Problemstellung der Multi-Label-Bewegungsklassifikation angewendet werden, da stets nur die eine Bewegungsklasse zurückgeliefert wird, deren HMM mit der höchsten Likelihood assoziiert ist (siehe Gleichung 3.4 und Abbildung 3.3). In diesem Abschnitt sollen daher Ansätze vorgestellt werden, die über eine Erweiterung der bereits vorgestellten Methoden eine Multi-Label-Klassifikation menschlicher Ganzkörperbewegung auf Basis von HMMs ermöglichen.

Das Forschungsfeld der Multi-Label-Klassifikation konnte ausgehend von frühen Arbeiten zur Dokumenten-/Textklassifikation während des letzten Jahrzehnts zunehmend an Bekanntheit gewinnen und Aufmerksamkeit aus anderen Disziplinen für Problemstellungen wie der Annotation von Videodaten, der automatischen Szenenklassifikation oder Kategorisierung von Musiktiteln bis hin zu Anwendungen der Bioinformatik wie der Proteinfunktionsvorhersage auf sich ziehen (M.-L. Zhang et al. 2014; Tsoumakas et al. 2007). Die existierende Literatur unterteilt Ansätze zur Multi-Label-Klassifikation hierbei in Problemtransformationsverfahren (*problem transformation methods*, PTV) und Algorithmenanpassungsverfahren (*algorithm adaption methods*, AAV) (M.-L. Zhang et al. 2014; Tsoumakas et al. 2007). Bei PTV wird das gegebene Problem der Multi-Label-Klassifikation auf ein oder mehrere etablierte Lernprobleme wie Multi-Class-Klassifikations- oder Regressionsprobleme zurückgeführt. AAV hingegen basieren auf der Erweiterung oder Anpassung eines bestehenden Lernverfahrens, sodass dieses im Anschluss direkt die Multi-Label-Daten handhaben kann. M.-L. Zhang

et al. (2014) beschreiben dies anschaulich so, dass bei PTV die Daten (durch eine Transformation) an einen bestehenden Algorithmus angepasst werden, während bei AAV umgekehrt der Algorithmus an die Daten angepasst wird. Diese beiden unterschiedlichen Herangehensweisen an das Problem der Multi-Label-Klassifikation legen die Grundlage für die drei im Folgenden vorgestellten Lösungsansätze hinsichtlich des Problems der HMM-basierten Multi-Label-Bewegungsklassifikation.

3.4.1 Problemtransformation mit Label-Potenzmengen-Methode

Die Label-Potenzmengen-Methode (*label powerset method*, LPM) stellt einen unkomplizierten und sehr einfach umzusetzenden Ansatz aus der Gruppe der PTV dar (M.-L. Zhang et al. 2014; Read et al. 2011). Auch wenn die nachfolgend diskutierten Nachteile einem Einsatz dieser Methode in der Praxis entgegenstehen, soll der Ansatz hier dennoch als einführendes Beispiel dienen. Grundlage der LPM ist die Rückführung des Multi-Label-Klassifikationsproblems auf ein Problem der Multi-Class-Klassifikation, indem jede in den Trainingsdaten vorkommende Label-Kombination als eigenständige disjunkte Klasse angesehen wird. Formal wird also die Potenzmenge der Labelmenge $\{L_1, \dots, L_M\}$ als Klassenmenge

$$C' = \mathcal{P}(\{L_1, \dots, L_M\}) = \{U \mid U \subseteq \{L_1, \dots, L_M\}\}$$

in einem Multi-Class-Klassifikationsproblem betrachtet. Analog zur Darstellung der Multi-Class-Bewegungsklassifikation in Abbildung 3.3 ist die Vorgehensweise hierfür in Abbildung 3.4 veranschaulicht.

Ein erheblicher Nachteil der LPM besteht darin, dass die Gesamtzahl der möglichen Kombinationen der Labels und damit der potentiell zu trainierenden und evaluierenden HMMs exponentiell mit der Zahl dieser Labels wächst. Auch wenn bei praktischen Anwendungen für Ganzkörperbewegungen im Rahmen dieser Arbeit nicht alle möglichen Labels sinnvoll kombiniert

werden können und daher in der Praxis kein exponentielles Wachstum auftritt, kann dennoch rasch eine Zahl an HMMs erforderlich werden, die den erforderlichen Rechenaufwand unverhältnismäßig erhöht. Noch schwerer wiegt allerdings die mit der Zunahme an HMMs verbundene Reduktion der Menge verfügbarer Trainingsdaten für jedes dieser HMMs, da bei der LPM jede Bewegungsaufnahme grundsätzlich nur zum Training eines einzelnen HMM verwendet werden kann. So werden beispielsweise mit {„Laufen“} gelabelte Bewegungen nicht als Trainingsbeispiele für das HMM der Label-Kombination {„Laufen“, „Kurve-Links“} genutzt. Durch die Tatsache, dass überhaupt nur bereits in den Trainingsdaten vorhandene Label-Kombinationen erkannt werden können, besteht außerdem eine inhärente Tendenz zum Overfitting (Read et al. 2011).

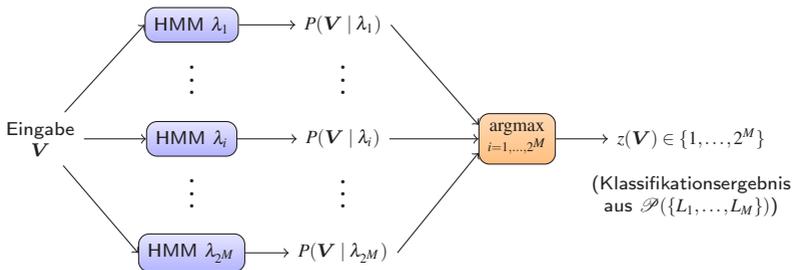


Abbildung 3.4: Multi-Label-Bewegungsklassifikation mit Label-Potenzmengen-Methode: Nur für die in den Trainingsdaten vorhandenen der möglichen 2^M Kombinationen der M Bewegungslabels werden HMMs trainiert und sie können erkannt werden.

3.4.2 Problemtransformation mit Binäre-Relevanz-Methode

Die Binäre-Relevanz-Methode (*binary relevance method*, BRM) stellt einen zweiten Ansatz aus der Gruppe der PTV dar, bei dem das Multi-Label-Klassifikationsproblem auf M unabhängige binäre Klassifikationsprobleme zurückgeführt wird, welche die Zugehörigkeit einer Bewegung zu jeweils genau einem der M Bewegungslabels beschreiben. Obwohl es sich bei der

BRM um einen auf den ersten Blick simplen Ansatz handelt, macht gerade die intuitive Handhabung, die Resistenz gegenüber Overfitting und ihr geringer Rechenaufwand diese Vorgehensweise und ihre Weiterentwicklungen interessant für verschiedene praktische Problemstellungen der Multi-Label-Klassifikation (Read et al. 2011).

Ausgangspunkt der BRM ist die Darstellung der Klassifikation einer Bewegung \mathbf{V} als ein binärer Vektor

$$\boldsymbol{\chi} = (\chi_1, \dots, \chi_M)^\top \in \{0, 1\}^M \quad \text{mit } \chi_i = \begin{cases} 1 & \text{falls } \mathbf{V} \text{ mit } L_i \text{ assoziiert ist,} \\ 0 & \text{sonst.} \end{cases}$$

Für die χ_i wird im Anschluss jeweils ein binärer Klassifikator K_i gelernt, welcher die Zugehörigkeit von \mathbf{V} zum Label L_i bestimmt. Jedem dieser binären Klassifikatoren stehen hierbei die unter den HMMs der Labels bestimmten Likelihoods einer gegebenen Bewegung zur Verfügung, wobei das HMM λ_i analog zur Vorgehensweise bei der Multi-Class-Bewegungsklassifikation (siehe Abschnitt 3.2.3) mit den Bewegungen trainiert wird, welche dem Label L_i (bzw. der Klasse C_i) zugeordnet sind.

Im einfachsten Fall wird bei der binären Klassifikation in K_i lediglich die Likelihood $P(\mathbf{V} \mid \lambda_i)$ mit einem Schwellwert θ_i verglichen, welcher entweder manuell vorgegeben oder aus den Trainingsdaten ermittelt wird. Abgesehen von der Schwierigkeit, geeignete Werte für alle θ_i bestimmen zu müssen, hat dieser Ansatz jedoch den für BRM oft beschriebenen Nachteil, dass Abhängigkeiten oder Korrelationen zwischen den Labels nicht berücksichtigt werden können⁷. Aus diesem Grund bietet es sich an, statt eines simplen Schwellwerts ein binäres Klassifikationsverfahren einzusetzen, welches als Eingabe direkt den Vektor (aus \mathbb{R}^M) der berechneten Likelihoods unter *allen* HMMs verarbeiten und somit Abhängigkeiten zwischen den Labels berück-

⁷ Derartige Abhängigkeit zeigen sich beispielsweise in Form von sich gegenseitig ausschließenden Bewegungslabels (z.B. „Langsam“ und „Schnell“) oder in Konstellationen, in denen ein Label ein weiteres Label nahelegt (z.B. „Treten“ impliziert „Fuß-Rechts“ oder „Fuß-Links“).

sichtigen kann. Diese Vorgehensweise ist in Abbildung 3.5 dargestellt. Für die Wahl des binären Klassifikationsverfahren bieten sich grundsätzlich alle etablierten und aus der Literatur bekannten Verfahren wie beispielsweise logistische Regression, SVMs, neuronale Netze, Bayessche Netze, Entscheidungsbäume usw. an.

Die zuvor für die LPM diskutierten Nachteile treten bei der hier vorgestellten BRM nicht auf: Da für jedes Label lediglich ein HMM trainiert wird, wächst die Zahl dieser und somit der Rechenaufwand für das HMM-Training wie im Multi-Class-Fall linear mit der Anzahl an Labels. Durch die reduzierte Zahl der HMMs und die Tatsache, dass eine Bewegung zum Training aller HMMs der assoziierten Labels verwendet werden kann, stehen wiederum entsprechend mehr Trainingsdaten für jedes einzelne HMM zu Verfügung. Schlussendlich erlaubt die resultierende Multi-Label-Klassifikation im Unterschied zur LPM auch die Erkennung und Ausgabe von bislang ungesesehenen Label-Kombinationen, welche nicht in den Trainingsdaten enthalten sind.

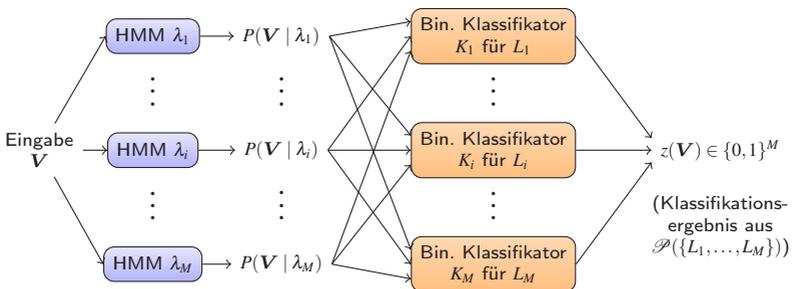


Abbildung 3.5: Multi-Label-Bewegungsklassifikation mit Binärer-Relevanz-Methode: Für jedes der M Bewegungslabels wird durch einen binären Klassifikator ausgehend von den HMM-Likelihoods die Zugehörigkeit einer zu klassifizierenden Bewegung V zu diesem Label bestimmt.

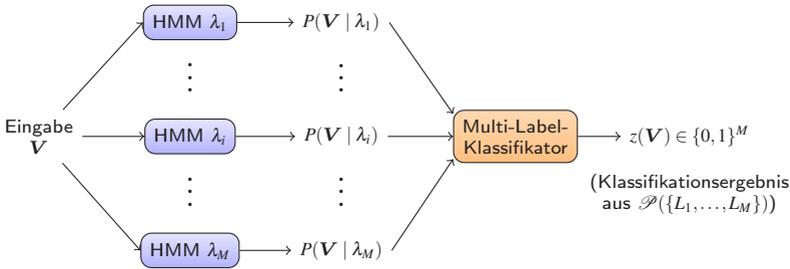


Abbildung 3.6: Multi-Label-Bewegungsklassifikation mit angepasstem Algorithmus: Ein Klassifikationsverfahren, welches zur direkten Anwendung auf Multi-Label-Klassifikationsaufgaben angepasst wurde, bestimmt die Zugehörigkeiten einer zu klassifizierenden Bewegung \mathbf{V} zu den M Bewegungslabels ausgehend von den HMM-Likelihoods.

3.4.3 Anpassung von Algorithmen

Neben Ansätzen basierend auf einer Problemtransformation existiert wie eingangs angesprochen mit AAV auch die Möglichkeit, ein bestehendes Lernverfahren durch Anpassungen oder Erweiterungen so zu adaptieren, dass dieses ohne die Notwendigkeit einer Problemtransformation direkt auf Multi-Label-Daten operieren kann. Derartige Anpassungen sind in der Literatur bereits für einige Lernverfahren beschrieben (Vens et al. 2008; Tsoumakas et al. 2007), beispielsweise für Entscheidungsbäume, eine Multi-Label-Variante des k-Nearest-Neighbor-Algorithmus (ML-KNN), den MMAC-Algorithmus (*Multi-Class Multi-Label Associative Classification*) und den AdaBoost-Algorithmus. Wird ein derart angepasster Algorithmus eingesetzt, kann auf Basis der Likelihoods $P(\mathbf{V} | \lambda_i)$ einer Bewegung unter den HMMs der Labels, wie in Abbildung 3.6 dargestellt, direkt eine Multi-Label-Bewegungsklassifikation erfolgen. Diese Vorgehensweise birgt gegenüber der BRM den Vorteil, dass die Vereinigung des gesamten Klassifikationsproblems in einer einzelnen Instanz eines Klassifikators (anstelle M binärer Klassifikatoren) potentiell die optimale Berücksichtigung von Abhängigkeiten oder Korrelationen zwischen den Bewegungslabels erlaubt, schränkt jedoch

die Wahl des Klassifikators auf an das Multi-Label-Problem angepasste Algorithmen ein. Im Rahmen dieser Arbeit werden hierfür mit dem angepassten CART-Algorithmus (Breiman et al. 1984) trainierte Entscheidungsbäume sowie aus mehreren unkorrelierten Entscheidungsbäumen bestehende *Random Forests* (Breiman 2001) als Multi-Label-Klassifikator evaluiert (siehe Abschnitt 3.5.2).

3.5 Evaluation und Ergebnisse

Dieser Abschnitt befasst sich mit einer Evaluation der im Rahmen dieses Kapitels vorgestellten Ansätze. Hierfür wird in Abschnitt 3.5.1 zunächst das in Abschnitt 3.3 eingeführte Verfahren zur Merkmalsauswahl auf einen Datensatz menschlicher Ganzkörperbewegungen angewendet, um geeignete niedrigdimensionale Merkmalsmengen für den Aufbau des Merkmalsvektors bei der Modellierung solcher Bewegungsdaten zu identifizieren. Nach einer Diskussion der Ergebnisse bildet diese niedrigdimensionale Repräsentation menschlicher Ganzkörperbewegung anschließend in Abschnitt 3.5.2 die Basis für eine Evaluation der in Abschnitt 3.4 vorgestellten HMM-basierten Ansätze zur Multi-Label-Bewegungsklassifikation.

Bei der Durchführung der Evaluation wurde auf eine umfassende Exploration des Hyperparameterraums, beispielsweise möglicher HMM-Topologien, verzichtet. Zum einen ergibt sich durch die hohe Zahl an Hyperparametern kombinatorisch eine überaus große Menge möglicher Parametrisierungen, was dementsprechend eine simultane Evaluation aller Hyperparameter rechnerisch unmöglich machen und eine geschichtete Vorgehensweise erfordern würde, deren Umsetzung weit über den Rahmen dieser Arbeit hinausgeht. Zum anderen finden sich in der existierenden Literatur (Kulić et al. 2007a; Takano et al. 2006) bereits zahlreiche Untersuchungen zu der Fragestellung, wie diese Hyperparameter bei einer Anwendung von HMMs für menschliche Ganzkörperbewegungsdaten gewählt werden sollten. Die in der nachfolgenden Evaluation gewählten Hyperparameter orientieren sich an diesen Arbei-

ten, indem – wie weiter oben bereits erwähnt – A und π gleichverteilt und ϕ mittels K-Means-Clustering initialisiert und zur Parameterreduktion die in ϕ enthaltenen Kovarianzmatrizen auf Diagonalmatrizen beschränkt werden. Zur Modellierung der HMMs werden 8 Zustände in einer Links-Rechts-Topologie mit $\Delta = 1$ eingesetzt und der zum Training verwendete Baum-Welch-Algorithmus wird bis zum Erreichen von $\Delta \text{Loglikelihoods} < 10^{-2}$, höchstens aber für zehn Iterationen ausgeführt.

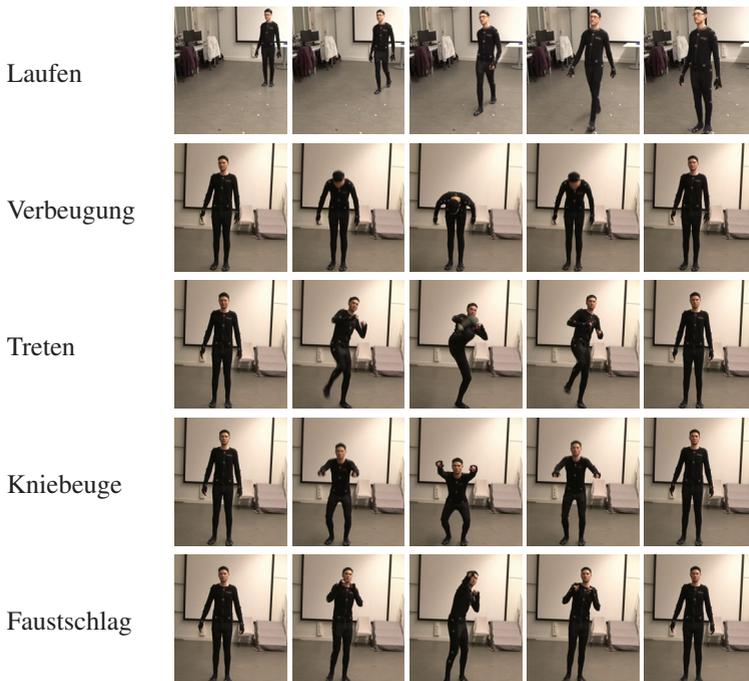


Abbildung 3.7: Manuell ausgewählte Schlüsselbilder exemplarischer Bewegungsaufnahmen des zur Merkmalsauswahl verwendeten Datensatzes (modifiziert nach Mandery et al. 2016b, ©2016 ISIF).

3.5.1 Merkmalsauswahl

Für die im Folgenden beschriebene und zuvor bereits in (Mandery et al. 2016b) publizierte Evaluation des in Abschnitt 3.3 vorgestellten Verfahrens zur Merkmalsauswahl wurde ein Datensatz bestehend aus 353 Bewegungsaufnahmen von 9 unterschiedlichen Probanden (6 männlich, 3 weiblich) zugrunde gelegt, welcher der in Kapitel 4 beschriebenen KIT-Bewegungsdatenbank entnommen wurde. Dieser Datensatz wurde mit der in Abschnitt 2.2.1 beschriebenen Methodik erfasst und beinhaltet insgesamt 23 unterschiedliche Klassen repräsentativer menschlicher Ganzkörperbewegungen, von Lauf- und Wurfbewegungen, sportlichen Aktivitäten wie Golf- oder Tennisschlägen bis hin zu Tanzen oder dem Spielen von Musikinstrumenten. Bei der Erfassung dieser Bewegungen wurde auf die Einbeziehung von Variationen geachtet und so beispielsweise die Bewegungsrichtung, bei einhändigen Bewegungen die ausführende Hand und bei periodischen Bewegungen die Anzahl der Wiederholungen variiert. Abbildung 3.7 zeigt exemplarisch manuell ausgewählte Schlüsselbilder aus fünf Bewegungsaufnahmen des Datensatzes, während sich in Anhang A eine vollständige Übersicht mit einer Aufstellung aller im Datensatz enthaltenen Bewegungsklassen findet, zusammen mit den erforderlichen Informationen zum Auffinden und Abrufen der zu diesen Bewegungsaufnahmen gehörenden Daten in der KIT-Bewegungsdatenbank. Für die Ausführung der durch Algorithmus 1 realisierten Merkmalsauswahl wurde insgesamt eine Rechenzeit von ca. acht Stunden auf einem handelsüblichen Quadcore-System⁸ beansprucht, wobei der Großteil dieser Zeitspanne für das HMM-Training benötigt wird, dessen Zeitaufwand für numerisch stabil berechenbare Dimensionalitäten je nach Merkmalsmenge zwischen etwa 15 und 90 Sekunden liegt.

⁸ Intel Core i7-4700MQ @ 2,4GHz, 16 GB RAM.

Tabelle 3.1: Jeweils beste gefundene Merkmalsmenge für die Dimensionalitäten 1 bis 15 (modifiziert nach Mandery et al. 2016b, ©2016 ISIF).

Dim.	Merkmalsmenge	F_1-Maß	KKR
1	marker_vel_norm	51,67 %	54,45 %
2	CoM_vel_norm, marker_vel_norm	82,28 %	82,20 %
3	angular_momentum_norm, CoM_vel_norm, marker_vel_norm	88,66 %	88,48 %
4	CoM_vel, joint_vel_norm	94,76 %	94,76 %
5	CoM_acc_norm, CoM_vel, marker_vel_norm	94,18 %	94,24 %
6	CoM_vel, CoM_vel_norm, marker_vel_norm, root_vel_norm	94,24 %	94,24 %
7	angular_momentum_norm, CoM_vel, joint_vel_norm, marker_vel_norm, root_vel_norm	94,27 %	94,24 %
8	CoM_vel, end_effectors_vel_norm, joint_vel_norm	95,77 %	95,81 %
9	CoM_vel, end_effectors_vel_norm, joint_vel_norm, root_vel_norm	95,28 %	95,29 %
10	CoM_acc_norm, CoM_vel, CoM_vel_norm, end_effectors_vel_norm, joint_vel_norm	95,75 %	95,81 %
11	CoM_pos, CoM_vel, end_effectors_vel_norm, joint_vel_norm	95,97 %	96,34 %
12	angular_momentum, CoM_vel, end_effectors_vel_norm, marker_vel_norm, root_acc_norm	96,32 %	96,34 %
13	CoM_vel_norm, end_effectors_vel	97,32 %	97,38 %
14	end_effectors_vel, marker_vel_norm, root_rot_norm	97,87 %	97,91 %
15	CoM_acc_norm, CoM_vel_norm, end_effectors_vel, root_vel_norm	97,89 %	97,91 %

Tabelle 3.1 zeigt die als Ergebnis der Merkmalsauswahl beste ermittelte Merkmalsmenge für alle Dimensionalitäten des Merkmalsvektors bis einschließlich 15. Zur Bewertung einer Merkmalsmenge wird, wie in Algorithmus 1 gezeigt, lediglich das F_1 -Maß herangezogen; aus Gründen der Vollständigkeit ist in Tabelle 3.1 dennoch ergänzend die jeweilige KKR aufgeführt, die für diese Anwendung der Multi-Class-Bewegungsklassifikation trotz der grundlegend unterschiedlichen Definition von F_1 -Maß und KKR (siehe Abschnitt 3.2.3) stark mit dem F_1 -Maß korreliert. Tabelle 3.1 ist unmittelbar zu entnehmen, dass wie vorab vermutet die bestmögliche Klassifikationsleistung bei den hier evaluierten niedrigdimensionalen Merkmalsmengen, bei denen Overfitting und numerische Probleme hochdimensionaler Merkmalsvektoren noch nicht zu Tage treten können, mit der Dimensionalität des eingesetzten Merkmalsvektors ansteigt. Zur Validierung der in Tabelle 3.1 präsentierten Resultate und damit indirekt auch der durch Algorithmus 1 beschriebenen Vorgehensweise wurde für Dimensionalitäten des Merkmalsvektors bis einschließlich 5 zusätzlich eine erschöpfende Suche durchgeführt⁹, im Rahmen derer alle potentiell denkbaren Merkmalsmengen dieser Dimensionalitäten evaluiert wurden und die von Algorithmus 1 ermittelten besten Merkmalsmengen bestätigt werden konnten.

Ein Vorteil der durch Algorithmus 1 beschriebenen Metaheuristik gegenüber einer klassischen Vorwärtsselektion besteht in dem parallelen Verfolgen mehrerer alternativer Merkmalsmengen anstelle eines strikt lokal optimalen Ansatzes, wodurch für jede betrachtete Dimensionalität nicht nur die jeweils beste Merkmalsmenge, sondern die $n = 10$ besten Merkmalsmengen berücksichtigt und angegeben werden können. Abbildung 3.8 zeigt wiederum für alle Dimensionalitäten des Merkmalsvektors bis einschließlich 15 die

⁹ Die Dimensionalität wurde auf 5 begrenzt, da mit einer Laufzeit von etwa einem halben Tag für die hierzu erforderliche Evaluation von 16055 Merkmalsmengen die Durchführung der erschöpfenden Suche rechnerisch gerade noch möglich war, während sie für höhere Dimensionalitäten aufgrund des zuvor angesprochenen exponentiellen Wachstum der Zahl möglicher Merkmalsmengen nicht mehr realisierbar ist.

F_1 -Maße dieser jeweils zehn besten Merkmalsmengen, wobei die Merkmalsmengen gemäß ihrem Rang innerhalb der jeweiligen Dimensionalität entlang der X-Achse aufgetragen sind. Insgesamt zeigt sich, dass die Klassifikationsleistung innerhalb der zehn besten Merkmalsmengen einer Dimensionalität beginnend ab fünf Merkmalsdimensionen nur noch unmerklich abnimmt, was darauf hindeutet, dass im Raum möglicher Merkmalsmengen eine größere Zahl vielversprechender Optionen existiert und sich verschiedene durch die Kombination geeigneter Merkmale gebildete Merkmalsmengen oft nur unwesentlich in ihrer Klassifikationsleistung unterscheiden.

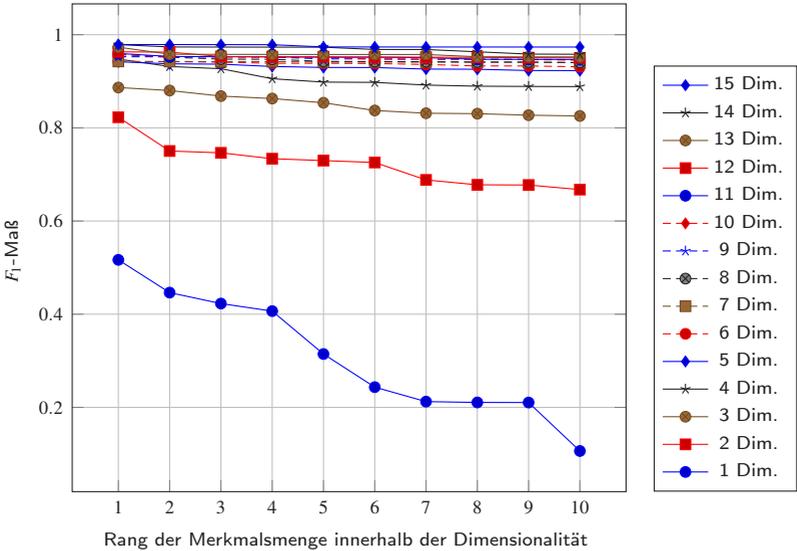


Abbildung 3.8: F_1 -Maße der jeweils zehn besten Merkmalsmengen für die Dimensionalitäten 1 bis 15 (modifiziert nach Mandery et al. 2016b, ©2016 ISIF). Die X-Achse repräsentiert den Rang einer Merkmalsmenge innerhalb ihrer jeweiligen Dimensionalität.

Neben der Identifikation vollständiger Merkmalsmengen erlaubt die Merkmalsauswahl zusätzlich auch eine Analyse einzelner Merkmale, indem betrachtet wird, ob bzw. wie oft diese Merkmale in den aus Algorithmus 1

resultierenden Merkmalsmengen Einsatz finden. Hierbei zeigt sich eine sehr starke Heterogenität zwischen den definierten Merkmalen, bei der einige Merkmale in einer Vielzahl an Merkmalsmengen enthalten sind und sich dementsprechend gut zur Modellierung menschlicher Ganzkörperbewegung zu eignen scheinen, während sich andere Merkmale überhaupt nicht in den erzeugten Merkmalsmengen finden. Insbesondere die Massenschwerpunkt-Merkmale, und hierbei speziell die Geschwindigkeit, stellen sich als sehr relevante Merkmale heraus, was sich darin äußert, dass in nahezu jeder der resultierenden Merkmalsmengen eines der beiden assoziierten Merkmale *CoM_vel* und *CoM_vel_norm* enthalten ist. Allgemein scheint es bei vektorwertigen Merkmalen in Form von Geschwindigkeits- oder Beschleunigungsvektoren in vielen Fällen ausreichend zu sein, die (skalare) euklidische Norm anstelle des ursprünglichen Vektors zu betrachten, was darauf schließen lässt, dass ein wesentlicher Anteil am Informationswert dieser Merkmale durch die betragsmäßige Größe des Merkmalswerts und weniger durch seine Richtung repräsentiert wird. Die auf dem Ganzkörper-Drehimpuls basierenden Merkmale *angular_momentum* und *angular_momentum_norm* werden nur in einer geringen Anzahl der resultierenden Merkmalsmengen verwendet und scheinen weniger vielversprechend für eine HMM-basierte Modellierung der Ganzkörperbewegung zu sein, als dies in Abschnitt 2.3.1 zunächst postuliert wurde. Eine mögliche Ursache hierfür könnte allerdings auch in dem zur Merkmalsauswahl eingesetzten Datensatz (siehe Anhang A) liegen, in welchem lediglich ein geringer Anteil der Bewegungen, beispielsweise Golf- und Tennisschlagbewegungen, ein hohes Maß an Ganzkörper-Bewegungsdynamik aufweist, während der Ganzkörper-Drehimpuls mutmaßlich gerade zur Modellierung solcher Bewegungen ein besonders gut geeignetes Merkmal darstellt.

Abschließend soll an dieser Stelle noch ein Vergleich der aus Algorithmus 1 resultierenden niedrigdimensionalen Merkmalsmengen mit klassischen, in existierenden Arbeiten zur menschlichen Ganzkörperbewegung verwendeten Merkmalen präsentiert werden. Wie bereits eingangs in Ab-

schnitt 3.3 diskutiert, wird in der überwiegenden Mehrzahl dieser Arbeiten keine Merkmalsauswahl vorgenommen, sondern lediglich ein einzelnes, manuell ausgewähltes Merkmal wie etwa Gelenkwinkel/-geschwindigkeiten oder Markerpositionen/-geschwindigkeiten zur Beschreibung der menschlichen Bewegung herangezogen. Tabelle 3.2 zeigt einen Vergleich der aus Algorithmus 1 resultierenden Merkmalsmengen für die Dimensionalitäten 4 und 8 (vgl. Tabelle 3.1) mit solchen klassischen Merkmalen hinsichtlich Dimensionalität, Klassifikationsleistung (F_1 -Maß und KKR), aber auch der insgesamt zum HMM-Training benötigten Zeitspanne kumuliert über alle drei Kreuzvalidierungsrunden. Das populäre, beispielsweise von Kulić et al. (2012) eingesetzte Merkmal der Gelenkwinkel (*joint_pos*) schneidet, auch bei der Betrachtung von Gelenkwinkelgeschwindigkeiten (*joint_vel*), gemessen an den ermittelten niedrigdimensionalen Merkmalsmengen bei der Multi-Class-Bewegungsklassifikation auf dem hier verwendeten Datensatz vergleichsweise schlecht ab¹⁰. Die Geschwindigkeiten und mehr noch die Positionen der Endeffektoren, unter Berücksichtigung der in Abschnitt 2.3.2 beschriebenen Normalisierungsschritte zur Repräsentation der kartesischen Positionen im lokalen Koordinatensystem des Menschen, schneiden hingegen – angesichts der im Datensatz zur Evaluation enthaltenen Bewegungstypen wenig überraschend – im Vergleich gut ab. Tatsächlich handelt es sich bei isolierter Betrachtung aller Merkmale (ohne Berücksichtigung ihrer Dimensionalität) bei *end_effectors_pos* gemessen am F_1 -Maß um das beste dieser Merkmale.

Eine Betrachtung der beiden im oberen Bereich von Tabelle 3.2 gezeigten, mit Algorithmus 1 erzeugten niedrigdimensionalen Merkmalsmengen zeigt, dass diese eine bessere Klassifikationsleistung als die meisten (4-dimensionale Merkmalsmenge) bzw. als alle (8-dimensionale Merkmals-

¹⁰Hierbei ist jedoch zu beachten, dass sich durch die Verwendung unterschiedlicher kinematischer Modelle und somit abweichender Gelenk-Definitionen sowie durch eine eventuell nachgelagerte Merkmalsextraktion, z.B. mit PCA, ein direkter Rückschluss auf den Einsatz dieser Merkmale in anderen Arbeiten methodologisch schwierig gestaltet.

menge) der zu Vergleichszwecken evaluierten klassischen Merkmale aufweisen, gleichzeitig aber durch ihre niedrigere Dimensionalität eine deutliche reduzierte Trainingszeit, verglichen mit *joint_pos* beispielsweise nur etwa die Hälfte, benötigen. Zusätzliche weitere Vorteile dieser niedrigdimensionaleren Darstellungen könnten beispielsweise in einer reduzierten Menge an benötigten Trainingsdaten oder einem reduzierten Risiko für Overfitting (Guyon et al. 2003), bestehen, was an dieser Stelle jedoch nicht abschließend beurteilt werden kann. Aufgrund dieser genannten Vorteile wurde die in Tabelle 3.2 gezeigte 8-dimensionale Merkmalsmenge, bestehend aus den Merkmalen *CoM_vel*, *joint_vel_norm*, *end_effectors_vel_norm*, in den weiteren Teilen dieser Arbeit zur Modellierung des Merkmalsvektors sowohl in der in Abschnitt 3.5.2 präsentierten Evaluation von Methoden der Multi-Label-Bewegungsklassifikation als auch für die in Abschnitt 4.4 vorgestellten Anwendung dieser Methoden für die datengetriebene Generierung von Klassifikationsvorschlägen in der KIT-Bewegungsdatenbank ausgewählt.

Tabelle 3.2: Vergleich von niedrigdimensionalen Merkmalsmengen (oberer Teil) mit manuell ausgewählten, häufig zur Beschreibung menschlicher Ganzkörperbewegung eingesetzten Merkmalen (unterer Teil) hinsichtlich Dimensionalität, Klassifikationsleistung und insgesamt benötigter Trainingszeit (modifiziert nach Mandery et al. 2016b, ©2016 ISIF).

Merkmalsmenge	Dim.	F_1-Maß	KKR	Tr.zeit
CoM_vel, joint_vel_norm	4	94,76 %	94,76 %	19 s
CoM_vel, joint_vel_norm, end_effectors_vel_norm	8	95,77 %	95,81 %	21 s
root_pos, root_rot	6	87,17 %	87,54 %	20 s
end_effectors_pos	12	95,22 %	95,29 %	25 s
end_effectors_vel	12	94,15 %	94,05 %	25 s
joint_pos	40	87,17 %	88,10 %	39 s
joint_vel	40	86,84 %	86,97 %	48 s
marker_pos	168	90,29 %	90,37 %	108 s
marker_vel	168	94,44 %	94,62 %	120 s

3.5.2 Multi-Label-Bewegungsklassifikation

Der folgende Abschnitt befasst sich mit einer Evaluation der in Abschnitt 3.4 vorgestellten HMM-basierten Ansätze zur Multi-Label-Bewegungsklassifikation. Die Implementierung dieser unterschiedlichen Ansätze, welche die Grundlage für die vorliegende Evaluation bildet, wurde im Rahmen einer Bachelorarbeit (Plappert 2015) auf Basis der Scikit-Learn-Bibliothek (Pedregosa et al. 2011) für maschinelles Lernen in der Programmiersprache Python (Python o.D.) vorgenommen. Konkret wurde die in Abschnitt 3.4.2 beschriebene BRM mit

1. logistischer Regression (Bishop 2006) und
2. SVMs (Steinwart et al. 2008)

als binären Klassifikatoren und das in Abschnitt 3.4.3 beschriebene AAV basierend auf

3. mit einem angepassten CART-Algorithmus trainierten Entscheidungsbäumen (Breiman et al. 1984) und
4. *Random Forests* (Breiman 2001), bestehend aus mehreren dieser Entscheidungsbäumen,

implementiert. Auf eine ausführliche Beschreibung dieser unterliegenden Klassifikatoren sei an dieser Stelle verzichtet und der Leser auf die angegebene Literatur verwiesen.

Für ihre Verwendung müssen diese Klassifikatoren allerdings jeweils parametrisiert werden, wobei diese (Hyper-)Parametrisierung problemspezifisch erfolgen muss, hier also für das Problem der Multi-Label-Bewegungsklassifikation ausgehend von den Likelihoods einer Bewegung unter den HMMs für die einzelnen Labels. Im Rahmen von (Plappert 2015) wurde in einer ersten Evaluation (mit einem kleineren Datensatz) bereits eine Exploration des Parameterraums durchgeführt, deren Ergebnisse für die Parametrisierung der Klassifikatoren in der folgenden Evaluation genutzt werden.

Die Implementierung der logistischen Regression, bei der die Wahrscheinlichkeitsverteilung einer binären Ausgabevariable mit Hilfe der logistischen Funktion abgeschätzt wird, basiert auf der in Scikit-Learn gekapselten LIBLINEAR-Bibliothek (Fan et al. 2008). Die Anwendung der logistischen Regression erfordert eine Tikhonov-Regularisierung, welche zur Verhinderung von Overfitting über einen zusätzlichen Term in der Kostenfunktion das Auftreten hoher Koeffizientenwerte bestraft. In der auf LIBLINEAR basierenden Implementierung in Scikit-Learn kann zwischen einer L1-Regularisierung mit der Summennorm $\|x\|_1 = \sum_{i=1}^n |x_i|$ und einer L2-Regularisierung mit der euklidischen Norm gewählt werden. Der skalare Koeffizient $C > 0$ ist hierbei invers zur Stärke der Regularisierung; niedrige Werte von C korrespondieren also mit einer stärkeren Regularisierung. Im Rahmen dieser Arbeit wird für die logistische Regression sowohl L1- als auch L2-Regularisierung jeweils mit $C \in \{10^{-3}, 10^{-2}\}$ evaluiert.

Auch die Implementierung der SVMs basiert auf einer Kapselung der LIBLINEAR-Bibliothek in Scikit-Learn. Bei SVMs handelt es sich um einen außerordentlich populären binären Klassifikator, bei welchem die Eingabedaten zunächst mit Hilfe einer Kernelfunktion in einen höherdimensionalen Raum transformiert werden, um dann in diesem Raum eine die beiden Klassen möglichst gut voneinander trennende Hyperebene zu schätzen. In der evaluierten Implementierung wird ein linearer Kernel und wie bei der logistischen Regression eine Tikhonov-Regularisierung zur Verhinderung von Overfitting eingesetzt. Die evaluierten Kombinationen der Hyperparameter werden hierbei analog zur logistischen Regression gewählt (L1- und L2-Regularisierung jeweils mit $C \in \{10^{-3}, 10^{-2}\}$).

Für die Anwendung der Entscheidungsbäume im Rahmen eines AAV müssen mit der Metrik zur Split-Bewertung und der maximalen Tiefe der trainierten Entscheidungsbäume zwei Hyperparameter festgelegt werden. Die Metrik zur Split-Bewertung wird während der Anwendung des CART-Algorithmus zum Training benötigt, wenn für die Festlegung der binären Fallunterscheidung in einem Baumknoten die Menge der an diesem Knoten

verfügbaren Trainingsdaten in zwei Teilmengen aufgespalten wird. Neben dem ursprünglich im CART-Algorithmus hierfür vorgesehenen Gini-Index steht mit der Transinformation auch ein informationstheoretisches Gütemaß bereit. Die maximale Baumtiefe setzt eine obere Schranke für die Anzahl der aufeinanderfolgenden Knoten in einem Entscheidungsbaum und trägt somit zum Verhindern von Overfitting bei. Aufgrund des verhältnismäßig schlechten Abschneidens der AAV in (Plappert 2015) wird im Rahmen dieser Evaluation lediglich die dort beste Parametrisierung mit Verwendung der Transinformation zur Split-Bewertung und einer maximalen Baumtiefe von 15 berücksichtigt.

Abschließend wurde mit einem Random Forest als Multi-Label-Klassifikator ein zweites AAV evaluiert. Bei Random Forests werden mehrere Entscheidungsbäume, die durch eine randomisierte Auswahl der während des Trainings betrachteten Merkmalsteilmengen dekorreliert werden, kombiniert, indem das mehrheitliche Klassifikationsergebnis der einzelnen Entscheidungsbäume die Ausgabe des gesamten Random Forest bestimmt. Zusätzlich zu den angesprochenen Hyperparametern der einzelnen Entscheidungsbäume, die wie zuvor gewählt werden, muss auch die Anzahl der insgesamt trainierten Entscheidungsbäume festgelegt werden. Im Rahmen dieser Evaluation wurde hierfür eine Zahl von 40 Entscheidungsbäumen gewählt.

Für die Durchführung der Evaluation wurde ein Datensatz bestehend aus 579 Bewegungsaufnahmen von 12 unterschiedlichen Probanden (8 männlich, 4 weiblich) zugrunde gelegt, bei dem es sich um eine Erweiterung des in (Plappert 2015) zur Evaluation verwendeten Datensatzes handelt. Wie in Abschnitt 3.5.1 besteht dieser Datensatz aus Aufnahmen unterschiedlicher menschlicher Ganzkörperbewegungen, welche mit der in Abschnitt 2.2.1 erläuterten Vorgehensweise erfasst und anschließend in der KIT-Bewegungsdatenbank abgelegt wurden. Im Unterschied zu dem in Abschnitt 3.5.1 eingesetzten Datensatz sind die Bewegungsaufnahmen hier nun aber jeweils mit mehreren von insgesamt 55 unterschiedlichen Bewegungslabels gekennzeichnet, die neben dem Typ der Bewegung auch

Modalitäten der Bewegungsausführung wie Richtungen, Geschwindigkeiten, beteiligte Extremitäten und ähnliches beschreiben. Die Label-Kardinalität (M.-L. Zhang et al. 2014) des Datensatzes als Zahl der durchschnittlich mit einer Bewegungsaufnahme assoziierten Labels beträgt 2,25. Anhang A enthält eine vollständige Übersicht über die im Datensatz enthaltenen Label-Kombinationen zusammen mit den jeweiligen IDs der mit jeder Kombination assoziierten Bewegungsexperimente in der KIT-Bewegungsdatenbank.

Tabelle 3.3: Ergebnisse der Evaluation unterschiedlicher Verfahren zur Multi-Label-Bewegungs-klassifikation. Die angegebene KKR bezieht sich auf den Anteil an Bewegungen, bei denen sämtliche Labels korrekt erkannt wurden.

Klassifikationsverfahren				
Typ	Klassifikator	Parametrisierung	F_1 -Maß	KKR
BRM	Logistische Regr.	L1-Reg. mit $C = 10^{-3}$	85,99 %	78,24 %
		L1-Reg. mit $C = 10^{-2}$	89,31 %	84,11 %
		L2-Reg. mit $C = 10^{-3}$	84,32 %	72,19 %
		L2-Reg. mit $C = 10^{-2}$	84,48 %	72,19 %
BRM	SVM	L1-Reg. mit $C = 10^{-3}$	89,84 %	80,83 %
		L1-Reg. mit $C = 10^{-2}$	90,72 %	81,52 %
		L2-Reg. mit $C = 10^{-3}$	83,70 %	69,78 %
		L2-Reg. mit $C = 10^{-2}$	82,82 %	70,12 %
AAV	Entscheidungsbaum	siehe Text	71,67 %	66,67 %
AAV	Random Forest	siehe Text	71,65 %	60,10 %

Analog zur Vorgehensweise in Abschnitt 3.5.1 erfolgt die Evaluation mit Hilfe einer 3-fachen stratifizierten Kreuzvalidierung, wobei zur Stratifizierung hier das von Sechidis et al. (2011) vorgestellte Verfahren für Multi-Label-Daten zum Einsatz kommt. Tabelle 3.3 zeigt die Ergebnisse der Evaluation in Form des F_1 -Maßes und der KKR für alle evaluierten Ansätze. Die angegebene KKR berücksichtigt hierbei als *Subset Accuracy* (M.-L. Zhang et al. 2014) nur Bewegungen, bei denen die prädizierte Label-Zugehörigkeit für sämtliche Labels korrekt war, und stellt somit im Vergleich zu anderen

zur Beurteilung von Multi-Label-Klassifikatoren eingesetzten Maßen wie beispielsweise dem *Hamming Loss* (M.-L. Zhang et al. 2014) ein sehr striktes Kriterium dar.

Eine Betrachtung der Ergebnisse in Tabelle 3.3 zeigt zunächst, dass keiner der evaluierten Ansätze offensichtlich ungeeignet für das Problem der Multi-Label-Bewegungsklassifikation ist und sogar bei der Betrachtung des am schlechtesten evaluierten Systems etwa 60 % der aus der Klassifikation resultierenden Label-Kombinationen mit der vorgegebenen Referenzklassifikation übereinstimmen. Allgemein liegt die Klassifikationsleistung für die auf der BRM basierenden Ansätze deutlich über der für die auf AAV basierenden Ansätze. Von den auf der BRM basierenden Ansätzen schnitt die Parametrisierung mit L1-Regularisierung und $C = 10^{-2}$ sowohl für die Verwendung von logistischer Regression als auch von SVMs als binärem Klassifikator jeweils am besten ab (in Tabelle 3.3 fett gedruckt). Während beim Vergleich dieser beiden besten Parametrisierungen das F_1 -Maß bei Verwendung der SVMs um 1,41 Prozentpunkte höher liegt, übertrifft der auf logistischer Regression basierende Multi-Label-Klassifikator diesen Ansatz hinsichtlich der KKR um 2,59 Prozentpunkte. Für die in Abschnitt 4.4 beschriebene praktische Anwendung von Multi-Label-Bewegungsklassifikation für die datengetriebene Generierung von Klassifikationsvorschlägen in der KIT-Bewegungsdatenbank wurde aus diesem Grund der Ansatz basierend auf der BRM und logistischer Regression ausgewählt und in Abschnitt 4.4 im Rahmen einer Benutzerstudie evaluiert.

3.6 Zusammenfassung von Kapitel 3

Dieses Kapitel befasste sich mit der Modellierung menschlicher Ganzkörperbewegung basierend auf Hidden Markov Models und den 29 in Kapitel 2 eingeführten Merkmalen, sowie mit der Anwendung der gelernten Modelle für das Problem der Multi-Label-Bewegungsklassifikation, bei dem eine vorgegebene Ganzkörperbewegung maschinell einer beliebigen Teil-

menge verfügbarer Bewegungslabels zugeordnet werden soll. Hiermit legt dieses Kapitel die Grundlage für die in Kapitel 4 beschriebene datengetriebene Erzeugung von Tagging-Vorschlägen in der dort behandelten Bewegungsdatenbank. Zunächst wurden die für die Modellierung von Ganzkörperbewegungen erforderlichen Grundlagen eingeführt, indem eine formale Definition von HMMs vorgenommen, der zum Modelltraining verwendete Baum-Welch-Algorithmus vorgestellt und die Parametrisierung der Modelle und des Trainingsalgorithmus diskutiert wurden. Zur Festlegung des Aufbaus des zur Bewegungsmodellierung eingesetzten Merkmalsvektors wurde ein Wrapper-Ansatz basierend auf einer im Rahmen dieser Arbeit entwickelten Metaheuristik vorgestellt, welcher durch die systematische Kombination von Merkmalen eine Identifikation von hochdeskriptiven niedrigdimensionalen Teilmengen der verfügbaren Merkmale ermöglicht. Im Zuge der Evaluation dieser Merkmalsauswahl konnte nachgewiesen werden, dass bei geschickter Auswahl von Merkmalen bereits eine geringe Zahl an Merkmalen bzw. an Dimensionen des Merkmalsvektors zum Erzielen einer guten Klassifikationsleistung genügt. So konnten bereits für 8-dimensionale Merkmalsvektoren Merkmalskombinationen identifiziert werden, die klassischen bei der Bewegungsklassifikation eingesetzten Merkmalsvektoren sowohl hinsichtlich ihrer Klassifikationsleistung mit einem F_1 -Maß von 95,77 % als auch hinsichtlich der zum Training benötigten Zeitdauer überlegen sind. Die niedrigdimensionale Repräsentation menschlicher Ganzkörperbewegung und ihre Modellierung mit HMMs bildete im Anschluss die Grundlage für eine Anwendung auf das Problem der Multi-Label-Bewegungsklassifikation. Für dieses Problem wurden unterschiedliche Lösungsansätze basierend auf Problemtransformationen bzw. auf dem Einsatz von angepassten Algorithmen diskutiert. Bei einer Evaluation mit einem Bewegungsdatensatz bestehend aus 579 Bewegungsaufnahmen von 12 Probanden und 55 Bewegungslabels wurde die Eignung der Ansätze mit unterschiedlichen Parametrisierungen untersucht, wobei das beste der evaluierten Systeme einen F_1 -Score von 89,31 % und eine Korrektklassifikationsrate von 84,11 % erzielen konnte.

4 Die KIT-Bewegungsdatenbank

Ein wesentlicher Beitrag der vorliegenden Arbeit ist die in ihrem Rahmen entstandene *KIT Whole-Body Human Motion Database*, im Folgenden als KIT-Bewegungsdatenbank bezeichnet, bei welcher es sich um eine großskalige Datenbank menschlicher Ganzkörperbewegung handelt. Diese Bewegungsdatenbank basiert auf dem MMM-Framework (siehe Abschnitt 2.1) und stellt zusätzlich zur menschlichen Bewegung auch Informationen über die beteiligten Probanden, z.B. anthropometrische Messungen, und über ebenfalls erfasste Umgebungsobjekte (siehe Abschnitt 2.2.2) bereit. Seit ihrer Entstehung wird diese Bewegungsdatenbank als zentrale Komponente für die Verwaltung menschlicher Bewegungsdaten am H²T eingesetzt und stellt als solche auch die in den anderen Teilen dieser Arbeit (Kapitel 3, 5 und 6) zur Evaluation verwendeten Bewegungsaufnahmen bereit. Neben dem internen Einsatz am H²T fand die Datenbank jedoch bereits auch Verwendung in verschiedenen EU-Forschungsprojekten und Arbeiten mehrerer externer Forschungsgruppen, beispielsweise bei Clever et al. (2016) und N. Chen et al. (2016). In diesem Kapitel sollen nun Aufbau, Struktur und Inhalte der KIT-Bewegungsdatenbank näher erläutert werden. Die Inhalte dieses Kapitels basieren dabei zum Teil auf (Mandery et al. 2015b), wo die KIT-Bewegungsdatenbank erstmals umfassend beschrieben wurde, und auf (Mandery et al. 2016c).

Der Aufbau dieses Kapitels gliedert sich wie folgt. Zunächst werden in Abschnitt 4.1 bekannte Ansätze zur Organisation von Ganzkörperbewegungen und existierende großskalige Datenbanken in der Robotik und verwandten Fachdisziplinen diskutiert. Abschnitt 4.2 beschreibt dann den Aufbau

der KIT-Bewegungsdatenbank hinsichtlich verfügbarer Datentypen, interner Softwarearchitektur und den nach außen bereitgestellten Schnittstellen. In Abschnitt 4.3 wird ein neuer Ansatz zur hierarchischen Klassifikation von Bewegungsdaten mit Hilfe eines *Motion Description Tree* vorgestellt, welcher in der KIT-Bewegungsdatenbank zur Organisation von Bewegungsaufnahmen zum Einsatz kommt. Die mit dem Motion Description Tree realisierte Klassifikation erfolgt prinzipiell manuell durch die Benutzer der Datenbank, welche hierbei jedoch durch Klassifikationsvorschläge auf Basis der in Kapitel 3 vorgestellten Ansätze zur Multi-Label-Bewegungsklassifikation unterstützt werden. Die Integration dieser Ansätze für die Generierung von Klassifikationsvorschlägen in der KIT-Bewegungsdatenbank und eine Benutzerstudie zur Evaluation werden in Abschnitt 4.4 beschrieben. Abschließend wird in Abschnitt 4.5 eine qualitative und quantitative Beschreibung der Inhalte der KIT-Bewegungsdatenbank vorgenommen.

4.1 Verwandte Arbeiten und existierende Datenbanken

Die systematische Erfassung und Analyse der menschlichen Bewegung ist eine unabdingbare Voraussetzung für eine große Zahl an Forschungsansätzen in der humanoiden Robotik, aber auch im Bereich der Computergrafik und des maschinellen Sehens. Viele dieser Ansätze erfordern größere Mengen an Bewegungsdaten und die Verfügbarkeit von Daten unterschiedlicher menschlicher Probanden, die unterschiedliche Parametrisierung menschlicher Aktionen beispielsweise für verschiedenartige Umgebungsbedingungen oder die Erfassung multipler Demonstrationen für jede Aktion. Entsprechend dieser Anforderungen existiert bereits eine Vielzahl an zum Teil öffentlich verfügbaren Datensammlungen unterschiedlicher Größen, welche Daten für derartige Forschungsfragestellungen beinhalten und im Folgenden vorgestellt werden. Soweit nicht anders erwähnt, kommen zur Erfassung der Bewegungsdaten stets markerbasierte Ansätze (vgl. Abschnitt 2.2.1) zum Einsatz. Als

generelle Beobachtung sei hier jedoch bereits vorangestellt, dass die meisten dieser Datensätze durch eine einzelne Forschungsgruppe mit der Bindung an ein spezifisches Forschungsprojekt erhoben wurden, wobei jeweils nur Daten gesammelt wurden, welche im Rahmen dieses Projekts von Interesse waren, und die Datensammlung nach kurzer Zeit mit Abschluss des Forschungsprojekts wieder eingestellt wurde. Die meisten der folgenden Datensätze haben daher sowohl gemessen an der Anzahl von Bewegungsaufnahmen als auch am Spektrum der erfassten menschlichen Bewegungen nicht den Umfang universeller großskaliger Bewegungsdatenbanken erreicht, wie ihn beispielsweise die CMU-Bewegungsdatenbank (CMUMoCapDB o.D.) und auch die in diesem Kapitel vorgestellte KIT-Bewegungsdatenbank aufweisen.

Eine der größten und bekanntesten Bewegungsdatenbanken ist die bereits erwähnte *CMU Graphics Lab Motion Capture Database* (CMUMoCapDB o.D.), welche eine Gesamtzahl von 2605 Bewegungsaufnahmen frei verfügbar bereitstellt und zu den am häufigsten in der Forschung verwendeten Bewegungsdatenbanken gehört. Mit verschiedenen Kategorien wie Fortbewegung, menschlicher Interaktion mit der Umgebung oder anderen Menschen und sportlichen Aktivitäten wird ein breites Repertoire menschlicher Ganzkörperbewegungen angeboten, die von insgesamt über 100 menschlichen Probanden aufgenommen wurden. Ein Nachteil der CMU-Bewegungsdatenbank stellt allerdings die nur lose Strukturierung der Daten dar, im Rahmen derer die Bewegungsaufnahmen lediglich gemäß einer zweistufigen Hierarchie kategorisiert werden (beispielsweise „*human interaction*“ → „*two subjects*“) und eine weitere Beschreibung der Bewegung nur anhand von nicht immer konsistent angegebenen Freitextanmerkungen erfolgt. Wie in Abschnitt 4.5 noch erläutert werden wird, wurden die Bewegungsdaten der CMU-Bewegungsdatenbank ebenfalls in die KIT-Bewegungsdatenbank integriert, womit diese Bewegungsaufnahmen inzwischen auf dieselbe Weise und im selben For-

mat wie die restlichen Bewegungsaufnahmen der KIT-Bewegungsdatenbank bereitstehen¹.

Eine andere sehr bekannte Bewegungsdatenbank stellt HDM05 (Müller et al. 2007) dar, welche als Kooperation mehrerer deutscher Forschungsinstitute für Fragestellungen der Bewegungsanalyse, -synthese und -klassifikation geschaffen wurde. Der Datensatz enthält 1457 Bewegungsaufnahmen mit einer Gesamtdauer von etwa 50 Minuten für 100 unterschiedliche Bewegungsklassen, wobei die einzelnen Aufnahmen durch die manuelle Unterteilung längerer Sequenzen (ca. 3 Stunden Bewegungsmaterial) gewonnen wurden. Im Vergleich zur CMU-Datenbank zeichnet sich HDM05 durch einen erhöhten Organisationsgrad aus, allerdings wurden nur Bewegungen von fünf menschlichen Probanden aufgenommen, wobei die Zahl verfügbarer Probanden für einige Bewegungsklassen sogar diesen Wert noch unterschreitet.

Basierend auf den Daten der CMU-Bewegungsdatenbank und HDM05 präsentieren Riaz et al. (2016) eine Organisation dieser Bewegungsdaten unter Verwendung einer relationalen Datenbankstruktur und demonstrieren, wie komplexe Abfragen innerhalb dieser Datenbankstruktur mit SQL formuliert werden können.

Ein weiterer Ansatz für eine methodisch klar strukturierte Erfassung von Bewegungsdaten stellt die von Guerra-Filho et al. (2012) vorgestellte *Human Motion Database* dar. Hierbei wird ein systematischer Ansatz zur Durchführung der Datenerhebung vorgeschlagen, welcher die Struktur der Datenbank und die zu erfassenden Bewegungsdaten festlegt. Die gesammelten Daten werden in fünf unterschiedliche Datensätze eingeteilt: Der Praxicon-Datensatz enthält Bewegungsaufnahmen von ca. 350 menschlichen Aktionen eines einzelnen Probanden, der Cross-Validation-Datensatz liefert Aufnahmen von 50 verschiedenen Probanden für eine Teilmenge von 70 dieser Aktionen und der Generalization-Datensatz besteht aus Aufnahmen, bei denen Parameter der Bewegung systematisch variiert wurden, z.B. die Posi-

¹ Allerdings fehlt eine Klassifikation der Bewegungen gemäß dem in der KIT-Bewegungsdatenbank verwendeten und in Abschnitt 4.3 beschriebenen *Motion Description Tree*.

tion eines Greifziels oder die Höhe des Stuhl bei Aufstehbewegungen. Die restlichen zwei Datensätze beinhalten Aufnahmen für zusammengesetzte Bewegungen und für die Interaktion zwischen zwei menschlichen Probanden. Des Weiteren beinhaltet die Arbeit von Guerra-Filho et al. (2012) einen Überblick über die wichtigsten Bewegungsdatenbanken hinsichtlich Inhalt, Struktur und Größe (auf dem Stand von 2012).

Die von Ma et al. (2006) vorgestellte *DB for Study of Gender, Identity and Emotion* wurde zur Untersuchung der in menschlicher Bewegung enthaltenen persönlichen Eigenschaften wie Geschlecht, Identität und emotionaler Stimmung sowie für Fragestellungen, wie diese Eigenschaften mit Hilfe menschlicher Bewegung dekodiert und enkodiert werden können, geschaffen. Der Datensatz enthält eine Gesamtzahl von 4080 Bewegungsaufnahmen, die aus 136 verschiedenen menschlichen Aktionen bestehen, welche jeweils von 30 Amateurschauspielern für vier verschiedene emotionale Stimmungen (neutral, wütend, glücklich, traurig) demonstriert wurden. Einen ähnlichen Zweck verfolgt die IEMOCAP-Datenbank (Busso et al. 2008), welche ebenfalls zur Emotionserkennung und der kombinierten Analyse von verbaler (sprachlicher) und nonverbaler (gestikulierter) Kommunikation geschaffen wurde. Enthalten sind Bewegungsdaten von zehn ausgebildeten Schauspielern, welche in fünf Sitzungen insgesamt zehn unterschiedliche emotionale Stimmungen wie Glück, Angst oder Wut demonstriert haben.

In der *Edinburgh CGVU Interaction Database* (CGVUInteractionDatabase o.D.) kommt die von Sandilands et al. (2013) beschriebene Kombination aus magnetischer Bewegungserfassung und dem Kinect-Sensor als RGB-D-Kamera zur Erfassung verschiedener Aufgaben der Objekt-Manipulation und -Interaktion zum Einsatz. Die magnetische Erfassung erlaubt bei diesem Ansatz einen Ausgleich des bei rein visuellen Ansätzen auftretenden Problems von Verdeckungen. Im Gegensatz zu anderen hier vorgestellten Datenbanken werden bei dieser Arbeit, genau wie bei der KIT-Bewegungsdatenbank, explizite Polygonmodelle für an der Interaktion beteiligte Objekte bereitgestellt und diese Objekte während der Bewegungsausführung erfasst.

Die HuMoD-Datenbank (Wojtusich et al. 2015) kombiniert Bewegungsaufnahmen des Menschen mit anthropometrischen Parametern wie Segmentlängen des menschlichen Körpers sowie Massen, Massenschwerpunkten und Trägheitsmomenten der Körpersegmente für die Untersuchung der Dynamik der muskelgetriebenen menschlichen Bewegung. Bei der Bewegungserfassung wurde ein Fokus auf die Bewegung des Unterkörpers gelegt und es wurden 13 periodische Bewegungen wie Laufen, Rennen und Kniebeugen für eine vorgegebene Zeitspanne von je einem männlichen und einem weiblichen Probanden aufgezeichnet.

Bei der von Kuehne et al. (2011) vorgestellten HMDB51 handelt es sich um eine Video-Datenbank, welche aus ca. 7000 manuell annotierten Bewegungsclips besteht und zur Evaluation der Leistungsfähigkeit und Robustheit von Verfahren zur Aktionserkennung im Bereich des maschinellen Sehens entwickelt wurde. Da diese Datenbank lediglich Video-Daten beinhaltet, stellt das Fehlen exakter Bewegungsdaten eine Hürde bei der Anwendung der Daten für die Bewegungsanalyse und -synthese in der Robotik dar. Eine weitere derartige Datenbank findet sich bei der *CMU Motion of Body (MoBo) Database* (Gross et al. 2001), die Videoaufnahmen für vier unterschiedliche Laufarten (langsames/schnelles Laufen, Aufwärtslaufen und Laufen mit einem Ball in den Händen) von 25 verschiedenen Probanden enthält.

Auch der nicht öffentlich verfügbare HumanEva-Datensatz (Sigal et al. 2010) wurde für die Evaluation von bildbasierten Verfahren entworfen, genauer für Verfahren zur Rekonstruktion der menschlichen Pose aus Videodaten. Hierfür wurden sechs Aktionen von jeweils vier menschlichen Probanden aufgenommen, die im Unterschied zu den anderen Datensätzen basierend auf markerbasierter Bewegungserfassung gemäß der Zielsetzung lose Alltagskleidung anstatt eines speziellen Anzugs zur Bewegungserfassung trugen, wodurch eine große Unsicherheit in den erfassten Bewegungsdaten entsteht. Ebenfalls der Evaluation von Verfahren zur Posen-Rekonstruktion dient der von Ionescu et al. (2014) vorgestellte Human3.6M-Datensatz, welcher eine Gesamtzahl von 3,6 Millionen menschlichen Posen von 11 professionellen

Schauspielern enthält, die von 4 hochauflösenden Kameras aufgenommen wurden. Zusätzlich werden Gelenkwinkel von einem Bewegungserfassungssystem, Tiefendaten von Time-of-Flight-Kameras und 3D-Laserscans der Probanden bereitgestellt.

Das Unternehmen *Eyes, JAPAN Co. Ltd.* (MoCapData o.D.) stellt sowohl kommerziell als auch teilweise kostenfrei unter einer Creative-Commons-Lizenz verfügbare Bewegungsaufnahmen bereit, wobei insgesamt ca. 5500 Aufnahmen angeboten werden. Ähnlich dazu werden auch von dem *Motion Capture (MoCap) Club* (MoCapClub o.D.) fertige Pakete von Bewegungsaufnahmen zum Download angeboten. Beide Angebote richten sich jedoch an Anwender aus dem Bereich der Computeranimation, bspw. für die Entwicklung von Videospielen, und sind von der Auswahl der Bewegungstypen (z.B. Tanz- und Kampfbewegungen) dementsprechend ausgelegt. Zusätzlich weisen die Datensätze eine lose Strukturierung auf und es finden sich keine Informationen beispielsweise zu den Probanden, dem Ablauf der Bewegungserfassung und der Platzierung der verwendeten Marker, was eine Anwendung der Daten zu Forschungszwecken erschwert.

Weitere Datensätze menschlicher Bewegung finden sich bei der *ICS Action Database* (ICS o.D.), in welcher jeder aufgenommene Frame manuell hinsichtlich seiner Zugehörigkeit zu einer von 25 vorab definierten Aktionen annotiert wurde, und bei der nicht öffentlich verfügbaren *Korea University Gesture Database* (Hwang et al. 2006), die 54 unterschiedliche menschlichen Gesten (eingeteilt in „normale“, „anormale“ und „Kommando-Gesten“) von 20 menschlichen Probanden enthält. Der Datensatz von *Human Identification at a Distance* (HumanIdentification o.D.) enthält Bewegungsaufnahmen von 20 Probanden zu Untersuchungen für die Identifikation menschlicher Probanden anhand ihrer Gangmuster. Mörtl et al. (2012) stellen einen Datensatz bestehend aus Positions- und Kraftdaten von 18 Mensch-Roboter- und 6 Mensch-Mensch-Interaktionsaufgaben vor, bei denen kooperativ ein großer Tisch an verschiedene Positionen bewegt wurde. Die Bewegungsdatenbank der *National University of Singapore* (NUSMoCap o.D.) enthält Aufnahmen

Tabelle 4.1 : Vergleich ausgewählter existierender Bewegungsdatenbanken (modifiziert nach Mandery et al. 2016c, ©2016 IEEE).

Name der Datenbank	Beschreibung	Bewegungsarten	Format(e)	# Prob.	# Aufn.
CMU Graphics Lab Motion Capture Database (CMUMoCapDB o.D.)	Sehr bekannte und weit verbreitet eingesetzte Bewegungsdatenbank	Weites Spektrum an Fortbewegung, Interaktion mit Umgebung oder anderen Menschen und Sportbewegungen	C3D, ASF/AMC, Videos	109	2605
HDM05 (Müller et al. 2007)	Längere Aufnahmesequenzen (insgesamt ca. 3h) wurden manuell in Bewegungsclips segmentiert	Vorrangig Fortbewegung und Sportbewegungen	C3D, ASF/AMC, teilw. Videos	5	1457
Human Motion Database (Guerra-Filho et al. 2012)	Systematischer Ansatz zur Auswahl der zu erfassenden Ganzkörperbewegungen	Fünf Datensätze (siehe Abschnitt 4.1), Praxicon-Datensatz: 350 Aktionen eines Probanden für "konkrete", aus einer lexikalischen Datenbank extrahierte Verben	n/a	Fünf Datensätze (siehe Abschnitt 4.1)	
DB for Study of Gender, Identity and Emotion (Ma et al. 2006)	Auswahl an Bewegungen, welche für vier verschiedene emotionale Stimmungen gespielt wurden	Laufen, Armbewegungen und Sequenzen davon	CSM (3ds, Max Char. Studio), PTD	30	4080
HEMOCAP (Basso et al. 2008)	Zehn emotionale Stimmungen mit zusätzlichen Sprachdaten und Transkriptionen, erfasste Marker nur an Gesicht, Kopf und Händen	Festgelegte und improvisierte Szenen	n/a	10	n/a
Edinburgh CGVU Interaction Database (CGVUInteractionDatabase o.D.)	Erfassung mittels Kombination aus magnetischer Bewegungserfassung und Kinect-RGB-D-Sensor	Manipulation und Interaktion mit Objekten	FBX, BVH	n/a	n/a
HuMoD Database (Wojtusich et al. 2015)	Fokus auf Unterkörper, zusätzlich Messungen mit EMG und Kraftplatten	Laufen, Rennen, einem Hindernis ausweichen, Kniebeuge, Treten und Springen	MAT (MATLAB-Format)	2	26
HMDB51 (Kuehne et al. 2011)	Reine Video-Datenbank aus mehreren Quellen für die Evaluation von Verfahren zur Aktionserkennung	51 festgelegte Aktionen, z.B. Kauen, Klatschen, Golfspielen oder Treten	Videos	n/a	6766

Name der Datenbank	Beschreibung	Bewegungsarten	Format(e)	# Prob.	# Aufn.
CMU Motion of Body (MoBo) Database (Gross et al. 2001)	11 Sekunden lange Bewegungsclips von Laufen auf einem Laufband	Vier unterschiedliche Laufarten	Bilder (PPM/TPG)	25	100
HumanEva Dataset (Sigal et al. 2010)	Nicht-öffentlicher Datensatz für die Evaluation von Verfahren zur Posenrekonstruktion	6 Aktionen: Laufen, Joggen, Werfen/Fangen, Gestikulation, Boxen, Kombination	n/a	4	56
Human3.6M (Ionescu et al. 2014)	Kombination von hochauflösenden Videoaufnahmen mit Bewegungserfassung und Time-of-Flight-Daten für die Evaluation von Verfahren zur Posenrekonstruktion	15 verschiedene Posen/Bewegungen, z.B. Laufen, Warten, Posieren und Sitzen	n/a	11	n/a
Mocapdata.com (MoCapData o.D.)	(Teilweise) kommerzielles Angebot von Eyes, JAPAN Co. Ltd.	Bewegungen für Anwendung in der Computerranimation, z.B. Sport und Tanzen	C3D, BVH und weitere (je nach Bewegung)	n/a	5604
Korea University Gesture (KUG) Database (Hwang et al. 2006)	Ganzkörperaufnahmen menschlicher Gesten	54 verschiedene Gesten, unterteilt in „normale“, „anormale“ und „Kommando-Gesten“	HTR, Videos	20	n/a
Human Identification at a Distance Database (HumanIdentification o.D.)	Aufnahmen zur Untersuchungen für die Identifikation menschlicher Probanden anhand ihrer Gangmuster	Laufbewegungen	MB (Maya), MOT (Text), Videos	20	n/a
NUS Motion Capture Database (NUSMoCap o.D.)	Ganzkörperbewegungen der National University of Singapore	Vorrangig Fortbewegung, sowie Tanz- und Kampfsportbewegungen	C3D, BVH, FBX und weitere	8	n/a
KIT Whole-Body Human Motion Database (Mandery et al. 2015b, 2016c)	Systematisch strukturierte Aufnahmen menschlicher Ganzkörperbewegungen inklusive der beteiligten Objekte, einheitliche Form der Bewegungsrepräsentation	Weites Spektrum verschiedenster Bewegungstypen (siehe Abschnitt 4.5)	C3D, MMM, Videos, manche: Beschleunigung, Kräfte	53	6701

verschiedener Bewegungen von insgesamt acht menschlichen Probanden und die *LAAS-CNRS Novela Motion Capture Database* (Novela o.D.) beinhaltet aufgenommene Tanzbewegungen eines einzelnen Probanden.

Ergänzend zeigt Tabelle 4.1 eine Übersicht über die wichtigsten der hier erwähnten Bewegungsdatenbanken, wobei zu Vergleichszwecken am Ende der Tabelle auch die im Rahmen dieser Arbeit entwickelte *KIT Whole-Body Human Motion Database* aufgeführt ist. Für jede Bewegungsdatenbank sind neben einer allgemeinen Beschreibung die vorrangig angebotenen Bewegungsarten, eine Liste der bereitgestellten Dateiformate und die Anzahl der verfügbaren Probanden und Bewegungsaufnahmen angegeben.

4.2 Aufbau der Bewegungsdatenbank

Der folgende Abschnitt soll einen Überblick über den Aufbau der entwickelten KIT-Bewegungsdatenbank liefern. Hierzu werden in Abschnitt 4.2.1 zunächst die verschiedenen in der Bewegungsdatenbank bereitgestellten Datentypen erläutert, die zum einen durch Benutzer in das System eingepflegt und zum anderen innerhalb der Bewegungsdatenbank aus bereits vorhandenen Daten erzeugt werden. Im Anschluss beschreibt Abschnitt 4.2.2 die Softwarearchitektur der Bewegungsdatenbank hinsichtlich ihres internen Aufbaus, bspw. das Datenbankschema der unterliegenden relationalen Datenbank. Abschließend werden in Abschnitt 4.2.3 die beiden öffentlich zugänglichen Schnittstellen, über welche auf die Bewegungsdatenbank zugegriffen werden kann, zusammen mit den damit verbundenen Fragestellungen der Rechteverwaltung diskutiert.

4.2.1 Datentypen

Wie in Abschnitt 2.2 bereits erwähnt, kann eine Erfassung der menschlichen Bewegung mit Hilfe unterschiedlicher Erfassungsprinzipien und Sensormodalitäten erfolgen, wobei systembedingt Rohdaten in jeweils unterschiedli-

chen Datenformaten entstehen. Während die hochgenauen Messungen eines markerbasierten passiv-optischen Bewegungserfassungssystems (vgl. Abschnitt 2.2.2) für alle Bewegungsaufnahmen der KIT-Bewegungsdatenbank zur Verfügung stehen, kommt darüber hinaus ein multimodaler Ansatz zur Anwendung, für den die Bewegung zeitsynchron mit weiteren unabhängigen Sensorsystemen wie beispielsweise Video-/Tiefenkameras, Beschleunigungs- oder Kraftsensoren erfasst werden kann. Als Folge daraus treten zusätzlich zu den Bewegungsdaten in der vereinheitlichenden MMM-Repräsentation Daten in zahlreichen weiteren (Roh-)Formaten auf, von denen die wichtigsten im Folgenden erläutert werden sollen.

Die Grundlage für die Verwaltung verfügbarer Bewegungsdaten der im Folgenden beschriebenen Datentypen stellt ein in der KIT-Bewegungsdatenbank angelegtes *Bewegungsexperiment* dar, das dort einen Datenbankeintrag im Sinne eines klar definierten Bewegungsablaufs unter denselben experimentellen Bedingungen bezeichnet, d.h. Bewegungsaufnahmen derselben Probanden und Umgebungsobjekte innerhalb einer Aufnahmesitzung. Bewegungsexperimente in der KIT-Bewegungsdatenbank werden mit Hilfe eindeutiger IDs (welche für die in dieser Arbeit verwendeten Bewegungsdaten auch an entsprechender Stelle angegeben werden, bspw. in Anhang A) sowie dem in Abschnitt 4.3 beschriebenen Klassifikationsschema auf Basis des *Motion Description Tree* gekennzeichnet. Jedes Bewegungsexperiment enthält in der Bewegungsdatenbank eine beliebige Anzahl von Dateien der im Folgenden erläuterten Datentypen für alle der, in der Regel drei bis zehn, mit diesem Bewegungsexperiment assoziierten Bewegungsaufnahmen.

Eine Übersicht über wichtige in der KIT-Bewegungsdatenbank verfügbare Datenformate und den Datenfluss zwischen ihnen ist in Abbildung 4.1 dargestellt. Zu unterscheiden ist zwischen Rohdaten, welche als Primärdaten durch einen Benutzer der Bewegungsdatenbank hochgeladen werden, und abgeleiteten Sekundärdaten, die durch automatische oder manuelle Verarbeitungsschritte innerhalb der Bewegungsdatenbank aus den hochgeladenen Primärdaten erzeugt werden. Zu den Primärdaten gehören vor allem

die aufgezeichneten Daten des markerbasierten Bewegungserfassungssystems, ergänzende Videoaufnahmen der Bewegungen und Daten weiterer Sensormodalitäten. Erstere werden in dem in Abschnitt 2.2.1 beschriebenen standardisierten C3D-Format bereitgestellt, welches gelabelte Markertrajektorien für die Marker des menschlichen Probanden und auch für die an Umgebungsobjekten angebrachten Marker enthält. Ergänzende Videoaufnahmen der Bewegung können im AVI-Format hochgeladen werden, wobei die Bewegungsdatenbank selbst (für die Erstellung von Vorschauvideos und für die Anonymisierung) durch Verwendung einer entsprechenden Softwarebibliothek ein breites Repertoire an Videocodecs unterstützt². Weitere Rohdaten zusätzlicher Sensoren wie beispielsweise Beschleunigungs- und Drehratenmessungen einer inertialen Messeinheit oder Daten von Kraftsensoren werden in einem sensorspezifischen Datenformat hochgeladen, bei den am KIT erfassten Sensordaten meist ein CSV-basiertes (*comma separated values*) Format.

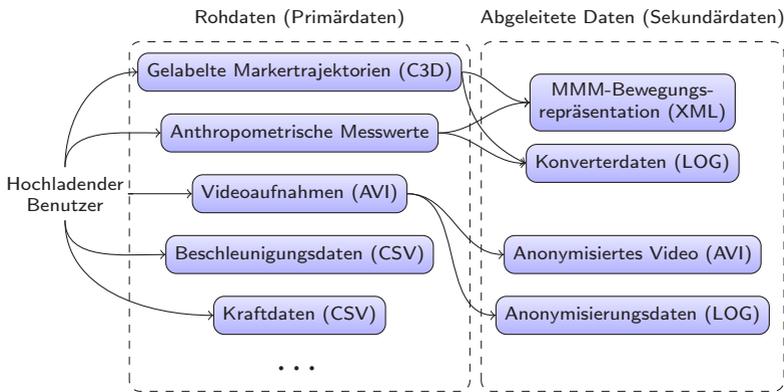


Abbildung 4.1: Wichtige in der KIT-Bewegungsdatenbank verfügbare Datentypen und der Datenfluss zwischen ihnen.

² Für das am H2T aufgenommene Videomaterial wird der MPEG4-v2-Codec verwendet und bewegungsspezifisch in den meisten Fällen die Audiospur entfernt.

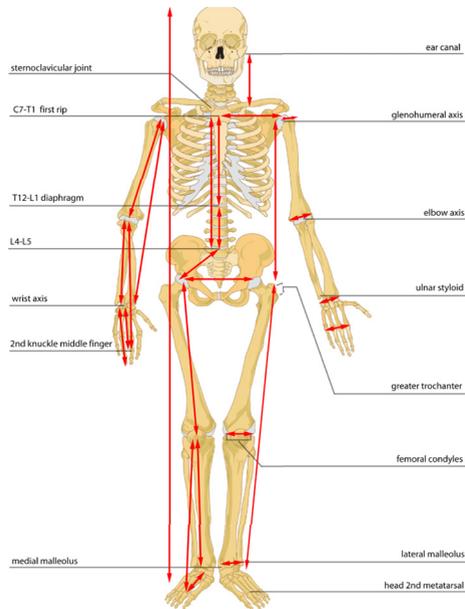


Abbildung 4.2: Satz anthropometrischer Messungen, der von menschlichen Probanden in der KIT-Bewegungsdatenbank erhoben wird (entnommen aus Mandery et al. 2015b, ©2015 IEEE).

Zusätzlich zu den hochgeladenen sensorisch erfassten Primärdaten werden durch den Benutzer Metadaten über ein aufgezeichnetes Bewegungsexperiment, den oder die menschlichen Probanden und über verwendete Umgebungsobjekte eingepflegt. Für ein Bewegungsexperiment beinhaltet dies eine Klassifikation der Bewegung gemäß dem in Abschnitt 4.3 beschriebenen Klassifikationsschema, welche ggf. auch durch die in Abschnitt 4.4 beschriebene Generierung von Tagging-Vorschlägen unterstützt werden kann. Für menschliche Probanden werden Name (nicht öffentlich sichtbar), Geschlecht, Alter, Körpergröße, Gewicht sowie Messungen von 25 anthropometrischen Parametern in der Datenbank erfasst. Eine Darstellung der anthropometrischen Parameter, die von Probanden manuell mit Hilfe eines Maßbands bestimmt werden, findet sich in Abbildung 4.2 und eine genauere Beschreibung

dieser zusammen mit den jeweiligen Anweisungen zu ihrer Bestimmung ist online³ abrufbar. Mögliche Anwendungsfelder dieser anthropometrischen Parameter finden sich bei ihrem Einsatz zur biomechanisch motivierten Analyse der menschlichen Bewegung und bei der Parametrisierung von Modellen des menschlichen Körpers.

Wie eingangs erwähnt, werden aus den durch Benutzer in die Bewegungsdatenbank hochgeladenen Primärdaten innerhalb der Datenbank weitere Sekundärdaten abgeleitet. Hierbei handelt es sich zum einen um eine normalisierte Darstellung der Bewegung, welche auf dem MMM-Framework (vgl. Abschnitt 2.1) basiert und mit dem in Abschnitt 2.2.2 vorgestellten Verfahren zur Rekonstruktion von Gelenkwinkeln aus gelabelten Markertrajektorien einer Bewegung und aus Körpergröße/-gewicht des Probanden zur Parametrisierung des MMM-Modells erzeugt wird. Diese normalisierte Repräsentation der Bewegung im XML-basierten MMM-Format, die zusätzlich zur Bewegung des menschlichen Probanden auch die Trajektorien der an der Aufnahme beteiligten Objekte enthält, stellt die Grundlage für die in Kapitel 5 und Kapitel 6 vorgestellten Anwendungen menschlicher Ganzkörperbewegung dar. Nach Abschluss der Konvertierung einer aufgezeichneten Bewegung in die MMM-Repräsentation wird zusätzlich zur Bewegung im MMM-Format auch eine *Konverter-Logdatei* geschrieben, welche das Ziel verfolgt, die Verifizier- und Reproduzierbarkeit des Konvertierungsprozesses sicherzustellen. Diese Logdatei enthält die verwendeten Versionen (Commit-Hashes der unterliegenden Versionsverwaltung) der am Konvertierungsprozess beteiligten Software-Pakete MMCore, MMTools sowie Simox, die Hashes der zum verwendeten MMM-Modell gehörenden Dateien und Informationen über die in der Bewegung enthaltenen Umgebungsobjekte. Neben ihrem informativen Wert für den Benutzer wird diese Logdatei insbesondere auch intern innerhalb der Bewegungsdatenbank verwendet, um bei Änderun-

³ https://motion-database.humanoids.kit.edu/anthropometric_table/

gen an Software- oder Modell-Komponenten eine Regenerierung betroffener Sekundärdaten aus den ursprünglichen Primärdaten veranlassen zu können.



Abbildung 4.3: Erstellung von anonymisierten Fassungen hochgeladener Videoaufnahmen mit Hilfe eines Anonymisierungswerkzeugs durch menschliche Unterstützung.

Zum anderen werden als zweite wichtige Gruppe von Sekundärdaten in der Bewegungsdatenbank anonymisierte Fassungen von hochgeladenen Videoaufnahmen erzeugt. Dieser Schritt ist erforderlich, da Videoaufnahmen, welche eine Identifikation menschlicher Probanden erlauben, aus naheliegenden Gründen nicht der Öffentlichkeit zugänglich gemacht werden sollen und dürfen, die Verfügbarkeit einer zur Bewegung zugehörigen Videoaufnahme jedoch ein wünschenswertes Kriterium ist, um über in der Oberfläche der Bewegungsdatenbank integrierte Vorschauvideos eine einfache Auswahl von Bewegungen zu ermöglichen. Die Anonymisierung von Videoaufnahmen erfolgt durch menschliche Unterstützung, indem ein Benutzer in einem speziell hierfür entwickelten Werkzeug die Einzelbilder des nicht anonymisierten Ausgangsvideos inspiziert und in selbst gewählten Schlüsselbildern die Gesichtsposition des menschlichen Probanden markiert⁴. Ein Beispiel hierfür ist in Abbildung 4.3 gezeigt. Im Anschluss wird durch das Anonymisierungswerkzeug eine anonyme Fassung der Videoaufnahme innerhalb der Bewegungsdatenbank erstellt, in welcher der Bereich um die markierten Gesichtspositionen mit einem Tiefpassfilter weichgezeichnet wird (für

⁴ Lösungen, welche bspw. basierend auf Algorithmen zur Gesichtserkennung ohne menschliche Unterstützung arbeiten könnten, verbieten sich an dieser Stelle alleine aufgrund des möglichen Schadens durch unvollständig anonymisierte Videoaufnahmen, die in der Bewegungsdatenbank frei verfügbar vorgehalten würden.

Einzelbilder zwischen zwei manuell annotierten Schlüsselbildern wird die Gesichtsposition linear interpoliert). Analog zur Erzeugung normalisierter Bewegungsdaten wird auch hier eine *Anonymisierungs-Logdatei* geschrieben, welche die manuell annotierten Gesichtspositionen für eine potentiell später erforderliche Regenerierung der anonymisierten Videodatei speichert.

4.2.2 Softwarearchitektur

An die Beschreibung der grundsätzlich in der KIT-Bewegungsdatenbank zu verwaltenden Datentypen soll sich in diesem Abschnitt nun eine Beschreibung des internen Aufbaus ihrer Implementierung anhand softwarearchitektonischer Kriterien anschließen. Zur Entwicklung der Bewegungsdatenbank kam die Programmiersprache Python (Python o.D.) in Version 2.7 zusammen mit dem Django-Framework (Django o.D.) in Version 1.10 zum Einsatz. Bei Python handelt es sich um eine dynamisch typisierte höhere Programmiersprache, die üblicherweise durch einen Interpreter ausgeführt wird und als eine der gegenwärtig populärsten Programmiersprachen mittlerweile ein sehr großes Ökosystem an verfügbaren Softwarebibliotheken und Frameworks aufweisen kann. Django stellt ein quelloffenes Framework für Python dar, welches insbesondere für die Entwicklung von datenbankgetriebenen Webanwendungen eine große Popularität erfährt. Es basiert auf dem Architektur- bzw. Entwurfsmuster des *Model View Controller* (MVC), das eine Unterteilung der Architektur in die drei Bestandteile Modell (Repräsentation der Daten), Präsentation (Darstellung der Daten für den Benutzer) und Steuerung (Annahme und Verarbeitung von Benutzeraktionen) vornimmt. Zusätzlich zu Django kamen bei der Entwicklung der KIT-Bewegungsdatenbank zahlreiche weitere Python-Softwarebibliotheken zum Einsatz, von denen hier mit der *Django Debug Toolbar*, *Django MPTT*, *Pillow*, *Psycopy2*, *PyParsing* und *NumPy* nur die wichtigsten erwähnt werden sollen.

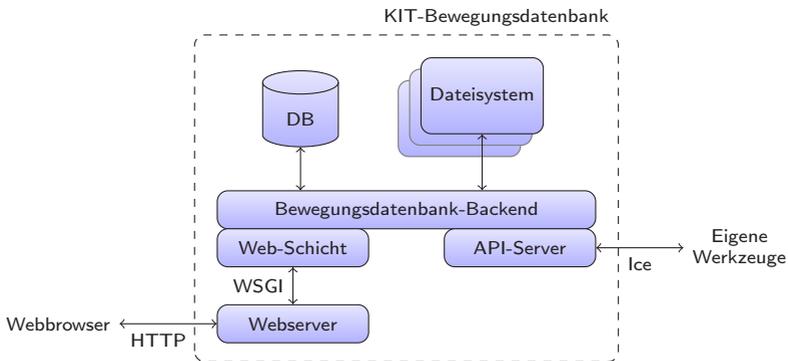


Abbildung 4.4: Stark abstrahierter Aufbau der KIT-Bewegungsdatenbank und bestehende Interaktionsmöglichkeiten mit dem Benutzer (modifiziert nach Mandery et al. 2015b, ©2015 IEEE).

Abbildung 4.4 zeigt den grundsätzlichen Aufbau der KIT-Bewegungsdatenbank und die Interaktionsmöglichkeiten mit dem Benutzer auf einer stark abstrahierten Ebene. Im Bewegungsdatenbank-Backend wird der in der Bewegungsdatenbank persistent gespeicherte Datenbestand verwaltet, womit das Backend gänzlich in die Modellkomponente des MVC-Musters einzuordnen ist. Zur Verwaltung der Daten kommt mit PostgreSQL (PostgreSQL o.D.) eine quelloffene relationale Datenbank zum Einsatz, welche unter Verwendung des im Django-Framework enthaltenen objektrelationalen Mappers abstrahiert wird, der das relationale Datenbankschema basierend auf der weiter unten diskutierten Klassenstruktur erzeugt. In dieser relationalen Datenbank werden die in Abschnitt 4.2.1 angesprochenen Nutzdaten verwaltet, wobei die tatsächliche Speicherung dieser Daten aus Performanzgründen außerhalb der relationalen Datenbank im Dateisystem erfolgt⁵. Auf dem Bewegungsdatenbank-Backend bauen mit dem API-Server und der Web-Schicht zwei unterschiedliche Komponenten auf, die mit den bei-

⁵ Um dem Problem möglicher Inkonsistenzen zwischen relationaler Datenbank und Dateisystem zu begegnen, wurden mehrere Hilfsanwendungen entwickelt, die als *Cronjob* periodisch die Konsistenz des Datenbestands überprüfen und ggf. wiederherstellen können.

den in Abschnitt 4.2.3 vorgestellten Schnittstellen korrespondieren und als Präsentations- und Steuerungskomponenten im Sinne des MVC-Musters einzuordnen sind. Über den API-Server ist mit einer Programmierschnittstelle ein direkter Zugriff auf den Datenbestand der Bewegungsdatenbank möglich. Die Web-Schicht hingegen dient der Implementierung einer webbasierten Oberfläche für die Bewegungsdatenbank und definiert basierend auf dem in Django 1.3 eingeführten Feature klassen-basierter Views sogenannte View-Klassen, wie etwa eine *ContentEditView*-Klasse zur Anzeige und Verarbeitung eines Webformulars zum Bearbeiten von Datenbankobjekten. Eine ebenfalls in der Web-Schicht definierte URL-Konfiguration beschreibt die Abbildung von URL-Mustern, die als reguläre Ausdrücke spezifiziert werden, auf diese View-Klassen⁶. Zur Generierung von Antworten auf Benutzeranfragen enthält die Web-Schicht des Weiteren 38 HTML-Templates (bspw. *list_motions.html* und *edit_files.html*), die durch die View-Klassen mit Hilfe des Template-Mechanismus von Django kontextspezifisch parametrisiert werden. Der Zugriff auf die Web-Schicht ist über einen Webserver möglich, welcher vom Webbrowser eines Benutzers gestellte HTTP-Anfragen beantwortet und hierzu unter Verwendung des *Web Server Gateway Interface* (WSGI) mit der Web-Schicht kommuniziert (siehe Abbildung 4.4). Als Webserver kommt in der KIT-Bewegungsdatenbank Apache 2.4 zum Einsatz, wobei dank der Standardisierung von WSGI ein Wechsel zu einer alternativen Serversoftware ohne Weiteres möglich wäre.

Wie zuvor erwähnt, wird zur Verwaltung des Datenbestands der im Django-Framework enthaltene objektrelationale Mapper zur Abbildung von Klassen in der Modell-Komponente auf das Schema einer unterliegenden relationalen Datenbank eingesetzt. Die wichtigsten dieser Klassen und ihre Attribute und Beziehungen sollen im Folgenden in einer vereinfachten Darstellung kurz beschrieben werden.

⁶ Beispielsweise kommt die zuvor erwähnte *ContentEditView*-Klasse zur Beantwortung von Anfragen auf URLs zum Einsatz, welche auf den folgenden regulären Ausdruck passen:

```
edit/(?P<table_name>motions|subjects|objects|projects)/(?P<object_id>\d+)/
```

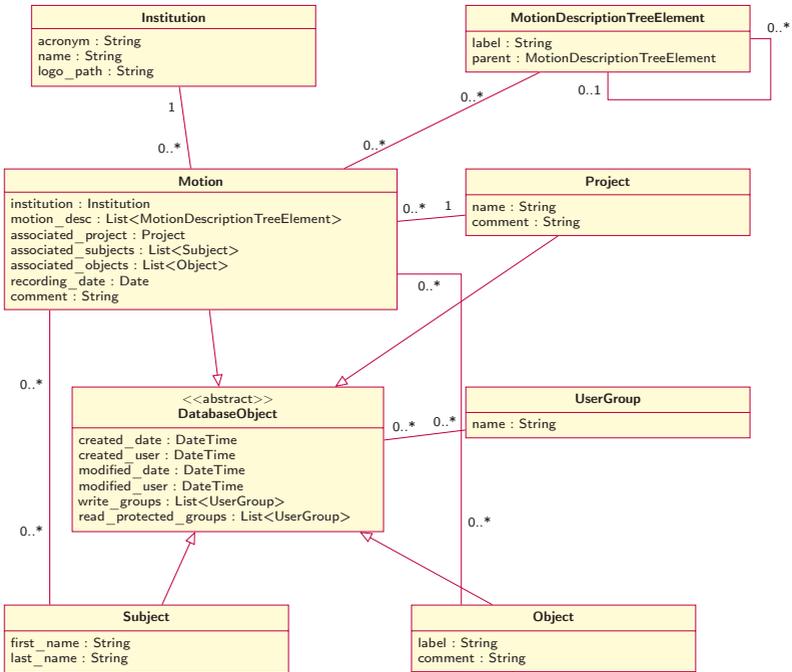


Abbildung 4.5: UML-Klassendiagramm für die an der Modellierung von in der Datenbank verfügbaren Datenbankobjekten beteiligten Klassen und ihren Beziehungen.

Abbildung 4.5 zeigt ein UML-Klassendiagramm⁷ für die an der Modellierung von in der Datenbank verfügbaren Datenbankobjekten beteiligten Klassen. Datenbankobjekte werden hierbei durch die vier Klassen *Motion* (Bewegungsexperiment), *Subject* (Proband), *Object* (Umgebungsobjekt) und *Project* (Forschungsprojekt) modelliert, die als eigenständige Entitäten innerhalb der Datenbank, im Unterschied zu anderen Klassen beispielsweise für hochgeladene Dateien, unabhängig vom restlichen Datenbestand unmittelbar durch die Benutzer bearbeitet werden können. Als gemeinsame Basisklasse für diese vier Typen von Datenbankobjekten ist *DatabaseObject* angelegt,

⁷ Siehe beispielsweise (Rumbaugh et al. 2004) für eine Beschreibung der Notation.

welche die für alle Datenbankobjekte verfügbaren Attribute enthält⁸. Hierbei handelt es sich zum einen um den Zeitpunkt und den assoziierten Benutzer der Erstellung und der letzten Änderung des Datenbankobjekts. Zum anderen werden Listen von Benutzergruppen für zwei unterschiedliche Rechtstufen, *Write Groups* und *Read Protected Groups*, gespeichert, welche für die Verwaltung von Zugriffsrechten von Bedeutung sind, die in Abschnitt 4.2.3 näher beschrieben wird. Abgeleitet von *DatabaseObject* repräsentiert *Subject* einen in der Bewegungsdatenbank erfassten menschlichen Probanden, wobei in Abbildung 4.5 aus Gründen der Übersichtlichkeit zahlreiche Attribute (beispielsweise die anthropometrischen Messwerte) nicht dargestellt werden. Die *Object*-Klasse enthält Informationen über in Bewegungsaufnahmen verwendete und mit Markern versehene Umgebungsobjekte (vgl. Abschnitt 2.2.2) und erlaubt beispielsweise die Verwaltung der für die Bewegungsrekonstruktion notwendigen 3D-Modelle dieser Umgebungsobjekte. Schlussendlich repräsentiert eine *Motion*-Instanz ein in der Datenbank verfügbares Bewegungsexperiment, welches wie zuvor erläutert aus einer beliebigen Anzahl an Demonstrationen für denselben Bewegungsablauf bestehen kann. Als solches ist eine *Motion*-Instanz mit genau einem Forschungsprojekt (*Project*-Klasse) und einer beliebigen Anzahl an Probanden und Umgebungsobjekten assoziiert. Des Weiteren verfügt jede *Motion*-Instanz über eine Angabe des Forschungsinstituts, an welchem das Bewegungsexperiment durchgeführt wurde, da in der KIT-Bewegungsdatenbank auch Daten externer Forschungsinstitute enthalten sind (mehr dazu in Abschnitt 4.5). Abschließend ist eine *Motion*-Instanz als Bewegungsexperiment mit einer beliebigen Anzahl an *MotionDescriptionTreeElement*-Objekten assoziiert, welche die Grundlage des in Abschnitt 4.3 beschriebenen *Motion Description Tree* zur Klassifikation menschlicher Bewegungen bilden.

⁸ Die Vererbungsrelation wird durch den objektrelationalen Mapper im relationalen Datenbankschema mit Hilfe einer separaten Datenbanktabelle für die Basisklasse modelliert, welche in den Datenbanktabellen der erbbenden Klassen mit einem Fremdschlüssel (*foreign key*) mit Einzigartigkeitsbeschränkung (*unique constraint*) referenziert wird.

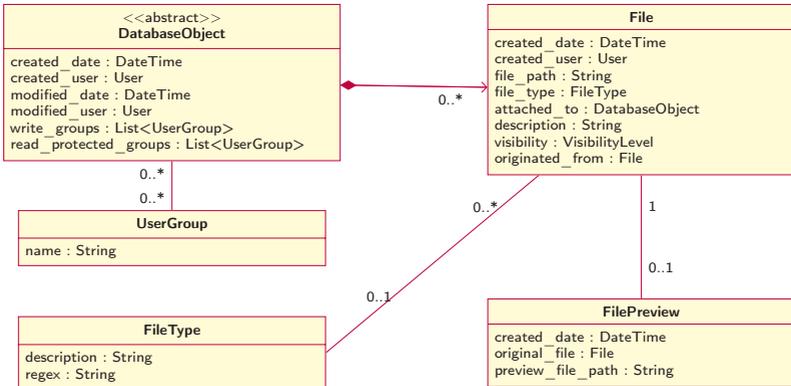


Abbildung 4.6: UML-Klassendiagramm für die an der Modellierung von hochgeladenen Dateien beteiligten Klassen und ihren Beziehungen.

In Abbildung 4.6 ist ein UML-Klassendiagramm für die an der Verwaltung hochgeladener Dateien beteiligten Klassen dargestellt. Eine hochgeladene Datei wird durch eine *File*-Instanz repräsentiert, die fest einem Datenbankobjekt, also Bewegungsexperiment, Proband, Umgebungsobjekt oder Projekt, zugeordnet ist, wohingegen für die Zahl der einem Datenbankobjekt zugeordneten Dateien keine Beschränkung besteht. Das *file_path*-Attribut der *File*-Klasse stellt die eingangs und in Abbildung 4.4 dargestellte Verknüpfung des Datenbankschemas mit dem Dateisystem dar, in welchem die eigentlichen Nutzdaten der gespeicherten Dateien abgelegt sind. Zur Organisation hochgeladener Dateien wird in der Bewegungsdatenbank manuell eine Liste möglicher Dateitypen gepflegt, die zum einen mit Hilfe regulärer Ausdrücke für den Dateinamen (beispielsweise zum Erkennen von Dateierendungen) automatisch erkannt, zum anderen aber auch manuell durch den Nutzer zugewiesen werden können. Diese Zuordnung von Dateien zu Dateitypen ermöglicht die Realisierung einer effizienten Suche nach bestimmten Dateitypen im Datenbestand. Über zusätzliche Attribute der *File*-Klasse kann eine Freitextbeschreibung (*description*-Attribut) für die hochgeladene Datei verwaltet und eine Sichtbarkeitsstufe für die Datei festgelegt werden, welche

für die in Abschnitt 4.2.3 beschriebene Verwaltung von Zugriffsrechten von Bedeutung ist. Außerdem besteht mit dem *originated_from*-Attribut die Möglichkeit, bei Sekundärdaten wie beispielsweise der MMM-Repräsentation einer Bewegung die ihrer Berechnung zugrunde liegende Ursprungsdatei zu referenzieren, um auf diese Weise den Verarbeitungsprozess reproduzierbar zu dokumentieren und bei Änderung der Ursprungsdatei eine Regenerierung der Sekundärdaten veranlassen zu können. Abschließend werden in der Bewegungsdatenbank automatisch größenoptimierte Vorschau-Ansichten hochgeladener Bild- und Videodateien erstellt, welche etwa den Benutzern der Weboberfläche ein Überblicken des Datenbestands erleichtern sollen. Erstellte Vorschau-Ansichten werden als *FilePreview*-Instanz repräsentiert, wobei jeder Datei höchstens eine Vorschau-Ansicht zugeordnet sein kann.

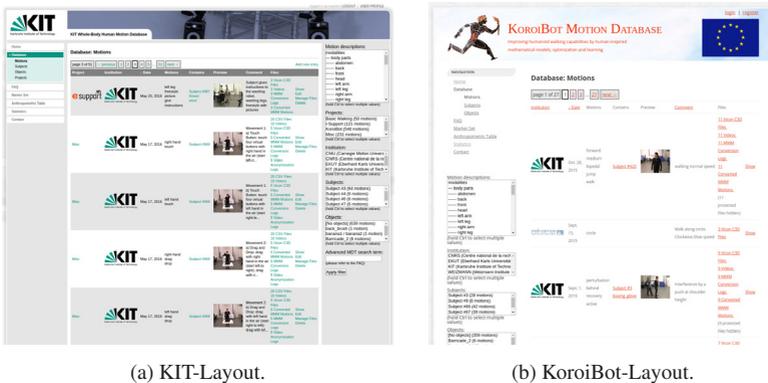
Zur Qualitätssicherung und Identifikation möglicher Probleme während der Entwicklung der KIT-Bewegungsdatenbank wurden Softwaretests angelegt. Die Implementierung dieser Tests basiert auf dem meist als *PyUnit* bezeichneten Standard-Framework der Programmiersprache Python für Komponententests, wobei nicht alle der implementierten Testfälle der Definition nach als Komponententests angesehen werden können, sondern zu einem großen Teil auch Testfälle implementiert wurden, die als Integrationstests das Zusammenspiel einzelner Komponenten validieren. Zur Durchführung der Tests auf einem bekannten und reproduzierbaren Datenbestand wurde eine *Test Fixture* erstellt, welche aus einem Datenbankdump im JSON-Format zusammen mit einem dazu konsistenten Datenbestand im Dateisystem besteht und vor Ausführung der Tests durch das Test-Framework automatisch geladen wird. Insgesamt wurde eine Gesamtzahl von 75 Testfällen implementiert, die sich als Integrationstests mit der Funktionalität sowohl der View-Klassen der Webschicht als auch des API-Servers befassen und zusätzlich direkt spezifische Funktionalitäten des Bewegungsdatenbank-Backends testen, beispielsweise die in Abschnitt 4.2.3 angesprochene Verwaltung von Zugriffsrechten sowie den Parser für die in Abschnitt 4.3 angesprochenen Suchausdrücke auf Basis des *Motion Description Tree*.

4.2.3 Schnittstellen

Wie im vorhergehenden Abschnitt erwähnt, werden Benutzern mit der webbasierten Oberfläche und der Programmierschnittstelle zwei verschiedene Schnittstellen für den Zugriff auf die KIT-Bewegungsdatenbank angeboten. Im Folgenden sollen diese Benutzerschnittstellen näher erläutert werden.

Die webbasierte Oberfläche ermöglicht, wie auch bei einigen anderen der in Abschnitt 4.1 beschriebenen Bewegungsdatenbanken, einen unkomplizierten Zugang zu den Inhalten der KIT-Bewegungsdatenbank ohne die Notwendigkeit der Installation von Client-Software. Eine Listen-Ansicht erlaubt das Durchsuchen des Datenbestands, wobei Bewegungsexperimente nach verschiedenen Kriterien wie der in Abschnitt 4.3 beschriebenen Bewegungsklassifikation, den assoziierten Forschungsprojekten, dem aufnehmenden Institut, den aufgenommenen Probanden sowie Objekten und auch mit Hilfe einer Freitextsuche auf Basis logischer Operatoren (siehe Abschnitt 4.3) gefiltert werden können. In einer Detail-Ansicht können anschließend die verfügbaren Informationen zu einem spezifischen Datenbankobjekt angezeigt, sowie die assoziierten Dateien einzeln oder als ZIP-Archiv heruntergeladen werden. Zusätzliche Unterseiten beinhalten statistische Informationen zum Inhalt der Bewegungsdatenbank, eine Dokumentation zur Definition des KIT Reference Marker Set (siehe Abschnitt 2.2.1) sowie der anthropometrischen Messungen (siehe Abschnitt 4.2.1) und Antworten auf „häufig gestellte Fragen“ (FAQ). Während ein Durchsuchen des öffentlich verfügbaren Datenbestands für anonyme Benutzer möglich ist, erfordert das Herunterladen gespeicherter Dateien die Registrierung eines Benutzerkontos, die direkt in der Weboberfläche möglich ist. Nach der weiter unten erläuterten Verwaltung von Zugriffsrechten befugte Benutzer können über die Weboberfläche neben lesenden Zugriffen auch Schreibvorgänge zur Verwaltung des Datenbestands durchführen, was bislang bei keiner der anderen in Abschnitt 4.1 beschriebenen existierenden Bewegungsdatenbanken mit webbasierter Benutzerschnittstelle realisiert wurde. Konkret ist das Anlegen und Löschen von

Datenbankobjekten (bspw. Bewegungsexperimente oder Probanden), das Ändern ihrer Metadaten (bspw. Aufnahme datum eines Bewegungsexperiments) und das Hochladen, Verwalten und Löschen dazugehöriger Dateien vorgesehen. Bei der Implementierung der webbasierten Oberfläche wurde wie zuvor beschrieben durch Verwendung des Django-Template-Mechanismus (Django o.D.) eine strikte Entkopplung der Webseitendarstellung von den dahinterliegenden Datenbankmodellen und der Programmlogik realisiert. So war es möglich, zwei unterschiedliche Gestaltungsvarianten für die KIT-Bewegungsdatenbank umzusetzen, wobei neben der Standard-Variante eine „Koroibot-Variante“ entworfen wurde, die unter einer separaten URL⁹ eine Teilmenge des Datenbestands in einer an das EU-Forschungsprojekt „Koroibot“ angepassten Layoutgestaltung zur Verfügung stellt. In Abbildung 4.7 ist mit der Listen-Ansicht verfügbarer Bewegungsexperimente eine der zentralen Seiten der webbasierten Oberfläche in beiden Gestaltungsvarianten dargestellt.



(a) KIT-Layout.

(b) Koroibot-Layout.

Abbildung 4.7: Webbasierte Oberfläche der KIT-Bewegungsdatenbank in den zwei unterschiedlichen Gestaltungsvarianten. Dargestellt ist die Listenansicht, welche eine Filterung und Sortierung von in der Datenbank verfügbaren Bewegungsexperimenten erlaubt.

⁹ <https://koroibot-motion-database.humanoids.kit.edu/>

```

idempotent File getFile(long fileId) throws
    InternalErrorException, InvalidParameterException;
idempotent FileList listFiles(long databaseObjectId) throws
    InternalErrorException, InvalidParameterException;
idempotent FileReader* getFileReader(long fileId) throws
    InternalErrorException, InvalidParameterException,
    NotAuthorizedException, TooManyOpenFilesException;

FileWriter* getFileWriter(long databaseObjectId, string
    fileName, string fileType, string description,
    VisibilityLevel visibility, optional(1) long
    originatedFromId) throws InternalErrorException,
    InvalidParameterException, NotAuthorizedException,
    TooManyOpenFilesException;
void deleteFile(long fileId) throws InternalErrorException,
    InvalidParameterException, NotAuthorizedException;

```

Abbildung 4.8: Exemplarischer Auszug aus den Slice-Definitionen zur Festlegung von Datenstrukturen und Methoden für die Programmierschnittstelle der KIT-Bewegungsdatenbank.

Als zweite Benutzerschnittstelle für den Zugriff auf die KIT-Bewegungsdatenbank steht eine Programmierschnittstelle (*application programming interface*, API) bereit, die auf dem Middleware-Framework Ice (Ice o.D.) aufbaut und das Durchsuchen des Datenbestands sowie einen umfassenden Lese-/Schreibzugriff ermöglicht. Die Bereitstellung einer derartigen Programmierschnittstelle repräsentiert eine Neuerung gegenüber existierenden Bewegungsdatenbanken, die eine derartige Schnittstelle nicht anbieten (und diese zum Teil auch gar nicht benötigen, da der geringe Umfang des jeweiligen Datenbestands im Gegensatz zur KIT-Bewegungsdatenbank ein vollständiges Herunterladen des gesamten Datenbestands zu dessen lokaler Verwendung erlaubt). Bei Ice handelt es sich um ein objektorientiertes Framework basierend auf *Remote Procedure Call* (RPC) zur Entwicklung von verteilten Anwendungen, welches ohne weitere Anpassungen am Code der KIT-Bewegungsdatenbank eine breite Unterstützung gängiger Plattformen und zahlreicher Programmiersprachen, beispielsweise C++, C#, Java, Python und Ruby, ermöglicht. Zur Verwendung von Ice ist es erforderlich,

dass Definitionen für die verwendeten Datenstrukturen und Methoden der Programmierschnittstelle in einem speziellen Format, genannt Slice (*Specification Language for Ice*), abgelegt werden, von dem ausgehend durch das Ice-Framework plattform-/sprachspezifische Anbindungen (*bindings*) generiert werden. Abbildung 4.8 zeigt einen exemplarischen Auszug aus diesen Slice-Definitionen für die KIT-Bewegungsdatenbank. Der Einsatzzweck der Programmierschnittstelle liegt in der Entwicklung von Anwendungen etwa zur Bewegungsanalyse, welche relevante Daten der KIT-Bewegungsdatenbank automatisiert identifizieren und maschinell verarbeiten sollen. Als Beispiel sei genannt, dass die Werkzeuge zur Erstellung der in Abschnitt 4.2.1 angesprochenen Sekundärdaten wie das in Abbildung 4.3 dargestellte Anonymisierungswerkzeug aus Gründen der Modularität nicht in den Code der KIT-Bewegungsdatenbank integriert wurden, sondern als externe Programme lediglich über die Programmierschnittstelle auf ihren Datenbestand zugreifen.

Eine zentrale Problemstellung bei der Konzeption der beiden beschriebenen Benutzerschnittstellen befasst sich mit der Frage, wie in einer großskaligen und projektübergreifenden Bewegungsdatenbank eine Verwaltung von Zugriffsrechten realisiert werden kann. Wie zuvor bereits erwähnt, erlaubt die KIT-Bewegungsdatenbank Benutzern, sich selbstständig ein Benutzerkonto zu registrieren. Ein solches Benutzerkonto stellt die Voraussetzung für das Herunterladen von Dateien, für den Zugriff auf geschützte Informationen nach Freischaltung und für Schreibzugriff auf die Bewegungsdatenbank dar. Zweifelsohne würde es ebenso wenig eine praktikable Lösung darstellen, jedem Benutzer uneingeschränkten Lese-/Schreibzugriff auf sämtliche verfügbaren Daten zu gewähren, wie individuellen Benutzern diesen Zugriff auf einzelne spezifische Einträge der Bewegungsdatenbank. Deshalb werden registrierte Benutzer manuell in eine beliebige Zahl von Benutzergruppen eingeordnet, die mit der Zugehörigkeit des Benutzers zu Forschungsinstituten oder -projekten korrespondieren. Für Datenbankobjekte, wie bspw. ein Bewegungsexperiment oder einen Probanden, wird dann durch den erstellenden Benutzer jeweils eine Auswahl der verfügbaren Gruppen als *Write*

Groups und *Read Protected Groups* festgelegt (vgl. Abschnitt 4.2.2), welche den Zugriff auf dieses Datenbankobjekt regeln: Nur Benutzer, die einer der *Read Protected Groups* angehören, können als geschützt markierte Dateien, zum Beispiel nicht anonymisierte Videos, einsehen und herunterladen (nicht geschützte Dateien sind hingegen öffentlich sichtbar und können durch jeden beliebigen registrierten Benutzer heruntergeladen werden). Benutzer, welche einer der *Write Groups* angehören, können zusätzlich das jeweilige Datenbankobjekt und alle dazugehörigen Dateien bearbeiten. Darüber hinaus erhalten auch der erstellende Benutzer und Administratorkonten uneingeschränkten Lese- und Schreibzugriff auf das jeweilige Datenbankobjekt.

4.3 Organisation von Bewegungsaufnahmen mit dem Motion Description Tree

Mit dem Aufbau einer großskaligen Bewegungsdatenbank eng verbunden ist die Fragestellung, wie Benutzer den verfügbaren Datenbestand effizient durchsuchen und für sie relevante Datensätze identifizieren können, d.h. auf welche Weise verfügbare Bewegungsaufnahmen organisiert und systematische Suchanfragen auf Basis dieser Organisation ermöglicht werden können. Angesichts der enormen Vielfalt an Ganzkörperbewegungen, die Menschen auszuführen im Stande sind, ist dies – selbst ohne Berücksichtigung von Interaktionen des Menschen mit seiner Umgebung – ein sehr schwieriges Problem, für welches zwar Ansätze, etwa von Riaz et al. (2016), Müller et al. (2006) und T. Yu et al. (2005), bislang jedoch keine umfassenden Lösungen bekannt sind. Bei denjenigen der in Abschnitt 4.1 genannten existierenden Bewegungsdatenbanken, welche wie die KIT-Bewegungsdatenbank auf eine umfassende Sammlung menschlicher Ganzkörperbewegungen abzielen (beispielsweise die CMU-Bewegungsdatenbank), wird dieses Problem dementsprechend auch nur unbefriedigend gelöst, indem Bewegungen grob in eine geringe Zahl vorgegebener Klassen unterteilt und ggf. mit einem Freitextfeld näher beschrieben werden. Der grundsätzlich hohen Mächtigkeit

einer Freitextbeschreibung stehen in diesem Fall die Inkonsistenzen, die bei der Klassifikation von Bewegungen durch unterschiedliche Benutzer zwangsläufig entstehen, und die Schwierigkeit einer systematischen Suche durch Mehrdeutigkeiten und Synonyme der natürlichen Sprache gegenüber. Bei den verbleibenden Bewegungsdatenbanken aus Abschnitt 4.1 ohne Fokus auf umfassende Ganzkörperbewegung tritt die Fragestellung der Datenorganisation hingegen in den Hintergrund, da bei der Datensammlung die erfassten Bewegungstypen bereits so weit verfeinert oder eingeschränkt wurden, dass weit spezifischere Organisationsschemata zum Einsatz kommen können, die nicht mehr dem Anspruch einer Beschreibung beliebiger menschlicher Ganzkörperbewegungen genügen müssen.

Für die Organisation der in der KIT-Bewegungsdatenbank erfassten Bewegungsexperimente wird ein Ansatz der hierarchischen Klassifikation verwendet, welcher auf in einer Baumstruktur, dem sogenannten *Motion Description Tree*, angeordneten *Motion Tags* beruht. Abbildung 4.9 zeigt einen Auszug aus der Struktur dieses Motion Description Tree, wobei der vollständige in der KIT-Bewegungsdatenbank realisierte Baum aus insgesamt 146 Knoten als Motion Tags besteht. Die Motion Tags beschreiben nicht nur den Typ einer menschlichen Ganzkörperbewegung, sondern auch Modalitäten der Bewegungsausführung wie beispielsweise die (subjektive) Geschwindigkeit, die Richtung bei gerichteten Bewegungen oder die ausführende Hand bei einhändigen Bewegungen. Wie zuvor beschrieben, findet die Klassifikation mit diesen Motion Tags hierbei auf Ebene der Bewegungsexperimente statt und zutreffende Tags werden analog zu existierenden Bewegungsdatenbanken zunächst manuell festgelegt. Ein wichtiger Unterschied im Vergleich zu klassischen Kategorierungsansätzen wie beispielsweise bei der CMU-Bewegungsdatenbank besteht allerdings darin, dass eine Bewegung nicht nur einer einzelnen Kategorie, sondern einer beliebigen Anzahl an Motion Tags des Motion Description Tree zugeordnet werden kann, wobei mit Stand vom 21.04.2017 im Durchschnitt 3,09 Motion Tags zur Klassifikation einer Bewegung verwendet werden. So kann etwa die Bewegung einer Person,

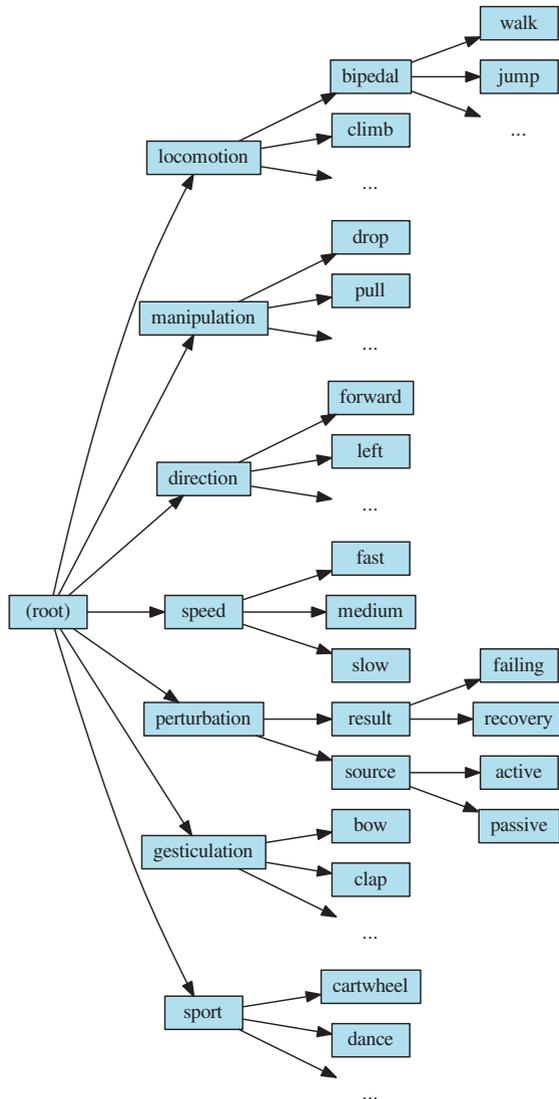


Abbildung 4.9: Auszug aus der Struktur des Motion Description Tree für die Klassifikation menschlicher Ganzkörperbewegung (entnommen aus Mandery et al. 2016c, ©2016 IEEE).

welche während des Abwärtslaufens einer Treppe stolpert und als Folge davon stürzt, wie folgt klassifiziert werden:

1. locomotion → bipedal → walk
2. speed → medium
3. direction → forward
4. direction → downward
5. perturbation → result → failing
6. perturbation → source → passive

Abbildung 4.10 zeigt drei weitere Beispiele für die Klassifikation von in der KIT-Bewegungsdatenbank verfügbaren Bewegungsaufnahmen gemäß dem Motion Description Tree.



- locomotion → go upstairs
- locomotion → go downstairs
- manipulation → hand over
- manipulation → screw



- manipulation → wipe
- speed → slow
- direction → leftward
- handedness → right hand



- poses → kneel
- poses → stand
- direction → upward
- handedness → left hand

Abbildung 4.10: Beispiele für die Klassifikation von Bewegungsaufnahmen gemäß dem Motion Description Tree.

Der Aufbau des Motion Description Tree basiert auf der lexikalischen Datenbank WordNet (Miller 1995), welche die zu seiner Konstruktion entscheidenden Hyponymie- und Hyperonymiebeziehungen zwischen Wörtern der englischen Sprache liefert. Hyponymie- und Hyperonymiebeziehungen beschreiben semantische Zusammenhänge zwischen den beiden in Beziehung stehenden Wörtern, wobei ein Hyponym als Unterbegriff mit einem engeren Bedeutungsfeld ein zweites Wort genauer präzisiert und ein Hypernym als Oberbegriff ein weiter gefasstes Bedeutungsfeld als ein zweites Wort aufweist und dieses somit beinhaltet. Die Kenntnis dieser semantischen Beziehungen konnte zur Konstruktion des Motion Description Tree genutzt werden, indem als Kindknoten eines Knoten (Ko-)Hyponyme innerhalb von WordNet gesucht wurden, z.B. das Wort „bow“ als Hyponym des Wortes „gesticulation“. Zur Erleichterung dieses Konstruktionsvorgangs wird der Motion Description Tree selbst innerhalb des Datenbankschemas verwaltet (vgl. Abbildung 4.5) und ist über die webbasierte Oberfläche bearbeitbar, womit im Laufe der Zeit bei zuvor nicht berücksichtigten Bewegungstypen wie z.B. knienden Bewegungsabläufen eine Erweiterung der Baumstruktur vorgenommen werden kann.

Wie bereits erwähnt, ermöglicht die Organisation von Bewegungsdaten auf Basis des Motion Description Tree in der KIT-Bewegungsdatenbank die Implementierung einer Suchfunktion, welche serverseitig ausgeführt und von den beiden in Abschnitt 4.2.3 beschriebenen Schnittstellen unterstützt wird. Mit der Suchfunktion können zum einen Bewegungsaufnahmen gefunden werden, die mit einem bestimmten Knoten oder einem gesamten Teilbaum des Motion Description Tree assoziiert sind. Zum Beispiel kann durch Suche nach dem Knoten „locomotion → bipedal → walk“ die Liste der verfügbaren Bewegungsexperimente nach zweibeinigen Laufbewegungen gefiltert werden. Zum anderen können derartige Suchanfragen nach Knoten bzw. Teilbäumen unter Verwendung der logischen Operatoren AND, OR und NOT verknüpft werden, womit sich die Möglichkeit zur Formulierung von beliebig komplexen Suchanfragen ergibt. So können durch Verwendung

der Suchanfrage „run OR (walk AND NOT(slow))“ etwa diejenigen Bewegungsaufnahmen gefunden werden, welche entweder eine Rennbewegung in beliebiger Geschwindigkeit oder eine nichtlangsame Laufbewegung zeigen.

Zusätzlich zur Klassifikation der Bewegungsdaten in der KIT-Bewegungsdatenbank gemäß dem Motion Description Tree wurden in einem die Bewegungsdatenbank ergänzenden und ebenfalls frei zugänglichen Datensatz, dem *KIT Motion-Language Dataset* (Plappert et al. 2016), für diese Daten darüber hinaus auch natürlichsprachliche Annotationen gesammelt. Der Zweck dieses Datensatzes liegt allerdings weniger in der Organisation der vorhandenen Bewegungsdaten, als vielmehr in seiner Anwendung als Benchmark-Datensatz für Forschungsfragestellungen der intuitiven Programmierung von Robotern, beispielsweise zur datengetriebenen Verknüpfung von Bewegungs- und Sprachmodellen (siehe hierzu auch Abschnitt 7.2). Die Sammlung textueller Bewegungsbeschreibungen erfolgte hierbei durch Crowdsourcing mit Hilfe eines zu diesem Zweck entwickelten, frei zugänglichen webbasierten *Motion Annotation Tool*¹⁰, das in Abbildung 4.11 dargestellt ist. Die Funktion dieser Webplattform ist, dem teilnehmenden Benutzer nach einer speziellen Heuristik¹¹ ausgewählte Bewegungsaufnahmen in einer dreidimensionalen Visualisierung anzuzeigen und eine natürlichsprachliche Beschreibung der dargestellten Bewegung als Satz der englischen Sprache abzufragen. Ziel war hierbei, für den Großteil der Bewegungsaufnahmen nicht nur eine einzelne natürlichsprachliche Beschreibung, sondern mehrere Beschreibungen unterschiedlicher Benutzer zu sammeln. Das zuvor angesprochene Problem von Inkonsistenzen bei natürlichsprachlichen Annotationen ist in diesem Fall nicht als Nachteil anzusehen, sondern in einem gewissen Umfang sogar unabdingbar, um die Vielfalt der natürlichen Sprache in adäquater Weise im

¹⁰<https://motion-annotation.humanoids.kit.edu/>

¹¹Die Heuristik basiert auf einer Verwendung der in Abschnitt 6.6.1 vorgestellten Perplexität für die Auswahl von Bewegungsaufnahmen, um dem Benutzer Bewegungen mit ungewöhnlichen oder bislang unzureichend im Datenbestand abgedeckten natürlichsprachlichen Beschreibungen bevorzugt zur Annotation zu präsentieren.

KIT Motion-Language Dataset abbilden zu können. Nähere Informationen zu der verwendeten Methodik und dem resultierenden Datensatz finden sich in (Plappert et al. 2016).

Abbildung 4.11: Webbasierte Oberfläche zum Sammeln natürlichsprachlicher Annotationen für menschliche Ganzkörperbewegungen durch Crowdsourcing.

4.4 Datengetriebene Generierung von Klassifikationsvorschlägen

Der in der KIT-Bewegungsdatenbank zum Einsatz kommende Ansatz zur Organisation von Bewegungsaufnahmen gemäß dem Motion Description Tree basiert grundsätzlich auf einer manuellen Klassifikation der im Datenbestand verfügbaren Bewegungsexperimente, wofür diese durch Benutzer mit einer beliebigen Anzahl der im Motion Description Tree definierten Motion Tags gekennzeichnet werden. Obwohl Inkonsistenzen bei der Klassifikation von Bewegungen gemäß dem Motion Description Tree durch verschiedene Benutzer ein geringeres Problem darstellen als bei den in zahlreichen

anderen Bewegungsdatenbanken verwendeten Freitextbeschreibungen (vgl. Abschnitt 4.1), können dennoch Probleme entstehen. Diese entstammen zum einen der Subjektivität von Motion Tags, die zu einer abweichenden Interpretation ihrer Bedeutungen oder auch der zu kennzeichnenden Bewegung und dadurch zu Inkonsistenzen zwischen unterschiedlich gekennzeichneten Bewegungsexperimenten führen können. Zum anderen können sich unvollständige Bewegungskennzeichnungen daraus ergeben, dass Benutzern ein vollständiger Überblick über den Motion Description Tree und mögliche Motion Tags fehlt¹² und somit für eine Bewegung offensichtlich relevante Tags einem Benutzer nicht bekannt sind und dementsprechend auch nicht hinzugefügt werden.

Zur Abschwächung der genannten Problematik wurde die KIT-Bewegungsdatenbank durch Integration des in Kapitel 3 vorgestellten Systems zur Multi-Label-Bewegungsklassifikation erweitert (auf Basis des dort am besten abschneidenden PTV basierend auf der BRM mit logistischer Regression, siehe Abschnitt 3.4.2). Zweck dieser Integration ist die automatische datengetriebene Generierung von Klassifikationsvorschlägen in der KIT-Bewegungsdatenbank, die in dem Hinzufügen oder Entfernen von Motion Tags zu Bewegungsexperimenten bestehen und dem Benutzer zur Überprüfung und eventuellen Umsetzung präsentiert werden. Da die Multi-Label-Bewegungsklassifikation hierbei selbst auf dem Datenbestand der KIT-Bewegungsdatenbank trainiert wird, ist einerseits eine automatische Adaption an neue Bewegungstypen und Motion Tags sichergestellt, jedoch wird andererseits auch eine bereits überwiegend korrekte Klassifikation des vorhandenen Datenbestands vorausgesetzt. Die Integration dieses Systems zur Multi-Label-Bewegungsklassifikation erfolgte im Rahmen einer Bache-

¹² Insbesondere in der frühen Entwicklungsphase der KIT-Bewegungsdatenbank trug hierzu auch der Umstand bei, dass nach Erweiterungen des Motion Description Tree Inkonsistenzen durch eine nur unvollständige Anpassung des bereits klassifizierten Datenbestands zunächst nicht beseitigt wurden.

lorarbeit (Langenstein 2016), deren Ergebnisse in diesem Abschnitt näher beschrieben werden.

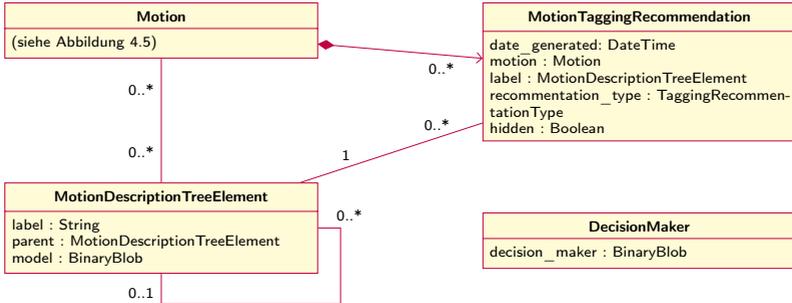


Abbildung 4.12: UML-Klassendiagramm für die an der Generierung und Verwaltung von Klassifikationsvorschlägen beteiligten Klassen und ihren Beziehungen.

Für die Umsetzung der Multi-Label-Bewegungserkennung und die Verwaltung der generierten Klassifikationsvorschläge ist es erforderlich, die Klassenstruktur in der Modell-Komponente, die wie in Abschnitt 4.2.2 beschrieben auf ein relationales Datenbankschema abgebildet wird, um zusätzliche Klassen und Attribute zu erweitern. Abbildung 4.12 zeigt ein UML-Klassendiagramm für die an der Generierung und Verwaltung von Klassifikationsvorschlägen beteiligten Klassen (vgl. auch Abbildung 4.5). In der *MotionDescriptionTreeElement*-Klasse enthält ein zusätzliches *model*-Attribut die gelernten Parameter des HMM, das im Rahmen der Multi-Label-Bewegungsklassifikation mit diesem Motion Tag assoziiert ist. Diese Parameter, genauer die Übergangswahrscheinlichkeitsmatrix A , die Anfangsverteilung π und die Emissionsverteilungen ϕ (vgl. Abschnitt 3.2), werden hierfür in einer mit Hilfe des *Pickle*-Moduls von Python erstellten binären Serialisierung persistent gespeichert. Die ebenfalls für die Multi-Label-Bewegungsklassifikation mit der BRM erforderlichen Parameter der nachgelagerten binären Klassifikatoren, hier Parameter der logistischen Regression, werden auf dieselbe Weise als Serialisierung in einer globalen *DecisionMaker*-Instanz gespeichert, von der zu jedem Zeitpunkt nur eine

existiert (*Singleton*-Entwurfsmuster)¹³. Generierte Klassifikationsvorschläge werden durch die *MotionTaggingRecommendation*-Klasse repräsentiert. Jeder Klassifikationsvorschlag ist mit genau einem Bewegungsexperiment und einem Motion Tag assoziiert und durch die Kombination dieser beiden Assoziationen eindeutig bestimmt. Das *recommendation_type*-Attribut kann momentan lediglich zwei mögliche Werte annehmen und wird zur Unterscheidung zwischen den beiden Vorschlagstypen für das Hinzufügen eines neuen Motion Tags zu einem Bewegungsexperiment bzw. für das Entfernen eines bereits vorhandenen Tags eingesetzt. Abschließend können mit Hilfe des *hidden*-Attributs Klassifikationsvorschläge, deren Umsetzung bei der Überprüfung durch den Benutzer abgelehnt wurde, als versteckt markiert werden. Diese Vorschläge werden in der webbasierten Oberfläche nicht mehr zur Überprüfung angezeigt, müssen jedoch gespeichert bleiben, um eine erneute Generierung und Anzeige des jeweiligen Vorschlags zu verhindern.

Die Generierung von Klassifikationsvorschlägen erfolgt periodisch und ist in drei Schritte untergliedert, welche über die zuvor beschriebene Erweiterung der Modell-Komponente voneinander entkoppelt sind.

Im ersten Schritt wird je ein HMM für jedes betrachtete Motion Tag des Motion Description Tree mit sämtlichen diesem Tag zugehörigen Bewegungsdaten der Datenbank trainiert (vgl. Abschnitt 3.2). Eine kleinere Anzahl an Teilbäumen des Motion Description Tree musste hierbei ausgeschlossen werden, wodurch die dort enthaltenen Motion Tags nicht für die Generierung von Klassifikationsvorschlägen zum Einsatz kommen. Hierzu gehören die Teilbäume unterhalb der Knoten „body parts“ (an der Bewegung beteiligte Körperteile) und „direction“ (Bewegungsrichtung), da die enthaltenen Motion Tags in einer sehr stark vom Bewegungstyp abhängigen Art und Weise verwendet werden, die durch den HMM-basierten Ansatz derzeit

¹³Obwohl die Parameter der binären Klassifikatoren bei der BRM jeweils eindeutig einem Label zugeordnet werden können, wurde diese globale Speicherung aller Parameter gewählt, um die Möglichkeit eines nachträglichen Wechsels zu anderen Ansätzen der Multi-Label-Bewegungsklassifikation, z.B. einem AAV, offen zu lassen.

nicht adäquat abgebildet werden kann. Ebenfalls ausgeschlossen wurde der Teilbaum „perturbation“ zur Beschreibung von Störungen wie etwa Stößen und deren Ursache, deren Bestimmung nur auf Basis von Bewegungsdaten kaum möglich ist, und mit „speed“ Angaben zur subjektiven Ausführungsgeschwindigkeit einer Bewegung, welche HMMs naturgemäß nur schlecht repräsentieren können. Die Wahl der HMM-Hyperparameter für das Training erfolgt, bis auf eine aus Effizienzgründen vorgenommene Reduktion der Anzahl versteckter Zustände auf 6, analog zu Kapitel 3¹⁴. Für den Aufbau des Merkmalsvektors zur Bewegungsrepräsentation wird die 8-dimensionale Merkmalsmenge, welche im Rahmen der in Abschnitt 3.5.1 beschriebenen Merkmalsauswahl ermittelt wurde, zugrunde gelegt. Diese besteht aus den in Abschnitt 2.3.1 definierten Merkmalen *CoM_vel*, *joint_vel_norm* und *end_effectors_vel_norm*, zu deren Berechnung die in Abschnitt 2.3.2 beschriebenen Verarbeitungsschritte zum Einsatz kommen.

Als zweiter Schritt erfolgt das Training der für die Multi-Label-Bewegungsklassifikation nach der BRM entscheidenden binären Klassifikatoren. Grundlage hierfür sind die für alle Bewegungsaufnahmen in der Datenbank ermittelten Likelihoods unter den im ersten Schritt trainierten HMMs. Wie bereits zuvor erwähnt, wurde in Übereinstimmung mit den Ergebnissen der Evaluation in Abschnitt 3.5.2 die dort vorgestellte logistische Regression mit L1-Regularisierung und $C = 10^{-2}$ als Lernverfahren für diese binäre Klassifikation gewählt¹⁵.

Abschließend erfolgt als dritter Schritt die eigentliche Generierung der Klassifikationsvorschläge auf Basis des zuvor trainierten Systems zur Multi-Label-Bewegungsklassifikation. Hierfür sei zunächst daran erinnert, dass

¹⁴ Links-Rechts-Topologie mit $\Delta = 1$, gleichverteilte Initialisierung der Zustandsübergangs- und der initialen Wahrscheinlichkeiten, Initialisierung der Mittelwertvektoren und Kovarianzmatrizen mit K-Means-Algorithmus, Beschränkung der Kovarianzmatrizen auf Diagonalmatrizen, 10 Iterationen des Baum-Welch-Algorithmus.

¹⁵ In (Langenstein 2016) fälschlicherweise an einer Stelle als L1-Regularisierung mit $C = 10^{-3}$ angegeben.

ein in der KIT-Bewegungsdatenbank verfügbares Bewegungsexperiment aus meist mehreren Bewegungsaufnahmen als Demonstrationen besteht, die Bewegungsklassifikation gemäß dem Motion Description Tree jedoch grundsätzlich auf der Ebene des Bewegungsexperiments und nicht der Bewegungsaufnahme erfolgt. Eine grundsätzliche Schwierigkeit besteht somit darin, dass Klassifikationsvorschläge für ein Bewegungsexperiment auf Basis der (sich eventuell widersprechenden) Ergebnisse der Multi-Label-Bewegungsklassifikation für seine einzelnen Bewegungsaufnahmen erzeugt werden müssen. Zur Generierung dieser Klassifikationsvorschläge für ein Bewegungsexperiment i werden zunächst für jedes seiner Bewegungsaufnahmen $j = 1, \dots, J$ die Likelihoods unter den HMMs für die M Motion Tags ermittelt. Anschließend werden ausgehend von diesen Likelihoods mit den binären Klassifikatoren der BRM für jede Bewegungsaufnahme die Zugehörigkeiten dieser Bewegungsaufnahme zu den verfügbaren Motion Tags bestimmt (siehe Abschnitt 3.4.2):

$$\mathbf{X}_{i,j} = (\chi_{i,j,1}, \dots, \chi_{i,j,M})^T \quad \text{mit } \forall m : \chi_{i,j,m} \in \{0, 1\},$$

wobei $\chi_{i,j,m}$ das Ergebnis der Multi-Label-Bewegungsklassifikation als Zugehörigkeit der Bewegungsaufnahme j zum Motion Tag m repräsentiert. Die Klassifikationsergebnisse $\mathbf{X}_{i,1}, \dots, \mathbf{X}_{i,J}$ für alle Bewegungsaufnahmen des Bewegungsexperiments i bilden zusammen mit \mathbf{Y}_i als Spaltenvektor der tatsächlichen Label-Zugehörigkeiten des Bewegungsexperiments i in der Bewegungsdatenbank die Grundlage für das Bestimmen von Klassifikationsvorschlägen:

$$\boldsymbol{\eta}_i = \frac{1}{J} \cdot \sum_{j=1}^J \mathbf{X}_{i,j} - \mathbf{Y}_i.$$

Wie bei $\mathbf{X}_{i,j}$ handelt es sich bei $\boldsymbol{\eta}_i$ um einen Spaltenvektor, dessen Komponenten mit jeweils einem Motion Tag korrespondieren. Diese Komponenten nehmen Werte zwischen -1 (Motion Tag ist bei Bewegungsexperiment gesetzt, wurde aber in keiner einzigen Bewegungsaufnahme erkannt) und

1 (Motion Tag wurde in allen Bewegungsaufnahmen erkannt, ist aber bei Bewegungsexperiment nicht gesetzt) an. Für jedes betrachtete Motion Tag wird, falls nicht schon vorhanden, ein Klassifikationsvorschlag zum Hinzufügen dieses Tags genau dann erzeugt, wenn der korrespondierende Eintrag in η_i einen Schwellwert von 0,6 überschreitet. Analog hierzu wird beim Unterschreiten von $-0,6$ ein Vorschlag zum Entfernen des Tags generiert.

Zur Überprüfung von erzeugten Klassifikationsvorschlägen durch die Benutzer der KIT-Bewegungsdatenbank enthält die webbasierte Oberfläche grafische Elemente zur Präsentation dieser Vorschläge. Die in Abbildung 4.7 dargestellte Listenansicht weist den Benutzer hierbei auf Bewegungsexperimente hin, für die bislang unbearbeitete Vorschläge zur Überprüfung vorliegen. Auf einer separaten Seite können diese Klassifikationsvorschläge eingesehen und für jeden Vorschlag eine von drei möglichen Handlungsoptionen ausgewählt werden. Im ersten Fall wird der Vorschlag umgesetzt, was zum Löschen dieses Vorschlags und zum Hinzufügen bzw. Entfernen des korrespondierenden Motion Tags zu bzw. von dem Bewegungsexperiment führt. Alternativ kann ein Vorschlag abgelehnt werden, wodurch der Datenbestand unverändert bleibt, jedoch der Vorschlag zukünftig versteckt und dem Benutzer nicht mehr angezeigt wird. Abschließend kann die Entscheidung über einen Klassifikationsvorschlag auch zurückgestellt werden, womit der jeweilige Vorschlag unverändert zur späteren Überprüfung beibehalten wird.

Eine Evaluation der mit dem beschriebenen System datengetrieben generierten Klassifikationsvorschläge wurde ebenfalls im Zuge der Bachelorarbeit (Langenstein 2016) vorgenommen. Hierfür wurde eine Benutzerstudie durchgeführt, in deren Verlauf fünf Testpersonen zufällig ausgewählte Bewegungsexperimente der KIT-Bewegungsdatenbank zusammen mit den erzeugten Klassifikationsvorschlägen gezeigt wurden. Aufgabe der Testpersonen war die Beurteilung der Vorschläge, wofür sie die Anweisung erhielten, diese in *korrekte*, *falsche* und, in Fällen von Mehrdeutigkeiten oder unscharf definierten Motion Tags, *inkonklusive* Vorschläge zu unterteilen. Des Weiteren wurden vermutete fehlende Motion Tags in der Bewegungsklassifikation

der KIT-Bewegungsdatenbank, für die jedoch kein Klassifikationsvorschlag generiert worden war, als *fehlende* Vorschläge gewertet. Die Ergebnisse der Benutzerstudie sind in Tabelle 4.2 dargestellt. Insgesamt wurden für die 115 evaluierten Bewegungsexperimente 34 Klassifikationsvorschläge generiert, von denen 15 (44,1 %) als korrekt und 17 als falsch beurteilt wurden. In Anbetracht der Tatsache, dass Vorschläge einer menschlichen Überprüfung unterliegen, handelt es sich bei 44,1 % um ein zufriedenstellendes Ergebnis, welches gleichzeitig den Nutzen der erzeugten Klassifikationsvorschläge demonstriert. Als eine mögliche Ursache für fehlerhafte Klassifikationsvorschläge lassen sich im Rahmen der Evaluation einerseits für wenige spezifische Motion Tags bereits vorhandene Inkonsistenzen im Datenbestand der Bewegungsdatenbank und damit in den Trainingsdaten für die Generierung der Vorschläge identifizieren. Auf der anderen Seite bezieht sich ein großer Teil der fehlerhaften Vorschläge auf den *Manipulation*-Teilbaum des Motion Description Tree, für dessen Motion Tags aufgrund der fehlenden Berücksichtigungen der Finger bei Bewegungsaufnahmen teilweise keine bessere Differenzierung feinmotorischer (Finger-)Bewegungen zu erwarten war.

Tabelle 4.2: Ergebnisse der Benutzerstudie für die datengetrieben generierten Klassifikationsvorschläge (modifiziert nach Langenstein 2016). Für jede der fünf Testpersonen ist die Aufteilung der evaluierten Vorschläge in korrekte, falsche sowie inkonklusive und die Anzahl der fehlenden Vorschläge angegeben.

Proband	# Evaluiert		Evaluation der Vorschläge				# Fehlende Vorschl.
	Bew.	Vorschl.	# korrekt	# falsch	# inkonk.	% korrekt	
1	20	2	2	0	0	100	0
2	20	9	3	5	1	33,33	2
3	20	8	4	3	1	50	3
4	20	5	2	3	0	40	6
5	35	10	4	6	0	40	8
Summe	115	34	15	17	2	44,1	19

4.5 Inhalt der Bewegungsdatenbank

Derzeit¹⁶ enthält die KIT-Bewegungsdatenbank eine Gesamtzahl von 9375 Bewegungsaufnahmen mit einer kumulativen Länge von 28,82 Stunden und stellt damit eine der weltweit größten öffentlich zugänglichen Datenbanken für menschliche Ganzkörperbewegungen dar (vgl. Abschnitt 4.1). Der folgende Abschnitt soll anhand qualitativer und quantitativer Kriterien eine Übersicht über den in der KIT-Bewegungsdatenbank verfügbaren Datenbestand liefern. Hierbei werden wie zuvor die beiden Begriffe des Bewegungsexperiments und der Bewegungsaufnahme unterschieden, wobei ein Bewegungsexperiment die wiederholte Durchführung eines identischen Bewegungsablaufs unter denselben experimentellen Bedingungen bezeichnet (siehe Abschnitt 4.2.1) und aus meist mehreren Bewegungsaufnahmen als Demonstrationen dieses Bewegungsablaufs besteht.

Tabelle 4.3: Bewegungsexperimente und -aufnahmen der KIT-Bewegungsdatenbank, aufgeschlüsselt nach beitragenden Instituten.

Institut	# Bew.exp.	# Aufnahmen
Karlsruher Institut für Technologie	787	4096
Eberhardt Karls Universität Tübingen	96	617
Weizmann Institute of Science	202	2701
Centre National de la Recherche Scientifique	4	158
Carnegie Mellon University	81	1803
Summe	1170	9375

Tabelle 4.3 zeigt eine nach den beitragenden Instituten aufgeschlüsselte Übersicht über die in der KIT-Bewegungsdatenbank enthaltenen Daten. Von den insgesamt 1170 in der Bewegungsdatenbank verfügbaren Bewegungsexperimenten wurden 787 im Bewegungserfassungslabor des H²T durchgeführt, welches in Abschnitt 2.2.1 näher beschrieben wurde (vgl. auch Abbildung 2.2). Weitere 302 Bewegungsexperimente wurden im Rahmen

¹⁶Die folgenden Angaben beziehen sich auf den Stand vom 21.04.2017.

einer Kooperation im EU-Forschungsprojekt KoroBot von der Eberhard Karls Universität Tübingen, dem Weizmann Institute of Science in Rehovot und dem Centre National de la Recherche Scientifique in Rennes beigetragen. Abschließend wurden auch Daten aus der in Abschnitt 4.1 vorgestellten CMU-Bewegungsdatenbank, einer weit verbreiteten und populären Datenbank menschlicher Ganzkörperbewegung, in die KIT-Bewegungsdatenbank integriert. Mit 1803 Bewegungsaufnahmen in 81 Bewegungsexperimenten wurden hierfür die Messungen des markerbasierten Bewegungserfassungssystems im C3D-Dateiformat für einen Großteil¹⁷ des CMU-Datenbestands importiert. Im Anschluss wurden die importierten C3D-Dateien analog zu den restlichen Bewegungsdaten in die MMM-Repräsentation überführt (vgl. Abschnitt 4.2.1), wobei das vom KIT Reference Marker Set abweichende CMU Marker Set berücksichtigt werden musste. Somit stehen die importierten CMU-Daten in der KIT-Bewegungsdatenbank dank der vereinheitlichten MMM-Repräsentation im selben Format wie der restliche Datenbestand bereit, wobei allerdings aufgrund von fehlenden Informationen innerhalb der CMU-Bewegungsdatenbank bei diesen Aufnahmen keinerlei Angaben über Umgebungsobjekte oder deren Bewegungen enthalten sind und die Klassifikation der Bewegung gemäß dem Motion Description Tree sowie diverse Metadaten wie etwa das Aufnahmedatum fehlen.

Viele der in der KIT-Bewegungsdatenbank bestehenden Bewegungsexperimente wurden im Rahmen eines Forschungsprojekts durchgeführt. Tabelle 4.4 zeigt für jedes Projekt die Gesamtzahl der in der Bewegungsdatenbank gespeicherten C3D-, MMM-, Video- und sonstigen Dateien. Abweichungen zwischen den angegebenen Werten eines Projekts für die ersten drei der genannten Dateitypen entstehen zum einen durch nicht in die MMM-Repräsentation konvertierbare C3D-Dateien, beispielsweise aufgrund der Verwendung eines abweichenden Marker Set bei der Vermarkierung des Probanden, und zum anderen durch fehlende bzw. mehrfach vorhandene Videoaufnahmen.

¹⁷Einige wenige Aufnahmen der CMU-Bewegungsdatenbank mussten aufgrund unvollständiger Markerdaten vom Import ausgeschlossen werden.

Tabelle 4.4: Bewegungsdaten der KIT-Bewegungsdatenbank, aufgeschlüsselt nach ihrer Zugehörigkeit zu Forschungsprojekten. Für jedes Forschungsprojekt ist die Gesamtzahl an C3D-Dateien (vgl. Abschnitt 2.2.1), Bewegungsdaten im MMM-Format (vgl. Abschnitt 2.2.2), Videoaufnahmen und sonstigen Dateitypen (z.B. Kraftmessdaten) angegeben. Die unter „Dauer“ angegebene Zeitdauer bezieht sich auf die kumulative Lauflänge der jeweiligen C3D-Dateien und die unter „Größe“ angegebene Datenmenge auf die Gesamtgröße aller zum Forschungsprojekt zugehörigen Dateien.

Projektname	# C3D	# MMM	# Video	# Sonstige	Dauer	Größe
KoroiBot ¹⁸	5309	2388	3974	4722	13,63 h	105,2 GiB
WALK-MAN ¹⁹	368	367	556	641	0,74 h	1,3 GiB
Xperience ²⁰	16	16	25	9	0,15 h	2,9 GiB
SecondHands ²¹	18	3	36	21	0,08 h	0,2 GiB
TimeStorm ²²	111	111	220	221	0,37 h	0,8 GiB
I-Support ²³	318	290	434	2431	3,05 h	1789,2 GiB ²⁴
Basic Walking ²⁵	434	434	20	872	0,73 h	0,5 GiB
Misc ²⁶	2801	2716	1304	3841	10,07 h	14,6 GiB
Summe	9375	6325	6569	12758	28,82 h	1914,7 GiB

Unter der angegebenen Anzahl der sonstigen Dateien sind jeweils alle weiteren verfügbaren Dateien subsumiert, etwa Messungen zusätzlicher Sensormodalitäten wie Beschleunigungs- oder Kraftsensoren und die in Abschnitt 4.2.1 angesprochenen Konverter- und Anonymisierungs-Logdateien. Außerdem

¹⁸ 7. Forschungsrahmenprogramm der EU, Förderkennzeichen 611909.

¹⁹ 7. Forschungsrahmenprogramm der EU, Förderkennzeichen 611832.

²⁰ 7. Forschungsrahmenprogramm der EU, Förderkennzeichen 270273.

²¹ Forschungsrahmenprogramm Horizont 2020 der EU, Förderkennzeichen 643950.

²² Forschungsrahmenprogramm Horizont 2020 der EU, Förderkennzeichen 641100.

²³ Forschungsrahmenprogramm Horizont 2020 der EU, Förderkennzeichen 643666.

²⁴ Videoaufnahmen von RGB-/Tiefendaten teilweise extern abgelegt und hier nicht inbegriffen.

²⁵ Bewegungsaufnahmen von zweibeiniger Fortbewegung ohne konkrete Zuordnung zu einem Forschungsprojekt.

²⁶ Diverse Bewegungsaufnahmen außer zweibeiniger Fortbewegung ohne konkrete Zuordnung zu einem Forschungsprojekt zusammen mit den aus der CMU-Bewegungsdatenbank importierten Daten.

ist für jedes Forschungsprojekt die kumulative Lauflänge der jeweiligen C3D-Dateien und die insgesamt gespeicherte Datenmenge angegeben. Bewegungsaufnahmen, die keinem Forschungsprojekt zugeordnet sind, werden in der KIT-Bewegungsdatenbank und in Tabelle 4.4 unter „Basic Walking“ und „Misc“ aufgeführt. „Basic Walking“ beinhaltet hierbei die frühesten in der Bewegungsdatenbank gespeicherten Aufnahmen zweibeiniger Laufbewegungen, die teilweise aus Vorarbeiten für das KoroiBot-Projekt zur Untersuchung des Ganzkörper-Drehimpulses beim menschlichen Laufen entstanden sind, aber von den im KoroiBot-Projekt durchgeführten Bewegungsstudien abweichen. „Misc“ umfasst hingegen sämtliche anderen nicht-projektbezogenen Bewegungsaufnahmen und somit auch den importierten Datenbestand der CMU-Bewegungsdatenbank.

Über die in Tabelle 4.4 angegebenen Dateien hinaus beinhaltet die KIT-Bewegungsdatenbank auch 233 Dateien, die nicht mit einem Bewegungsexperiment, sondern stattdessen mit einem der derzeit 68 eingetragenen Umgebungsobjekte assoziiert sind, für welche wie zuvor bereits erwähnt mit dem in Abschnitt 2.2.2 erläuterten Verfahren rekonstruierte Bewegungen zur Verfügung gestellt werden. Bei diesen mit Umgebungsobjekten assoziierten Dateien handelt es sich demzufolge um die für die Bewegungsrekonstruktion erforderlichen 3D-Modelle der Objekte im Blender- und Simox-Format und ergänzende Daten wie bspw. annotierte Bilder zur Dokumentation der jeweiligen Marker Sets.

Tabelle 4.5: Statistische Verteilung von Alter, Körpergröße und Gewicht für die in der KIT-Bewegungsdatenbank verfügbaren Probanden (aufgrund unvollständiger Daten ohne Berücksichtigung der CMU-Probanden).

	Mittelwert/Standardabw.	Minimum	Maximum
Alter	26,0 Jahre \pm 7,3 Jahre	15 Jahre	62 Jahre
Körpergröße	1,76 m \pm 0,08 m	1,55 m	1,96 m
Gewicht	70,4 kg \pm 10,4 kg	44 kg	93 kg

Insgesamt enthält die KIT-Bewegungsdatenbank, ausgenommen die importierten Bewegungsaufnahmen der CMU, Daten von 92 menschlichen Probanden, von denen 68 männlich und 24 weiblich sind. Tabelle 4.5 zeigt die statistische Verteilung von Alter, Körpergröße und Gewicht für diese Probanden. Durch den Import der CMU-Bewegungsdaten wurden in der KIT-Bewegungsdatenbank 81 weitere Probanden erfasst, wobei jedoch nicht jeder dieser Probanden mit einer unterschiedlichen realen Person korrespondiert²⁷. Aufgrund fehlender Informationen in der CMU-Bewegungsdatenbank sind zu diesen Probanden keine weiteren Daten vorhanden – es fehlen insbesondere Geschlecht, die in Tabelle 4.5 aufgeführten Daten und die in Abschnitt 4.2.1 angesprochenen anthropometrischen Messungen. Die zur Bewegungsrekonstruktion mit dem in Abschnitt 2.2.2 vorgestellten Verfahren erforderliche Körpergröße wurde für diese Probanden deshalb im Zuge des Datenimports heuristisch auf Basis der kartesischen Positionen der Kopfmarker aus den C3D-Dateien approximiert.

Einen wesentlichen Anteil am Datenbestand der KIT-Bewegungsdatenbank haben Bewegungsaufnahmen zweibeiniger Fortbewegung und Aufnahmen von *Push Recovery*, welche die Ausweichbewegungen des Menschen bei Stößen während des Stehens oder Laufens zeigen. Zu diesen Kategorien von Bewegungsaufnahmen gehören insbesondere die im Rahmen des KoroiBot-Projekts entstandenen Aufnahmen, die zusätzlich in der in Abbildung 4.7 gezeigten speziell angepassten KoroiBot-Gestaltungsvariante der webbasierten Benutzeroberfläche öffentlich zugänglich gemacht wurden. Eine Übersicht über die in der KIT-Bewegungsdatenbank verfügbaren Aufnahmen zweibeiniger Fortbewegung ist in Tabelle 4.6 dargestellt. Zusätzlich zu einfachen Laufbewegungen in unbeschränkten Umgebungen, d.h. auf ebenerdigem Boden, sind ebenso Bewegungsaufnahmen von Laufbewe-

²⁷ Dieselbe reale Person kann bei mehrfachen Bewegungsaufnahmen an unterschiedlichen Tagen in der CMU-Bewegungsdatenbank als neuer Proband ausgewiesen sein. Da weiterführende Informationen hierzu nicht verfügbar sind, bleibt somit unklar, wie viele unterschiedliche Personen tatsächlich aufgenommen wurden.

gungen in beschränkten Umgebungen verfügbar, bei denen die menschliche Fortbewegung bspw. auf einer Treppe oder einem Balken erfasst wurde. Bei diesen Bewegungsaufnahmen wird durch die Rekonstruktion der Bewegung beteiligter Umgebungsobjekte eine vergleichsweise exakte Bestimmung der Kontaktpunkte des Menschen mit seiner Umgebung ermöglicht, was eine wesentliche Grundlage für die in Abschnitt 5.3 beschriebene Erkennung von Support-Kontakten darstellt. Bei Bewegungstypen, welche wie das Laufen auf einer Treppe die Möglichkeit einer Variation bieten (bspw. mit/ohne Geländer, aufwärts/abwärts, vorwärts/rückwärts), wurde im Rahmen der Bewegungserfassung versucht, diese Variationsmöglichkeiten experimentell abzubilden. Abbildung 4.13 zeigt einige Beispiele für derartige in der KIT-Bewegungsdatenbank verfügbare Aufnahmen zweibeiniger Fortbewegung.

Tabelle 4.6: Aufnahmen zweibeiniger Fortbewegung in der KIT-Bewegungsdatenbank, aufgeschlüsselt nach unterschiedlichen Bewegungstypen.

Bewegungstyp	# Aufn.
Geradliniges Laufen/Sprinten auf flachem Boden	1220
Laufen einer Links-/Rechtskurve	164
Laufen eines vorgegebenen Pfads (Kreis/Ellipse/Achter usw.)	1241
Rückwärtslaufen	79
Laufen mit Drehung auf der Stelle	421
Laufen auf einer Rampe (auf-/abwärts)	253
Laufen auf einer Treppe (auf-/abwärts)	259
Laufen auf einem Balken	284
Laufen auf einer Wippe	66
Laufen auf unregelmäßigen Trittsteinen	72
Laufen auf einer nachgiebigen Matratze	29
Laufen mit Ausweichen vor statischen Hindernissen	88
Laufen mit Ausweichen vor einem beweglichen Hindernis	158
Überschreiten einer gedachten Spalte	107
Ausweichbewegungen bei Stößen während des Stehens/Laufens	476

Weitere in der KIT-Bewegungsdatenbank verfügbare Bewegungsklassen umfassen Aufnahmen von Manipulationsaufgaben einschließlich *Loko-Manipulation* (Kombination von Fortbewegung mit Manipulation), Aufnahmen von Gestikulations-, Musik- und Sportbewegungen sowie Multi-Probanden-Aufnahmen. Insbesondere bei den Aufnahmen von Manipulationsaufgaben wie bspw. Wischen oder Rühren ist, genau wie bei den Aufnahmen von Fortbewegung in beschränkten Umgebungen, die durch die KIT-Bewegungsdatenbank bereitgestellte Bewegungsrekonstruktion für die Umgebungsobjekte von zentraler Bedeutung. Ein Beispiel für eine derartige Manipulationsaufgabe mit vier Umgebungsobjekten und die durch die KIT-Bewegungsdatenbank erzeugte MMM-Repräsentation ist in Abbildung 2.8 gezeigt. Mit Musik- und Sportbewegungen sind Bewegungsaufnahmen beispielsweise vom Spielen verschiedener (teilweise imaginärer) Musikinstrumente wie Trommeln oder Gitarre und Aufnahmen von Tritten, Kniebeugen, Radschlagen, verschiedenen Tänzen usw. in der KIT-Bewegungsdatenbank enthalten. Abschließend sind auch Multi-Probanden-Aufnahmen, bei denen simultan die Bewegungen zweier Probanden erfasst werden, Teil des Datenbestands, beispielsweise von Mensch-Mensch-Interaktionsaufgaben wie der Übergabe von Objekten, Waschbewegungen oder Händeschütteln.

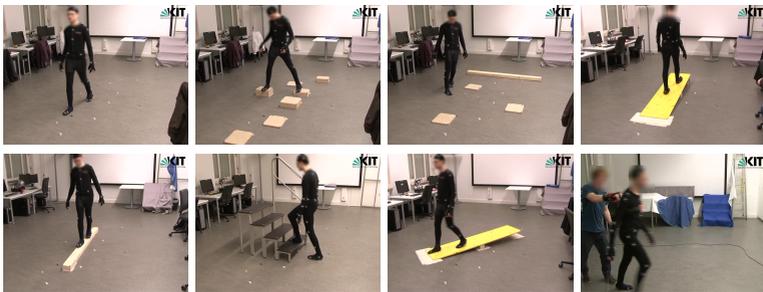


Abbildung 4.13: Verschiedene Bewegungstypen in Aufnahmen zweibeiniger Fortbewegung der KIT-Bewegungsdatenbank.

4.6 Zusammenfassung von Kapitel 4

In diesem Kapitel wurde die *KIT Whole-Body Human Motion Database* eingeführt, welche als großskalige Datenbank menschlicher Ganzkörperbewegungen die Grundlage für die Analyse solcher Bewegungen und für die datengetriebene Bewegungsgenerierung in den nachfolgenden Teilen dieser Arbeit legt. In der Bewegungsdatenbank werden hierfür nicht nur eine normalisierte Darstellung der menschlichen Bewegung mit der in Kapitel 2 vorgestellten MMM-Repräsentation berücksichtigt, sondern auch die Bewegungen beteiligter Objekte und Umgebungselemente, um eine Analyse der Interaktion zwischen Mensch und Umgebung zu ermöglichen. Neben einer Beschreibung der internen Struktur der Bewegungsdatenbank anhand softwarearchitektonischer Kriterien wurden die Benutzerschnittstellen diskutiert, wobei neben einer webbasierten Oberfläche eine Programmierschnittstelle einen maschinellen Zugriff auf den Datenbestand mit einer großen Anzahl unterschiedlicher Programmiersprachen erlaubt. Weiterhin wurde der zur systematischen Organisation von Bewegungsexperimenten in der Bewegungsdatenbank vorgeschlagene *Motion Description Tree* eingeführt, eine hierarchische Struktur möglicher *Motion Tags* zur Beschreibung menschlicher Ganzkörperbewegungen aus einer lexikalischen Datenbank. Der Einsatz dieses *Motion Description Tree* ermöglicht dabei nicht nur eine Festlegung disjunkter Bewegungstypen, sondern zusätzlich eine flexible Spezifikation von Modalitäten der Bewegungsausführungen wie Richtung oder Geschwindigkeit und bietet die Möglichkeit zur Formulierung komplexer Suchabfragen auf Basis logischer Operatoren. Die Zuweisung verfügbarer *Motion Tags* zu den Bewegungsexperimenten erfolgt durch Benutzer, die hierbei durch maschinell erzeugte Tagging-Vorschläge unterstützt werden. Für die Generierung dieser Tagging-Vorschläge wurde eine Integration der in Kapitel 3 vorgestellten Ansätze zur Multi-Label-Bewegungsklassifikation in die Bewegungsdatenbank beschrieben und eine Benutzerstudie präsentiert, welche eine stichprobenartige Validierung der erzeugten Tagging-Vorschläge zum

Ziel hatte. Abschließend wurde der gegenwärtige Inhalt der Bewegungsdatenbank mit 9375 Bewegungsaufnahmen und einer Gesamtlaufzeit von über 28 Stunden hinsichtlich qualitativer und quantitativer Kriterien dargestellt und hierfür die vorhandenen Bewegungsexperimente unterschiedlichen Bewegungstypen, den beitragenden Instituten und Forschungsprojekten zugeordnet.

5 Analyse menschlicher Bewegung anhand einer Ganzkörperposen-Taxonomie

Ein Ziel der vorliegenden Arbeit ist es, durch Analyse und Verständnis der menschlichen Ganzkörperbewegung einen Beitrag zur Generierung von Multi-Kontakt-Bewegungen bei humanoiden Robotern zu leisten. Dieses Kapitel widmet sich hierzu einer statistischen Analyse von Aufnahmen menschlicher Ganzkörper- und vor allem Multi-Kontakt-Bewegungen zur Untersuchung von beobachtbaren Aspekten dieser Bewegungen hinsichtlich der genutzten Support-Kontakte des Menschen mit seiner Umgebung. Grundlage dieser Analyse ist die am H^2T entwickelte Ganzkörperposen-Taxonomie (GKP-Taxonomie), die eine Darstellung von Multi-Kontakt-Bewegungen als Sequenz der in dieser GKP-Taxonomie definierten Support-Posen erlaubt. Hierfür wird in diesem Kapitel ein Verfahren für die automatische Segmentierung menschlicher Bewegungsaufnahmen hinsichtlich dieser Support-Posen vorgestellt. Die daraus resultierende segmentierte Darstellung menschlicher Bewegung als Posen-Sequenz bildet wiederum die Grundlage für das in Kapitel 6 einzuführende Verfahren zur datengetriebenen Generierung derartiger Posen-Sequenzen für Multi-Kontakt-Bewegungen humanoider Roboter in neuen Problemstellungen.

Dieses Kapitel ist wie folgt gegliedert. Nach einer Diskussion verwandter Arbeiten in Abschnitt 5.1 wird in Abschnitt 5.2 zunächst die am H^2T entwickelte GKP-Taxonomie vorgestellt und definiert, welche ein Kernelement für die in diesem Kapitel und in Kapitel 6 präsentierten Arbeiten bildet. Darauf

aufbauend wird in Abschnitt 5.3 das im Rahmen dieser Arbeit entwickelte Segmentierungsverfahren erläutert. Abschließend werden in Abschnitt 5.4 die Ergebnisse einer Evaluation präsentiert, bei der das vorgestellte Segmentierungsverfahren zunächst validiert wird und im Anschluss für eine Analyse von Aufnahmen menschlicher Ganzkörperbewegungen hinsichtlich der Auftretenshäufigkeit von Support-Posen, ihrer Transitionen sowie der assoziierten Zeitdauern und für eine partielle Validierung der GKP-Taxonomie zum Einsatz kommt.

5.1 Verwandte Arbeiten

Der folgende Abschnitt befasst sich mit einem in unterschiedliche Themenfelder gegliederten Überblick über verwandte Arbeiten. Nach einer Diskussion der bereits angesprochenen, als Inspiration für die GKP-Taxonomie dienenden Greiftaxonomien werden existierende (nicht auf Taxonomien basierende) Ansätze zur Dimensionalitätsreduktion menschlicher Ganzkörperposen und -bewegungen behandelt. Im Anschluss hieran werden Untersuchungen zur Nutzung von (Support-)Kontakten durch den Menschen und Methoden zu einer Erkennung solcher Kontakte in Bewegungsdaten vorgestellt. Abschließend werden wichtige existierende Verfahren zur Segmentierung menschlicher Bewegungen erläutert. Verwandte Arbeiten zur Generierung von Multi-Kontakt-Posen und -Bewegungen werden hingegen in Abschnitt 6.1 diskutiert.

Grifftaxonomien: Wie einleitend und in Abschnitt 1.1 bereits angesprochen, beruht die Motivation zum Einsatz der GKP-Taxonomie auf Arbeiten im Bereich des Greifens, wo gezeigt werden konnte, dass sich Griffe trotz der humanoiden Händen innewohnenden Hochdimensionalität (21 Bewegungsfreiheitsgrade im Fall der menschlichen Hand), mit Techniken wie PCA gut in einem niedrigdimensionalen Raum darstellen lassen (Ciocarlie et al. 2009; Santello et al. 1998; Bläsing et al. 2013) und wo seit Beginn

der Neunzigerjahre mit großem Erfolg Taxonomien zur Beschreibung und Kategorisierung menschlicher und robotischer Griffe eingesetzt werden. Die Betrachtung dieser Griff-taxonomien, bei denen der Versuch unternommen wird, mögliche Griffe im hochdimensionalen Raum auf einige fundamentale Griff-typen abzubilden, ist interessant, da durch eine Dualität des Greif- und Ganzkörperstabilitätsproblems (mehr dazu in Abschnitt 5.2) aus den existierenden Taxonomien auch Ansätze und Kriterien zur Formulierung der GKP-Taxonomie abgeleitet werden können.

Feix et al. (2016) liefern eine aktuelle Übersichtsarbeit über 22 Griff-taxonomien, von denen im Folgenden nur die wichtigsten genannt werden sollen. Ein grundlegendes Kriterium zur Einteilung menschlicher Griffe in diesen Taxonomien stellt hierbei die – schon von Napier (1956) beschriebene – Unterscheidung zwischen Kraftgriffen (*power grasps*) und Präzisionsgriffen (*precision grasps*) dar, die von einem Großteil der Taxonomien, manchmal ergänzt um dazwischenliegenden *intermediate grasps*, realisiert wird. Die von Cutkosky (1989) vorgeschlagene Taxonomie berücksichtigt darüber hinaus die Größe des zu greifenden Objekts, während beispielsweise in der Taxonomie von Kamakura (Bernardin et al. 2005), ursprünglich vorgestellt in (Kamakura 1989), eine weitere Unterteilung der Griffe nach dem Zweck des Griffes und den in Kontakt stehenden Handflächen vorgenommen wird. Bullock et al. (2013) präsentieren eine Taxonomie, welche Bewegungen der Hand bei der Klassifikation von Griffen berücksichtigt und so die Beschreibung von In-Hand-Manipulationsstrategien ermöglicht. Auf Basis der existierenden Taxonomien wurde von Feix et al. (2009) der Versuch unternommen, Griffe aus zahlreichen existierenden Griff-taxonomien systematisch in einer vereinheitlichenden Taxonomie darzustellen, was in der in (Feix et al. 2016) vorgestellten GRASP-Taxonomie (benannt nach dem Akronym des Forschungsprojekts) mit einer Gesamtzahl von 33 Griff-typen mündete.

Dimensionalitätsreduktion von menschlichen Ganzkörperbewegungen:

Das Ziel und die Motivation für den Einsatz der GKP-Taxonomie im Rahmen

dieser Arbeit besteht in einer Vereinfachung und Dimensionalitätsreduktion für die Darstellung menschlicher Ganzkörperposen bzw. -bewegungen, welche im Anschluss die Grundlage für die Analyse von Multi-Kontakt-Bewegungen (in diesem Kapitel) bzw. für die Generierung neuer Multi-Kontakt-Bewegungen (in Kapitel 6) legt. In der Literatur finden sich verschiedene Arbeiten, die ebenfalls eine derartige Dimensionalitätsreduktion, allerdings ohne den Einsatz einer Taxonomie, bezwecken und sich von der hier eingesetzten Vorgehensweise durch einen rein datengetriebenen Ansatz unterscheiden, der im Unterschied zu dem Einsatz der GKP-Taxonomie meist keine intuitive semantische Interpretation der vereinfachten Bewegungsrepräsentation zulässt.

Die bereits erwähnte PCA wurde als populäre klassische Methode der Dimensionalitätsreduktion in verschiedenen Arbeiten zur Vorverarbeitung menschlicher Ganzkörperbewegungen eingesetzt, beispielsweise bei Yamane et al. (2011), Lin et al. (2014) und Calinon et al. (2007). In (Forner-Cordero et al. 2005) wird auf Basis einer PCA eine Untersuchung von Mehrgelenkbewegungen hinsichtlich der Koordinaten beteiligter Gelenke bei der Bewegungsausführung durchgeführt.

In (N. Chen et al. 2015) werden Techniken des Deep Learning (*Autoencoder*) eingesetzt, um eine 50-dimensionale Repräsentation menschlicher Ganzkörperbewegungen in einen niedrigdimensionalen latenten Raum zu transformieren und dort Dynamic Movement Primitives (DMPs) aus Bewegungsaufnahmen der CMU-Bewegungsdatenbank von Laufen, Joggen, usw. zu lernen. Dieser Ansatz wird in (N. Chen et al. 2016), auch unter Verwendung von Daten der KIT-Bewegungsdatenbank, auf *Variational Autoencoder* erweitert, bei denen eine Maximierung der Unabhängigkeit zwischen den Dimensionen des latenten Raums die Generalisierungsfähigkeit erhöht und somit die Abbildung sehr unterschiedlicher Bewegungstypen in diesem Raum ermöglicht. Colomé et al. (2014) befassen sich ebenfalls mit der Dimensionalitätsreduktion von Bewegungsprimitiven, allerdings mit probabilistischen

Techniken und für eine andere Modellierung dieser Bewegungsprimitive als *Probabilistic Movement Primitives* (ProMPs).

Im Kontext der Generierung von Roboterbewegungen finden sich zahlreiche weitere Anwendungen für Dimensionalitätsreduktion. Suárez et al. (2015) untersuchen beispielsweise die beidseitigen 12-dimensionalen Handbewegungen eines Menschen mit einer PCA und setzen die resultierenden Synergien (Korrelationen zwischen den Bewegungsfreiheitsgraden) zur Planung von menschenähnlichen Roboterbewegungen in dem resultierenden niedrigdimensionalen Raum mit zwei (Montageaufgabe) bzw. sieben (freie Bewegungen) Dimensionen ein. In (García et al. 2015) wird dieser Ansatz für eine Anwendung des RRT-Algorithmus (*rapidly-exploring random tree*) zur Bewegungsplanung erweitert und hierfür eine PCA mit Gelenkwinkelgeschwindigkeiten durchgeführt. Luck et al. (2014) hingegen kombinieren Dimensionalitätsreduktion mit Ansätzen des bestärkenden Lernens (*reinforcement learning*), wodurch mit einer vergleichsweise kurzen Trainingsperiode mit einem NAO-Roboter neue Ganzkörperbewegungen zum Balancieren auf einem Bein gelernt werden können. Eine weitere derartige Kombination von Techniken zur Dimensionalitätsreduktion (hier PCA) mit bestärkendem Lernen findet sich bei Curran et al. (2016).

Abschließend wurde die Dimensionalitätsreduktion menschlicher Bewegung auch im Bereich der Computergrafik umfassend behandelt. Safonova et al. (2004) stellen einen auf quadratischer Programmierung und der CMU-Bewegungsdatenbank basierenden Ansatz zur Dimensionalitätsreduktion menschlicher Ganzkörperbewegung vor, der mit Lauf- und akrobatischen Bewegungen evaluiert wurde. Von Grochow et al. (2004) wurden in der Computergrafik für eine Lösung des IK-Problems *Gaussian Process Latent Variable Models* (GPLVMs) eingeführt, welche im Anschluss auch in verschiedenen Arbeiten der Robotik zur Dimensionalitätsreduktion für das Lernen von DMPs (Bitzer et al. 2009) und für bestärkendes Lernen (Bitzer et al. 2010) Verwendung fanden.

Untersuchungen zu Support-Kontakten bei Menschen: Obwohl Menschen, wie in Kapitel 1 exemplarisch beschrieben, zur Erhöhung ihrer Stabilität oftmals auf die Nutzung zusätzlicher Support-Kontakte über die Füße hinaus zurückgreifen, existieren nur verhältnismäßig wenige Untersuchungen zur Nutzung verfügbarer Support-Kontakte durch den Menschen (Nori et al. 2014). Existierende Arbeiten entstammen hierbei dem Bereich der biomechanischen Forschung oder der Neurowissenschaften und betrachten lediglich spezifische Posen in speziellen Aufgabenstellungen, wie beispielsweise Reich-/Greifbewegungen und Übergänge zwischen stehenden und sitzenden Posen. So wurden in (Babič et al. 2014) die Auswirkungen eines zusätzlichen Hand-Supports auf die Stabilität eines Menschen während des Auftretens von translatorischen durch eine Steward-Plattform erzeugten Störungen und die vom Menschen auf den als Support-Kontakt dienenden Griff ausgeübten Kräfte in Abhängigkeit von der Richtung der auftretenden Störung und von verschiedenen Griff-Positionen untersucht. Eine ähnliche Untersuchung findet sich in (Čamernik et al. 2016) und es konnte gezeigt werden, dass bei der Verwendung von Support-Kontakten mit der Hand während des Auftretens einer Störung signifikant geringere Abweichungen des Druckmittelpunkts (*center of pressure*) beim menschlichen Probanden beobachtet werden können. Ebenfalls eine Untersuchung zur Auswirkung von Hand-Supports auf die Stabilität, bei der wie zuvor die Abweichungen des Druckmittelpunkts und zusätzlich Muskelaktivierungen mittels Elektromyographie gemessen wurden, findet sich in (Sarraf et al. 2014), wobei hier jedoch eine linear bewegliche Plattform als Simulation eines fahrenden Verkehrsmittels wie etwa Bus oder U-Bahn zum Einsatz kommt. Johannsen et al. (2012) konnten zeigen, dass ein zusätzlicher Finger- oder Schulter-Kontakt beim Menschen das Schwanken eines Probanden reduziert und dieser Effekt sich bereits beobachten lässt, wenn es sich nur um eine schwache Berührung mit sehr geringen wirkenden Kräften handelt.

Erkennung von (Support-)Kontakten: Ein wesentliches Kernelement des in diesem Kapitel vorgestellten Verfahrens zur Segmentierung menschlicher Bewegungsaufnahmen hinsichtlich Support-Posen ist eine Erkennung von Support-Kontakten, welche die Geschwindigkeiten von Körpersegmenten und ihren Abstand zu Umgebungsobjekten berücksichtigt. Zur Erkennung von Kontakten des Menschen wurde am H²T von Wächter et al. (2015) bereits ein verwandtes Verfahren vorgestellt, das ebenfalls auf einer markerbasierten Erfassung des Menschen und beteiligter Umgebungsobjekte aufbaut. Statt einer Erkennung von Support-Kontakten zielt dieses Verfahren jedoch auf eine Erkennung von Kontakten für Manipulationsbewegungen ab, sodass anstelle des gesamten menschlichen Körpers nur die Hände berücksichtigt werden müssen und für die Kontakte keine raumfeste Position mehr angenommen werden kann. Aus diesem Grund werden in diesem Verfahren zwar, wie in dieser Arbeit, die berechneten räumlichen Relationen zwischen dem Menschen und seiner Umgebung berücksichtigt, nicht jedoch die Geschwindigkeiten von Körpersegmenten. Auch das in (Lengagne et al. 2012) vorgestellte Verfahren beruht auf einer markerbasierten Erfassung der menschlichen Bewegung, jedoch werden hier keine Informationen über die Umgebung benötigt und die Positionen der beteiligten Umgebungsobjekte nicht erfasst. Stattdessen kann aus einer Beschreibung der Geschwindigkeiten und Beschleunigungen aller Körpersegmente als minimale orientierte *Bounding Box* direkt bestimmt werden, welche Körpersegmente mutmaßlich als Umgebungskontakte in Frage kommen und unter Verwendung eines dynamischen Modells anschließend die minimale Teilmenge dieser Kontakte, welche zum Erfüllen der Stabilitätsanforderungen genügt, bestimmt werden. In (Karčnik 2003) wurde eine Methode zur Erkennung von Fuß-Boden-Kontakten basierend auf an den Schuhen angebrachten Markern und einer einfachen Schwellwertheuristik für die Positionen und Geschwindigkeiten dieser Marker vorgestellt.

Weitere Arbeiten zur Erkennung menschlicher Kontakte beruhen anstelle markerbasierter Bewegungserfassung auf markerloser Bewegungserfassung

bzw. der Verarbeitung von Videodaten. Pham et al. (2016) rekonstruieren die in Multi-Kontakt-Posen wirkenden Kontaktkräfte des Menschen mit Techniken des Deep Learning (rekurrente neuronale Netze mit *Long Short-Term Memory*) aus einer markerlosen Bewegungserfassung mit inertialen Messeinheiten, wobei bestehende Kraftmessungen als Referenzdaten zum Training dienen. Ein modellfreier Ansatz für Daten eines Stereokamerasystems (Aksoy et al. 2011) bzw. für RGB-D-Daten eines Kinect-Sensors (Aksoy et al. 2015) ermittelt Kontakte aus Berührungen bzw. Überlappungen zwischen erkannten Bildsegmenten der menschlichen Hand und der beteiligten Objekte und verwendet die erkannten Kontakte im Anschluss zur semantischen Beschreibung einer Szene und von darin durchgeführten Manipulationsaktionen. Der von E. Yu et al. (2006) beschriebene Ansatz zur Detektion menschlicher Kontakte basiert auf visuellen Daten, in welchen stabile Kontaktpunkte gesucht werden, die sich einerseits durch eine stationäre Position über mehrere aufeinanderfolgende Frames und andererseits durch ihre äußere Lage bezogen auf die erkannte Kontur des menschlichen Körpers auszeichnen. Brubaker et al. (2009) setzen Videodaten zur Ermittlung von Boden-Kontakten für eine Schätzung der Bodenreaktionskräfte (*ground reaction forces*) bei athletischen Bewegungen ein. Abschließend finden sich weitere Ansätze zur Beschreibung und Modellierung von Kontakten in menschlichen Bewegungen im Bereich der Computergrafik (Kapadia et al. 2016; Liu et al. 2010).

Segmentierung menschlicher Bewegungen: Als Bewegungssegmentierung¹ bezeichnet man die zeitliche Unterteilung einer Bewegung in aufeinanderfolgende disjunkte und, je nach Aufgabenstellung, beispielsweise mit Bewegungsprimitiven oder mit semantisch definierten Aktivitäten des Menschen assoziierte Bewegungssegmente. Zwar ist nach Kenntnisstand

¹ Der Begriff *Segmentierung* wird mehrdeutig in unterschiedlichen Problemstellungen verwendet, beispielsweise auch für Bildsegmentierung im Bereich des maschinellen Sehens. Im Kontext dieser Arbeit bezeichnet *Segmentierung* jedoch stets die temporale Segmentierung menschlicher Bewegungsaufnahmen als Zeitreihendaten.

des Autors in der Literatur kein Verfahren zur Segmentierung menschlicher Bewegungen bekannt, welches wie das im Rahmen dieser Arbeit entwickelte Verfahren auf den verwendeten Support-Kontakten des Menschen basiert, jedoch wurden in der Robotik in Bereichen wie dem Imitationslernen, der Rehabilitationsrobotik oder der Mensch-Roboter-Interaktion und auch in anderen Disziplinen wie der Computergrafik und der biomechanischen Forschung eine mannigfaltige Zahl anderer Verfahren zur automatischen Segmentierung entwickelt. Einige der wichtigsten dieser Segmentierungsverfahren sollen im Folgenden vorgestellt werden.

In der bereits angesprochenen Arbeit von Wächter et al. (2015) wird ein zweistufiges Segmentierungsverfahren vorgestellt, welches auf einer semantischen Ebene Bewegungssegmente in Manipulationsbewegungen auf Basis von Mensch-Objekt-Kontakten (siehe oben) erkennt und diese Segmente in einer zweiten Ebene anhand von Bewegungscharakteristika weiter unterteilt. Die semantische (obere) Ebene dieser hierarchischen Segmentierung ähnelt dem in dieser Arbeit vorgestellten Segmentierungsverfahren hinsichtlich der Verwendung von Mesh-Modellen beteiligter Umgebungsobjekte zur Berechnung der Distanzen zwischen dem Menschen und diesen Objekten, betrachtet jedoch wie zuvor erwähnt nur die Hände und keine Ganzkörperbewegungen. Zudem werden auf der semantischen Segmentierungsebene die Geschwindigkeiten der Körpersegmente nicht berücksichtigt.

Mehrere Segmentierungsverfahren basieren auf den in Abschnitt 3.2.1 vorgestellten HMMs. In (Janus et al. 2005) werden Bewegungsdaten in einem gleitenden Fenster betrachtet und mit dem Viterbi-Algorithmus optimale Zustandssequenzen für ein HMM bestimmt, wobei die ermittelten Zustandsübergänge des HMM, an denen sich die modellierte Emissionsverteilung für die 20 betrachteten Gelenkwinkel ändert, als Segmentgrenzen interpretiert werden. Kulić et al. (2009) stellen eine Segmentierung vor, bei der gelernte Bewegungsprimitive als HMMs modelliert und mit der Kullback-Leibler-Distanz als Ähnlichkeitsmaß hierarchisch organisiert werden. Beobachtete Bewegungen können anschließend basierend auf diesem Ähnlichkeitsmaß

existierenden Bewegungsprimitiven zugeordnet oder als neue Bewegungsprimitive erkannt und gelernt werden.

Neben ihrem Einsatz zur Verarbeitung von Bewegungsdaten werden Techniken der Dimensionalitätsreduktion oft auch unmittelbar zur Segmentierung herangezogen. So wird von Barbič et al. (2004) ein auf PCA basierendes Segmentierungsverfahren vorgestellt, bei dem jeweils eine PCA für einen gefensterten Ausschnitt aus den beobachteten Bewegungsdaten durchgeführt und eine festgelegte Zahl der besten Hauptkomponenten beibehalten wird. Nachfolgende Frames werden nun ebenfalls in den resultierenden niedrigdimensionalen Raum transformiert und eine Segmentgrenze wird festgelegt, sobald der durch diese Transformation entstehende Rekonstruktionsfehler im hochdimensionalen Raum einen gewissen Schwellwert überschreitet. Darüber hinaus werden in (Barbič et al. 2004) auch eine Anwendung von probabilistischer PCA (PPCA) zur Segmentierung und eine Bewegungssegmentierung basierend auf Gauß-Mischmodellen vorgestellt, bei denen die Grenzen zwischen den einzelnen Normalverteilungen des Mischmodells die Segmentgrenzen festlegen. Weitere Arbeiten zur Segmentierung mit Dimensionalitätsreduktionstechniken basieren auf *Unsupervised Kernel Regression* (Steffen et al. 2009) oder auf dem Isomap-Algorithmus (Jenkins et al. 2002).

Lin et al. (2014) interpretieren die Segmentierung einer Bewegung als Problem der binären Klassifikation, bei dem ausgehend von Gelenkwinkeln oder Geschwindigkeiten jeder Datenpunkt einer Bewegung als Segmentgrenze oder als innerhalb eines Segments liegender Punkt klassifiziert wird. Nach einer (optionalen) Vorverarbeitung der Bewegungsdaten mit einer PCA werden hierfür vier verschiedene binäre Klassifikationsverfahren, u.a. SVMs, evaluiert, deren Parameter direkt aus vom Menschen erstellten Referenzsegmentierungen gelernt werden können. In (Zhao et al. 2013) wird ein Merkmal zur Beschreibung von Bewegungen basierend auf dem paarweisen Abstand zwischen Gelenken eingeführt und aus dem Clustering von manuell segmentierten Bewegungen mit Einträgen für jeweils eine Merkmalsdimension eine Datenbank beobachteter Sequenzen (als *Template Dictionary* bezeichnet)

erstellt, die eine Zuordnung neuer Bewegungsdaten erlaubt. Lan et al. (2015) fassen das Segmentierungsproblem als Problem der Textklassifikation auf, wofür die Frames einer Bewegung durch Posen aus einer durch hierarchisches Clustering generierten Menge an Schlüsselposen (*mo-vocabulary* als Vokabular für Bewegungen) ersetzt werden. Basierend auf den in einem gleitenden Fenster vorkommenden Posen wird anschließend mit der für Textklassifikation populären *Latent Dirichlet Allocation* (LDA) die Teilbewegung *mo-topics* (Themenfelder im Sinne einer Textklassifikation) zugeordnet und eine Segmentgrenze bei signifikanten Änderungen in der Verteilung dieser *mo-topics* erkannt. Vögele et al. (2014) definieren ein Ähnlichkeitsmaß für Frames einer Bewegung und bestimmen eine *Sparse Self-Similarity Matrix* der Ähnlichkeiten zwischen allen Frames. Innerhalb dieser Matrix werden nach dem Entfernen der Hauptdiagonale Techniken des *Region Growing* zur Identifikation isolierter Abschnitte eingesetzt, welche die Bewegungssegmente repräsentieren. In (Santos et al. 2015) werden Bewegungsdaten durch ein gleitendes Fenster mit adaptiver Größe unterteilt, ein Merkmalsvektor mit einer diskreten Fourier-Transformation aus den Beschleunigungen der Körperteile bestimmt und für die beobachteten Merkmalswerte mit einem Expectation-Maximation-Algorithmus ein Gauß-Mischmodell gelernt. Andere Ansätze, beispielsweise in (Fod et al. 2002), verwenden Schwellwerte für Geschwindigkeiten bzw. Nulldurchgänge der Geschwindigkeit (*zero-velocity crossings*) als vergleichsweise einfaches Segmentierungskriterium.

Über die genannten Arbeiten hinaus existieren zahlreiche weitere Ansätze zur Bewegungssegmentierung, welche beispielsweise auf DMPs (Meier et al. 2011) oder auf den von Laban et al. (2011) entwickelten Laban-Bewegungsstudien (Bouchard et al. 2007) basieren. Eine ausführliche Übersichtsarbeit zu Verfahren für die Segmentierung in Bewegungsprimitive, die weitere hier nicht genannte Segmentierungsverfahren vorstellt, findet sich in (Lin et al. 2016).

5.2 Taxonomie menschlicher Ganzkörper-Support-Posen

In der Vergangenheit wurden die in Abschnitt 5.1 beschriebenen Griff-taxonomien mit großem Erfolg für Fragestellungen des Greifens wie der autonomen Griffplanung und -auswahl, aber auch bei verwandten Problemstellungen wie beispielsweise der Konstruktion oder Leistungsevaluation neuer Roboterhände eingesetzt. Im Folgenden soll die am H²T entwickelte GKP-Taxonomie vorgestellt und erläutert werden, welche den Versuch unternimmt, den Ansatz von Griff-taxonomien auf Ganzkörper-Support-Posen zu übertragen, und als nach Kenntnisstand des Autors erste derartige Taxonomie statischer humanoider Ganzkörper-Support-Posen die Grundlage für die im Rahmen dieser Arbeit vorgestellten Beiträge in diesem Kapitel und in Kapitel 6 darstellt. Die Einführung der GKP-Taxonomie in diesem Abschnitt basiert dabei teilweise auf ihrer erstmaligen Vorstellung in (Borràs et al. 2015) und auf (Asfour et al. 2018; Borràs et al. 2017).

Die Motivation für die Etablierung der GKP-Taxonomie nach dem Vorbild von Griff-taxonomien findet sich zum einen in der Erkenntnis, dass das Verständnis von Multi-Kontakt-Bewegungen zunächst ein Verständnis und eine Beschreibung der sie prägenden Ganzkörperposen erfordert, und zum anderen in einer Dualität zwischen dem Greifproblem und dem Problem der Ganzkörperstabilität (Asfour et al. 2018), die in Abschnitt 1.1 bereits kurz erwähnt wurde. So kann das Problem der Ganzkörperstabilität als duales Greifproblem interpretiert werden, bei welchem ein humanoider Roboter durch die geschlossenen kinematischen Ketten seiner Kontaktpunkte in dem ihn „greifenden“ Raum stabilisiert werden muss, und es können demzufolge die im Bereich des Greifens angewandten Methoden eingesetzt werden, um beispielsweise die Stabilität eines derartigen „Griffes“ zu untersuchen (Farnioli et al. 2015). Analog zu einer Griff-taxonomie, welche mögliche Griff-typen für eine gegebene Handkinematik beschreibt, werden in der GKP-Taxonomie „Ganzkörper-Griff-typen“ als stabile Multi-Kontakt-Posen des gesamten Kör-

pers, die Support-Posen, beschrieben und, ebenfalls analog zu Griff-taxonomien, ein hochdimensionaler und somit rechnerisch schwer zu bewältigender Raum (der möglichen Ganzkörperposen mit mehreren Dutzend Bewegungsfreiheitsgraden) auf eine überschaubare Zahl einiger zu betrachtender Posentypen abgebildet. Über die reine Auswahl von Support-Posen als duales Problem zur Griffauswahl hinaus erlaubt die resultierende GKP-Taxonomie allerdings auch die Beschreibung möglicher Übergänge zwischen diesen Support-Posen, denen (anders als bei Griff-taxonomien) bei der Beschreibung und Generierung von Bewegungen als Sequenz dieser Posen eine wichtige Bedeutung zukommt.

5.2.1 Definition möglicher Support-Kontakte

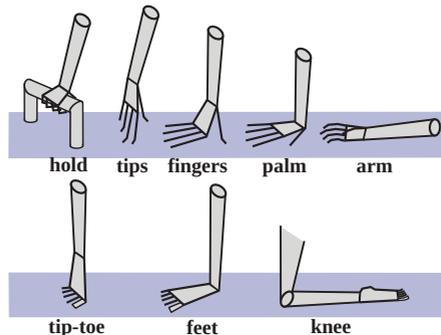


Abbildung 5.1: Mögliche Typen von Support-Kontakten für Arme und Beine (entnommen aus Borràs et al. 2015, ©2015 IEEE).

Eine wichtige Grundlage für die Definition von Support-Posen und darauf aufbauend der GKP-Taxonomie ist das Konzept des Support-Kontakts (im Folgenden auch kurz als *Support* bezeichnet), welcher einen Kontakt des Körpers zu seiner Umgebung darstellt, der zur Stabilität des Körpers beiträgt. Für die Konstruktion der GKP-Taxonomie wurde versucht, mögliche Vorkommen dieser Support-Kontakte in unterschiedliche Typen zu unterteilen und hierfür die in Abbildung 5.1 gezeigte Kategorisierung möglicher

Kontakttypen für die Arme und Beine als Extremitäten des Körpers vorgeschlagen (Torso-Kontakte werden weiter unten separat thematisiert). Zur Reduktion der Anzahl zu betrachtender Kontakttypen wurden im Anschluss (unter Berücksichtigung des Leistungsvermögens heutiger humanoider Roboter) die eher akrobatisch orientierten und auch vom Menschen im Alltag selten eingesetzten Kontakttypen *tips*, *fingers* und *tip-toe* für die Definition von Support-Posen ausgeschlossen.

Konkret werden für die Definition von Support-Posen also die folgenden der in Abbildung 5.1 gezeigten Typen von Support-Kontakten berücksichtigt:

- Support-Kontakte der Arme:
 1. Festhalte-Support durch Greifen der Umgebung, z.B. Greifen eines Geländers (Typ *hold*),
 2. Flächenkontakt mit der Handfläche zum Abstützen oder Anlehnen (Typ *palm*),
 3. Linienkontakt mit dem Unter-/Oberarm (Typ *arm*).
- Support-Kontakte der Beine:
 1. Flächenkontakt mit der Fußsohle (Typ *feet*),
 2. Linienkontakt mit dem Unterschenkel und Knie (Typ *knee*).

Formal wird ein Support-Kontakt k als Tupel

$$k = \{l, m, c, n\}$$

beschrieben mit dem den Kontakt herstellenden Körperteil l , dem Kontakttyp $m \in \{hold, palm, arm, feet, knee\}$, den Koordinaten des Kontakts c und der Normalen an der Kontaktoberfläche n . $K = \{k_1, \dots, k_n\}$ bezeichnet die Gesamtmenge aller Support-Kontakte zu einem gegebenen Zeitpunkt.

Es soll an dieser Stelle nochmals betont werden, dass sich die Definition von Support-Kontakten nur auf Kontakte des Körpers mit seiner Umgebung bezieht, welche durch die an ihnen wirkenden Kräfte zur Stabilität beitragen

(und nicht etwa auf reine Manipulationskontakte mit einem in der Hand gehaltenen Objekt). Durch die Angabe dieser Kontakte in K sind, ungeachtet von konkreten Gelenkwinkeln oder der Massen einzelner Körpersegmente, die *Struktur* der maßgeblichen Dynamikgleichungen und Stabilitätskriterien (Bretl et al. 2008) und somit beispielsweise die Rechenvorschriften eines entsprechenden Reglers eindeutig bestimmt, was die Voraussetzung zur Definition einer Support-Pose darstellt.

5.2.2 Definition von Support-Posen aus Support-Kontakten

Die Definition einer Support-Pose leitet sich von der Menge aller Support-Kontakte K ab und betrachtet die Anzahl dieser Support-Kontakte $|K|$ sowie die eingesetzten Körperteile l_i und die Kontakttypen m_i , wobei sich mögliche Support-Posen aus den unterschiedlichen kombinatorischen Optionen ergeben. Aufgrund der zuvor beschriebenen Zielsetzung, eine Definition von Support-Posen anhand der Struktur der herrschenden Dynamikgleichungen vorzunehmen, sind die konkreten Koordinaten eines Kontakts c_i für die Bestimmung der Support-Pose hingegen nicht von Interesse (wohl aber für die Bestimmung der Konfigurationsposen in Abschnitt 6.3), ebensowenig die n_i .

Für die festgelegte Zahl von $c_a = 3$ Typen an Arm-Supports (*hold*, *palm* und *arm*) und $c_l = 2$ Typen an Bein-Supports (*feet* und *knee*), ergibt sich kombinatorisch eine Zahl von

$$\sum_{n_a=0}^2 \binom{c_a + n_a - 1}{n_a} \cdot \sum_{n_l=1}^2 \binom{c_l + n_l - 1}{n_l} = (1 + 3 + 6) \cdot (2 + 3) = 50$$

möglichen Support-Posen mit $n_a \in \{0, 1, 2\}$ Arm-Supports und $n_l \in \{1, 2\}$ Bein-Supports². Da die Körperseite der in einer Support-Pose genutzten

² Support-Posen ohne Bein-Supports (z.B. hängende Posen) werden hier nicht weiter betrachtet, da ihr Einsatz weit außerhalb der Fähigkeiten gängiger humanoider Roboter (und vieler Menschen) liegt.

Support-Kontakte hier nicht berücksichtigt wird, beinhalten die resultierenden Support-Posen immer alle entsprechend seitensymmetrischen Fälle für Arm- bzw. Bein-Supports.

Von den 50 möglichen Support-Posen wurden manuell solche mit komplexen und offenkundig ungebräuchlichen Kombinationen von Support-Kontakten entfernt, beispielsweise alle $4 \cdot 3 = 12$ Support-Posen, die einen *knee*-Kontakt mit einem Support des Typs *arm* verbinden. Fünf weitere entfernte Support-Posen repräsentieren Kombinationen mit einem Fuß-Support und Kontakten des *arm*-Typs sowie Kombinationen, welche sowohl die Kontakttypen *hold* als auch *arm* beinhalten. Für die zweibeinigen Support-Posen mit einem oder auch zwei Supports des Typs *arm* hingegen wurden $2 + 1 = 3$ zusätzliche Posen eingefügt, sodass dieser *arm*-Kontakt unterschiedlich jeweils als Unterarm- oder Oberarm-Kontakt interpretiert werden kann (siehe hierzu z.B. SP 3.6 und SP 3.7 in Abbildung 5.2). Die resultierenden 36 Support-Posen bilden die Grundlage der GKP-Taxonomie.

Abschließend sei kurz erwähnt, dass es sich bei dieser Definition der Support-Posen als „Griffe“ im Sinne einer Griff-taxonomie um *non-prehensile grasps*³ handelt, deren Stabilität von der angenommenen Richtung der Schwerkraft (nach unten) abhängt.

5.2.3 Taxonomie zur Beschreibung von Support-Posen und ihren Transitionen

Abbildung 5.2 zeigt die resultierende GKP-Taxonomie mit einer Gesamtzahl von 46 Support-Posen, horizontal unterteilt in je 18 stehende und kniende Support-Posen sowie 10 ergänzende Ruheposen. Die insgesamt 36 stehenden und knienden Support-Posen entsprechen hierbei dem Ergebnis der vorangegangenen kombinatorischen Analyse von möglichen Support-Kontakten.

³ Bei den zuvor angesprochenen Griff-taxonomien wird eine Unterscheidung zwischen *prehensile grasps* und *non-prehensile grasps* vorgenommen, wobei bei den (häufiger vorkommenden) *prehensile grasps* das betrachtete Objekt fest von der Hand gegriffen wird, sodass Umgebungskräfte wie Gravitation keine Berücksichtigung finden.

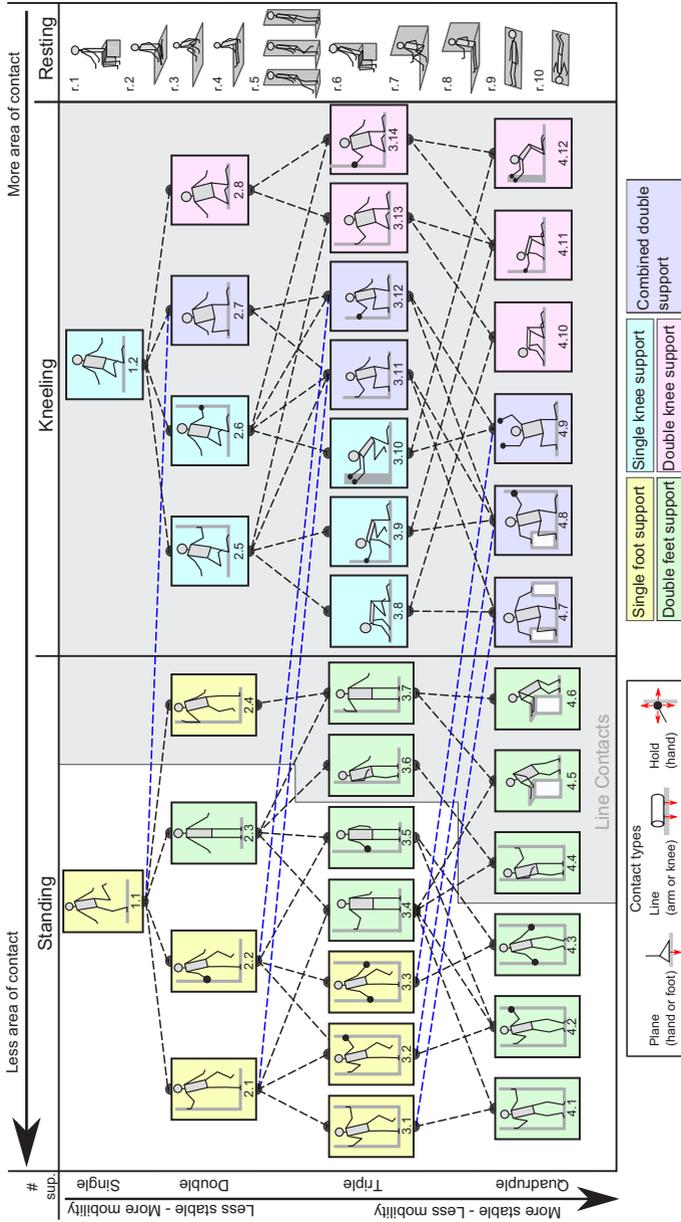


Abbildung 5.2: Am H²T entworfene Taxonomie menschlicher Ganzkörper-Support-Posen (entnommen aus Borràs et al. 2015, ©2015 IEEE).

Jede in der GKP-Taxonomie dargestellte Support-Pose beinhaltet hierbei alle die den jeweiligen Support-Bedingungen entsprechenden Posen und insbesondere, wie zuvor beschrieben, auch die entsprechend zum jeweiligen Piktogramm seitensymmetrischen Fälle (d.h. symmetrisch bezogen auf die Medianebene des Körpers).

In der Terminologie einer Griff-taxonomie korrespondieren die 36 stehenden und knienden Support-Posen mit Präzisionsgriffen, da Support-Kontakte hier nur durch Extremitäten, nicht aber durch den Torso selbst hergestellt werden⁴. Demgegenüber stehen die 10 manuell festgelegten Ruheposen, die zusätzlich zu den kombinatorisch ermittelten Support-Posen hinzugefügt wurden und durch Flächenkontakte mit dem Torso ein hohes, von Pose *r.1* bis *r.10* weiter steigendes Maß an Stabilität aufweisen. Aufgrund der komplexen Übergänge zwischen den Ruheposen und den restlichen Support-Posen und ihrer geringen Bedeutung für Fortbewegungsaufgaben sollen diese Posen im verbleibenden Teil dieser Arbeit jedoch nicht weiter betrachtet werden.

Die farbliche Kennzeichnung der Support-Posen in Abbildung 5.2 repräsentiert den Typ der jeweils verwendeten Fuß-Supports (wie erwähnt existieren in der GKP-Taxonomie keine Support-Posen ohne Fuß-Support) und die Support-Posen innerhalb der grau hinterlegten Fläche beinhalten Linienkontakte, womit sich insgesamt in der GKP-Taxonomie eine von links nach rechts zunehmende Kontaktfläche zeigt. Zusammen mit der vertikalen Anordnung der Support-Posen in Abbildung 5.2, basierend auf der jeweiligen Anzahl an Support-Kontakten, ergibt sich somit eine Zunahme der Stabilität der Support-Posen von der linken oberen Ecke (wenige Support-Kontakte, geringe Kontaktfläche) hin zur rechten unteren Ecke (viele Support-Kontakte, große Kontaktfläche) der GKP-Taxonomie. Nachdem gezeigt werden konnte, dass es sich bei Stabilität und Beweglichkeit – zumindest bei menschlichen Ganzkörperbewegungen – um sich gegenseitig in einem gewissen Sinne wi-

⁴ Manche Griff-taxonomien sehen neben Präzisions- und Kraftgriffen auch sogenannte *intermediate grasps* als eine dritte Griffkategorie vor, welcher im Sinne der GKP-Taxonomie Support-Posen mit Knie- oder Ellbogen-Supports zugeordnet werden könnten.

dersprechende Anforderungen handelt (H. J. Huang et al. 2011), erlaubt diese Anordnung in der GKP-Taxonomie eine intuitive Darstellung der Support-Posen hinsichtlich dieses Zielkonflikts.

Die in Abbildung 5.2 dargestellten Kanten repräsentieren mögliche Transitionen zwischen Support-Posen⁵ unter der Annahme, dass genau ein Support-Kontakt entweder hinzukommt oder gelöst wird und sich die Zahl der Support-Kontakte somit stets um eins verändert. Obwohl diese Einschränkung durch eine drastische Reduktion der kombinatorisch möglichen Transitionen zwischen Support-Posen und damit letztendlich der zu modellierenden Übergänge (siehe hierzu Kapitel 6) für eine Anwendung in der Robotik wünschenswert scheint, handelt es sich dennoch zunächst um eine Vermutung, die in Abschnitt 5.4 mit menschlichen Bewegungsdaten eingehend untersucht und dabei in sehr vielen Fällen, jedoch nicht immer bestätigt wird.

Basierend auf der Definition möglicher Support-Posen-Transitionen in der GKP-Taxonomie kann eine gegebene Ganzkörperbewegung als Abfolge von Support-Posen in einem Teilgraph der GKP-Taxonomie dargestellt werden, was in Abbildung 5.3 exemplarisch für Aufwärtssteigen auf einer Treppe mit Verwendung des Geländers gezeigt ist. Umgekehrt ist es jedoch auch möglich, die Planung komplexer Support-Posen-Transitionen als Wegfindungsproblem in dem durch mögliche Transitionen beschriebenen Graphen zu betrachten. Als Beispiel soll hier der Übergang von der stehenden SP 2.3 zur SP 2.6, die eine kniende Pose mit einem Festhalte-Kontakt der Hand repräsentiert, dienen. Für diesen Übergang lassen sich in der GKP-Taxonomie, neben beliebig vielen weiteren Pfaden, die folgenden zwei möglichen Abfolgen von Support-Posen identifizieren:

- Zunächst wird ein Fuß vom Boden gelöst (Übergang zu SP 1.1) und durch Übergang zur halb-knienden SP 2.7 ein Boden-Support mit dem Knie hergestellt. Im Anschluss wird der noch auf dem Boden stehen-

⁵ Blaue Kanten stellen Übergänge zwischen stehenden und knienden Support-Posen dar.

de Fuß angehoben (Übergang zu SP 1.2) und durch Herstellen des Festhalte-Kontakts mit der Hand die gewünschte Ziel-SP 2.6 erreicht.

- Alternativ wird zunächst im Stand der Festhalte-Kontakt hergestellt (Übergang zu SP 3.5) und erst dann der Fuß vom Boden angehoben (Übergang zu SP 2.2). Im Anschluss erfolgt ein Übergang zur halbknienden SP 3.12 mit Hand-Support und durch Anheben des auf dem Boden verbliebenen Fußes wird die Ziel-SP 2.6 erreicht.

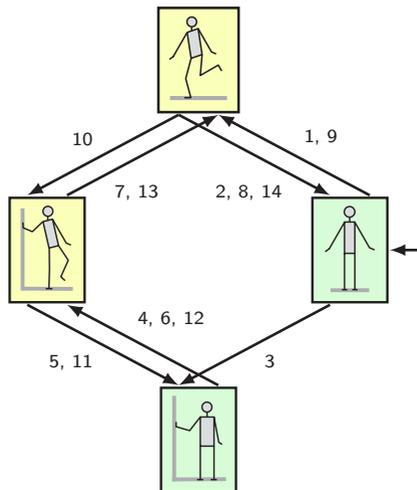


Abbildung 5.3: Repräsentation einer beispielhaften Ganzkörperbewegung (Treppe mit Geländer aufwärts steigen) als mit dem in Abschnitt 5.3 vorgestellten Segmentierungsverfahren ermittelte Abfolge von Support-Posen der GKP-Taxonomie. Die Beschriftung der Kanten kennzeichnet die Reihenfolge der Support-Posen-Transitionen.

Über dieses einführende Beispiel hinaus befasst sich Kapitel 6 mit einem Ansatz zur systematischen datengetriebenen Generierung derartiger Posen-Sequenzen basierend auf einem durch die Beobachtung des Menschen gelernten statistischen Modell.

Abschließend soll an dieser Stelle daran erinnert werden, dass für die Definition der Support-Posen in der GKP-Taxonomie die Position der Kon-

taktunkte c_i und ihre Normalen n_i nicht berücksichtigt wurden. Dies impliziert, dass sich weder die kinematische Konfiguration des Körpers noch die geometrische Gestalt der Umgebung in diesen Support-Posen widerspiegeln. Dennoch dürfen kinematische Variationen innerhalb einer Support-Pose, wie sie in Abbildung 5.4 beispielhaft für die SP 4.1 und SP 4.10 dargestellt sind, bei der Anwendung der GKP-Taxonomie in Kapitel 6 zur Generierung von Multi-Kontakt-Bewegungen selbstverständlich nicht ignoriert werden, da die kinematische Start- und End-Konfiguration einer Support-Posen-Transition die konkrete Ausführung des jeweiligen Übergangs maßgeblich beeinflusst. In Abschnitt 6.3 wird aus diesem Grund eine weitergehende datengetriebene Unterteilung der Support-Posen in Konfigurationsposen vorgestellt.

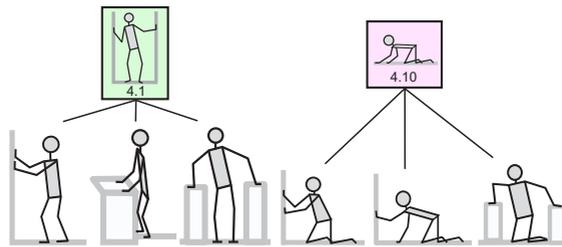


Abbildung 5.4: Beispielhafte kinematische Instanzen von SP 4.1 und SP 4.10 für unterschiedliche Umgebungsgeometrien (entnommen aus Borràs et al. 2015, ©2015 IEEE).

5.3 Segmentierung von Bewegungsaufnahmen zur Extraktion von Support-Posen

Der folgende Abschnitt befasst sich mit einem im Rahmen dieser Arbeit entwickelten und in (Mandery et al. 2015a) erstmals vorgestellten Verfahren zur Segmentierung menschlicher Ganzkörperbewegung, welches auf dem in Abschnitt 5.2 vorgestellten Konzept des Support-Kontakts bzw. der Support-Pose aufbaut. Dieses Segmentierungsverfahren erlaubt für eine gegebene Bewegungsaufnahme die Extraktion der verwendeten Support-Posen und

dementsprechend die Darstellung der Bewegungsaufnahme als Sequenz dieser Support-Posen. Mit Ausnahme einer fehlenden Unterscheidung zwischen *hold*- und *palm*-Kontakten können hierbei alle in Abschnitt 5.2 verwendeten Kontakttypen (*hold*, *palm*, *arm*, *feet* und *knee*) bzw. die mit diesen Kontakttypen korrespondierenden Support-Posen erkannt werden (*hold*-Kontakte werden im Folgenden als Kontakte des *palm*-Typs behandelt).

Da durch die fehlende Unterscheidung von *hold*- und *palm*-Kontakten jeder Kontakttyp eindeutig einem Körperteil zugeordnet werden kann, sollen die resultierenden Support-Posen im verbleibenden Teil dieser Arbeit der Einfachheit halber zusätzlich zu ihrer Bezeichnung in der GKP-Taxonomie auch anhand der jeweiligen Anzahl beteiligter Support-Kontakte benannt werden. Beispielsweise könnte SP 3.4 entsprechend dieser Notation auch als *2F-1H* (2 Fuß-Supports, 1 Hand-Support) bezeichnet werden.

Ausgangspunkt für die Segmentierung einer Ganzkörperbewegung ist die markerbasierte Erfassung dieser Bewegung mit der in Abschnitt 2.2.1 beschriebenen Methodik und ihre Darstellung im MMM-Format, da das bei der Beschreibung einer Bewegung in diesem Format eingesetzte und in Abschnitt 2.1 vorgestellte MMM-Referenzmodell als das für die Erkennung von Support-Kontakten maßgebliche Modell des menschlichen Körpers herangezogen wird. Eine wesentliche Anforderung stellt zudem dar, dass bei der Bewegungserfassung zusätzliche Marker auf den vom Menschen zur Herstellung von Support-Kontakten eingesetzten Umgebungsobjekten, ausgenommen den Boden, angebracht werden. Wie in Abschnitt 2.2.2 beschrieben, wird dadurch mit dem MMM-Framework eine Rekonstruktion der Objekt-Posen auf Basis von erstellten Objektmodellen ermöglicht. Diese Modelle müssen neben den Positionen der virtuellen Marker ebenfalls über ein, in der Regel manuell erstelltes, Mesh-Modell verfügen, welches die Geometrie des erfassten Umgebungsobjekts möglichst akkurat nachbilden soll. Durch den Einsatz geeigneter Bewegungsaufnahmen aus der KIT-Bewegungsdatenbank sind die drei genannten Voraussetzungen – die sensorische Erfassung von Umgebungsobjekten, die Verfügbarkeit von Modellen dieser Objekte und

die Repräsentation der erfassten Bewegung im MMM-Format – automatisch erfüllt.

Im Kern basiert das vorgestellte Segmentierungsverfahren auf einer Analyse der Geschwindigkeiten menschlicher Körperteile und ihrer Distanz zu Umgebungsobjekten. Die für die Bestimmung von Support-Kontakten berücksichtigten Körperteile sind hierbei Hände, Füße, Knie und Ellbogen. Für jeden dieser Körperteile wurden mit dem jeweiligen Körperteil korrespondierende Segmente des MMM-Referenzmodells ausgewählt⁶. Jedes dieser Modellsegmente verfügt über jeweils ein Mesh-Modell, welches im MMM-Framework zur Visualisierung des Modells eingesetzt wird, aber auch für die bei der Segmentierung notwendigen Distanzberechnungen benötigt wird. Um Verwechslungen mit durch die Segmentierung erzeugten Bewegungssegmenten zu vermeiden, werden Segmente des MMM-Referenzmodells im Folgenden als Körpersegmente (*body segments*) bezeichnet.

Algorithmus 2 zeigt die Vorgehensweise zur Berechnung von geglätteten Geschwindigkeiten für die berücksichtigten Körpersegmente in allen Frames einer Bewegungsaufnahme. Für jedes Körpersegment werden hierzu zunächst mit Hilfe der Vorwärtskinematik des MMM-Referenzmodells und des zentralen Differenzenquotienten (siehe Gleichung 2.2) die Geschwindigkeitsvektoren in jedem Frame bestimmt. Im nächsten Schritt werden ausgehend von den mit einem gleitenden Mittelwert ($\theta_{\text{smoothwin}}$) geglätteten Geschwindigkeitsvektoren die skalaren Geschwindigkeiten der Körpersegmente berechnet. Diese Vorgehensweise hat gegenüber einer direkten Berechnung und Glättung von skalaren Geschwindigkeiten den Vorteil, dass Rauschen in den kartesischen Positionen von Körpersegmenten in aufeinanderfolgenden

⁶ Für die Hand werden im MMM-Referenzmodell die Segmente der Handfläche $\{L/R\}W_joint$ und alle Fingersegmente $\{Left/Right\}FingerJoint\{1-5\}\{1-3\}y_joint$ berücksichtigt, für den Fuß die Segmente des Hinterfußes $\{L/R\}Ay_joint$ und des Vorderfußes $\{L/R\}Fx_joint$. Mit dem Knie sind die Segmente $\{L/R\}Ksegment_joint$ und mit dem Ellbogen die Segmente $\{L/R\}Esegment_joint$ assoziiert.

Frames ausgeglichen werden kann, das ansonsten zu einem systematischen Fehler in Form von zu hohen berechneten Geschwindigkeiten führen würde.

Algorithmus 2 Berechnung von geglätteten Geschwindigkeiten der Körpersegmente

```
1: function COMPUTEBODYSEGMENTVELOCITIES(motion)
2:   res  $\leftarrow$  Map()
3:   for all s  $\in$  bodySegments do
4:     // Bestimme Geschw.vektoren mit zentralem Diff.quotienten
5:     // (siehe Gleichung 2.2)
6:     velVecs  $\leftarrow$  List()
7:     for frame  $\leftarrow$  1 to motion.frameCount do
8:       Bestimme Pos. von s in frame (und benachb. Frames)
          mit Vorwärtskinematik
9:       v  $\leftarrow$  Geschwindigkeitsvektor von s in frame
10:      Hänge v an velVecs an
11:    end for
12:
13:    // Bestimme geglättete (skalare) Geschwindigkeiten
14:    speeds  $\leftarrow$  List()
15:    for frame  $\leftarrow$  1 to motion.frameCount do
16:      start  $\leftarrow$  max(frame -  $\frac{1}{2} \cdot \theta_{\text{smoothwin}}$ , 1)
17:      end  $\leftarrow$  min(frame +  $\frac{1}{2} \cdot \theta_{\text{smoothwin}}$ , motion.frameCount)
18:      Hänge  $\|(\sum_{i=start}^{end} \text{velVecs}[i]) / (end - start + 1)\|$  an speeds an
19:    end for
20:    res[s]  $\leftarrow$  speeds
21:  end for
22:  return res
23: end function
```

In Algorithmus 3 ist die ebenfalls erforderliche Berechnung der Distanz zwischen Körpersegmenten und den in der Bewegungsaufnahme enthaltenen Umgebungsobjekten dargestellt. Die Vorgehensweise beruht auf der durch Simox (Vahrenkamp et al. 2012) bereitgestellten Distanzberechnung, welche intern wiederum die PQP-Bibliothek (*Proximity Query Package*) zur dreiecksgenauen Abstandsberechnung zwischen den Mesh-Modellen einsetzt.

Algorithmus 3 Berechnung von Distanzen zwischen Körpersegmenten und der Umgebung

```

1: function COMPUTEBODYSEGMENTDISTANCES(motion)
2:   res  $\leftarrow$  Map()
3:   for all s  $\in$  bodySegments do
4:     sd  $\leftarrow$  List()
5:     skipCheck  $\leftarrow$  0
6:     for frame  $\leftarrow$  1 to motion.frameCount do
7:       minDist  $\leftarrow$   $+\infty$ 
8:       if skipCheck = 0 then
9:         for all o  $\in$  motion.objects do
10:          // Distanzberechnung zwischen Mesh-Modellen
11:          // basierend auf Simox
12:          dist  $\leftarrow$  Distanz zwischen MMM-Segment s und
13:          // Objekt o in frame
14:          if dist < minDist then
15:            minDist  $\leftarrow$  dist
16:          end if
17:        end for
18:        if minDist >  $\delta_{\text{dist}}(s)$  then
19:          skipCheck  $\leftarrow$  (dist -  $\delta_{\text{dist}}(s)$ ) /  $\theta_{\text{maxvel}}$ 
20:        end if
21:      else
22:        // Distanzberechnung nicht erforderlich
23:        skipCheck  $\leftarrow$  skipCheck - 1
24:      end if
25:      Hänge minDist an sd an
26:    end for
27:    res[s]  $\leftarrow$  sd
28:  return res
29: end function

```

Da diese dreiecksgenaue Abstandsberechnung mit einem nicht vernachlässigbaren Rechenaufwand verbunden ist und zur Bestimmung der minimalen Distanz eines Körpersegments zur Umgebung eine Berechnung der Abstände dieses Körpersegments zu allen Umgebungsobjekten erforderlich ist, steigt der insgesamt benötigte Rechenaufwand insbesondere bei einer größeren Zahl an Umgebungsobjekten oder beim Vorhandensein von Umgebungsobjekten mit einem geometrisch komplexen Mesh-Modell sehr stark an. Aus diesem Grunde wird in Algorithmus 3 eine Heuristik zur Reduktion benötigter Distanzberechnungen eingesetzt, die beim Überschreiten des für die Bestimmung von Support-Kontakten relevanten segmentspezifischen Distanzschwellwerts δ_{dist} (siehe unten) die Distanzberechnung für das jeweilige Körpersegment in einer gewissen Zahl an Folgeframes deaktiviert (*skipCheck*-Variable). Zur Bestimmung der zu überspringenden Folgeframes wird eine maximale, als vom Menschen unerreichbar angenommene Geschwindigkeit von Körpersegmenten θ_{maxvel} zugrunde gelegt, wobei erst beim Überschreiten dieser Geschwindigkeit Fehler bei der Bestimmung von Support-Kontakten durch entfallene Distanzberechnungen auftreten können.

Algorithmus 4 zeigt schließlich das Segmentierungsverfahren selbst, das für jeden Frame einer gegebenen Bewegungsaufnahme die korrespondierende Support-Pose zurückliefert. Hierfür werden zunächst Umgebungsobjekte wie beispielsweise eine vom Menschen gegriffene und bewegte Tasse, deren Position im Zeitablauf der Bewegungsaufnahme um mehr als θ_{prune} von ihrer Ausgangsposition abweicht, aus der Liste der berücksichtigten Umgebungsobjekte entfernt, da für diese Objekte keine der in Abschnitt 6.2 näher erläuterten Support- und Lehn-Affordanzen mehr angenommen wird⁷.

⁷ Da die Erkennung von Affordanzen nicht im Fokus dieser Arbeit stand, wurde diese simple Heuristik gewählt, welche für die bislang in der Evaluation eingesetzten Umgebungsobjekte zufriedenstellende Ergebnisse liefert. In zukünftigen Erweiterungen des Verfahrens könnten hier jedoch bereits etablierte Verfahren zur Erkennung von Affordanzen (siehe hierzu Abschnitt 6.2) ihren Einsatz finden und so auch eine Erkennung von Support-Kontakten mit mobilen Umgebungsobjekten, z.B. bei der Benutzung eines Skateboards, ermöglichen.

Algorithmus 4 Support-Posen-Segmentierung von Ganzkörperbewegungen

```

1: function SUPPORTPOSESEGMENTATION(motion)
2:   for all object  $\in$  motion.objects do
3:     for all frame  $\in$  object.frames do
4:       if  $\|frame.position - object.frames[0].position\| > \theta_{prune}$  then
5:         // Objekt nicht berücksichtigen (siehe Text)
6:         Entferne object aus motion.objects
7:       end if
8:     end for
9:   end for
10:  Füge Boden (Ebene mit  $z = 0$ ) zu motion.objects hinzu
11:  sv  $\leftarrow$  COMPUTEBODYSEGMENTVELOCITIES(motion)
12:  sd  $\leftarrow$  COMPUTEBODYSEGMENTDISTANCES(motion)
13:
14:  for all s  $\in$  bodySegments do
15:    suppCount[s]  $\leftarrow$  0 // Frames mit ununterbrochenem Kontakt
16:  end for
17:  res  $\leftarrow$  List()
18:  for i  $\leftarrow$  1 to motion.frameCount do
19:    suppSegments  $\leftarrow$  Set()
20:    for all s  $\in$  bodySegments do
21:      end  $\leftarrow$   $\min(i + \max(\theta_{segwin} - suppCount[s] - 1, 0),$ 
22:        motion.frameCount)
23:      isSupp  $\leftarrow$  true
24:      for j  $\leftarrow$  i to end do
25:        if sd[s][j]  $>$   $\delta_{dist}(s)$  or sv[s][j]  $>$   $\delta_{vel}$  then
26:          isSupp  $\leftarrow$  false, breche Schleife ab
27:        end if
28:      end for
29:      if isSupp then
30:        Füge s zu suppSegments hinzu
31:        suppCount[s]  $\leftarrow$  suppCount[s] + 1
32:      else
33:        suppCount[s]  $\leftarrow$  0
34:      end if
35:    end for
36:    Bestimme Supp.-Pose sp aus Supp.-Kontakten suppSegments
37:    Hänge sp an res an
38:  end for
39:  return res
40: end function

```

Im Anschluss wird der Boden, welcher sich durch das Setzen des Koordinatenursprungs bei der Kalibrierung des Bewegungserfassungssystem als Ebene mit $z = 0$ darstellt und deshalb nicht direkt lokalisiert werden muss, zur Liste der Umgebungsobjekte hinzugefügt und es werden wie zuvor beschrieben die Geschwindigkeiten und Umgebungsdistanzen der betrachteten Körpersegmente berechnet, auf deren Basis anschließend dann die Support-Kontakte und daraus die Support-Posen bestimmt werden.

Zur Feststellung eines Support-Kontakts für das Körpersegment s müssen sowohl ein Distanz- als auch ein Geschwindigkeitskriterium erfüllt sein: Einerseits muss der Abstand von s zum nächstgelegenen berücksichtigten Umgebungsobjekt niedriger sein als ein segmentspezifischer Schwellwert $\delta_{\text{dist}}(s)$. Andererseits muss die Geschwindigkeit von s unterhalb des Schwellwerts δ_{vel} liegen, womit die Annahme modelliert wird, dass bei einem *idealen* Support-Kontakt der sechsdimensionale Geschwindigkeitsvektor an der Kontaktfläche Null sein sollte⁸. Zur Erhöhung der Robustheit müssen beide Bedingungen außerdem nicht nur im jeweiligen Frame, sondern innerhalb eines gleitenden Fensters der Größe θ_{segwin} erfüllt sein, was in Algorithmus 4 durch $\text{suppCount}[s]$ als Zähler der zum jeweiligen Zeitpunkt bereits vergangenen Frames mit ununterbrochenem Support-Kontakt am Körpersegment s realisiert wird. Nach der Bestimmung der Support-Kontakte in einem Frame wird aus diesen die korrespondierende gültige⁹ Support-Pose der GKP-Taxonomie ermittelt. Flächenkontakte vom Typ *arm* bzw. *knee* werden hierbei implizit beim Vorhandensein eines Ellbogen- bzw. Knie-Supports

⁸ In der Praxis kann eine Geschwindigkeit von Null für das Körpersegment allein schon deshalb nicht beobachtet werden, weil bei den betrachteten Segmenten des MMM-Referenzmodells der für die Geschwindigkeitsberechnung relevante Ursprung dieser Segmente nicht exakt mit der Fläche des Support-Kontakts zusammenfällt, sondern innerhalb des Körpersegments liegt.

⁹ Hierzu sei auf Abschnitt 5.2 bzw. Abbildung 5.2 verwiesen: Nicht aus jeder möglichen Kombination von Support-Kontakten lässt sich eine in der GKP-Taxonomie vorgesehene Support-Pose erzeugen. So sieht die GKP-Taxonomie beispielsweise weder Support-Posen ohne Fuß-Supports, z.B. hängende Posen, noch eine singuläre Support-Pose mit null Kontakten vor, welche beispielsweise während Spring- oder Rennbewegungen auftreten könnte.

angenommen, während Kontakte vom Typ *hold* – alleine aufgrund der bei der Bewegungserfassung nicht berücksichtigten Finger – wie eingangs erwähnt nicht bzw. nur als *palm*-Kontakte erkannt werden können.

Die zur Durchführung des Segmentierungsverfahrens erforderlichen Parameterwerte wurden empirisch ermittelt. Für die Geschwindigkeits- und Distanzschwellwerte ergaben sich folgende Werte:

$$\delta_{\text{vel}} = 200 \text{ mm s}^{-1}, \quad \delta_{\text{dist}}(\text{Fuß}) = 15 \text{ mm}, \quad \delta_{\text{dist}}(\text{Hand}) = 15 \text{ mm}, \\ \delta_{\text{dist}}(\text{Knie}) = 35 \text{ mm}, \quad \delta_{\text{dist}}(\text{Ellbogen}) = 30 \text{ mm},$$

wobei der Wert von δ_{dist} für alle zum selben Körperteil gehörenden Segmente des MMM-Referenzmodells identisch ist. Für die vier den Ablauf der Segmentierung betreffenden θ -Parameter wurden folgende Werte ermittelt:

$$\theta_{\text{prune}} = 300 \text{ mm}, \quad \theta_{\text{maxvel}} = 10 \text{ m s}^{-1}, \quad \theta_{\text{smoothwin}} = 10, \quad \theta_{\text{segwin}} = 5.$$

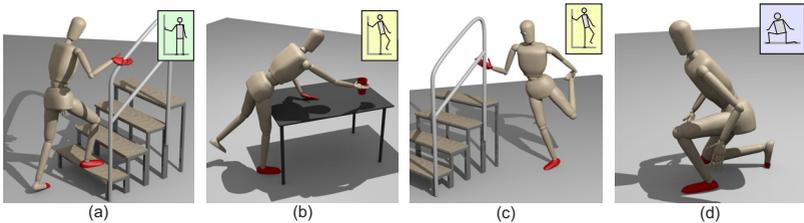


Abbildung 5.5: Bei der Segmentierung ermittelte Support-Posen für manuell ausgewählte Schlüsselbilder exemplarischer Ganzkörperbewegungen (entnommen aus Mandery et al. 2015a, ©2015 IEEE).

Zur Veranschaulichung des vorgestellten Segmentierungsverfahrens sind in Abbildung 5.5 manuell ausgewählte Schlüsselbilder aus vier exemplarischen Ganzkörperbewegungen zusammen mit den jeweiligen durch das Segmentierungsverfahren ermittelten Support-Posen gezeigt. Abbildung 5.6 zeigt hingegen die aus der Segmentierung einer einzelnen Ganzkörperbewegung, nämlich Aufwärtssteigen auf einer Treppe mit Verwendung des

Geländers, resultierende Abfolge von Support-Posen, wobei die extrahierten Support-Posen jeweils zum Zeitpunkt ihres ersten Auftretens eingezeichnet sind. Selbst bei der Betrachtung eines solchen isolierten Beispiels ist bereits deutlich zu erkennen, dass die längsten Segmente der Bewegung, in Abbildung 5.6 blau dargestellt, grundsätzlich mit der Schwungphase des Laufzyklus und somit mit Ein-Fuß-Support-Posen assoziiert sind (mehr dazu später). Über dieses exemplarische Beispiel hinaus befasst sich Abschnitt 5.4.1 im Evaluationsteil dieses Kapitels mit einer systematischen Validierung des Segmentierungsverfahrens durch einen Vergleich der erzeugten Sequenzen von Support-Posen mit einer manuell festgelegten Referenzsegmentierung.

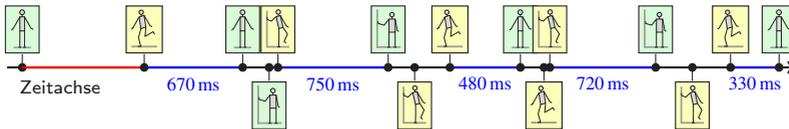


Abbildung 5.6: Aus der Segmentierung einer exemplarischen Bewegung (Treppe mit Geländer aufwärts steigen) resultierende Abfolge von Support-Posen (modifiziert nach Mandery et al. 2015a, ©2015 IEEE). Die blauen Segmente korrespondieren mit der Schwungphase des Laufzyklus innerhalb von Ein-Fuß-Support-Posen (mit oder ohne Hand-Support). Das rote Segment stellt die Zeitspanne vom Aufnahmebeginn bis zum Start der Bewegung bzw. zur ersten Posen-Transition dar und wird im Rahmen dieser Arbeit nicht ausgewertet.

5.4 Evaluation und Ergebnisse

Nach der Vorstellung der GKP-Taxonomie und dem darauf aufbauenden Segmentierungsverfahren für menschliche Ganzkörperbewegungen sollen in diesem Abschnitt die teilweise bereits in (Mandery et al. 2015a) publizierten Ergebnisse einer datengetriebenen Evaluation vorgestellt werden, welche sich zum einen mit einer Validierung des Segmentierungsverfahrens per se und zum anderen mit einer statistischen Analyse segmentierter menschlicher Multi-Kontakt-Bewegungen befasst. Für diese Analyse werden ausgehend von segmentierten Bewegungsaufnahmen die Auftretenshäufigkeiten spezi-

fischer Support-Posen und der Transitionen zwischen ihnen sowie die mit diesen Transitionen assoziierten Zeitauern untersucht. Die hieraus resultierenden Ergebnisse erlauben im Anschluss die Darstellung beobachteter Support-Posen-Transitionen als automatisch generierten Graphen und tragen schlussendlich zu einer partiellen Validierung der GKP-Taxonomie bei.

Der für die Evaluation eingesetzte Bewegungsdatensatz besteht aus insgesamt 121 Aufnahmen menschlicher Ganzkörperbewegung, welche mit der in Abschnitt 2.2.1 beschriebenen Methodik zur markerbasierten Bewegungserfassung aufgezeichnet und ausgehend von der durch die KIT-Bewegungsdatenbank bereitgestellten MMM-Repräsentationen dieser Bewegungen segmentiert wurden. Die 121 Bewegungsaufnahmen untergliedern sich hierbei in 19 Bewegungstypen der vier Kategorien Fortbewegung (Lokomotion), Fortbewegung mit Manipulation (Loko-Manipulation), Balancieren und kniende Bewegungen. Wie durch das Einbeziehen kniender Bewegungen bereits ersichtlich wird, wurde bei der Auswahl des Datensatzes versucht, ein möglichst vielfältiges Spektrum an menschlichen Bewegungsabläufen mit einem Fokus auf Multi-Kontakt-Bewegungen abzudecken, ungeachtet der tatsächlichen Eignung oder praktischen Relevanz dieser Bewegungsabläufe für die Bewegungsgenerierung bei heutigen humanoiden Robotern. Tabelle 5.1 zeigt eine vollständige Übersicht über die berücksichtigten Bewegungstypen und in Abbildung 5.7 sind manuell ausgewählte Schlüsselbilder einiger exemplarischer Bewegungsaufnahmen aus dem Datensatz dargestellt (siehe auch Abbildung 4.13 für Aufnahmen mit Balken und Stößen). Für einen detaillierten Ansicht der im Datensatz enthaltenen Bewegungen sei der Leser auch auf die in der KIT-Bewegungsdatenbank verfügbaren Videoaufnahmen verwiesen, auf die unter Verwendung der ebenfalls in Tabelle 5.1 angegebenen IDs der Bewegungsexperimente zugegriffen werden kann.

Tabelle 5.1: Zur Evaluation eingesetzter Bewegungsdatensatz, zusammen mit den Ergebnissen für die Validierung des in Abschnitt 5.3 vorgestellten Verfahrens zur Support-Posen-Segmentierung (modifiziert nach Mandery et al. 2015a, ©2015 IEEE). Angegeben ist, jeweils normiert auf eine Bewegungsaufnahme, die Gesamtzahl an extrahierten Support-Posen und die Anzahl der falschen sowie fehlenden Support-Posen beim Vergleich mit der Referenzsegmentierung. Die angegebenen IDs korrespondieren mit den Bewegungsexperimenten in der KIT-Bewegungsdatenbank unter der URL <https://motion-database.humanoids.kit.edu/details/motions/<ID>/>.

Bewegungstyp [ID(s) in KIT-Bew.datenb.]	Aufn.	# SP (Ø je Aufn.)		
		Ges.	Falsch	Fehl.
Fortbewegung (Lokomotion)				
Treppe mit Geländer hinab [385, 412, 460]	10	12,3	0,1 ¹⁰	2,5 ¹¹
Tr. mit Geländer hinauf [383, 410, 456, 606]	19	17,05	0,26 ¹⁰	0,16 ¹¹
Tr. mit Gel. hinauf/drehen/hinab [415, 463]	7	29,714	0,143 ¹⁰	4 ¹¹
Hindernis ausweichen mit Hand-Supp. [520]	5	13,2	0,4 ¹⁰	2,2 ¹¹
Über Balken mit Geländer laufen [521]	5	19,4	0,2	0
Fortbewegung mit Manipulation (Loko-Manipulation)				
Kiste treten mit Hand-Support [600]	6	12,5	0,33 ¹⁰	1,167 ¹¹
Lehnen um Tasse auf Tisch zu stellen [598]	6	15,33	0,17 ¹²	0
Lehnen um Tasse auf Tisch greifen [523]	5	5	0	0
Lehnen um Tasse aus Luft greifen [597]	7	15	0,14 ¹⁰	0
Lehnen um Tisch zu wischen [599]	6	12,5	0,5	0
Zweihändiges Tragen von Kiste [529]	6	13,833	0,833 ¹³	0,667 ¹¹
Abstützen um Tasse aufzuheben [527]	3	4,67	0,67 ¹⁴	0
Balancieren				
Abstützen nach Stoß von hinten [531]	5	6,2	0	0
Abstützen nach Stoß von links [530]	9	9,3	0,11	0
Schuhsohle ansehen mit Hand-Support [604]	2	11	0	0
Gleichgew.verlust mit Hand-Support [607]	5	10,8	0	0,2
Zweihändiges Abstützen auf Tisch [601]	4	16,25	1,25 ¹⁰	0
Kniende Bewegungen				
Aus dem Stand hinknien [515]	4	8	0	0
Aus Knien aufstehen [516, 517]	7	7,857	0	0
Gesamt:	121	239,94	5,11	10,89
			2,13 %	4,53 %



Abbildung 5.7: Manuell ausgewählte Schlüsselbilder einiger der im Bewegungsdatensatz enthaltenen Bewegungsaufnahmen (siehe Tabelle 5.1).

5.4.1 Validierung des Segmentierungsverfahrens

Der erste Teil der Evaluation befasst sich mit einer Validierung des in Abschnitt 5.3 beschriebenen Segmentierungsverfahrens als solchem, indem die Funktionsfähigkeit und Robustheit des Verfahrens durch einen Vergleich resultierender Segmentierungen mit manuell festgelegten Referenzsegmentierungen untersucht wird. Dieser Validierung kommt insofern eine entscheidende Bedeutung zu, als die Segmentierung menschlicher Bewegungsaufnahmen mit diesem Verfahren die Grundlage nicht nur für alle weiteren Ergebnisse der hier präsentierten Evaluation, sondern auch für das in Kapitel 6 vorgestellte Verfahren zur datengetriebenen Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen bildet.

Im Zuge der Validierung wurde für jede der 121 Bewegungsaufnahmen die aus der Segmentierung der MMM-Repräsentation resultierende Abfolge von Support-Posen durch manuellen Vergleich mit der Visualisierung dieser MMM-Repräsentation und der korrespondierenden Videoaufzeichnung überprüft und die Zahl an *falschen* und *fehlenden* Support-Posen bestimmt.

¹⁰Falsche Support-Posen korrespondieren mit im Bewegungsablauf zwischenzeitlich fälschlicherweise nicht erkannten Hand-Supports.

¹¹Fehlende Support-Posen korrespondieren mit fälschlicherweise nicht erkannten Doppel-Fuß-Supports während Laufbewegungen.

¹²Falsche Support-Posen korrespondieren mit fälschlicherweise erkannten Fuß-Supports.

¹³Falsche Support-Posen korrespondieren mit im Bewegungsablauf zwischenzeitlich fälschlicherweise nicht mehr erkannten Fuß-Supports.

¹⁴Falsche Support-Posen korrespondieren mit fälschlicherweise erkannten Hand-Supports.

Als *falsch* wurde eine vom Segmentierungsverfahren ausgegebene Support-Pose bewertet, wenn diese Pose nicht ebenfalls durch den Menschen an der jeweiligen Stelle der Bewegungsaufnahme identifiziert wurde, während eine *fehlende* Support-Pose eine vom Menschen vorgegebene Pose kennzeichnet, die aber vom Segmentierungsverfahren nicht an dieser Stelle erkannt wurde.

Tabelle 5.1 zeigt, jeweils auf eine Bewegungsaufnahme normiert, für jeden der 19 Bewegungstypen die Gesamtzahl der erkannten Support-Posen und die jeweilige Anzahl an falschen und fehlenden Support-Posen, sowie die aggregierten Summen der Werte aus den einzelnen Bewegungstypen. Hierbei zeigt sich, dass mit Hilfe des Segmentierungsverfahrens die überwiegende Mehrheit der Support-Posen korrekt erkannt werden konnten und bei einem großen Teil der Bewegungsaufnahmen, bzw. teilweise sogar bei allen Aufnahmen eines Bewegungstyps, weder falsche noch fehlende Support-Posen auftraten. Insgesamt wurden etwa 2 % der durch die Segmentierung extrahierten Support-Posen als falsch bewertet, während ca. 4,5 % der vom Menschen vorgegebenen Support-Posen als fehlende Posen nicht vom Segmentierungsverfahren erkannt werden konnten.

Für die beobachteten falschen bzw. fehlenden Support-Posen können zwei Hauptursachen identifiziert werden (siehe hierzu auch die in Tabelle 5.1 referenzierten Fußnoten):

1. Den Großteil der fehlenden, d.h. nicht erkannten Support-Posen repräsentieren Doppel-Fuß-Supports (mit oder ohne ergänzende Hand-Supports). Die Ursache hierfür liegt allerdings nicht in einer intrinsischen Schwierigkeit der Erkennung dieser Posen, sondern darin, dass bei vielen der betrachteten menschlichen Bewegungsabläufen diese Pose, verglichen mit anderen Support-Posen, nur sehr kurz auftritt¹⁵. So besteht ein typischer Doppel-Fuß-Support beim Laufen nur für wenige Frames, was eine zuverlässige Erkennung der gleichzeitigen

¹⁵Der am Anfang und Ende einer Bewegungsaufnahme stets vorliegende lang andauernde Doppel-Fuß-Support kann aus diesem Grund auch sehr zuverlässig erkannt werden.

Kontaktsituation für beide Füße innerhalb dieses kurzen Zeitraums entsprechend anspruchsvoll gestaltet. Neben einer generellen Erhöhung von $\delta_{\text{dist}}(\text{Fuß})$ – mit dem entsprechenden Risiko von mehr fälschlicherweise erkannten Fuß-Supports – könnte zur Verringerung dieses Problems das Segmentierungsverfahren um eine auf Vorwissen basierende Heuristik erweitert werden, die beispielsweise bei aufeinanderfolgenden Support-Wechseln der Füße mit veränderten Schwellwerten gezielt nach einem dazwischenliegenden Doppel-Fuß-Support sucht. Angesichts der relativ geringen Tragweite des Problems und zur Wahrung der Allgemeinheit des angewandten Verfahrens wurde im Rahmen dieser Arbeit hierauf jedoch verzichtet.

2. Hand-Supports können in seltenen Fällen nicht konsistent erkannt werden, wodurch bei Unterbrechungen in der Erkennung eines länger andauernden Hand-Supports dazwischenliegende falsche Support-Posen (ohne diesen Hand-Support) eingefügt werden. Augenscheinliche Ursache dieses Problems ist zumeist die fehlende Berücksichtigung der Hand-Pose bei der markerbasierten Bewegungserfassung (siehe Tabelle 2.2), wodurch für die Gelenkwinkel der Finger lediglich die ausgestreckte Nullpose angenommen wird. Neben dem offensichtlichen Lösungsansatz, beispielsweise mit Hilfe von Datenhandschuhen eine akkurate Erfassung der Hand-Pose zu ermöglichen, kann dieses Problem – unter Inkaufnahme möglicher fälschlicherweise erkannter Hand-Supports – durch die Wahl größerer Werte für $\delta_{\text{dist}}(\text{Hand})$ verringert werden.

Wie in Tabelle 5.1 ersichtlich, sind diese beiden Fehlerquellen für fast alle der ca. 2 % bzw. 4,5 % als falsch bzw. fehlend bewerteten Support-Posen verantwortlich und es treten darüber hinaus auf dem hier betrachteten Bewegungsdatensatz praktisch keine weiteren Fehler in den extrahierten Support-Posen-Sequenzen auf. Zusammenfassend kann daher festgestellt werden, dass das validierte Verfahren für die Segmentierung menschlicher Ganzkör-

perbewegung hinsichtlich ihrer Support-Posen und somit für die nachfolgende Analyse geeignet ist.

5.4.2 Darstellung von Support-Posen-Transitionen als Graph

Die bei der Segmentierung extrahierten Sequenzen von Support-Posen erlauben eine Darstellung der beobachteten Transitionen zwischen den Support-Posen als gerichteten Graphen, der datengetrieben aus den analysierten Bewegungsaufnahmen erstellt werden kann. Die Visualisierung dieses Graphen ermöglicht anschließend einen intuitiven Vergleich der aus der Segmentierung resultierenden Support-Posen und ihrer Transitionen mit den Vorgaben der GKP-Taxonomie und bietet so die Möglichkeit zur Validierung der bei der Definition dieser GKP-Taxonomie unterstellten Prämissen. Mit dem in dieser Arbeit vorgestellten Segmentierungsverfahren kann diese Validierung allerdings zwangsläufig nur partiell erfolgen, da aufgrund der bestehenden Limitationen keine Support-Kontakte der Typen *hold* und *palm* unterschieden werden können (siehe Abschnitt 5.3).

Abbildung 5.8 zeigt den für die analysierten 121 Bewegungsaufnahmen erzeugten Graphen. Zur Erstellung des Graphen wurden die vollständigen Posen-Sequenzen einschließlich der Start- und End-Support-Posen berücksichtigt, wodurch sich jede betrachtete Ganzkörperbewegung als im Doppel-Fuß-Support (SP 2.3) startender und endender, möglicherweise Zyklen beinhaltender, Pfad in diesem Graphen darstellen lässt. Wie zu erkennen ist, werden in den betrachteten Ganzkörperbewegungen 11 der durch die GKP-Taxonomie definierten Support-Posen erreicht, gleichzeitig wurde durch das Segmentierungsverfahren in den Bewegungsdaten jedoch keine der für die Definition der GKP-Taxonomie ausgeschlossenen Support-Posen (siehe Abschnitt 5.2) identifiziert.

Die Beschriftung der Kanten in Abbildung 5.8 gibt die Auftretenshäufigkeiten der jeweiligen Support-Posen-Transitionen an. Schwarze Kanten

korrespondieren hierbei mit Transitionen, welche durch die GKP-Taxonomie vorgesehen sind, rote Kanten mit nicht vorgesehenen Transitionen. Unmittelbar ersichtlich ist, dass Transitionen zwischen SP 1.1 und SP 2.1, zwischen SP 1.1 und SP 2.3 und zwischen SP 2.1 und SP 3.4 sehr häufig beobachtet werden können, während Transitionen zwischen den häufig besuchten SP 2.3 und SP 3.4 vergleichsweise selten (59 bzw. 64 mal) festgestellt werden. Wie nachfolgend zu sehen ist, wird dies durch die geringe in SP 2.3 verbrachte Zeitdauer verursacht.

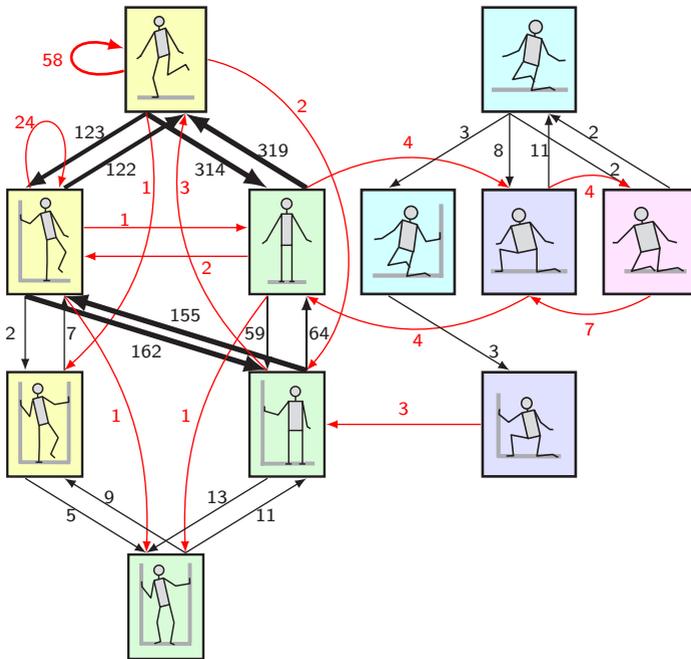


Abbildung 5.8: Durch Segmentierung der analysierten Bewegungsaufnahmen erzeugter gerichteter Graph für die Transitionen zwischen Support-Posen (entnommen aus Mandery et al. 2015a, ©2015 IEEE). Die Beschriftung einer Kante gibt die Häufigkeit der jeweiligen Transition an, wobei rote Kanten von der GKP-Taxonomie nicht vorgesehene Übergänge signalisieren.

Bei der Betrachtung der roten Kanten, welche durch die GKP-Taxonomie nicht vorgesehene Support-Posen-Transitionen repräsentieren, zeigt sich, dass der überwiegende Teil dieser Transitionen bei reflexiven Transitionen in SP 1.1 und SP 2.1 auftritt, die mit dem bereits in Abschnitt 5.4.1 identifizierten Problem der nicht erkannten Doppel-Fuß-Supports für ca. 4,5 % der Support-Posen korrespondieren. Ohne Betrachtung dieser reflexiven Transitionen treten die meisten der (wenigen) verbleibenden roten Kanten vor allem bei Übergängen innerhalb der knienden Support-Posen und zwischen stehenden und knienden Support-Posen auf. Für diese knienden Bewegungsabläufe zeigt sich in Videoaufzeichnungen der Bewegungen ein nicht stationärer, auf dem Boden „rutschender“ Fuß-Support, welcher den Annahmen für Support-Posen in Abschnitt 5.2 widerspricht und so innerhalb der GKP-Taxonomie nicht vorgesehene Support-Posen-Transitionen ermöglicht¹⁶.

Selbstverständlich wird die Struktur des in Abbildung 5.8 dargestellten Graphen maßgeblich von den analysierten Bewegungsaufnahmen beeinflusst, die aus diesem Grund möglichst vielfältig und verschiedenartig gewählt wurden. Dennoch werden insbesondere die knienden Support-Posen nur selten erreicht und sollen daher im Rahmen der folgenden statistischen Analyse nicht mehr weiter betrachtet werden.

5.4.3 Häufigkeit spezifischer Support-Posen und ihrer Transitionen

Der folgende Abschnitt befasst sich mit einer statistischen Analyse zur Auftretenshäufigkeit spezifischer Support-Posen und ihrer Transitionen. Hierfür kommt weiterhin der in Tabelle 5.1 dargestellte Bewegungsdatensatz zum Einsatz, es werden jedoch wie zuvor erwähnt die knienden Bewegungsaufnahmen für den verbleibenden Teil der Evaluation ausgenommen. Von

¹⁶Ein Beispiel hierfür findet sich bei der Transition von der (dunkelblauen) SP 2.7 zur (rosafarbenen) SP 2.8, bei der in vier Fällen der auf dem Boden stehende Fuß überhaupt nicht angehoben, sondern unter Beibehaltung des Supports nach hinten geschoben wird, was einen unmittelbarer Übergang zur SP 2.8 ermöglicht.

den 121 Bewegungsaufnahmen verbleiben somit 110 Bewegungsaufnahmen aus den drei Kategorien Lokomotion, Loko-Manipulation und Balancieren. Die zum Beginn und Ende jeder Bewegungsaufnahme auftretenden Doppel-Fuß-Supports (SP 2.3) verfügen als erste Support-Pose vom Beginn einer Bewegungsaufnahme bis zur ersten Support-Posen-Transition (siehe hierzu auch das rote Segment in Abbildung 5.6) bzw. als letzte Support-Pose bis zum Ende der Bewegungsaufnahme über eine variable und von der konkreten Bewegung unabhängige Zeitdauer und werden im Rahmen der verbleibenden Evaluation daher nicht weiter betrachtet.

Tabelle 5.2: Relative Auftretenshäufigkeiten spezifischer Support-Posen-Transitionen, kniende Posen ausgeschlossen (modifiziert nach Mandery et al. 2015a, ©2015 IEEE).

Start-SP (<i>in Tax.</i>)	End-SP						Gesamt
	1F	1F-1H	2F	2F-1H	2F-2H	1F-2H	
1F (1.1)	4,38 %	9,30 %	22,90 %	0,15 %	–	0,08 %	36,81 %
1F-1H (2.1)	9,15 %	1,81 %	0,08 %	12,24 %	0,08 %	0,15 %	23,51 %
2F (2.3)	16,02 %	0,15 %	×	3,48 %	0,08 %	–	19,73 %
2F-1H (3.4)	0,23 %	11,72 %	4,61 %	×	0,98 %	–	17,54 %
2F-2H (4.1)	–	–	–	0,83 %	×	0,68 %	1,51 %
1F-2H (3.1)	–	0,53 %	–	–	0,38 %	×	0,91 %

Insgesamt wurden 1323 Transitionen zwischen den sechs vorkommenden stehenden Support-Posen erkannt. Tabelle 5.2 zeigt die relative Auftretenshäufigkeit der beobachteten Support-Posen-Transitionen, jeweils bezogen auf die Gesamtzahl aller Transitionen. Die letzte Spalte von Tabelle 5.2 gibt hierbei die relative Häufigkeit des Vorkommens der jeweiligen Support-Pose als Start-Pose einer Transition und damit gleichzeitig die Sortierung der Support-Posen (beginnend mit der häufigsten) an. „–“ kennzeichnet mögliche Support-Posen-Transitionen in Tabelle 5.2, welche überhaupt nicht festgestellt werden konnten, und „×“ reflexive Transitionen auf der Diago-

nalen, die nicht beobachtet wurden, aber auch in der GKP-Taxonomie nicht vorgesehen sind¹⁷.

Bei den am häufigsten beobachteten Support-Posen-Transitionen handelt es sich, wenig überraschend, um die Übergänge zwischen Einzel- und Doppel-Fuß-Supports, SP 1.1 → SP 2.3 (22,90 %) und SP 2.3 → SP 1.1 (16,02 %). Diese Support-Posen-Transitionen entsprechen einer normalen Laufbewegung ohne Hand-Supports und die mit ihnen assoziierten Übergänge bzw. ihre Modellierung, z.B. mit Hilfe von Bewegungsprimitiven, sind dementsprechend gut erforscht. Weitere häufige Transitionen stellen die Übergänge zwischen SP 1.1 und SP 2.1 dar (mit 9,30% bzw. 9,15% nahezu gleich häufig), die das Herstellen bzw. Lösen eines zusätzlichen Hand-Supports bei genau einem bestehenden Fuß-Support repräsentieren. Die ansonsten äquivalenten, aber von einem Doppel-Fuß-Support ausgehenden Transitionen treten hingegen mit 3,48 % bzw. 4,61 % deutlich seltener auf.

5.4.4 Zeitdauer von Support-Posen-Transitionen

Über reine Auftretenshäufigkeiten hinaus sind ebenfalls die für Support-Posen-Transitionen benötigten Zeitdauern und ihre Varianzen von Interesse, beispielsweise für eine Abschätzung, inwieweit sich unterschiedliche Ansätze zur Modellierung von Bewegungsprimitiven für die Realisierung dieser Transitionen eignen. Der letzte Teil dieser Evaluation befasst sich daher mit einer statistischen Analyse der mit spezifischen Support-Posen-Transitionen assoziierten Zeitdauern.

Tabelle 5.3 zeigt die für spezifische Support-Posen-Transitionen benötigten Zeitdauern. Der erste Wert jeder Tabellenzeile gibt hierbei, analog zu Tabelle 5.2, den Anteil der in der jeweiligen Transition verbrachten Zeit an der kumulierten Zeitdauer aller Support-Posen-Transitionen an. Der zweite

¹⁷Die in der Diagonalen von Tabelle 5.2 genannten Fälle SP 1.1 → SP 1.1 und SP 2.1 → SP 2.1 entstehen durch fälschlicherweise nicht erkannte dazwischenliegende Doppel-Fuß-Supports, entsprechend den reflexiven roten Kanten in Abbildung 5.8.

Wert hingegen repräsentiert den Quotienten aus diesem Werts und der korrespondierenden Auftretenshäufigkeit aus Tabelle 5.2 und stellt somit ein Maß für die „relative Länge“ einer Support-Posen-Transition im Vergleich zu anderen Transitionen dar. Die letzte Spalte zeigt wiederum die kumulierten Werte aller Transitionen mit der jeweiligen Start-Support-Pose.

Tabelle 5.3: Benötigte Zeitdauern für spezifische Support-Posen-Transitionen, kniende Posen ausgeschlossen (modifiziert nach Mandery et al. 2015a, ©2015 IEEE). Angegeben ist sowohl die Zeitdauer als Anteil an den insgesamt segmentierten Bewegungsdaten (erster Wert), als auch die bezogen auf die Auftretenshäufigkeit der jeweiligen Support-Posen-Transition (siehe Tabelle 5.2) normierte Zeitdauer (zweiter Wert).

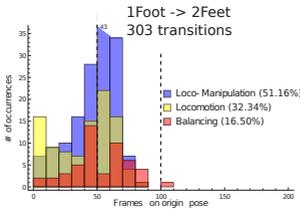
Start-SP (in Tax.)	End-SP						Gesamt
	1F	1F-1H	2F	2F-1H	2F-2H	1F-2H	
1F (1.1)	5,69 % 123 %	7,90 % 85 %	25,56 % 112 %	0,26 % 173 %	–	0,04 % 50 %	39,44 % 107 %
1F-1H (2.1)	13,64 % 149 %	2,26 % 124 %	0,03 % 38 %	16,59 % 136 %	0,02 % 25 %	0,02 % 13 %	32,57 % 139 %
2F (2.3)	10,05 % 63 %	0,04 % 27 %	×	2,23 % 64 %	0,06 % 75 %	–	12,38 % 62 %
2F-1H (3.4)	0,07 % 30 %	4,38 % 37 %	5,31 % 115 %	×	0,15 % 15 %	–	9,92 % 57 %
2F-2H (4.1)	–	–	–	1,22 % 147 %	×	0,75 % 110 %	1,97 % 130 %
1F-2H (3.1)	–	1,27 % 240 %	–	–	2,45 % 645 %	×	3,72 % 409 %

Basierend auf Tabelle 5.3 lassen sich Support-Posen-Transitionen grob in lange Transitionen, bei denen der zweite angegebene Wert größer als 100 % ist, und kurze Transitionen unterteilen. Lange Transitionen sind hierbei vor allem mit zur Fortbewegung eingesetzten Einzel-Fuß-Supports assoziiert, während zur Fortbewegung eingesetzte Doppel-Fuß-Supports meistens mit kürzeren Support-Posen-Transitionen einhergehen. Dies zeigt sich dadurch,

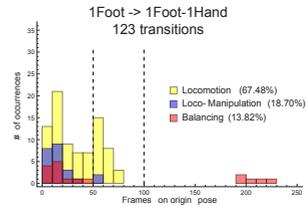
dass die kürzesten Support-Posen-Transitionen ausgehend von Doppel-Fuß-Supports (SP 2.3 mit 62 % und SP 3.4 mit 57 %) beobachtet werden, während die einen beidhändigen Support beinhaltende SP 4.1 vorwiegend bei Manipulationsaufgaben¹⁸ zum Einsatz kommt und entsprechend für eine längere Zeitdauer (130 %) beibehalten wird. In einer Untersuchung von Winter (1984) zu langsamen und schnellen menschlichen Gangbewegungen lag die im Doppel-Fuß-Support, also in der Transition SP 2.3 → SP 1.1, zugebrachte Zeit zwischen 11 und 19 Frames, während im Einzel-Fuß-Support (SP 1.1 → SP 2.3) 38 bis 52 Frames verbracht wurden. Dieses Verhältnis von etwa 3:1 bestätigt sich beim Blick auf reine Laufbewegungen, wird in Tabelle 5.3 jedoch durch die Verwendung der jeweiligen Support-Posen als Bestandteil komplexerer Transitionen mit Hand-Supports und durch die Verwendung des Doppel-Fuß-Supports in langdauernden Support-Posen-Transitionen zur (Loko-)Manipulation etwas vermindert.

Die in Abbildung 5.9 dargestellten Histogramme erlauben durch die farbliche Aufschlüsselung der unterschiedlichen Bewegungskategorien eine detaillierte Untersuchung der Zeitauern für die acht relevantesten Support-Posen-Transitionen. Die mit Lokomotion assoziierten Support-Posen-Transitionen, für die sich in Abbildung 5.9a, Abbildung 5.9b und Abbildung 5.9f eine bimodale Verteilung der Zeitauern zeigt, sind hierbei in gelber Farbe dargestellt. In den ersten beiden Fällen kann für diese bimodale Verteilung eine Mischverteilung zweier Normalverteilungen mit den Parametern $N(\mu = 11,76, \sigma = 8,60)$ bzw. $N(\mu = 53,89, \sigma = 11,35)$ im Fall von Abbildung 5.9a und $N(\mu = 15,99, \sigma = 9,68)$ bzw. $N(\mu = 55,85, \sigma = 8,72)$ im Fall von Abbildung 5.9b geschätzt werden (bei einem Konfidenzniveau von 96,9 % bzw. 94,4 %). Dies deutet darauf hin, dass die in Abbildung 5.9a dargestellten Transition SP 1.1 → SP 2.3 sowohl die Rolle einer langen Transition mit durchschnittlich 54 Frames, als auch die einer kurzen Transition

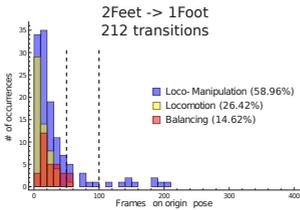
¹⁸Die im Datensatz enthaltenen Bewegungsaufnahmen für Fortbewegung boten keine Möglichkeit für beidseitige Hand-Supports, sodass SP 4.1 hier nicht im Kontext von Fortbewegungsaufgaben auftreten kann.



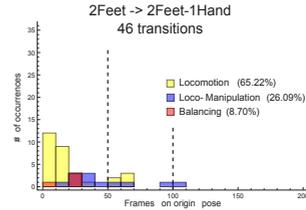
(a) 1F (SP 1.1) → 2F (SP 2.3).



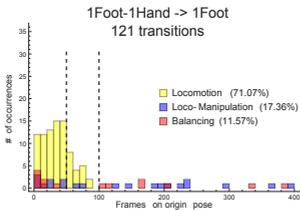
(b) 1F (SP 1.1) → 1F-1H (SP 2.1).



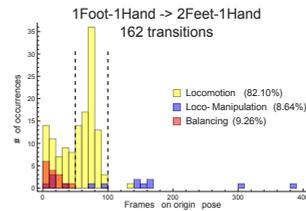
(c) 2F (SP 2.3) → 1F (SP 1.1).



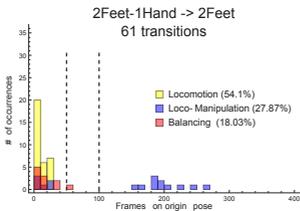
(d) 2F (SP 2.3) → 2F-1H (SP 3.4).



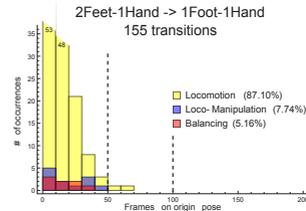
(e) 1F-1H (SP 2.1) → 1F (SP 1.1).



(f) 1F-1H (SP 2.1) → 2F-1H (SP 3.4).



(g) 2F-1H (SP 3.4) → 2F (SP 2.3).



(h) 2F-1H (SP 3.4) → 1F-1H (SP 2.1).

Abbildung 5.9: Verteilung der Zeitdauern als Histogramme mit einer Klassengröße von 100 ms (entsprechend 10 Frames) für spezifische Support-Posen-Transitionen (entnommen aus Mandery et al. 2015a, ©2015 IEEE). Aufgrund der unterschiedlichen X-Skalierung der Histogramme kennzeichnen zwei gestrichelte Linien jeweils eine Zeitdauer von 0,5 s bzw. 1 s.

mit etwa 12 Frames übernehmen kann (ähnlich bei der in Abbildung 5.9b dargestellte Transition SP 1.1 → SP 2.1). Auch für die in Abbildung 5.9f dargestellte Transition SP 2.1 → SP 3.4 lassen sich zwei Normalverteilungen mit $N(\mu = 36,47, \sigma = 26,59)$ bzw. $N(\mu = 73,11, \sigma = 8,32)$ schätzen, allerdings deutet das Konfidenzniveau von nur 80 % auf eine unzureichende Datenmenge für diese Support-Posen-Transition hin. Dennoch ist auch hier ersichtlich, dass diese Support-Posen-Transition sowohl als lange wie auch als kurze Transition angesehen werden kann. Bei den verbleibenden Histogrammen zeigen sich für Lokomotion lediglich kurze Transitionen, so beispielsweise in Abbildung 5.9c (SP 2.3 → SP 2.1) für den mit Laufbewegungen assoziierten Übergang vom Doppel- zum Einzel-Fuß-Support.

Die Zeitdauern der zu Loko-Manipulation gehörenden Support-Posen-Transitionen sind in den Histogrammen in blauer Farbe dargestellt. Es ist zu beachten, dass bei den meisten dieser Bewegungen die eigentliche Manipulation durch lediglich eine Transition repräsentiert wird, während der überwiegende Teil der Transitionen der Fortbewegung dient. Ähnlich wie bei Lokomotion zeigt Abbildung 5.9a deshalb lange und Abbildung 5.9c vorwiegend kurze Support-Posen-Transitionen, allerdings sind in Abbildung 5.9c ebenfalls einige lange Transitionen zu erkennen. Diese langen Transitionen sind spezifisch mit der Manipulationskomponente der Loko-Manipulation assoziiert, wie beispielsweise mit dem Aufheben einer Kiste im Doppel-Fuß-Support. Vergleichbare lange, ebenfalls mit Manipulation assoziierte Support-Posen-Transitionen zeigen sich auch für andere Übergänge, beispielsweise in Abbildung 5.9e, Abbildung 5.9f und Abbildung 5.9g.

Bewegungen des Balancierens sind als dritte Kategorie in den Histogrammen in roter Farbe dargestellt und überwiegend mit kurzen Support-Posen-Transitionen assoziiert, da diese Bewegungen, beispielsweise nach Stößen, oftmals eine hohe Geschwindigkeit aufweisen. In Abbildung 5.9a und Abbildung 5.9c finden sich wie zuvor Transitionen für den auch in diesen Bewegungsaufnahmen meist enthaltenen Anteil des Laufens ohne Hand-Support. Die ca. 200 Frames langen Support-Posen-Transitionen in Abbil-

dung 5.9b hingegen sind mit Balancieren auf einem Bein assoziiert und stellen die Zeitdauer vom Herstellen des Einzel-Fuß-Supports bis zum Verlust des Gleichgewichts dar, was im Anschluss eine Stabilisierung durch das Herstellen eines Hand-Supports erfordert. Dieser Hand-Support wird danach bis zum Erreichen von Stabilität beibehalten und anschließend wieder gelöst, was zusammen mit den Bewegungsaufnahmen vom Typ „Schuhsole ansehen mit Hand-Support“ in den in Abbildung 5.9e gezeigten langen Support-Posen-Transitionen resultiert.

Abschließend soll noch erwähnt werden, dass bei der Betrachtung von Sequenzen kurzer Support-Posen-Transitionen oftmals wiederkehrende „Cluster“ beobachtbar sind, welche lediglich durch ihre Start- und End-Support-Posen charakterisiert werden können und eine beliebige Permutation der enthaltenen kurzen Transitionen ermöglichen. Eine ausführliche Untersuchung dieser Strukturen geht jedoch über den Rahmen dieser Arbeit hinaus und findet sich in (Borràs et al. 2017).

5.5 Zusammenfassung von Kapitel 5

Dieses Kapitel befasste sich mit einer Analyse menschlicher Ganzkörperbewegungen, die auf einer Betrachtung der vom Menschen genutzten Support-Kontakte mit seiner Umgebung und einer darauf aufbauenden heuristischen Segmentierung von Bewegungsaufnahmen basiert. Hierzu wurde die am H^2T entwickelte und in (Borràs et al. 2015) eingeführte Ganzkörperposen-Taxonomie vorgestellt, welche die wesentliche Grundlage für die Beschreibung von Ganzkörperbewegungen als Abfolge charakteristischer Support-Posen und für die darauf beruhende Segmentierung von Bewegungsaufnahmen darstellt. Die Taxonomie ermöglicht hierbei eine durch Arbeiten aus dem Robotergreifen inspirierte Komplexitätsreduktion für die Beschreibung möglicher Multi-Kontakt-Posen und der Transitionen zwischen ihnen, indem der hochdimensionale Raum möglicher Posen mit mehreren Dutzend Bewegungsfreiheitsgraden auf eine geringe Zahl fest-

gelegter Support-Posen reduziert wird. Zur Einführung der Taxonomie in diesem Kapitel wurde zunächst das Konzept eines Support-Kontakts mit verschiedenen möglichen Kontakttypen beschrieben und darauf aufbauend als Kombinationen möglicher Support-Kontakte die Support-Posen festgelegt, von denen insgesamt 46 zur Definition der Taxonomie genutzt wurden. Im Anschluss wurde das im Rahmen dieser Arbeit entwickelte Verfahren zur heuristischen Segmentierung von Bewegungsaufnahmen vorgestellt, welches auf der in Kapitel 2 eingeführten MMM-Repräsentation menschlicher Ganzkörperbewegungen einschließlich der beteiligten Umgebungsobjekte aufbaut. Für die Bewegungssegmentierung werden die Geschwindigkeiten der betrachteten Körperteile Hände, Füße, Knie und Ellbogen sowie ihre Distanz zu berücksichtigten Umgebungsobjekten in einem gleitenden Fenster betrachtet und hieraus zu jedem Zeitpunkt einer Bewegung die genutzten Support-Kontakte des Menschen bestimmt, denen wiederum die jeweilige korrespondierende Support-Pose der Taxonomie zugeordnet werden kann. Zur Evaluation wurden 121 Bewegungsaufnahmen einschließlich der notwendigen Mesh-Modelle beteiligter Umgebungsobjekte aus der KIT-Bewegungsdatenbank eingesetzt, wobei diese Bewegungsaufnahmen in die vier Kategorien Fortbewegung (Lokomotion), Fortbewegung mit Manipulation (Loko-Manipulation), Balancieren und kniende Bewegungen unterteilt wurden. Zunächst wurde das vorgeschlagene Segmentierungsverfahren validiert, indem resultierende Segmentierungen mit manuellen, durch den Menschen festgelegten Referenzsegmentierungen verglichen wurden. Hierbei konnte gezeigt werden, dass ca. 98 % der aus der Segmentierung resultierenden Support-Posen korrekt waren und nur ca. 4,5 % der durch die Referenzsegmentierung vorgegebenen Posen nicht vom Segmentierungsverfahren identifiziert werden konnten, und es wurden mögliche Ursachen für die bei der Segmentierung entstehenden Fehler diskutiert. Im weiteren Teil der Evaluation wurden die aus der Segmentierung resultierenden Posen-Sequenzen statistisch untersucht. Die Darstellung der segmentierten menschlichen Ganzkörperbewegungen als Abfolge charakteristischer

Support-Posen ermöglichte dabei zum einen eine Analyse der Auftretenshäufigkeit spezifischer Posen und ihrer Transitionen sowie der jeweils benötigten Zeitdauern und legte hiermit die Grundlage für eine partielle Validierung der zugrunde liegenden Taxonomie. Zum anderen liefern die aus der Segmentierung resultierenden Posen-Sequenzen die erforderlichen Trainingsdaten für den datengetriebenen Ansatz zur Generierung von Posen-Sequenzen in Kapitel 6.

6 Datengetriebene Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen

Während zweibeiniges Laufen bei humanoiden Robotern (Kajita et al. 2010, 2003; Engelsberger et al. 2013) zusammen mit der Behandlung dabei auftretender Störungen (Pratt et al. 2006) zunehmend gut verstanden wird, stellt die Beschreibung und Generierung der in Kapitel 1 angesprochenen Multi-Kontakt-Bewegungen noch immer ein komplexes und in der humanoiden Robotik nicht zufriedenstellend gelöstes Problem dar. Die Beobachtung der menschlichen Bewegung bildet zusammen mit der in Kapitel 5 eingeführten GKP-Taxonomie und dem darauf basierenden Segmentierungsverfahren eine Möglichkeit zur niedrigdimensionalen Darstellung von Multi-Kontakt-Bewegungen und damit zur Komplexitätsreduktion dieses Problems.

Vor diesem Hintergrund widmet sich das folgende Kapitel daher einem neuartigen Ansatz, der basierend auf von menschlichen Bewegungen gelerten, durch Techniken aus der natürlichen Sprachverarbeitung inspirierten statistischen Modellen für die in Kapitel 5 eingeführte Bewegungsrepräsentation die Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen humanoider Roboter ermöglicht. Wie bei der Erörterung verwandter Arbeiten in Abschnitt 6.1 noch ausführlicher diskutiert wird, greift der hier beschriebene Ansatz eine in der Literatur etablierte Verfahrensweise – meist als *contacts-before-motion approach* bezeichnet – auf, nach der das Problem der Bewegungsgenerierung in zwei Schritte unterteilt wird: In einem ersten Schritt wird zunächst eine Abfolge von Multi-Kontakt-Posen, oft auch

als *Stance Poses* bezeichnet, gesucht, die im zweiten Schritt zu einer kontinuierlichen Trajektorie verbunden wird. Der hier beschriebene Ansatz zur Generierung von Posen-Sequenzen stellt im Sinne dieser Verfahrensweise eine Lösung des ersten Schritts für Fortbewegungen (Lokomotion) dar und wurde erstmals in (Mandery et al. 2016a) veröffentlicht.

Der Aufbau dieses Kapitels gliedert sich wie folgt. Zunächst werden in Abschnitt 6.1 existierende Arbeiten zur Generierung von Multi-Kontakt-Bewegungen in Form von Kontakt-Posen oder Trajektorien und zu Lern-/Planungsansätzen in der Robotik vorgestellt. Im Anschluss daran wird in Abschnitt 6.2 das in diesem Kapitel behandelte Problem der Multi-Kontakt-Lokomotion vorgestellt und formalisiert. Abschnitt 6.3 befasst sich mit einer weitergehenden Unterteilung der durch die GKP-Taxonomie vorgegebenen Support-Posen in Konfigurationsposen und der Übertragung der resultierenden Posen auf die Kinematik humanoider Roboter. Abschnitt 6.4 führt dann mit einem N-Gramm-Transitions- und einem Translationsmodell zwei für das in diesem Kapitel präsentierte Verfahren zentrale statistische Modelle ein, bevor in Abschnitt 6.5 der Planungsalgorithmus zur Generierung von Posen-Sequenzen vorgestellt wird. Abschließend werden in Abschnitt 6.6 die Ergebnisse einer Evaluation des vorgestellten Verfahrens präsentiert, die sich mit einer Exploration des Parameterraums und einer Betrachtung der für exemplarische Szenarien geplanten Posen-Sequenzen befasst.

6.1 Verwandte Arbeiten

Eng verknüpft mit der humanoiden Robotik ist die Frage, wie menschenähnliche Ganzkörperbewegungen für Roboter mit einer anthropomorphen Gestalt erzeugt werden können. In der Vergangenheit wurde, auch am H²T beispielsweise in den EU-Forschungsprojekten KoroiBot und WALK-MAN, hierfür in der wissenschaftlichen Gemeinschaft eine Vielzahl von Ansätzen und Verfahren entwickelt. Die Generierung von Multi-Kontakt-Bewegungen ist hierbei erst seit wenigen Jahren verstärkt in den Fokus der Forschung

gerückt (Padois et al. 2017), wobei eine umfassende Lösung des Problems, wie in Kapitel 1 bereits erwähnt, bislang nicht in greifbarer Nähe liegt. Im Folgenden sollen die wichtigsten existierenden Arbeiten zur Generierung von Ganzkörperbewegungen und insbesondere von Multi-Kontakt-Bewegungen vorgestellt werden.

Fast alle der in der Robotik zu dieser Fragestellung vorgeschlagenen und im Folgenden diskutierten Verfahren nehmen, anders als der in dieser Arbeit beschriebene Ansatz, keine Berücksichtigung des durch Bewegungserfassung verfügbaren Vorwissens über die Ausführung von Ganzkörperbewegungen durch den Menschen vor. Dies kann dazu führen, dass die erzeugten Trajektorien oftmals unnatürlich bzw. „robotisch“ wirken. Ein weiterer Nachteil besteht insbesondere bei den auf Optimierungsansätzen basierenden Verfahren in der für ihre Ausführung benötigten Rechenzeit, die selbst für manchmal vereinfachte Modelle fast immer über den Erfordernissen einer Online-Anwendung liegt und teilweise etwa eine (Escande et al. 2013) oder mehrere Stunden (Bouyarmane et al. 2009; Lengagne et al. 2013) umfasst.

Direkte Erzeugung von Ganzkörperbewegungen: Für „einfache“ Ganzkörperbewegungen wie beispielsweise Greif-, Lehn- oder Hinsetzbewegungen, bei denen keine oder nur einzelne, vorher bekannte Kontaktänderungen auftreten, kann die Erzeugung stabiler Bewegungstrajektorien als nicht-lineares Optimierungsproblem, z.B. als Problem der quadratischen Programmierung, formuliert und mit hierfür entwickelten *Solvern* in einem Schritt gelöst werden. Von Saab et al. (2013) wurde, basierend auf früheren Arbeiten in (Saab et al. 2011a,b), ein generisches Framework für die Erzeugung von Multi-Kontakt-Bewegungen mit einer Kaskade von (hierarchischen) quadratischen Programmen vorgestellt, welches sowohl die Einhaltung von Gelenkwinkel- und Drehmomentbeschränkungen als auch die dynamische Stabilität des Roboters sicherstellt. Zur Evaluation wurden offline zwei Bewegungen in Form einer Schwing- und einer Hinsetzbewegung für den humanoiden Roboter HRP-2 generiert und ihre Durchführbarkeit auf dem realen

Roboter demonstriert. In (Posa et al. 2014) wird sequentiell quadratische Programmierung eingesetzt, um die Auswahl der zu verwendenden Kontakte und die resultierende Trajektorie in einem einzigen Optimierungsproblem zu formulieren. Der Ansatz wird in verschiedenen vergleichsweise einfachen Szenarien evaluiert, von denen das komplexeste die Generierung von Lauftrajektorien für den FastRunner-Roboter mit 13 Bewegungsfreiheitsgraden und einer Laufgeschwindigkeit von über 30 km/h darstellt. Gerade bei dieser Aufgabenstellung mit den damit verbundenen sehr häufigen Kontaktwechseln kann der integrierte Ansatz der gleichzeitigen Bestimmung von Kontakten und Trajektorie seine volle Stärke ausspielen. Auch Erez et al. (2012) befassen sich mit Optimierungstechniken für die Erzeugung von Lauftrajektorien, allerdings werden die Kontakte und die resultierende Trajektorie hier nicht gemeinsam bestimmt, sondern zunächst auf Basis einer (vorläufigen) Trajektorie die zu verwendenden Kontakte festgelegt. Für einen simulierten humanoiden Roboter (bestehend aus sphärischen Segmenten) mit 31 Bewegungsfreiheitsgraden konnten mit diesem Ansatz Lauftrajektorien erzeugt werden, die mit einer synchronen Schwingung eines Arm und des gegenüberliegenden Beins Charakteristika der menschlichen Laufbewegung aufweisen. In der Arbeit von Tassa et al. (2012) aus derselben Forschungsgruppe wird hingegen die Generierung von Aufstehbewegungen für einen simulierten Roboter mit 22 Bewegungsfreiheitsgraden gezeigt. Schulman et al. (2014) stellen mit *TrajOpt* ein freies Softwareframework für die Erzeugung kollisionsfreier Bewegungstrajektorien auf Basis sequentiell konvexer Optimierung vor, dessen Anwendung in unterschiedlichen Szenarien bis hin zur Generierung statisch stabiler Laufbewegungen für den Atlas-Roboter mit 34 Bewegungsfreiheitsgraden demonstriert wird. Zwei weitere Beispiele für die Generierung von (einfachen) Ganzkörperbewegungen mit Hilfe von Optimierungsverfahren finden sich bei Miossec et al. (2006) mit einer Minimierung der zur Bewegungsausführung benötigten Energie und bei Suleiman et al. (2007) mit einer Glättung und Stabilisierung vorgegebener Bewegungen bei einer Minimierung der auftretenden Drehmomente und Kräfte. Beide

Verfahren wurden mit dem humanoiden Roboter HRP-2 evaluiert, wobei in (Miossec et al. 2006) eine Kick-Bewegung erzeugt und in (Suleiman et al. 2007) eine vorgegebene Laufbewegung geglättet wurde, in welcher der Roboter während des Laufens seinen Schwerpunkt in Richtung des Bodens verlagert und gleichzeitig einen in der Hand gehaltenen Stab aufrichtet.

Weitere Verfahren zur Generierung ähnlicher wie der zuvor beschriebenen Ganzkörperbewegungen basieren statt auf klassischen Optimierungstechniken auf randomisierten Verfahren wie etwa Varianten des RRT-Algorithmus, mit denen direkt Lösungen für kollisionsfreie Ganzkörpertrajektorien in hochdimensionalen Räumen gefunden werden können. Solche Verfahren finden sich beispielsweise bei Dalibard et al. (2009) und Kuffner et al. (2001) für Ganzkörperbewegungen zum Reichen und Greifen in Umgebungen mit Hindernissen, wobei (Support-)Kontakte hier allerdings nicht explizit modelliert werden. Während diese Verfahren kollisionsfreie, kinematisch korrekte Trajektorien erzeugen können, werden dynamische Aspekte der Bewegung wie Stabilität in diesen Arbeiten nicht bzw. nur in einem nachgelagerten Filter-Schritt (Kuffner et al. 2001) betrachtet. Zur Generierung von Laufbewegungen unter Nebenbedingungen wie dem Tragen eines Objekts und dem Ausweichen von Hindernissen wurde in (Yoshida et al. 2005) die Kombination von klassischen Ansätzen des Laufens basierend auf einem *Walking Pattern Generator* (Kajita et al. 2003) mit Techniken der Bewegungsplanung vorgeschlagen, wobei sich derartige Ansätze aufgrund der mit dem ZMP verbundenen Einschränkungen nicht ohne Weiteres auf die nicht koplanaren (d.h. nicht in einer Ebene liegenden) Kontakte bei Multi-Kontakt-Bewegungen übertragen lassen. In (Sentis et al. 2011) wird eine Fallstudie zur Generierung „extremer Fortbewegungstechniken“, bei denen ausgehend von einem Wand-Support die Sprossen einer Leiter erklommen werden, präsentiert, wobei erforderliche Abschnitte der zu planenden Bewegung manuell vorgegeben werden und die resultierende Trajektorie mit Hilfe numerischer Integration erzeugt wird.

Planung von Kontakten für Multi-Kontakt-Bewegungen: Wie zuvor beschrieben, stellt die direkte Erzeugung kontinuierlicher Trajektorien für humanoide Roboter z.B. mit Optimierungsverfahren ein sehr anspruchsvolles Problem dar und es wurden bislang keine erfolgreichen Anwendungen solcher Verfahren für komplexe Multi-Kontakt-Fortbewegungsaufgaben beschrieben, bei denen weder die zu verwendenden Support-Kontakte noch die Trajektorie vorab bekannt sind und beide gleichzeitig geplant werden sollen. Zur Vereinfachung des Problems wurde eine Unterteilung vorgeschlagen, die als *contacts-before-motion approach* (Bouyarmane et al. 2012; Hauser et al. 2005), gelegentlich auch *stances-before-motion*, bezeichnet wird und im ersten Schritt zunächst die Planung einer Sequenz von Kontakt-Änderungen bzw. Kontakt-Posen (oft als *Stance Poses* bezeichnet) vorsieht, welche in einem zweiten Schritt anschließend zu einer kontinuierlichen Trajektorie verbunden werden¹. Beide Teilprobleme können unabhängig voneinander gelöst werden, wobei für beide Teilschritte jeweils verschiedene Verfahren vorgeschlagen wurden. Das im Rahmen dieses Kapitels vorgestellte Verfahren stellt mit der Generierung von Posen-Sequenzen im Sinne dieser Herangehensweise an das Problem ebenfalls eine Lösung für das erste Teilproblem dar. Im Folgenden sollen weitere in der Literatur bekannte Arbeiten zur Lösung des ersten Teilproblems für Multi-Kontakt-Bewegungen diskutiert werden.

Bouyarmane et al. (2012) stellen ein Verfahren zur Planung der Kontakte für Fortbewegung und Manipulationsbewegungen vor, das auf einer gierigen Bestensuche basiert und hierfür zur Identifikation geeigneter statisch stabiler *Stance Poses* wiederholt das als Optimierungsproblem formulierte kombinierte IK- und Stabilitätsproblem lösen muss, bei dem unilaterale und bilaterale Kontakte (Festhalte-Kontakte) in beliebigen Positionen berücksich-

¹ In Fällen wie dem Laufen auf flachem Grund, in denen sich die Planung möglicher Kontakte weniger anspruchsvoll darstellt, ist auch die Umkehrung dieser Vorgehensweise als *motion-before-contacts approach* denkbar (Hauser et al. 2005), bei dem zunächst die Bewegung für ein vereinfachtes Robotermodell ohne Berücksichtigung von Kontakten und dann die für die Bewegung notwendigen Kontakte bestimmt werden.

tigt werden können. Die aus der Ausführung des Verfahrens resultierende Bewegung ergibt sich anschließend, ähnlich wie in dieser Arbeit, als Abfolge von Kontaktänderungen (Herstellen bzw. Lösen jeweils genau eines Support-Kontakts). Die einzelnen Komponenten dieses Gesamtsystems wurden hierbei zuvor bereits in (Bouyarmane et al. 2010) für die Lösung des IK-Problems zur Erzeugung statisch stabiler Kontakt-Posen mit beliebigen Umgebungskontakten und in (Bouyarmane et al. 2011b) für den verwendeten Planungsalgorithmus (einschließlich einer Anwendung für Kooperationsaufgaben mehrerer Roboter) vorgestellt und evaluiert. In (Bouyarmane et al. 2011a) wird von denselben Autoren zudem eine Erweiterung auf Support-Kontakte mit deformierbaren Umgebungsobjekten vorgestellt, bei der die Formulierung des IK-Problems mit Techniken der Finite-Elemente-Methode zur Berücksichtigung solcher elastischer Umgebungsobjekte ergänzt wird. Escande et al. (2013) demonstrieren sehr eindrucksvoll die Fähigkeiten des in (Bouyarmane et al. 2012) beschriebenen Ansatzes, indem Bewegungen für den HRP-2-Roboter in drei komplexen Szenarien geplant werden: Das Hochklettern auf einer Leiter, Kriechen durch einen engen Tunnel sowie eine Hinsetzbewegung auf einen Stuhl mit einem zusätzlichen Tisch als Hindernis, während derer ein Becher gehalten wird. Ähnlich wird in (Vaillant et al. 2016) eine Bewegung für das Hochsteigen auf einer Leiter mit HRP-2 generiert und im Anschluss durch die Ausführung der geplanten Bewegung auf dem realen Roboter validiert.

Ein weiterer Ansatz zur Generierung von Stance Poses für Multi-Kontakt-Bewegungen findet sich bei Hauser et al. (2005). Für vorgegebene mögliche Kontaktpunkte in der Umgebung werden hier sampling-basierte Techniken zur Generierung geeigneter Konfigurationen eingesetzt. Da beim Einsatz derartiger randomisierter Algorithmen der Großteil der erzeugten Posen mindestens einer der Nebenbedingungen wie Kollisionsfreiheit oder den Kontakt-/Stabilitätsbedingungen nicht genügt, wird in der Arbeit mit *Iterative Constraint Enforcement* ein Verfahren eingeführt, welches diese Nebenbedingungen direkt beim Sampling möglicher Konfigurationen berücksichtigt und

so durch eine Reduktion des Anteils erzeugter ungültiger Konfigurationen die benötigte Rechenzeit deutlich verringern kann. Der vorgeschlagene Ansatz wurde für HRP-2 mit dem Erklimmen einer 0,5 m hohen Stufe, Fortbewegung auf (aus fraktalem Rauschen generierten) unebenem Terrain und dem Besteigen einer Leiter in drei unterschiedlichen Szenarien evaluiert. Bei dem in (Pryor et al. 2016) beschriebenen Verfahren wird die Generierung von Kontaktfolgen für Fortbewegungsaufgaben, ähnlich wie in dieser Arbeit, als Suchproblem in einem Graphen angesehen, das dort mit Hilfe einer Erweiterung des A*-Algorithmus, nämlich *Anytime Nonparametric A**, gelöst wird. Mit jedem Übergang im Graphen wird hierbei genau ein Support-Kontakt hinzugefügt, gelöst oder dessen Pose verändert, wobei für die zu jedem Zeitschritt geltenden Support-Kontakte statische Stabilität gefordert wird. Die Eingabe des Verfahrens basiert hierbei allerdings nicht wie bei den zuvor vorgestellten Arbeiten auf einem manuell festgelegten Planungsproblem bzw. einer vorgegebenen Szenerie, sondern besteht in den Affordanzen der Umgebung (siehe Abschnitt 6.2), welche aus einer sensorischen Erfassung dieser extrahiert werden.

Erzeugung von Ganzkörperbewegungen aus Kontakten: Nach der Bestimmung der Kontakte im ersten Schritt eines *contacts-before-motion approach* kann im nächsten Schritt die verbindende Bewegungstrajektorie erzeugt werden, wofür ebenfalls verschiedene Techniken vorgeschlagen wurden. Bouyarmane et al. (2011c) formulieren dieses Problem als Optimierungsproblem in Form eines quadratischen Programms, welches die unterschiedlichen Nebenbedingungen wie die dynamische Stabilität der Bewegung und die Berücksichtigung der Begrenzungen für Drehmomente und für die an den Kontakten wirkenden Kräfte als gewichtete Zielfunktionen berücksichtigt (statt eine strikte Hierarchisierung dieser Anforderungen vorzunehmen). Ein Vorteil dieses Verfahrens besteht darin, dass anstelle von statisch stabilen auch dynamisch stabile Bewegungen erzeugt werden können, allerdings wird die Freiheit der erzeugten Trajektorien von Selbst- und Umge-

bungskollisionen durch das Verfahren nicht garantiert und muss anderweitig sichergestellt werden, in (Bouyarmane et al. 2011c) beispielsweise mit Hilfe einer manuell vorgegebenen Heuristik. Die Evaluation des Verfahrens findet statt anhand von fünf unterschiedlichen Bewegungen wie Treppensteigen und Hinsetzen für HRP-2. Auch bei Lengagne et al. (2013) ergeben sich, ausgehend von Kontaktfolgen des in (Bouyarmane et al. 2012) vorgestellten Verfahrens, Bewegungstrajektorien aus der Lösung eines Optimierungsproblems. Konkret wird das Problem als ein semi-infinites Programm formuliert, bei dem Werte für eine begrenzte Anzahl zu optimierender Variablen in Form der Gelenkwinkel für die kontinuierlichen Nebenbedingungen jedes Zeitschritts gefunden werden müssen, und auch hier genügen die erzeugten Bewegungen der Anforderung dynamischer Stabilität. In der Arbeit wird, neben weiteren Bewegungen, eine Hinsetzbewegung für HRP-2 demonstriert, für welche die Stance Poses zuvor mit dem in (Bouyarmane et al. 2012) beschriebenen Verfahren erzeugt wurden.

Neben Optimierungsverfahren eignen sich auch klassische in der Robotik bekannte Ansätze zur Bewegungsplanung für die Erzeugung von Trajektorien in einem *contacts-before-motion approach*. In der weiter oben schon angesprochenen Arbeit von Hauser et al. (2005) werden *Probabilistic Roadmaps* zur Planung der verbindenden Trajektorie eingesetzt. In (Bouyarmane et al. 2009) kommt hingegen eine Bewegungsplanung basierend auf Potentialfeldern für die Identifikation von kollisionsfreien Pfaden in einem *Guide Space* zum Einsatz, der eine Einschränkung des Freiraums hinsichtlich der bestehenden Kontaktbedingungen und der Anforderungen statischer Stabilität darstellt. Die Evaluation erfolgt hier mit HRP-2 und den mit einer früheren Arbeit derselben Forschungsgruppe (Escande et al. 2009) erzeugten Stance Poses für eine Aufstehbewegung aus einer sitzenden Konfiguration an einem Tisch, sowie für weitere komplexe Bewegungen wie Kriechen in einem Tunnel, für welche die benötigten Kontakte nicht geplant, sondern manuell vorgegeben wurden.

Erzeugung von Ganzkörperbewegungen aus Bewegungsprimitiven:

Ein weiteres vielversprechendes Konzept zur Beschreibung der Transitionen zwischen vorgegebenen Kontakt-Posen stellen Bewegungsprimitive dar. Obwohl sich viele Arbeiten mit DMPs (Ijspeert et al. 2002, 2013) und ihren Erweiterungen beschäftigt haben und gezeigt werden konnte, wie diese zur Modellierung komplexer Bewegungen für humanoide Roboter eingesetzt (Ude et al. 2010), verkettet (Nemec et al. 2012; Kulvicius et al. 2012) und aus multiplen Beobachtungen des Menschen gelernt (Matsubara et al. 2011) werden können, existieren kaum Arbeiten, in welchen DMPs zur Beschreibung von Ganzkörperbewegungen eingesetzt werden. Das Kernproblem hierbei besteht darin, dass sich DMPs nur schlecht für eine Anwendung in hochdimensionalen Räumen eignen, weshalb zur Repräsentation von Ganzkörperbewegungen zunächst eine Dimensionalitätsreduktion erforderlich ist (N. Chen et al. 2015, 2016). Für einen Einsatz von HMMs (siehe Kapitel 3) zur Modellierung und Verkettung von Ganzkörper-Bewegungsprimitiven finden sich hingegen mehrere Anwendungen, beispielsweise bei Kulić et al. (2012), Takano et al. (2015) und Inamura et al. (2004), wobei auch in diesen Arbeiten Kontakte oder die daraus resultierende Frage nach der Stabilität der erzeugten Bewegungen nicht berücksichtigt werden.

In (Hauser et al. 2008) werden mögliche Bewegungsprimitive in einem zeitaufwendigen Offline-Verfahren vorab berechnet, wobei in der Arbeit auch die (nicht realisierte) Möglichkeit des Lernens der Primitive aus der Beobachtung des Menschen erwähnt wird. Bewegungsprimitive werden hierbei als Pfade im Konfigurationsraum definiert und beschreiben, ähnlich wie in zuvor diskutierten Arbeiten, jeweils das Herstellen bzw. Lösen genau eines Kontakts. Im Anschluss können die verfügbaren Primitive in einem probabilistischen sampling-basierten Bewegungsplanungsalgorithmus zur Beschleunigung der Planung und der Verbesserung der Qualität geplanter Bewegungen dienen, indem mit einer Heuristik zunächst ein geeignetes Bewegungsprimativ ausgewählt und anschließend in einem wachsenden Bereich um das ausgewählte Primativ gesampelt wird. Im Sinne dieses Ansatzes wur-

den in (Y. Zhang et al. 2013) manuell sieben Bewegungsprimitive für das Ersteigen einer Leiter vorgegeben und ihre Verkettung zur Erzeugung einer Bewegung für den Roboter Hubo-II+ demonstriert. In (Settimi et al. 2016) werden Bewegungsprimitive mit einem als *P-Search** bezeichneten sampling-basierten Algorithmus kombiniert und die aus der Kombination der Primitive möglicherweise auftretenden Instabilitäten aufgelöst. Zur Evaluation werden hier Bewegungen für verschiedene Aufgaben der Loko-Manipulation mit dem WALK-MAN-Roboter generiert. Cagnetti et al. (2015) beschreiben Bewegungsprimitive (ähnlich zu dem in Abschnitt 6.4.2 vorgestellten Translationsmodell) anhand der mit ihnen assoziierten Verlagerungen des Ganzkörper-Massenschwerpunkts, wobei in Vorarbeiten (Cagnetti et al. 2014) zunächst stattdessen die Verlagerung der Fußpositionen betrachtet wurde. Mit einem randomisierten Planungsalgorithmus können im Anschluss Ganzkörperbewegungen als kollisionsfreie Trajektorien im Konfigurationsraum für eine Abfolge der gelernten Primitive generiert werden, was mit einer Simulation des NAO-Roboters evaluiert wurde.

Stabilität und Regelung von Multi-Kontakt-Posen: Weitere Arbeiten befassen sich vorrangig mit den dynamischen Eigenschaften und der Stabilität von einzelnen Multi-Kontakt-Posen oder von Übergängen zwischen zwei vorgegebenen Posen und der Konzeption geeigneter Regelungen zu deren Ausführung. So wurden beispielsweise Stabilitätskriterien für Multi-Kontakt-Posen als Erweiterung des ZMP-Konzepts (Caron et al. 2017), mit einer *Dynamic Balance Matrix* (Nikolić et al. 2017) oder einem *Contact Wrench Cone* (Hirukawa et al. 2006) vorgeschlagen. Collette et al. (2008, 2007) präsentieren eine auf quadratischer Programmierung basierende Stabilitätsregelung für Multi-Kontakt-Posen mit beliebigen uni- und auch bilateralen Kontakten. Auch in (Salini et al. 2011a) wird quadratische Programmierung eingesetzt, um die benötigten Drehmomente für die gleichzeitige Ausführung mehrerer gewichteter Aufgaben unter Kontaktbedingungen zu bestimmen. Weitere Ansätze zur Regelung von Multi-Kontakt-Posen finden sich etwa bei

Khatib et al. (2014, 2008) und Sentis et al. (2010), sowie bei Werner et al. (2016) in Kombination mit einem Gesamt-Framework für die Erzeugung von Ganzkörperbewegungen mit einer Erweiterung des RRT-Algorithmus und einer Anwendung auf dem Roboter TORO.

Ansätze aus der Computergrafik: Zusätzlich zur Robotik existieren auch in der Computergrafik verschiedene Ansätze zur Generierung von Ganzkörper- oder Multi-Kontakt-Bewegungen für animierte Charaktere, die dort meist als *contact-rich motions* bezeichnet werden. Mordatch et al. (2012) stellen die Erzeugung solcher Bewegungen auf Basis eines kontakt-invarianten Optimierungsverfahrens vor, bei dem das Vorhandensein möglicher Kontakte durch Hilfsvariablen beschrieben wird. Diese Hilfsvariablen werden im Optimierungsproblem als konstant innerhalb von Bewegungsabschnitten angenommen, deren Anzahl und Länge manuell vorgegeben wird, was eine implizite Mit-Optimierung der Kontakte einer Bewegung und somit die gleichzeitige Bestimmung der notwendigen Kontakte und der kontinuierlichen Trajektorie auch für komplexe Bewegungsabläufe wie Klettern erlaubt. In (Kapadia et al. 2016) werden aus der Beobachtung menschlicher Bewegungen mögliche durch die Umgebungskontakte beschriebene *Motion Skills* des Menschen extrahiert und durch eine Analyse der Umgebungsgeometrie Oberflächen identifiziert, an denen die gelernten Skills angewendet werden können. Liu et al. (2010) schätzen die in Multi-Kontakt-Posen wirkenden Kontaktkräfte für komplexe Bewegungsabläufe wie etwa akrobatische Bewegungen aus erfassten Bewegungsdaten (siehe Abschnitt 5.1) und können unter Verwendung von randomisiertem Sampling eine Übertragung (*Retargeting*) dieser Bewegungen auf andere Kinematiken vornehmen, was mit einem ASIMO-ähnlichen Robotermodell evaluiert wurde. Zwei weitere Verfahren der Computergrafik finden sich bei Kang et al. (2014), wo mit Hilfe einer Datenbank menschlicher Posen automatisch geeignete Kontaktpunkte für den Menschen in einer gegebenen Umgebungsgeometrie erzeugt werden, und bei Al-Asqhar et al. (2013), wo ein Verfahren zur Echtzeit-Adaption

eines animierten Charakters an die gegebene räumliche Form der Umgebung vorgestellt wird. Einschränkungen der in der Computergrafik beschriebenen Ansätze bestehen meist in einer nur rudimentären Berücksichtigung von physikalischen oder dynamischen Parametern und entsprechend stark vereinfachten (oder gar nicht erst existenten) Modellen, da das primäre Ziel dieser Verfahren immer die Charakteranimation darstellt – beispielsweise wird für die Dynamik in (Mordatch et al. 2012) angenommen, dass lediglich der Torso über eine Masse verfügt.

Ansätze des Lernens und Planens: Abschließend kann der in dieser Arbeit vorgestellte datengetriebene Ansatz zur Generierung von Posen-Sequenzen auch mit existierenden Ansätzen des Lernens bzw. der Planung in Verbindung gebracht werden, die in der Robotik in vielfältiger Weise Anwendung finden. Klassische Ansätze der Planung, welche beispielsweise von Botvinick et al. (2012), Petrick et al. (2002) und Richter et al. (2010) beschrieben wurden, basieren auf einer symbolischen Weltbeschreibung, die eine Definition von innerhalb des Modells erlaubten Aktionen als den Weltzustand verändernde Operatoren ermöglicht. Die Modellierung des Weltzustands und möglicher Aktionen kann hierbei auf verschiedenste Arten erfolgen, die von Mitteln der symbolischen Logik, z.B. Modallogik bei Petrick et al. (2002), bis hin zu *Object-Action Complexes* bei Wächter et al. (2013) reichen. Konkrete Anwendungen solcher Planungstechniken zur Generierung von Ganzkörperbewegungen finden sich bei (Philippsen et al. 2009) mit reaktivem symbolischen Planen für Ganzkörper-Manipulationsaufgaben und (Salini et al. 2011b), wo mit unscharf (*fuzzy*) definierten Regeln in einem High-Level-Planer Aktionen für die in (Salini et al. 2011a) vorgestellte und zuvor bereits erwähnte Regelung ausgewählt werden. Auch in dem in dieser Arbeit in Abschnitt 6.5 eingeführten Planungsalgorithmus stellt sich die Planung als ein Suchproblem dar, bei dem im hochdimensionalen Raum möglicher Aktionssequenzen mit Hilfe von Heuristiken möglichst effizient diejenige Sequenz gefunden werden soll, die ausgehend von einem gege-

benen Startzustand den gewünschten Zielzustand erreicht und hierbei eine gegebene Zielfunktion minimiert. Ansätze, welche auf einer Modellierung von Aktionsfolgen auf Basis von (gelernten) Grammatiken aufbauen, finden sich beispielsweise bei Yang et al. (2015), Summers-Stay et al. (2012) und Pastra et al. (2012) für Manipulationsaktionen, bei Krüger et al. (2010) für Aktionsprimitive und bei Dantam et al. (2013) für die Repräsentation und Verifikation von Roboterregelungen. Wie in Abschnitt 6.4.1 weiter ausgeführt wird, sind derartige Ansätze durch die in diesem Kapitel vorgestellte linguistische Sichtweise auf das Problem der Bewegungsgenerierung ebenfalls mit dem hier vorgestellten Ansatz verknüpft.

6.2 Definition der Problemstellung

Wie in Kapitel 1 erläutert, kann die bei Multi-Kontakt-Bewegungen realisierte Verwendung zusätzlicher Kontakte über die Füße hinaus zu einer Erhöhung der Stabilität beitragen und zudem die Verwendung von Bewegungsstrategien wie *Lehnen* erlauben, die bei ausschließlicher Nutzung von Fußkontakten ansonsten von vornherein ausgeschlossen wären. Eine mögliche Kategorisierung von Multi-Kontakt-Bewegungen in Bewegungstypen (vgl. Abschnitt 5.4.1) unterteilt diese hierbei in Fortbewegungen (Lokomotion), Ganzkörper-Manipulationsbewegungen (z.B. Anheben einer Kiste), die als Loko-Manipulation bezeichnete Kombination aus beidem (z.B. Tragen eines Stuhls oder Drehen eines großen Ventils) und sonstige Bewegungstypen (z.B. balancierende oder akrobatische Bewegungen). Obwohl zahlreiche Aspekte des im Rahmen dieses Kapitels vorgestellten Ansatzes, wie beispielsweise die in Abschnitt 6.3 eingeführten Konfigurationsposen und das in Abschnitt 6.4.1 eingeführte N-Gramm-Modell zur Beschreibung der Transitionen zwischen ihnen, von einem konkreten Bewegungstyp unabhängig sind und lediglich die Verfügbarkeit einer ausreichenden Menge an geeigneten Trainingsdaten erfordern, soll für die in diesem Kapitel beschriebene Generierung von Posen-Sequenzen als einziger Bewegungstyp

(Multi-Kontakt-)Lokomotion betrachtet werden. In diesem Abschnitt werden die hierfür getroffenen Annahmen und somit die Formalisierung einer gegebenen Fortbewegungsaufgabe näher erläutert.

Für die betrachteten Fortbewegungsaufgaben wird angenommen, dass die Fortbewegung ebenerdig und in einer geraden Linie zwischen der bekannten Start- und Zielposition erfolgt. Komplexere Fortbewegungen mit Richtungsänderungen oder etwa auf Treppen und Leitern sollen somit nicht betrachtet werden. Mit der Einschränkung auf geradlinige Fortbewegung lässt sich die Position eines Menschen oder Roboters während der Ausführung solcher Bewegungen als Punkt auf der geradlinigen Verbindung zwischen Start- und Zielposition der Fortbewegung beschreiben, wodurch die mit den Posen einer Posen-Sequenz assoziierte Position als (im Zeitverlauf monoton steigende) eindimensionale Translation auf dieser geradlinigen Verbindung angesehen werden kann. Die Annahme geradliniger Fortbewegung bzw. die Beschreibung der jeweiligen Position als skalarer Wert stellt eine Vereinfachung des Problems dar, welche eine Reduktion der erforderlichen Trainingsdaten insbesondere für das in Abschnitt 6.4.2 eingeführte Translationsmodell ermöglicht, ist aber für die grundsätzliche Funktionsweise des beschriebenen Verfahrens nicht notwendigerweise erforderlich (siehe auch Abschnitt 7.2).

Die erforderliche Beschreibung der Umgebung für eine Fortbewegungsaufgabe basiert auf der Position von Umgebungsobjekten und den mit ihnen assoziierten Affordanzen (*affordances*). Das von Gibson (1979) zunächst in der Psychologie eingeführte Konzept von Affordanzen bezeichnet hierbei alle mit einem Objekt verknüpften Aktionen für einen Menschen oder einen humanoiden Roboter (für einen Stuhl beispielsweise *Sitzen*, *Hochheben* usw.), wobei für eine Generierung von Multi-Kontakt-Bewegungen naturgemäß Ganzkörper-Affordanzen wie die Lehn-Affordanz einer vertikalen Wand oder die Support-Affordanz eines Geländers von besonderem Interesse sind. Zur Ermittlung von Affordanzen aus der sensorischen Beobachtung der Umwelt existierende Ansätze ermöglichen inzwischen neben einer Bestimmung von Greifaffordanzen mit Hilfe visueller (Popović et al. 2010; Detry

et al. 2011) oder haptischer Sensormodalitäten (Bierbaum et al. 2009) auch die Ermittlung von Affordanzen für die gesamte einen Roboter umgebende Szenerie aus RGB-D-Daten (Kaiser et al. 2015b). In der jüngeren Vergangenheit wurden zahlreiche weitere Ansätze etwa von Pryor et al. (2016) vorgestellt, vor allem aber im Rahmen der in Kapitel 1 bereits erwähnten *DARPA Robotics Challenge*² beispielsweise von Fallon et al. (2015), Radford et al. (2015) und Kohlbrecher et al. (2015). Auch das am H²T von Kaiser et al. (2015b) vorgestellte Verfahren zur Extraktion von Affordanzen im ArmarX-Framework (Vahrenkamp et al. 2015) basierend auf den RGB-D-Daten eines Kinect-Sensors wurde unlängst in (Kaiser et al. 2015a, 2016) für eine Extraktion der im Kontext dieser Arbeit besonders interessanten Ganzkörper-Affordanzen wie Support- und Lehn-Affordanzen erweitert, was in Abbildung 6.1 exemplarisch dargestellt ist.

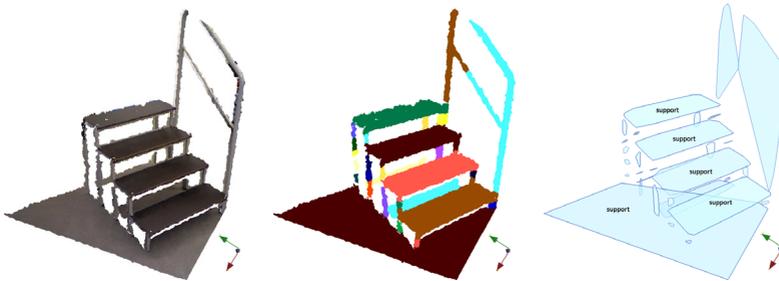


Abbildung 6.1: Automatische Extraktion geometrischer Primitive für die Affordanzen der Umgebung aus RGB-D-Daten mit dem in (Kaiser et al. 2016) beschriebenen Verfahren.

Nach der Bestimmung von Umgebungsaffordanzen und der Lage der mit ihnen assoziierten geometrischen Primitive im Raum können auf Basis der Annahme geradliniger Bewegung aus diesen Affordanzen direkt mögliche Support-Kontakte des Roboters für jeden Punkt auf der gedachten Gerade zwischen Start- und Zielpunkt der Fortbewegung bestimmt werden. Im

² Dort allerdings meist für das vereinfachte Problem von *Affordance Templates*, bei dem mögliche Objekte einer Szene bereits bekannt sind und vorab modelliert wurden.

Rahmen dieser Arbeit wird hierbei angenommen, dass die mit einer Fortbewegungsaufgabe assoziierten Umgebungsaffordanzen mit Hilfe eines der zuvor erwähnten Verfahren und die daraus resultierenden möglichen Support-Kontakte für jeden Punkt der Fortbewegung bereits bestimmt wurden.

Eine Fortbewegungsaufgabe wird im Rahmen dieser Arbeit daher wie folgt formalisiert:

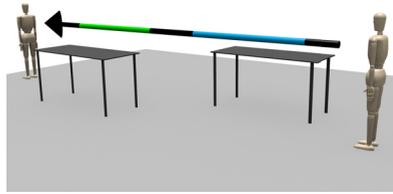
1. Die zurückzulegende Distanz d beschreibt die Länge der geradlinigen Wegstrecke zwischen dem Start- und Endpunkt der Fortbewegung.
2. $SK(x)$ gibt die möglichen Support-Kontakte (linker/rechter Fuß/Hand) gemäß den Lehn-/Support-Affordanzen der Umgebung als Funktion der bereits zurückgelegten Wegstrecke $x \in [0, d]$ an.
3. k_{start} und k_{end} legen die gewünschte Start- und End-Konfiguration der zu generierenden Bewegung als eine der im nachfolgenden Abschnitt 6.3 eingeführten Konfigurationsposen fest³.

Abbildung 6.2 zeigt eine derartige Definition einer exemplarischen Fortbewegungsaufgabe für eine Distanz von $d = 8$ m, die in Abbildung 6.2a angegebenen erlaubten Support-Kontakte abhängig von der zurückgelegten Wegstrecke und für die Wahl eines neutralen Doppel-Fuß-Supports für k_{start} und k_{end} . In Abbildung 6.2b ist eine visuelle Repräsentation dieser Fortbewegungsaufgabe dargestellt, wobei der Pfeil die Bewegungsrichtung und das blaue bzw. grüne Pfeilsegment verfügbare Support-Affordanzen (repräsentiert durch einen Tisch) für die rechte bzw. linke Hand signalisieren.

³ Während bei einer Anwendung des beschriebenen Verfahrens auf einem realen Roboter die Start-Konfiguration in der Regel durch die momentane Konfiguration des Roboters vorgegeben ist, erscheint es in manchen Fällen sinnvoll, *gar keine* End-Konfiguration für die zu generierende Posen-Sequenz festzulegen. Obwohl dieser Fall im Rahmen dieser Arbeit nicht explizit behandelt wird, wäre er durch eine Entfernung der entsprechenden Nebenbedingung in dem in Abschnitt 6.5 beschriebenen Planungsalgorithmus trivial realisierbar.

Distanz	Support-Kontakte
0 m bis 1 m	Füße
1 m bis 3 m	Füße, rechte Hand
3 m bis 4 m	Füße
4 m bis 6 m	Füße, linke Hand
6 m bis 8 m	Füße

(a) Erlaubte Support-Kontakte (SK).



(b) Visualisierung (entnommen aus Mandery et al. 2016a, ©2016 IEEE).

Abbildung 6.2: Exemplarische Definition einer Fortbewegungsaufgabe.

6.3 Unterteilung von Support-Posen hinsichtlich kinematischer Konfigurationen

Die Support-Posen der GKP-Taxonomie werden, wie in Abschnitt 5.2 erläutert, anhand der für die Struktur geltender Dynamikgleichungen maßgeblichen Support-Kontakte definiert, wodurch jede Support-Pose eindeutig mit den Rechenvorschriften eines Reglers für Bewegungen innerhalb dieser Pose assoziiert werden kann. Die kinematische Konfiguration des Körpers, repräsentiert beispielsweise durch die Positionen der Kontaktpunkte c_i , wird bei dieser Definition jedoch nicht berücksichtigt, weswegen mit einer gegebenen Support-Pose folglich zunächst keine konkreten Gelenkwinkel assoziiert werden können. Daher lassen sich durch unterschiedliche, von der geometrischen Gestalt der Umgebung (mit-)bestimmte Gelenkwinkel zahlreiche verschiedene Instanzen derselben Support-Pose bilden, die sich kinematisch teilweise stark unterscheiden können (siehe hierzu auch Abbildung 5.4). Obwohl diese kinematischen Instanzen keine Bedeutung für die Realisierung einer Regelung haben, ist dennoch offensichtlich, dass der Wahl geeigneter kinematischer Konfigurationen eine entscheidende Bedeutung für die Generierung von Multi-Kontakt-Bewegungen als Posen-Sequenzen zukommt, da zur Ausführung von Transitionen zwischen zwei aufeinanderfolgenden Posen letztendlich immer eine zulässige Trajektorie zum Übergang von der Start- zur Ziel-Konfiguration im Gelenkwinkelraum gefunden werden muss.

Der folgende Abschnitt befasst sich daher mit einer Unterteilung der Support-Posen in sogenannte Konfigurationsposen (KP)⁴, welche mögliche kinematische Ganzkörperkonfigurationen für die jeweiligen Support-Pose vorgeben und auf Basis einer Analyse der in menschlichen Bewegungen auftretenden Posen festgelegt werden. Diese Konfigurationsposen werden im Rahmen dieser Arbeit mit einem numerischen Suffix für die assoziierte Support-Pose gekennzeichnet, z.B. *1F-1H.7* für die siebte Konfigurationspose der Support-Pose *1F-1H* (SP 2.1). Zusätzlich wird in diesem Abschnitt eine auf dem MMM-Framework basierende Übertragung der erzeugten Konfigurationsposen vom menschlichen Körper, repräsentiert durch das MMM-Referenzmodell, auf die Kinematik humanoider Roboter vorgestellt, die aufgrund der kinematischen Unterschiede zwischen Mensch und Roboter für eine Anwendung der Konfigurationsposen zur Generierung von Bewegungen für humanoide Roboter erforderlich ist.

Als einführendes Beispiel für das Konzept von Konfigurationsposen soll zunächst eine einfache Laufbewegung ohne Nutzung von Hand-Supports betrachtet werden, die sich in der GKP-Taxonomie als alternierende Sequenz von SP 1.1 (1F) und SP 2.3 (2F) darstellt. Zur Beschreibung einer solchen Laufbewegung bietet sich eine Unterteilung der Einzel-Fuß-Support-Posen in zwei seitensymmetrische Konfigurationsposen *1F.1* und *1F.2* an, in denen der angehobene Fuß jeweils frei neben dem anderen, den Boden-Support herstellenden Fuß schwingt. Die Doppel-Fuß-Support-Posen lässt sich hingegen in drei Konfigurationsposen unterteilen: Eine Konfigurationspose *2F.1* bildet mit parallelen Fuß-Supports den Fall einer stehenden Konfiguration zum Beginn und Ende einer Laufbewegung ab, während sich bei den verbleibenden zwei Konfigurationsposen *2F.2* und *2F.3* jeweils einer der beiden Fuß-Supports vor dem anderem gelegen (anterior) befindet. Werden mögliche Transitionen zwischen diesen Konfigurationsposen für Laufbewegungen betrachtet, zeigt sich, dass von den zwölf kombinatorisch insgesamt möglichen Transitionen

⁴ Konfigurationsposen zur weiteren Unterteilung von Support-Posen wurden erstmals in (Mandery et al. 2016a) eingeführt, dort allerdings noch als *Shape Poses* bezeichnet.

zwischen Konfigurationsposen zur Beschreibung einer zyklischen Laufbewegung ohne den Übergang von/zu einer stehenden Konfigurationspose *2F.1* zum Beginn/Ende einer solchen Bewegung nur vier erforderlich sind⁵, während auch bei Berücksichtigung der stehenden Konfigurationsposen lediglich acht Transitionen tatsächlich auftreten können⁶.

Die Beantwortung der Frage, welche Transitionen zwischen Konfigurationsposen erforderlich oder überhaupt sinnvoll sind, ist (über einfache Laufbewegungen hinaus) allerdings nicht trivial und insbesondere nicht wie bei Support-Posen unmittelbar durch die Betrachtung der die Posen unterscheidenden Support-Kontakte möglich. Stattdessen hängt diese Fragestellung von komplexeren Aspekten wie der Ähnlichkeit der Konfigurationsposen hinsichtlich ihrer kinematischen Konfigurationen und hinsichtlich der Positionen ihrer gemeinsamen Support-Kontakte ab, weswegen Transitionen zwischen Konfigurationsposen im Rahmen des in diesem Kapitel vorgestellten Verfahrens mit Hilfe eines aus Bewegungsdaten gelernten statistischen Modells beschrieben werden (mehr hierzu in Abschnitt 6.4).

Werden abseits reiner Laufbewegungen auch die für Multi-Kontakt-Bewegungen entscheidenden Posen mit Nutzung der Arme betrachtet, steigt die Zahl möglicher Konfigurationsposen auch für Support-Posen ohne Hand-Supports stark an, da auch in diesen Support-Posen unterschiedliche Armkonfigurationen für die Durchführung von Transitionen von/zu anderen Posen mit Hand-Supports erforderlich sind. Berücksichtigt man beispielsweise für die drei zuvor beschriebenen Konfigurationsposen der SP 2.3 je drei mögliche kinematische Konfigurationen für jeden der beiden Arme, ergibt sich bereits eine Gesamtzahl an $3 \cdot 3 \cdot 3 = 27$ theoretisch möglichen Konfigurationsposen (die jedoch nicht zwangsläufig alle auftreten müssen). Abbildung 6.3 veranschaulicht dies durch die Darstellung einiger dieser Konfigurationsposen.

⁵ Die Transitionen von KP *2F.2*, in welcher der rechte Fuß vor dem linken Fuß steht, zu *1F.1*, in welcher der linke Fuß frei neben dem rechten Fuß-Support schwingt, und dann zu *2F.3*, sowie die weiteren zwei seitensymmetrischen Transitionen von KP *2F.3* zu *1F.2* und dann *2F.2*.

⁶ Zusätzlich die Transitionen zwischen KP *2F.1* und *1F.1* bzw. *1F.2* in beiden Richtungen.

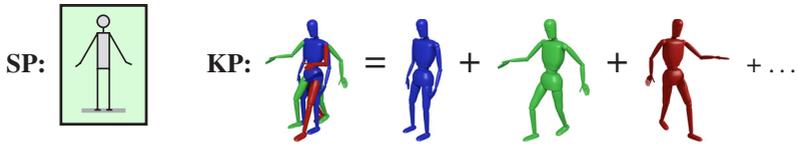


Abbildung 6.3: Veranschaulichung des Konzepts von Konfigurationsposen durch die Darstellung exemplarischer Optionen für die Konfiguration der Arme und Beine innerhalb der SP 2.3 (modifiziert nach Mandery et al. 2016a, ©2016 IEEE).

Durch eine Untersuchung menschlicher Ganzkörperbewegungen, die mit dem in Abschnitt 5.3 vorgestellten Verfahren hinsichtlich der auftretenden Support-Posen segmentiert wurden, können Konfigurationsposen für die in Abschnitt 5.4 erkannten Support-Posen identifiziert werden, wobei kniende Posen hier nicht weiter betrachtet werden sollen. Zur Definition der Konfigurationsposen wurden zunächst für jede dieser Support-Posen (SP 1.1, 2.1, 2.3, 3.1, 3.4 und 4.1) manuell mögliche Positionen aller vier Extremitäten in jeweils anterioren, posterioren oder körpernahen Konfigurationen (nahe der Frontalebene) bestimmt. Für jede sich aus diesen Konfigurationsoptionen kombinatorisch ergebende Konfiguration des MMM-Referenzmodells wurde anschließend im analysierten Bewegungsdatensatz für Multi-Kontakt-Lokomotion, der in Abschnitt 6.6 näher beschrieben wird, nach Vorkommen gesucht. Hierbei wurde für jede aus der Segmentierung einer Bewegungsaufnahme resultierende Support-Posen-Transition die kinematische Konfiguration zum Zeitpunkt in der Mitte zwischen den zwei aufeinanderfolgenden Änderungen von Support-Kontakten, d.h. zwischen dem Zeitpunkt des erstmaligen Auftretens der zu unterteilenden Support-Pose und dem Zeitpunkt des Verlassens dieser Pose mit der Erkennung der nachfolgenden Support-Pose, als repräsentative Konfiguration der jeweiligen Transition betrachtet. Die resultierenden Konfigurationsposen repräsentieren schließlich alle betrachteten Konfigurationen, für welche bei der Analyse des Bewegungsdatensatzes mindestens ein Vorkommen identifiziert werden konnte.

Tabelle 6.1: Unterteilung von Support-Posen in unterschiedliche Kombinationen von Support-Kontakten, zusammen mit der jeweiligen Zahl der Vorkommen dieser Support-Posen und der Anzahl der extrahierten Konfigurationsposen (modifiziert nach Mandery et al. 2016a, ©2016 IEEE).

Support-Pose	Support-Kontakte	# Vorkommen	# KP
1F (SP 1.1)	r. Fuß	275	16
	l. Fuß	384	15
1F-1H (SP 2.1)	r. Fuß, r. Hand	191	9
	r. Fuß, l. Hand	123	7
	l. Fuß, r. Hand	130	6
	l. Fuß, l. Hand	170	6
2F (SP 2.3)	beide Füße	662	14
1F-2H (SP 3.1)	r. Fuß, beide Hände	89	8
	l. Fuß, beide Hände	111	8
2F-1H (SP 3.4)	beide Füße, r. Hand	247	8
	beide Füße, l. Hand	244	6
2F-2H (SP 4.1)	beide Füße, beide Hände	187	8
Gesamt:		2813	111

Tabelle 6.1 zeigt eine Übersicht über die im Rahmen dieser Untersuchung extrahierten Konfigurationsposen für die zuvor angesprochenen sechs Support-Posen, die zudem für Support-Posen mit genau einem Hand- oder genau einem Fuß-Kontakt nach den seitensymmetrischen Fällen für die Wahl von Support-Kontakten untergliedert ist. Für jeden der angegebenen Fälle ist die Zahl der aus der Segmentierung resultierenden Vorkommen dieser Support-Posen sowie die Anzahl der extrahierten Konfigurationsposen angegeben. Insgesamt wurden für die 2813 beobachteten Vorkommen der sechs Support-Posen 111 unterschiedliche Konfigurationsposen bestimmt.

Die Definition von Konfigurationsposen ermöglicht anschließend eine Erweiterung des in Abschnitt 5.3 beschriebenen Segmentierungsverfahrens, indem nach der erfolgten Segmentierung einer Bewegungsaufnahme jede der

resultierenden Support-Posen einer der für die jeweilige Support-Pose definierten Konfigurationspose zugeordnet wird. Analog zu der Vorgehensweise bei der Festlegung der Konfigurationsposen wird als maßgebliche Konfiguration zur Klassifikation einer Support-Pose hierbei nicht der Zeitpunkt des initialen Auftretens dieser Pose, sondern der zeitlich gesehen mittlere Punkt zwischen Etablierung und Verlassen der jeweiligen Pose betrachtet. Für die Zuordnung dieser Konfiguration zu einer der verfügbaren Konfigurationsposen wird, anstelle eines Vergleichs der Posen im hochdimensionalen Gelenkwinkelraum, der 12-dimensionale Vektor bestehend aus den vier Positionen der Hände und Füße im lokalen Koordinatensystem des MMM-Referenzmodells bestimmt und diejenige Konfigurationspose ausgewählt, bei welcher der euklidische Abstand für diesen 12-dimensionalen Vektor zwischen der zu klassifizierenden Konfiguration und der Konfigurationspose minimal wird.

Mit der Zuordnung der Support-Posen einer Bewegung zu jeweils einer Konfigurationspose lässt sich eine segmentierte Bewegungsaufnahme im Anschluss nicht nur als Support-Posen-Sequenz, sondern mit der Darstellung als Konfigurationsposen-Sequenz auch in einer Form beschreiben, die neben den vom Menschen zur Bewegungsausführung genutzten Support-Kontakten auch Informationen über die kinematische Ausführung der jeweiligen Bewegung enthält. Diese Darstellung beobachteter Bewegungen als Konfigurationsposen-Sequenz bildet die Grundlage für die in Abschnitt 6.4 beschriebene statistische Modellierung von Posen-Transitionen und damit für das gesamte im Rahmen dieses Kapitels vorgestellte Verfahren zur Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen.

Es ist allerdings zu beachten, dass die Definition von Konfigurationsposen im Gegensatz zu den ihnen zugrunde liegenden Support-Posen von einem konkreten kinematischen Modell abhängig ist. Durch die Extraktion der zuvor definierten Konfigurationsposen aus der Segmentierung menschlicher Bewegungsaufnahmen beziehen sich diese Posen daher zunächst auf das MMM-Referenzmodell und geben Gelenkwinkel für die mit diesem Modell

assozierten Gelenke vor. Um eine Anwendung der Konfigurationsposen für die Generierung von Bewegungen eines humanoiden Roboters zu ermöglichen, ist es daher erforderlich, diese Konfigurationsposen unter bestmöglicher Beibehaltung ihrer kinematischen Ausgestaltung auf die Zielkinematik des jeweiligen humanoiden Roboters zu übertragen. Hierfür kommt das in Abschnitt 2.4 beschriebene Verfahren basierend auf nicht-linearer Optimierung zum Einsatz, das im Rahmen dieser Arbeit innerhalb des MMM-Frameworks entwickelt wurde und zur Übertragung von Konfigurationsposen unabhängig voneinander für jede dieser Posen durchgeführt wird. Abbildung 6.4 zeigt exemplarisch Konfigurationsposen zweier Support-Posen für das MMM-Referenzmodell und ihre mit diesem Verfahren durchgeführte Übertragung auf die Kinematik des humanoiden Roboters ARMAR-4 (Asfour et al. 2013).

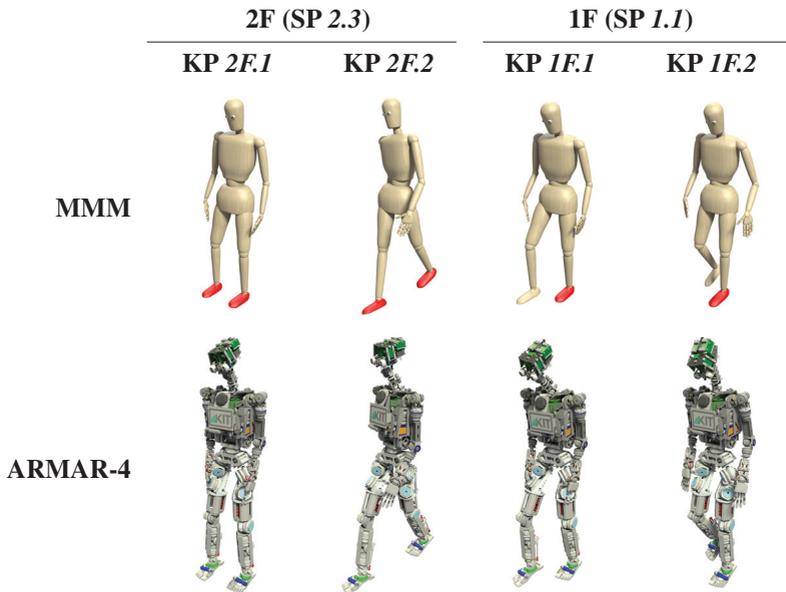


Abbildung 6.4: Exemplarisch ausgewählte Konfigurationsposen verschiedener Support-Posen für das MMM-Referenzmodell und ihre Übertragung auf den humanoiden Roboter ARMAR-4.

6.4 Statistische Modellierung von Posen-Transitionen

Der im Rahmen dieses Kapitels vorgestellte Ansatz zur Bewegungsgenerierung basiert auf zwei statistischen Modellen, die mit Hilfe des in Abschnitt 5.3 beschriebenen Verfahrens zur Segmentierung menschlicher Ganzkörperbewegungen und der resultierenden Darstellung von Bewegungen als Abfolge der in Abschnitt 6.3 eingeführten Konfigurationsposen datengetrieben aus der Beobachtung des Menschen gelernt werden. Bevor in Abschnitt 6.5 der eingesetzte Planungsalgorithmus erläutert wird, soll daher zunächst eine Beschreibung dieser beiden Modelle in diesem Abschnitt vorangestellt werden.

6.4.1 N-Gramm-Modell für Transitionswahrscheinlichkeiten

Im Zuge der Einführung von Konfigurationsposen in Abschnitt 6.3 wurde bereits festgestellt, dass die Identifikation von möglichen Transitionen zwischen diesen Konfigurationsposen nicht wie bei den Support-Posen der GKP-Taxonomie (dort basierend auf der Annahme, dass sich die Start- und Ziel-Support-Pose einer Transition durch genau einen Support-Kontakt unterscheiden) rein definitorisch erfolgen kann. Dies bedeutet zum Beispiel, dass nicht jede Kombination von Konfigurationsposen zweier Support-Posen, für welche die GKP-Taxonomie eine Transitionsmöglichkeit vorgibt, eine zur Bewegungsgenerierung sinnvolle Transition darstellt (wenngleich natürlich zumindest ein Paar solcher Konfigurationsposen für diese beiden Support-Posen existieren sollte). So würde etwa der Übergang aus einer Doppel-Fuß-Pose, in welcher der rechte vor dem linken Fuß steht, zu einer Doppel-Fuß-Pose mit Hand-Support, in der die Fuß-Supports ihre Positionen getauscht haben, durch den Wechsel der Fußpositionen der Annahme von während des Übergangs fortbestehenden Fuß-Supports widersprechen. Eine systematische Untersuchung, ob ein Übergang zwischen zwei gegebenen

Konfigurationsposen problemlos, problematisch oder gar unmöglich ist, stellt sich hierbei allerdings als komplexe Fragestellung verschiedener Aspekte wie der kinematischen Ähnlichkeit der Posen und der Stabilität möglicher Transitionstrajektorien dar, und es liegt daher nahe, mögliche Transitionen zwischen Konfigurationsposen ähnlich wie die Posen selbst durch eine Analyse von Bewegungsdaten aus der Beobachtung des Menschen zu lernen.

Zur Modellierung möglicher Transitionen zwischen Konfigurationsposen wird im Folgenden eine linguistische Sichtweise auf das Problem vorgeschlagen: Indem Konfigurationsposen als *Wörter* in einem *Satz*, der als Sequenz dieser *Wörter* die Bewegung repräsentiert, angesehen werden, stellt sich das Problem des Findens möglicher Transitionen zwischen ihnen als ein zu Problemstellungen der natürlichen Sprachverarbeitung analoges Problem dar, wo die Wahrscheinlichkeit natürlichsprachlicher Sätze bewertet bzw. für eine gegebene Wortfolge mögliche Folgewörter bestimmt werden sollen.

Diese Sichtweise ähnelt den in Abschnitt 6.1 angesprochenen Arbeiten in der Robotik, bspw. von Krüger et al. (2010) und Dantam et al. (2013), in denen versucht wird, regelbasierte *Grammatiken* zur Beschreibung von Bewegungen, Bewegungsprimitiven oder Sensorwerten zu lernen. Ein Problem derartiger regel-/grammatikbasierter Systeme besteht jedoch darin, dass diese Ansätze mit wachsender Unsicherheit und Problemkomplexität oftmals scheitern, weil sich das Identifizieren und Formulieren zugrunde liegender allgemeingültiger Regeln zunehmend schwieriger gestaltet. In Bereichen der natürlichen Sprachverarbeitung, beispielsweise bei automatischer Spracherkennung oder maschineller Übersetzung, wurden daher schon vor langer Zeit regelbasierte durch rein statistische Ansätze ersetzt (Rosenfeld 2000), wobei N-Gramm-Sprachmodelle und ihre Erweiterungen den bekanntesten und populärsten Modelltyp darstellen. Durch den Einsatz solcher N-Gramm-Modelle für die Modellierung von Transitionswahrscheinlichkeiten zwischen Konfigurationsposen wird im Rahmen dieser Arbeit versucht, diese Vorgehensweise auch für die Erzeugung von Posen-Sequenzen zum Zweck der Bewegungsgenerierung anzuwenden. Während sich durchaus Beispiele für

den Einsatz von N-Gramm-Modellen auf nicht-natürlichsprachlichen Problemstellungen in der Literatur finden (Wu et al. 2007; Chu et al. 2009), handelt es sich bei dem hier und zuvor in (Mandery et al. 2016a) beschriebenen Ansatz um den nach Kenntnisstand des Autors erstmaligen Einsatz von N-Gramm-Modellen für eine Modellierung der Abfolge von Ganzkörperposen des Menschen oder humanoider Roboter.

Die Aufgabe eines statistischen Sprachmodells besteht darin, die Auftretenswahrscheinlichkeit eines Worts w_i , gegeben alle vorhergehenden Wörter (w_1, \dots, w_{i-1}) , zu beschreiben, wodurch das Modell im stochastischen Sinne eine (bedingte) Wahrscheinlichkeitsverteilung über mögliche Folgewörter festlegt. Bei einem N-Gramm-Modell wird die zusätzliche Annahme getroffen, dass diese bedingte Wahrscheinlichkeit lediglich von den $n - 1$ vorhergehenden Wörtern $(w_{i-n+1}, \dots, w_{i-1})$ abhängt. n wird hierbei analog zur Ordnung einer Markow-Kette (siehe Abschnitt 3.2.1) auch als die Ordnung des N-Gramm-Modells bezeichnet. Im Sinne des hier beschriebenen N-Gramm-Transitionsmodells repräsentieren die im Folgenden als k_i bezeichneten *Wörter* die Konfigurationsposen und ein Satz (k_1, \dots, k_N) als Konfigurationsposen-Sequenz eine Bewegung. Das N-Gramm-Modell beschreibt dann die bedingte Wahrscheinlichkeit

$$P(k_i \mid (k_{i-n+1}, \dots, k_{i-1})) \quad (6.1)$$

der Beobachtung einer Konfigurationspose k_i , gegeben die vorhergehenden Konfigurationsposen $(k_{i-n+1}, \dots, k_{i-1})$. Durch die Multiplikation der bedingten Wahrscheinlichkeiten aller Konfigurationsposen in einer Bewegung kann anschließend die durch das Modell geschätzte Auftretenswahrscheinlichkeit der mit dieser Bewegung assoziierten Konfigurationsposen-Sequenz (k_1, \dots, k_N) berechnet werden:

$$P((k_1, \dots, k_N)) = \prod_{i=1}^N P(k_i \mid (k_{i-n+1}, \dots, k_{i-1})). \quad (6.2)$$

Ein Problem bei der Schätzung von N-Gramm-Modellen stellt die Spärlichkeit verfügbarer Trainingsdaten dar, da die Anzahl kombinatorisch möglicher N-Gramme prinzipiell exponentiell mit der Ordnung des Modells ansteigt. Obwohl mit 111 verfügbaren Konfigurationsposen die Vokabulargröße in dem hier beschriebenen N-Gramm-Modell deutlich unter derer natürlicher Modelle liegt, ist dennoch nicht davon auszugehen, dass alle möglichen N-Gramme als eine Folge von $n - 1$ zulässigen Transitionen zwischen Konfigurationsposen in den Trainingsdaten vertreten sind. Eine einfache Maximum-Likelihood-Schätzung, bei welcher die durch Gleichung 6.1 beschriebenen Wahrscheinlichkeiten durch eine Zählung der relativen Häufigkeiten von N-Grammen in den Trainingsdaten angenähert werden, würde solchen N-Grammen und somit allen diese Teilsequenzen von Konfigurationsposen beinhaltenden Konfigurationsposen-Sequenzen dementsprechend in der Folge eine Wahrscheinlichkeit von Null zuweisen.

Zur Lösung dieses Problems kommen bei N-Gramm-Modellen üblicherweise Glättungstechniken (*smoothing techniques*) zum Einsatz, welche die anhand einer beschränkten Menge von Trainingsdaten gelernten Modelle glätten, indem eine Reduktion der Wahrscheinlichkeiten für alle in den Trainingsdaten enthaltenen N-Gramme vorgenommen wird (weswegen man die Vorgehensweise auch als *Discounting* bezeichnet). Die hierdurch freiwerdende Wahrscheinlichkeitsmasse für eine Folge vorhergehender Konfigurationsposen ($k_{i-n+1}, \dots, k_{i-1}$) kann anschließend verwendet werden, um auch in den Trainingsdaten nicht beobachteten Folgewörtern k_i eine Wahrscheinlichkeit größer Null zuzuweisen. Als Folge hieraus enthält ein N-Gramm-Modell nicht nur N-Gramme der Länge n , sondern ebenfalls der Längen $n - 1$, $n - 2$, usw. bis hin zu den (unbedingten) Wahrscheinlichkeiten einzelner Wörter als Unigramme, sodass beim Nichtauffinden eines N-Gramms im Modell iterativ N-Gramme niedrigerer Ordnung abgefragt werden können (sogenanntes *Backing-Off*). S. F. Chen et al. (1999) beschreiben eine eingehende Untersuchung verschiedener Techniken zur Glättung von N-Gramm-Modellen, von denen im Rahmen dieser Arbeit Good-Turing-Glättung (Good 1953), modifi-

zierte Kneser-Ney-Glättung (S. F. Chen et al. 1999) und Witten-Bell-Glättung (Witten et al. 1991) berücksichtigt und evaluiert werden.

Die für die Schätzung des N-Gramm-Modells erforderliche Repräsentation der Trainingsdaten besteht in einer textuellen Darstellung der aus der Segmentierung menschlicher Bewegungen resultierenden Konfigurationsposen-Sequenzen. Basierend auf dieser Repräsentation kann anschließend mit dem weit verbreiteten und (für nicht-kommerzielle Nutzung) frei verfügbaren SRILM-Toolkit (Stolcke 2002) das N-Gramm-Modell geschätzt und in dem ebenfalls textbasierten ARPA-Back-Off-Dateiformat gespeichert werden. Abbildung 6.5 zeigt einen beispielhaften Auszug aus dieser textbasierten Repräsentation eines gelernten Trigramm-Modells ($n = 3$), in welchem für die angegebenen Trigramme lediglich die (logarithmierte) Wahrscheinlichkeit am Beginn einer Zeile, für die Bigramme jedoch zusätzlich auch der im Rahmen der Glättung des Modells ermittelte Back-Off-Faktor am Ende der Zeile angegeben wird.

```
\2-grams:
  :
-0.4253142 LeftFootRightFoot_1 LeftFoot_1 0.05318153
-1.742332 LeftFootRightFoot_1 LeftFoot_2 -0.06390977
  :
\3-grams:
  :
-0.2152434 LeftFootRightFoot_1 LeftFoot_1 LeftFootRightFoot_2
-0.5245459 LeftFootRightFoot_1 LeftFoot_1 LeftFootRightFoot_5
  :
```

Abbildung 6.5: Auszug aus der Repräsentation eines gelernten Trigramm-Modells im ARPA-Back-Off-Dateiformat.

6.4.2 Modellierung der räumlichen Translation bei Posen-Transitionen

Mit dem in Abschnitt 6.4.1 vorgestellten N-Gramm-Transitionsmodell wurde ein Modell eingeführt, das mögliche Transitionen zwischen Konfigurationsposen aus segmentierten Bewegungsaufnahmen des Menschen lernen kann. Jedoch findet in diesem Modell keine Berücksichtigung der aus Konfigurationsposen-Sequenzen resultierenden räumlichen Fortbewegung statt, welche für die im Rahmen dieses Kapitels betrachteten Fortbewegungsaufgaben allerdings gerade von entscheidender Bedeutung ist: Mit ausschließlicher Verwendung des N-Gramm-Transitionsmodells kann zwar eine Modellierung von kinematisch zulässigen Übergängen zwischen durch die GKP-Taxonomie vorgegebenen Support-Posen realisiert werden, nicht jedoch eine Lösung für Fortbewegungsaufgaben gefunden werden, bei denen die erzeugte Konfigurationsposen-Sequenz als Nebenbedingung darüber hinaus auch das Erreichen der Zielposition der Fortbewegung sicherstellen muss. Daher wird zur Berücksichtigung der Fortbewegung ein weiteres statistisches Modell benötigt, das nachfolgend eingeführt werden soll.

Zur Berücksichtigung räumlicher Fortbewegung in dem in diesem Kapitel eingeführten Verfahren wird angenommen, dass sich die totale durch eine Fortbewegung realisierte Translation des Menschen im Raum als Summe von Translationen darstellen lässt, die mit den diese Fortbewegung repräsentierenden Transitionen zwischen Konfigurationsposen assoziiert sind. Basierend darauf wird ein Translationsmodell als zweites statistisches Modell aus menschlichen Bewegungsdaten gelernt, welches diese mit spezifischen Transitionen assoziierte räumliche Fortbewegung als Translation des Ganzkörper-Massenschwerpunkts beschreibt. Die durch dieses Translationsmodell für Transitionen zwischen Konfigurationsposen vorgegebenen Translationswerte können anschließend in dem in Abschnitt 6.5 vorgestellten Planungsverfahren als Abschätzung für das Ausmaß der mit einer generierten Konfigurationsposen-Sequenz assoziierten Fortbewegung in Richtung der

Zielposition einer Fortbewegungsaufgabe eingesetzt werden⁷. Die mit der Transition von der Konfigurationspose k zu j im Translationsmodell assoziierte Translation des Ganzkörper-Massenschwerpunkts wird hierbei als $T(k, j) \in \mathbb{R}$ bezeichnet. Aufgrund der im Rahmen dieser Arbeit getroffenen Annahme geradliniger Fortbewegung (siehe Abschnitt 6.2), wird T hierfür zur Vereinfachung lediglich als reellwertige Funktion (statt als Funktion des \mathbb{R}^3 bzw. \mathbb{R}^6 bei Berücksichtigung von Rotation) definiert, welche die räumliche Fortbewegung entlang der Gerade zwischen dem Start- und Endpunkt der jeweiligen Bewegung beschreibt.

Für die Festlegung der Funktionswerte von T auf Basis menschlicher Bewegungsdaten werden mögliche Kombinationen zweier Konfigurationsposen k und j unabhängig voneinander betrachtet. Hierbei wird für jede spezifische Konfigurationsposen-Kombination (k, j) die mit der jeweiligen Transition verknüpfte Translation im Translationsmodell als arithmetisches Mittel der Translationen aller Vorkommen dieser Transition in den Trainingsdaten bestimmt:

$$T(k, j) = \frac{\sum_{i \in M(k, j)} \left\| \mathbf{c}_{\text{msp}}(t_{\text{end}}(i)) - \mathbf{c}_{\text{msp}}(t_{\text{start}}(i)) \right\|_2}{|M(k, j)|}, \quad (6.3)$$

wobei $M(k, j)$ die Menge aller Vorkommen der Transition (k, j) in den Trainingsdaten, $\mathbf{c}_{\text{msp}}(t)$ den Ganzkörper-Massenschwerpunkt des Menschen zum Zeitpunkt t und $t_{\text{start}}(i)$ bzw. $t_{\text{end}}(i)$ die durch die Segmentierung bestimmten Zeitpunkte des Beginns (als erstmaliges Auftreten der Pose k) bzw. des Endes (als erstmaliges Auftreten der Pose j) der i -ten Transition zwischen den Konfigurationsposen in den Bewegungsdaten bezeichnet.

⁷ Durch das Lernen der Translationen vom Menschen wird angenommen, dass ein humanoider Roboter, auf dessen Kinematik Konfigurationsposen wie in Abschnitt 6.3 beschrieben übertragen wurden, in der Lage ist, die entsprechenden vom Menschen gelernten Translationen ebenfalls zu realisieren. Sollten die Fähigkeiten des Roboters aufgrund eingeschränkter Gelenkwinkelgrenzen oder einer geringeren Körpergröße hierfür nicht ausreichend sein, müssen diese Translationen entsprechend adaptiert, z.B. linear skaliert, werden.

6.5 Planungsalgorithmus zur Generierung von Posen-Sequenzen

Der folgende Abschnitt befasst sich mit einem auf den zuvor eingeführten statistischen Modellen basierenden Planungsalgorithmus zur Generierung von Konfigurationsposen-Sequenzen für das in Abschnitt 6.2 eingeführte Problem der Multi-Kontakt-Lokomotion, das hier nach der zuvor beschriebenen linguistischen Sichtweise als Wegfindungsproblem unter Nebenbedingungen in dem durch Transitionen zwischen Konfigurationsposen aufgespannten Baum interpretiert wird. Eine gegebene Fortbewegungsaufgabe wird hierbei, wie in Abschnitt 6.2 dargestellt, durch die zurückzulegende Distanz d , die erlaubten Support-Kontakte SK als Funktion der bereits zurückgelegten Distanz und die Start- und Ziel-Konfigurationsposen k_{start} und k_{end} der zu planenden Bewegung festgelegt.

Formal stellt sich das hier beschriebene Planungsproblem als Suche nach der durch

$$\operatorname{argmax}_K (P(K) - \text{Penalty}(K)) \quad (6.4)$$

beschriebenen Konfigurationsposen-Sequenz $K = (k_1, \dots, k_N)$ einer beliebigen Länge N im Raum aller möglichen Konfigurationsposen-Sequenzen dar. P bezeichnet hierbei die durch das N-Gramm-Transitionsmodell vorgegebene logarithmierte Wahrscheinlichkeit der Sequenz K (siehe Gleichung 6.2). Die später genauer beschriebene Penalty-Funktion hingegen repräsentiert einen zusätzlichen Term, welcher eine Bewertung resultierender Konfigurationsposen-Sequenzen hinsichtlich (nicht-)verwendeter Support-Möglichkeiten vornimmt und dadurch dazu beiträgt, dass Sequenzen, die durch eine möglichst effiziente Nutzung erlaubter Support-Kontakte mutmaßlich eine höhere Stabilität aufweisen, vom Planungsverfahren bevorzugt werden.

Das durch Gleichung 6.4 beschriebene Optimierungsproblem unterliegt hierbei den folgenden Nebenbedingungen, welche die spezifischen Anforderungen einer vorgegebenen Fortbewegungsaufgabe modellieren:

1. Die aus dem Translationsmodell (siehe Gleichung 6.3) bestimmte kumulative Translation der Konfigurationsposen-Sequenz K muss zur Abdeckung der Distanz d zwischen Start- und Zielposition der Fortbewegung geeignet sein:

$$\sum_{i=1}^{N-1} T(k_i, k_{i+1}) \geq d.$$

2. Jede der in K enthaltenen Konfigurationsposen muss der durch die Fortbewegungsaufgabe vorgegebenen Beschränkung verfügbarer Support-Kontakte entsprechen. Dies bedeutet, dass jede Konfigurationspose k_i lediglich die für ihre jeweilige Translation erlaubten Support-Kontakte

$$SK \left(\sum_{j=1}^{i-1} T(k_j, k_{j+1}) \right)$$

beinhalten darf.

3. Die Translation jeder Teilsequenz (k_u, \dots, k_v) , $v > u$ in K , innerhalb derer alle Konfigurationsposen k_u, \dots, k_j einen gemeinsamen Support-Kontakt s beinhalten, muss unterhalb eines körperteilspezifischen Schwellwerts $\theta_{\max\text{supp}}(s)$ liegen:

$$\sum_{i=u}^{v-1} T(k_i, k_{i+1}) \leq \theta_{\max\text{supp}}(s).$$

Die in Abschnitt 5.2 beschriebenen und bei der Analyse menschlicher Bewegungsaufnahmen betrachteten Support-Kontakte repräsentieren stationäre Kontakte, für die angenommen wird, dass sich der kontakt-herstellende Körperteil während des Bestehens des Kontakts an einer

ortsfesten Umgebungsposition befindet. Dies bedeutet im Umkehrschluss, dass mit diesen Support-Kontakten kinematisch lediglich ein gewisses Maß an Translation realisiert werden kann, bevor der Kontakt gelöst und ggf. neu hergestellt werden muss. So wird beispielweise angenommen, dass es nicht möglich ist, ein Geländer ohne Lösen des Kontakts gleitend über eine Translation von mehreren Metern als Support zu nutzen. Die angegebene Nebenbedingung stellt sicher, dass die erzeugten Konfigurationsposen-Sequenzen dieser Einschränkung genügen, obwohl der hier beschriebene Ansatz keine explizite Modellierung solcher kinematischer Beschränkungen vornimmt.

4. Die Start- und End-Konfigurationspose der geplanten Sequenz müssen der durch die Fortbewegungsaufgabe vorgegebenen Start- und End-Pose entsprechen:

$$k_1 = k_{\text{start}}, \quad k_N = k_{\text{end}}.$$

Nach der formalen Darstellung des durch Gleichung 6.4 und die Nebenbedingungen (1) bis (4) gegebenen Planungsproblems soll nun der im Rahmen dieser Arbeit entwickelte Planungsalgorithmus zu dessen Lösung erläutert werden. Wie eingangs bereits erwähnt, kann das Finden der gemäß Gleichung 6.4 optimalen Konfigurationsposen-Sequenz als Wegfindungsproblem in dem durch Transitionen zwischen Konfigurationsposen aufgespannten Baum angesehen werden, in dem ausgehend von der Wurzel als Startpunkt jede Kante das Anhängen einer Konfigurationspose repräsentiert. Jeder Knoten in diesem Baum ist eindeutig mit einer möglichen Konfigurationsposen-Sequenz assoziiert und soll im Folgenden auch als *Pfad* im Sinne des Wegfindungsproblems bezeichnet werden. Eine herkömmliche Breiten- oder Tiefensuche zur Traversierung der Baumstruktur verbietet sich hier schon angesichts des hohen Verzweigungsgrads entsprechend der Anzahl aller möglichen Nachfolge-Konfigurationsposen sowie der Länge erzeugter

Konfigurationsposen-Sequenzen für typische Fortbewegungsaufgaben im mittleren zweistelligen Bereich. Eine gierige (*greedy*) Suche hingegen, bei welcher in jedem Zeitschritt der mit der jeweiligen Konfigurationsposen-Sequenz assoziierte Knoten expandiert und durch seinen besten Kindknoten (gemäß Gleichung 6.4) ersetzt wird, wäre hingegen rechnerisch prinzipiell möglich, erscheint aber angesichts des strikt lokalen Ansatzes, der das Finden einer global optimalen Konfigurationsposen-Sequenz unwahrscheinlich macht, wenig erfolgversprechend. Im Rahmen dieser Arbeit wird daher ein informiertes, d.h. heuristisches, Suchverfahren mit Pruning vorgeschlagen, das parallel zahlreiche unterschiedliche Knoten (als Konfigurationsposen-Sequenzen) als mögliche Hypothesen verfolgt und in jedem Zeitschritt den jeweils besten expandiert, wobei ein Pruning aussichtsloser Hypothesen den zur Durchführung erforderlichen Rechenaufwand deutlich reduziert⁸.

Algorithmus 5 zeigt den im Folgenden erläuterten Planungsalgorithmus als Pseudocode. In *activePaths* werden während der Ausführung des Verfahrens alle zum jeweiligen Zeitpunkt berücksichtigten Pfade als Knoten des Baums, die bislang weder expandiert noch geprunt wurden, verwaltet. Die Implementierung von *activePaths* als binärer Max-Heap garantiert hierbei, dass entsprechend der Verwendung von *activePaths* in Algorithmus 5 als Vorrangwarteschlange sowohl das Finden und Entfernen des Pfads mit der jeweils besten Bewertung (Zeile 8) als auch das Einfügen eines neuen Pfads beim Expandieren eines Pfads (Zeile 28) bzw. beim Pruning (Zeile 16) in $\mathcal{O}(\log n)$ erfolgen kann. Zu Beginn wird *activePaths* mit dem Pfad entsprechend einer lediglich aus der Konfigurationspose k_{start} bestehenden Sequenz initialisiert (Zeile 2), wodurch das Einhalten von Nebenbedingung (4) hinsichtlich der vorgegebenen Start-Pose für alle erzeugten Konfigurationsposen-Sequenzen zwangsläufig sichergestellt ist.

⁸ Diese Vorgehensweise ähnelt stark der in Systemen der automatischen Spracherkennung und maschinellen Übersetzung oftmals eingesetzten *Beam Search* (X. Huang et al. 2001, Kapitel 12).

Algorithmus 5 Planung von Posen-Sequenzen (modifiziert nach Mandery et al. 2016a, ©2016 IEEE)

```
1: activePaths ← Heap()
2: Füge Pfad(startPose) in activePaths ein
3: i ← 0
4: loop
5:   i ← i + 1
6:
7:   // Entferne besten Pfad aus Heap
8:   bestPath ← EXTRACTMAX(activePaths)
9:
10:  // Reduktion der zu betrachtenden Pfade durch Pruning
11:  if (i mod  $\theta_{\text{pruneperiod}}$ ) = 0 then
12:    pruneDist ← bestPath.distance -  $\theta_{\text{prunethresh}}$ 
13:    newPaths ← Heap()
14:    for all path ∈ activePaths do
15:      if path.distance ≥ pruneDist then
16:        Füge path in newPaths ein
17:      end if
18:    end for
19:    activePaths ← newPaths
20:  end if
21:
22:  if activePath.distance ≥ distance and
    bestPath.endPose = endPose then
23:    return bestPath // Lösung gefunden
24:  else
25:    expandedPaths ← EXPANDPATH(bestPath)
26:    for all path ∈ expandedPaths do
27:      Bewerte path gemäß Gleichung 6.4
28:      Füge path in activePaths ein
29:    end for
30:  end if
31: end loop
```

Algorithmus 6 Expandieren eines Pfads

```

1: function EXPANDPATH(path)
2:   newPaths  $\leftarrow$  List()
3:
4:   // Betrachte alle 111 möglichen Konfigurationsposen
5:   for all k  $\in$  allCPs do
6:     newP  $\leftarrow$  path + k
7:     dist  $\leftarrow$   $\sum_{i=1}^{|newP|-1} T(newP[i], newP[i+1])$ 
8:
9:     if  $\exists c \in k.supportContacts : c \notin SK(dist)$  then
10:       Nebenbedingung (2) verletzt, breche Schleifendurchlauf ab
11:     end if
12:
13:     for all sc  $\in$  k.supportContacts do
14:       // Bestimme die früheste KP in path, ab welcher der
15:       // Support-Kontakt sc existiert
16:       firstPose  $\leftarrow$   $|newP| - 1$ 
17:       while sc  $\in$  path[firstPose].supportContacts do
18:         firstPose  $\leftarrow$  firstPose - 1
19:       end while
20:       if  $\sum_{i=firstPose}^{|newP|-1} T(newP[i], newP[i+1]) > \theta_{maxsupp}(sc)$  then
21:         Nebenbedingung (3) verletzt, breche äußeren
           Schleifendurchlauf ab
22:       end if
23:     end for
24:
25:     Hänge newP an newPaths an
26:   end for
27:   return newPaths
28: end function

```

In jeder Iteration des Planungsalgorithmus wird der jeweils beste aktive Pfad (*bestPath*) aus *activePaths* entnommen und, falls dieser Pfad noch keine vollständige Lösung für die gegebene Fortbewegungsaufgabe darstellt, expandiert (Zeile 25), wobei durch das Anhängen möglicher Konfigurationen an diesen Pfad weitere Pfade entstehen. Wie in Algorithmus 6 gezeigt, werden hierbei nur solche Konfigurationsposen berücksichtigt, die sowohl der durch Nebenbedingung (2) vorgegebenen Beschränkung möglicher Support-Kontakte als auch der durch Nebenbedingung (3) vorgegebenen Begrenzung der unter Beibehaltung eines Support-Kontakts maximal möglichen Translation genügen. Für jeden der neu entstehenden Pfade wird in Algorithmus 5 eine Bewertung gemäß Gleichung 6.4 vorgenommen (Zeile 27), wobei zur Berechnung der hier berücksichtigten Penalty-Funktion für jede Konfigurationspose die Anzahl der am jeweiligen Punkt der Fortbewegung möglichen, aber nicht genutzten Support-Kontakte ermittelt wird. Der Wert der Penalty-Funktion ergibt sich im Anschluss als Produkt aus der kumulativen Zahl nicht verwendeter Support-Kontakte und θ_{penalty} . Nach ihrer Bewertung werden neue Pfade in *activePaths* eingefügt (Zeile 28).

Zur Reduktion des Rechenaufwands durch eine Einschränkung des Suchraums wird in Algorithmus 5 ein heuristisches Pruning eingesetzt (Zeilen 10 bis 20), welches gesteuert durch den Parameter $\theta_{\text{pruneperiod}}$ periodisch in *activePaths* enthaltene aussichtslose Pfade verwirft. Hierzu werden alle Pfade, die eine durch ihre Sequenz von Transitionen zwischen Konfigurationsposen vorgegebene Translation unterhalb des derzeit besten Pfads abzüglich eines Schwellwerts $\theta_{\text{prunethresh}}$ aufweisen, entfernt und *activePaths* wird durch einen neuen Heap mit den verbleibenden Pfaden ersetzt.

Algorithmus 5 terminiert (Zeilen 22 und 23), sobald eine gültige Lösung für die gegebene Fortbewegungsaufgabe gefunden wurde, welche die durch Nebenbedingung (1) geforderte kumulative Translation und gleichzeitig die

durch Nebenbedingung (4) vorgegebene End-Konfigurationsposen erfüllt⁹. Da durch das Anhängen einer zusätzlichen Konfigurationspose beim Expandieren eines Pfads die durch das N-Gramm-Transitionsmodell vorgegebene Wahrscheinlichkeit $P(K)$ nur abnehmen kann¹⁰, der Wert von $\text{penalty}(K)$ aber nur zunehmen kann, nimmt die Bewertung des durch *bestPath* gegebenen Pfads während der Ausführung monoton ab. Es kann daher garantiert werden, dass mit dem Terminieren des Algorithmus keine besseren Lösungen für die gegebene Fortbewegungsaufgabe mehr gefunden werden können¹¹.

6.6 Evaluation und Ergebnisse

Der folgende Abschnitt befasst sich mit einer Evaluation des in diesem Kapitel vorgestellten Verfahrens zur Generierung von Posen-Sequenzen. Die Ergebnisse dieser Evaluation wurden hierbei teilweise bereits in (Mandery et al. 2016a) veröffentlicht.

Die Evaluation beruht auf einem Bewegungsdatensatz, welcher aus 137 Aufnahmen von Laufbewegungen mit zusätzlichen Hand-Supports besteht und Trainingsdaten für die in Abschnitt 6.4 beschriebenen statistischen Modelle für Posen-Transitionen bereitstellt. Dieser Datensatz untergliedert sich in acht Bewegungsexperimente, von denen zwei normale Laufbewegungen

⁹ Wie zuvor beschrieben, garantiert die gewählte Vorgehensweise zwangsläufig die Einhaltung der Nebenbedingungen (2) und (3) und der durch Nebenbedingung (4) vorgegebenen Start-Konfigurationsposen für alle bei der Ausführung von Algorithmus 5 berücksichtigten Pfade.

¹⁰ Alle durch das N-Gram-Modell in Gleichung 6.1 vorgegebenen bedingten Auftretenswahrscheinlichkeiten von Konfigurationsposen sind kleiner Eins und beim Anhängen einer Konfigurationspose an eine vorgegebene Konfigurationsposen-Sequenz wird zur Bestimmung der Wahrscheinlichkeit der resultierenden neuen Sequenz unter dem N-Gramm-Modell die Wahrscheinlichkeit der vorherigen Konfigurationsposen-Sequenz mit einer solchen bedingten Auftretenswahrscheinlichkeit kleiner Eins multipliziert.

¹¹ Selbstverständlich kann infolge des Pruning dennoch nicht ausgeschlossen werden, dass eine bessere Lösung existiert (die aber nicht mehr gefunden werden kann). Im Rahmen der in Abschnitt 6.6 beschriebenen experimentellen Evaluation wurde dieser Effekt bei der Variation der Pruning-Parameter für die letztendlich gewählten Parameterwerte jedoch nicht beobachtet.

ohne zusätzliche Hand-Supports und sechs Multi-Kontakt-Laufbewegungen mit Hand-Supports repräsentieren. In den sechs Bewegungsexperimenten mit Hand-Supports standen dem menschlichen Probanden seitlich zur Fortbewegungsrichtung mit Geländern und Tischen verschiedene mit Support-Affordanzen assoziierte Umgebungsobjekte zur Verfügung. In drei dieser Experimente wurde die Fortbewegung des Menschen zudem auf einen Holzbalken verlagert, der durch eine zusätzliche Einschränkung möglicher Fußkontakte die Stabilität der Bewegung reduziert und somit zur Verwendung von Hand-Supports durch den Menschen beiträgt. Für jedes Bewegungsexperiment wurden jeweils zehn Aufnahmen für die Hin- und Rückrichtung der Fortbewegung durchgeführt¹², wodurch ein hinsichtlich der Verwendung von linken und rechten Hand-Supports annähernd seitensymmetrischer Bewegungsdatensatz resultiert.

Tabelle 6.2 zeigt eine Übersicht über die eingesetzten Bewegungsdaten und in Abbildung 6.6 sind ergänzend manuell ausgewählte Schlüsselbilder für einige der enthaltenen Bewegungsaufnahmen dargestellt. Wie in den anderen Teilen dieser Arbeit sei der Leser zusätzlich auch auf die mit den in Tabelle 6.2 angegebenen IDs zugänglichen Einträge in der KIT-Bewegungsdatenbank mit den dort verfügbaren Videoaufnahmen verwiesen.

Im Rahmen der folgenden Evaluation findet zunächst in Abschnitt 6.6.1 eine Exploration des Parameterraums für das N-Gramm-Transitionsmodell statt, um geeignete Parameterwerte für den Einsatz dieses Modells zu ermitteln. Im Anschluss werden in Abschnitt 6.6.2 drei beispielhafte Fortbewegungsaufgaben betrachtet und die für diese Szenarien generierten Posen-Sequenzen untersucht.

¹²Drei der insgesamt 140 Bewegungsaufnahmen mussten infolge von Problemen bei der Aufnahme bzw. Nachbearbeitung nachträglich ausgeschlossen werden.

Tabelle 6.2: Zur Evaluation eingesetzte Bewegungsaufnahmen als Trainingsdaten für die in Abschnitt 6.4 beschriebenen statistischen Modelle. Die angegebenen IDs korrespondieren mit den Bewegungsexperimenten in der KIT-Bewegungsdatenbank unter der URL <https://motion-database.humanoids.kit.edu/details/motions/<ID>/>.

Bewegungstyp	ID	Objekte mit Support-Aff.	# Aufn.
Laufen	395	–	9
	396	–	9
	677	Geländer rechts	10
		Geländer links	10
	679	Tisch rechts	10
		Tisch links	10
	681	Tisch rechts, Geländer links	10
		Tisch links, Geländer rechts	10
Laufen auf Balken	678	Geländer rechts	10
		Geländer links	10
	705	Tisch rechts, Geländer links	10
		Tisch links, Geländer rechts	10
	724	Tisch rechts, Geländer links	10
		Tisch links, Geländer rechts	9
Gesamt:			137



Abbildung 6.6: Manuell ausgewählte Schlüsselbilder aus einigen der in Tabelle 6.2 aufgelisteten Bewegungsaufnahmen.

6.6.1 Parametrisierung des N-Gramm-Transitionsmodells

Der Einsatz des in Abschnitt 6.4.1 beschriebenen N-Gramm-Modells zur statistischen Modellierung von Transitionen zwischen Konfigurationsposen erfordert die Festlegung der Werte zweier Parameter: Zum einen muss die Ordnung n des Modells festgelegt werden, welche die maximale Länge der im Modell betrachteten N-Gramme und damit die bei der Abschätzung bedingter Wahrscheinlichkeiten maximal betrachtete Anzahl von $n - 1$ vorhergehenden Konfigurationsposen festlegt. Im Bereich der natürlichen Sprachverarbeitung werden heutzutage meist maximal Tetra- oder Pentagramme ($n \in \{4, 5\}$) betrachtet, wobei mit wachsendem n prinzipiell zwar die Mächtigkeit des N-Gramm-Modells zunimmt, gleichzeitig aber aufgrund der exponentiell wachsenden Zahl möglicher N-Gramme eine größere Menge an Trainingsdaten benötigt wird. Zum anderen muss zusätzlich zur Wahl von n ein Verfahren zur Glättung des Modells gewählt werden, wofür im Kontext dieser Arbeit die drei in Abschnitt 6.4.1 angesprochenen Glättungstechniken, nämlich Good-Turing-Glättung, modifizierte Kneser-Ney-Glättung und Witten-Bell-Glättung, Berücksichtigung finden.

Zur Exploration des Parameterraums wurden alle Kombinationen von $n \in \{2, 3, 4, 5, 6\}$ und den drei genannten Glättungstechniken im Rahmen einer 5-fachen Kreuzvalidierung (siehe hierzu auch Abschnitt 3.3) evaluiert, bei der in jeder von fünf Runden ein N-Gramm-Modell mit vier Fünfteln der verfügbaren segmentierten Bewegungsdaten trainiert und mit dem verbleibenden Fünftel bewertet wird. Die Bewertung erfolgt hierbei durch die auf einer mit Testdaten gemessenen Annäherung der Kreuzentropie

$$\hat{H}(T) = -\frac{1}{M} \sum_{i=1}^M \log_2 P(t_i)$$

aufbauenden Perplexität (Jelinek et al. 1977)

$$PP(T) = 2^{\hat{H}(T)},$$

wobei P die durch das Modell vorgegebene Wahrscheinlichkeitsverteilung repräsentiert. Diese Perplexität stellt ein bei der Arbeit mit N-Gramm-Modellen weit verbreitetes Maß dafür dar, wie gut ein gelerntes Modell zu den durch die M Sätze $T = \{t_1, \dots, t_M\}$ gegebenen Testdaten passt und wie gut es diese vorhersagen kann, und beschreibt anschaulich den durchschnittlichen Verzweigungsfaktor (Rosenfeld 2000) bei der Vorhersage des nächsten Wortes¹³, wobei niedrigere Werte auf ein die Trainingsdaten besser beschreibendes Modell hindeuten. Im Zuge dieser Evaluation wird die Perplexität mit dem zuvor bereits angesprochenen SRILM-Toolkit (Stolcke 2002) berechnet.

Tabelle 6.3: Berechnete Perplexität des N-Gramm-Transitionsmodells für verschiedene Ordnungen (Werte von n) und Glättungstechniken (modifiziert nach Mandery et al. 2016a, ©2016 IEEE).

Ordnung	Glättung		
	Good-Turing	Mod. Kneser-Ney	Witten-Bell
2	4,7125	4,7108	4,4668
3	4,1451	4,6747	3,9049
4	4,1759	5,1635	3,7684
5	4,2496	5,7179	3,7493
6	4,3471	6,2004	3,7511

Tabelle 6.3 zeigt die Perplexität für alle evaluierten Parametrisierungen des N-Gramm-Transitionsmodells. Generell zeigt sich bei Betrachtung der unterschiedlichen Glättungstechniken, dass der Einsatz der Witten-Bell-Glättung stets mit der niedrigsten Perplexität assoziiert ist (wobei Good-Turing-Glättung wiederum besser als modifizierte Kneser-Ney-Glättung abschneidet). Die beste Perplexität wurde für die Kombination von Witten-Bell-Glättung mit einer Ordnung von $n = 5$ festgestellt (in Tabelle 6.3 fett gedruckt). Diese Kombination stellt daher die Parametrisierung dar, welche im weiteren Verlauf der folgenden Evaluation ausschließlich zum Einsatz kommt.

¹³ Informationstheoretisch kann dies als Informationsgehalt einer gegebenen Konfigurationssequenz bei Kenntnis des gelerntes N-Gramm-Modells interpretiert werden.

6.6.2 Generierung von Posen-Sequenzen für verschiedene Beispielszenarien

In diesem Abschnitt soll durch die Betrachtung von drei exemplarischen Fortbewegungsaufgaben die Eignung des in diesem Kapitel eingeführten Verfahrens zur Generierung komplexer Posen-Sequenzen für diese Szenarien untersucht werden. Die drei evaluierten Fortbewegungsaufgaben unterscheiden sich hierbei sowohl in der zurückzulegenden Distanz als auch in der Verfügbarkeit von Hand-Supports. Konkret soll in Szenario 1 eine Wegstrecke von 6 m zurückgelegt werden, wobei auf dem Teilstück von 1 m bis 4 m durch eine Support-Affordanz auf der rechten Seite Support-Kontakte mit der rechten Hand ermöglicht werden. Ähnlich beschreibt Szenario 2 das Zurücklegen von 8 m mit der Möglichkeit rechtshändiger Support-Kontakte von 1 m bis 3 m sowie linkshändiger Support-Kontakte von 4 m bis 6 m und Szenario 3 das Zurücklegen von 6 m, wobei von 2 m bis 4 m Support-Kontakte sowohl mit der linken als auch mit der rechten Hand möglich sind. Als Start- und End-Pose wird in allen drei Szenarien die neutrale Doppel-Fuß-Konfigurationspose angenommen, in der sich die Füße parallel auf dem Boden und die Arme seitlich des Körpers befinden (blaue Pose in Abbildung 6.3). In Abbildung 6.7 ist eine Visualisierung der durch diese Szenarien beschriebenen Fortbewegungsaufgaben dargestellt.

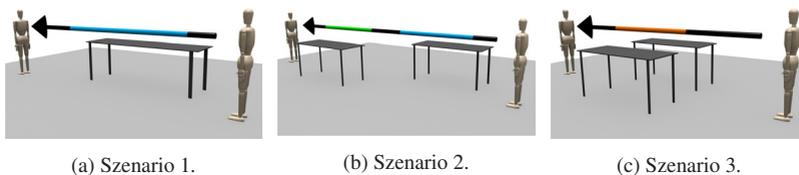


Abbildung 6.7: Visualisierung der drei evaluierten exemplarischen Fortbewegungsaufgaben (entnommen aus Mandery et al. 2016a, ©2016 IEEE). Die dargestellten Tische repräsentieren angenommene Support-Affordanzen der Umgebung, für welche die farblich hervorgehobenen Pfeilsegmente die resultierenden erlaubten Support-Kontakte der Hände kennzeichnen (blau: rechte Hand, grün: linke Hand, orange: beide Hände).

Gemäß der in Abschnitt 6.2 eingeführten Formalisierung der Problemstellung sind die betrachteten Fortbewegungsaufgaben demnach beschrieben durch

$$d_1 = 6 \text{ m}, \quad d_2 = 8 \text{ m}, \quad d_3 = 6 \text{ m},$$

$$SK_1(x) = \begin{cases} \{1. \text{ Fuß, r. Fuß, r. Hand}\} & \text{falls } 1 \text{ m} \leq x \leq 4 \text{ m}, \\ \{1. \text{ Fuß, r. Fuß}\} & \text{sonst,} \end{cases}$$

$$SK_2(x) = \begin{cases} \{1. \text{ Fuß, r. Fuß, r. Hand}\} & \text{falls } 1 \text{ m} \leq x \leq 3 \text{ m}, \\ \{1. \text{ Fuß, r. Fuß, l. Hand}\} & \text{falls } 4 \text{ m} \leq x \leq 6 \text{ m}, \\ \{1. \text{ Fuß, r. Fuß}\} & \text{sonst,} \end{cases}$$

$$SK_3(x) = \begin{cases} \{1. \text{ Fuß, r. Fuß, l. Hand, r. Hand}\} & \text{falls } 2 \text{ m} \leq x \leq 4 \text{ m}, \\ \{1. \text{ Fuß, r. Fuß}\} & \text{sonst,} \end{cases}$$

und

$$k_{\text{start},\{1,2,3\}} = k_{\text{end},\{1,2,3\}} = \text{KP } 2FI.$$

Die Werte für die in Abschnitt 6.5 beschriebenen Parameter des Planungsalgorithmus wurden empirisch wie folgt gewählt:

$$\theta_{\text{prunepriod}} = 1000, \quad \theta_{\text{prunethresh}} = 0,5 \text{ m},$$

$$\theta_{\text{maxsupp}}(s) = \begin{cases} 1 \text{ m} & \text{für Handsegmente } s, \\ \infty & \text{sonst,} \end{cases} \quad \theta_{\text{penalty}} = -2.$$

Die generierten Konfigurationsposen-Sequenzen wurden zusätzlich mit dem in Abschnitt 6.3 beschriebenen Ansatz auf den humanoiden Roboter ARMAR-4 (Asfour et al. 2013) übertragen, einen am H²T entwickelten humanoiden Roboter, der mit einer Größe von 1,70 m zu einer dem Menschen ähnlichen theoretischen Nutzung der Support-Affordanzen von Geländern, Tischen usw. in der Lage ist. Aufgrund seiner anthropomorphen Abmessungen

und kinematischen Parameter wurde angenommen, dass die in Abschnitt 6.3 beschriebene mögliche Adaption der Translationswerte $T(k, j)$ hierbei nicht erforderlich ist. Trotz Übertragung der resultierenden Posen-Sequenzen auf ARMAR-4 sollen für die Visualisierung im Folgenden bei kleineren Abbildungen dennoch die ursprünglichen Konfigurationsposen des MMM-Referenzmodells dargestellt werden, die aufgrund dessen einfarbiger Textur eine deutlich bessere Unterscheidbarkeit aufweisen.

Tabelle 6.4: Durchgeführte Zahl an Iterationen für die Ausführung von Algorithmus 5 und die hierfür benötigte Laufzeit.

	Iterationen	Laufzeit
Szenario 1	5171	1,71 s
Szenario 2	7046	3,19 s
Szenario 3	14052	8,17 s

Tabelle 6.4 zeigt die bei der Ausführung von Algorithmus 5 durchgeführte Zahl an Iterationen zusammen mit der benötigten Laufzeit auf dem eingesetzten System¹⁴ für alle drei Szenarien. Zusätzlich ist in Abbildung 6.8 die Anzahl der jeweils aktiven Pfade zu verschiedenen Zeitpunkten während der Ausführung von Algorithmus 5 und die durch den jeweils besten Pfad realisierte Translation dargestellt. Hierbei ist sowohl der Effekt des periodischen Pruning alle 1000 Frames (bestimmt durch $\theta_{\text{pruneperiod}}$) deutlich ersichtlich, als auch die bei Szenario 3 durch die Möglichkeit von vier Support-Kontakten entstehende kombinatorische Explosion möglicher Pfade, welche durch die Vielzahl zu berücksichtigender (aber sehr ähnlicher) Pfade zu einem entsprechend langsamen Vorankommen der Planung (Plateau der rechten blauen Kurve) führt, das erst nach dem Pruning der meisten dieser Pfade endet. Abbildung 6.9 visualisiert für Szenario 1 den jeweils besten Pfads als unfertige Konfigurationsposen-Sequenz zu verschiedenen Zeitpunkten der Planung (entsprechend der rechten roten Kurve in Abbildung 6.8).

¹⁴Intel Core i7-4700MQ @ 2,4 GHz, 16 GB RAM.

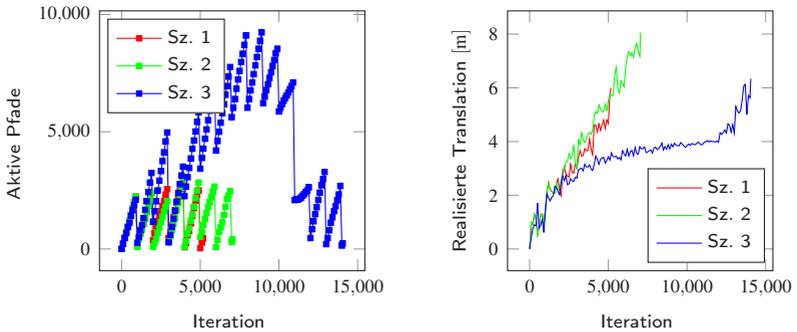


Abbildung 6.8: Anzahl aktiver Pfade (links) zu verschiedenen Zeitpunkten während der Ausführung von Algorithmus 5 und die durch den jeweils besten Pfad realisierte Translation (rechts).

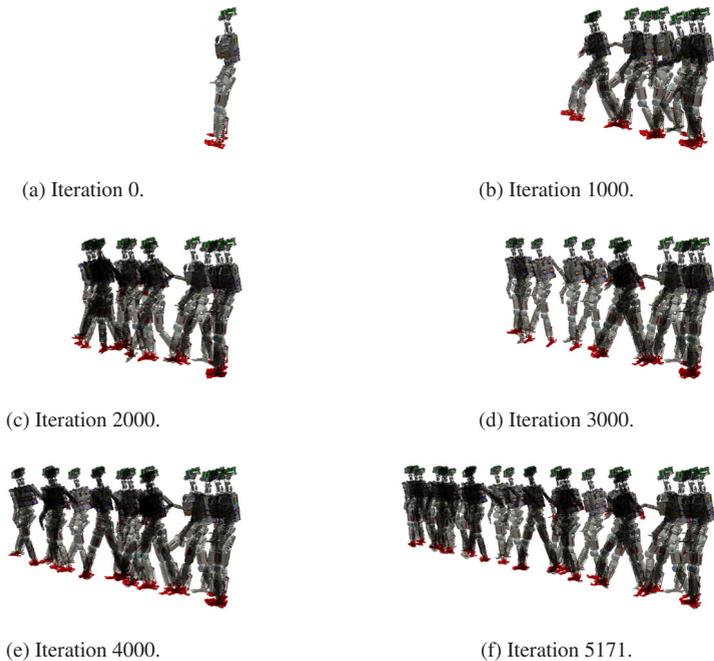


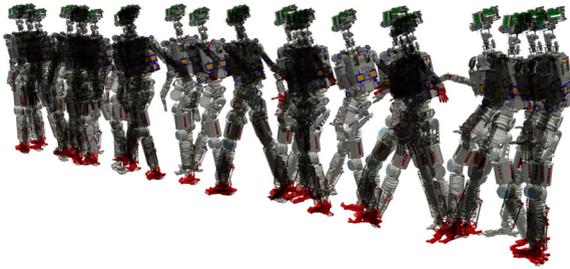
Abbildung 6.9: Jeweils beste Konfigurationsposen-Seqenz zu verschiedenen Zeitpunkten während der Ausführung von Algorithmus 5 für Szenario 1.

Die Laufzeit von Algorithmus 5 lag, trotz des prototypischen Charakters der vorliegenden Implementierung in der Programmiersprache Python, selbst für das komplexeste Szenario 3 deutlich unter 10 s. Somit konnte gezeigt werden, dass der evaluierte Ansatz über – im Vergleich zu anderen Ansätzen für die Generierung von Multi-Kontakt-Posen – vorteilhafte Laufzeitanforderungen verfügt, die in zukünftigen Einsatzgebieten auch eine (Re-)Generierung von Posen-Sequenzen in Echtzeit, beispielsweise als Eingabe für eine Regelung basierend auf quadratischer Programmierung (siehe Abschnitt 7.2), ermöglichen könnte.

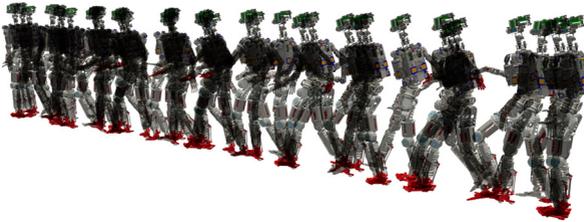
Abbildung 6.10 zeigt eine überlagerte Visualisierung der generierten Konfigurationsposen-Sequenzen für alle drei Szenarien aus der durch Abbildung 6.7 vorgegebenen Perspektive. In Abbildung 6.11 ist hingegen eine separate Darstellung der einzelnen Konfigurationsposen abgebildet, wobei die Posen horizontal entsprechend der jeweiligen mit ihnen assoziierten Translationen positioniert und aus Gründen der Übersichtlichkeit unter Verwendung der für das MMM-Referenzmodell definierten Konfigurationsposen dargestellt werden. Zur besseren Kennzeichnung der jeweils als Support-Kontakt verwendeten Körpersegmente wird die in Abschnitt 6.3 eingeführte Notation hierbei auf die Angabe der beteiligten Körpersegmente erweitert (LF/RF: linker/rechter Fuß, LH/RH: linke/rechte Hand, numerische Suffixe unterscheiden wie in Abschnitt 6.3 Konfigurationsposen derselben Support-Pose). Abschließend zeigt Tabelle 6.5 eine tabellarische Auflistung aller Transitionen zwischen Konfigurationsposen in der für Szenario 1 geplanten Sequenz, bei der für jede Transition $k_{i-1} \rightarrow k_i$ die mit ihr assoziierte Translation $T(k_{i-1}, k_i)$, die Wahrscheinlichkeit $P(k_i | (k_{i-n+1}, \dots, k_{i-1}))$ im N-Gramm-Transitionsmodell (hier: $n = 5$), der Wert der Penalty-Funktion für nicht-genutzte Support-Kontakte (siehe Abschnitt 6.5) und zusätzlich die insgesamt durch die Konfigurationsposen-Subsequenz (k_1, \dots, k_i) realisierte Translation als erreichte Fortbewegungsdistanz angegeben ist.

Bei einer Betrachtung der geplanten Konfigurationsposen-Sequenzen ist zunächst festzuhalten, dass die zusätzlich verfügbaren Hand-Supports in

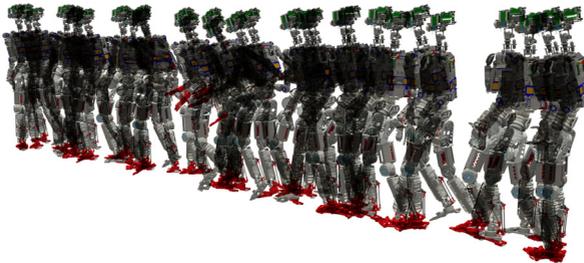
allen drei Szenarien für die überwiegende Wegstrecke ihrer Verfügbarkeit genutzt wurden. Im Hinblick auf die in den geplanten Sequenzen enthaltenen Transitionen zwischen Konfigurationsposen zeigt sich, dass alle generierten Transitionen gültige Übergänge (entsprechend den gestrichelten Kanten in Abbildung 5.2) zwischen Support-Posen der in Abschnitt 5.2 definierten GKP-Taxonomie darstellen. Die generierten Transitionen zwischen Konfigurationsposen erscheinen sowohl hinsichtlich der kinematischen Übergänge zwischen aufeinanderfolgenden Posen als auch hinsichtlich der mit den Transitionen assoziierten Translationen plausibel und es lassen sich keine bedeutenden kinematischen Abweichungen zwischen zwei aufeinanderfolgenden Konfigurationsposen identifizieren, die einer Ausführung der Posen-Sequenz im Wege stehen könnten. Hierzu zeigt Tabelle 6.6 die durchschnittlichen Distanzen zwischen den Positionen jeweils eines Fußes bzw. einer Hand in aufeinanderfolgenden Konfigurationsposen aller drei Szenarien. Es ist ersichtlich, dass in keinem der drei Szenarien die Hände bzw. Füße um durchschnittlich mehr als 27,3 cm bewegt werden müssen, um einen Übergang zur nachfolgenden Konfigurationspose zu realisieren, und dass diese Distanz zudem im dritten Szenario, in dem durch die verfügbaren Support-Affordanzen mehr Hand-Supports genutzt werden können, weiter sinkt. Bei einer Betrachtung der geplanten Bewegungen ist insbesondere zu erkennen, dass in den einem Hand-Support vorhergehenden Posen die Konfiguration des Arms bereits in Antizipation des Kontakts in die entsprechende Richtung bewegt wird (analog zeigt sich für die einem Hand-Support nachfolgenden Posen, dass die Ausrichtung des Arms in nachfolgenden Posen zunächst beibehalten wird). Dies demonstriert die Eignung des N-Gramm-Transitionsmodells in Kombination mit den in Abschnitt 6.3 eingeführten Konfigurationsposen zur Generierung kinematisch schlüssiger Posen-Sequenzen. In Bewegungsabschnitten, in denen über die Füße hinaus keine weiteren Support-Kontakte möglich waren, zeigt sich in den resultierenden Posen-Sequenzen dementsprechend der normale, bereits in Abschnitt 6.3 beschriebene Laufzyklus (in Tabelle 6.5 z.B. nach dem Zurücklegen einer Distanz von 4 m zu sehen).



(a) Szenario 1.



(b) Szenario 2.



(c) Szenario 3.

Abbildung 6.10: Überlagerte Visualisierung der generierten Konfigurationsposen-Sequenzen für ARMAR-4.

Tabelle 6.5: Generierte Konfigurationsposen-Sequenz für Szenario 1 (modifiziert nach Mandery et al. 2016a, ©2016 IEEE). Angegeben sind alle Transitionen zwischen Konfigurationsposen mit der jeweiligen Start-/Ziel-Pose, der mit der Transition assoziierten Translation, der Transitionswahrscheinlichkeit, dem Wert der Penalty-Funktion und der insgesamt bis einschließlich zur Ziel-Pose realisierten Fortbewegungsdistanz.

	Start-KP	Ziel-KP	Transl.	Tr.w.keit	Penalty	Distanz
1	LFRF.1	→ LF.1	0,07 m	37 %	-2	0,07 m
2	LF.1	→ LFRF.2	0,25 m	60 %	0	0,32 m
3	LFRF.2	→ RF.4	0,09 m	17 %	-2	0,41 m
4	RF.4	→ LFRF.10	0,55 m	31 %	0	0,96 m
5	LFRF.10	→ LFRFRH.2	0,04 m	79 %	0	1,00 m
6	LFRFRH.2	→ LFRH.4	0,11 m	79 %	-2	1,11 m
7	LFRH.4	→ LF.3	0,38 m	79 %	-4	1,49 m
8	LF.3	→ LFRF.5	0,41 m	1 %	-2	1,91 m
9	LFRF.5	→ LFRFRH.5	0,10 m	54 %	0	2,01 m
10	LFRFRH.5	→ RFRH.2	0,09 m	5 %	-2	2,10 m
11	RFRH.2	→ LFRFRH.1	0,50 m	30 %	0	2,61 m
12	LFRFRH.1	→ LFRH.4	0,07 m	42 %	-2	2,67 m
13	LFRH.4	→ LF.3	0,38 m	8 %	-4	3,05 m
14	LF.3	→ LFRH.1	0,37 m	2 %	-2	3,42 m
15	LFRH.1	→ LF.3	0,58 m	15 %	-2	4,00 m
16	LF.3	→ LFRF.4	0,09 m	87 %	0	4,10 m
17	LFRF.4	→ RF.2	0,12 m	22 %	-2	4,22 m
18	RF.2	→ LFRF.3	0,50 m	55 %	0	4,72 m
19	LFRF.3	→ LF.1	0,11 m	42 %	-2	4,82 m
20	LF.1	→ LFRF.2	0,25 m	93 %	0	5,08 m
21	LFRF.2	→ RF.2	0,11 m	93 %	-2	5,19 m
22	RF.2	→ LFRF.3	0,50 m	93 %	0	5,69 m
23	LFRF.3	→ LF.2	0,10 m	50 %	-2	5,80 m
24	LF.2	→ LFRF.1	0,21 m	74 %	0	6,00 m

Tabelle 6.6: Mittelwert und Standardabweichung der Positionsveränderung von jeweils einem Fuß bzw. einer Hand zwischen aufeinanderfolgenden Konfigurationsposen.

	Fuß	Hand
Szenario 1	271,7 mm ± 215,9 mm	271,9 mm ± 187,8 mm
Szenario 2	261,1 mm ± 221,9 mm	273,3 mm ± 189,4 mm
Szenario 3	244,0 mm ± 190,5 mm	237,7 mm ± 165,4 mm

Zusammenfassend konnte gezeigt werden, dass mit dem in diesem Kapitel vorgestellten Verfahren vielversprechende Ergebnisse für die drei exemplarisch evaluierten Fortbewegungsaufgaben erzielt werden konnten und sich der beschriebene Ansatz für eine Generierung komplexer Sequenzen von Multi-Kontakt-Posen für die Fortbewegung mit Hand-Supports eignet. Dessen ungeachtet stellt der hier eingeführte Ansatz im Sinne der eingangs erwähnten Verfahrensweise (*contacts-before-motion approach*) lediglich den ersten von zwei erforderlichen Schritten für eine tatsächliche Nutzung derartiger Multi-Kontakt-Bewegungen dar, weswegen sich die hier generierten Posen-Sequenzen auch und vor allem an ihrem zukünftigen Einsatz zur Bewegungsgenerierung auf einem realen humanoiden Roboter messen lassen müssen (siehe Abschnitt 7.2).

6.7 Zusammenfassung von Kapitel 6

In diesem Kapitel wurde ein neuartiger Ansatz zur datengetriebenen Bewegungsgenerierung vorgestellt, der die Erzeugung von Posen-Sequenzen für Multi-Kontakt-Bewegungen humanoider Roboter aus der Beobachtung und Analyse menschlicher Ganzkörperbewegungen ermöglicht. Der vorgestellte Ansatz baut dabei auf der in Kapitel 5 eingeführten Bewegungsrepräsentation auf, indem eine generierte Multi-Kontakt-Bewegung als Abfolge charakteristischer Konfigurationsposen dargestellt wird, und bildet somit den ersten Teilschritt einer etablierten, als *contacts-before-motion approach* bezeichne-

ten Herangehensweise für das Problem der Bewegungsgenerierung in der Robotik. Die 111 Konfigurationsposen, welche die Grundlage des vorgestellten Ansatzes verkörpern, entstammen hierbei einer Unterteilung der in Kapitel 5 eingeführten Support-Posen und werden aus einer Analyse menschlicher Bewegungsdaten gewonnen, an die sich eine Übertragung der resultierenden Gelenkwinkel des menschlichen Körpers auf die Kinematik humanoider Roboter anschließt. Zur Modellierung von Sequenzen der Konfigurationsposen wurde eine linguistische Sichtweise auf das Problem der Bewegungsgenerierung eingeführt, in welcher die eine Bewegung bildenden Posen als *Wörter* zu dem die Bewegung repräsentierenden *Satz* verknüpft werden. Basierend auf dieser Sichtweise wurde mit dem Einsatz eines N-Gramm-Modells ein etablierter Typ von Sprachmodellen aus der Computerlinguistik vorgeschlagen, um bedingte Transitionswahrscheinlichkeiten zwischen Konfigurationsposen statistisch zu beschreiben. Mit dem in Kapitel 5 eingeführten Verfahren segmentierte menschliche Bewegungsaufnahmen bilden die Trainingsdaten für das Lernen dieses N-Gramm-Modells und eines weiteren Translationsmodells zur Beschreibung der mit Posen-Transitionen assoziierten räumlichen Fortbewegung. Die gelernten statistischen Modelle ermöglichen im Anschluss die Planung von Posen-Sequenzen für Bewegungsaufgaben, wofür ein durch Methoden aus der natürlichen Sprachverarbeitung inspirierter heuristischer Suchalgorithmus mit Pruning vorgestellt wurde. Für die abschließende Evaluation des vorgestellten Ansatzes wurde das Problem der geradlinigen Multi-Kontakt-Fortbewegung betrachtet und mögliche Support-Kontakte mit der Umgebung für jede Position zwischen dem gegebenen Start- und Zielpunkt als bekannt angenommen. Zunächst wurden im Rahmen einer Exploration des Parameterraums durch eine Betrachtung der Perplexität als Maß für die Modellgüte geeignete Parameterwerte für die Anwendung des N-Gramm-Modells identifiziert und als Parametrisierung des Modells die Ordnung $n = 5$ und der Einsatz von Witten-Bell-Glättung festgelegt. Im Anschluss wurden drei exemplarische Szenarien für den humanoiden Roboter ARMAR-4 vorgegeben und es konnte demonstriert werden, dass sich der entwickelte Ansatz

für eine Generierung von Posen-Sequenzen in diesen Szenarien eignet. In allen drei Szenarien konnte die Nutzung möglicher Hand-Supports in den geplanten Posen-Sequenzen gezeigt und die kinematische Plausibilität der erzeugten Posen-Sequenzen durch eine Betrachtung der Distanzen zwischen den Positionen der Endeffektoren in aufeinanderfolgenden Posen validiert werden. Die zur Planung der Posen-Sequenzen benötigte Rechenzeit betrug trotz der derzeit prototypischen Implementierung in den drei betrachteten Szenarien lediglich wenige Sekunden und lag damit erheblich unter der Laufzeit bestehender Planungsverfahren, die Sequenzen von Multi-Kontakt-Posen ohne eine Berücksichtigung menschlicher Bewegungsdaten generieren.

7 Zusammenfassung und Ausblick

Das Ziel der vorliegenden Arbeit war die Entwicklung eines neuartigen Verfahrens zur datengetriebenen Generierung von Multi-Kontakt-Bewegungen für humanoide Roboter, bei denen neben den Füßen Kontakte weiterer Körperteile wie etwa der Hände mit der Umgebung zu einer Erhöhung der Stabilität bei der Ausführung von Fortbewegungsaufgaben genutzt werden. Grundlage für das entwickelte Verfahren bildet die Analyse von Aufnahmen menschlicher Ganzkörperbewegungen, welche hierfür in einer im Rahmen dieser Arbeit entwickelten großskaligen Bewegungsdatenbank organisiert und in einer vereinheitlichenden Repräsentation bereitgestellt werden. Die Segmentierung dieser Bewegungsaufnahmen und ihre Darstellung als Sequenz von Posen einer Ganzkörperposen-Taxonomie ermöglicht eine Abstraktion und Komplexitätsreduktion der Bewegungsrepräsentation und stellt die Basis für das Lernen eines statistischen Modells möglicher Transitionen zwischen diesen Ganzkörperposen dar. Die Generierung von Multi-Kontakt-Bewegungen für humanoide Roboter als Abfolge der sie beschreibenden Ganzkörperposen kann im Anschluss als Optimierungsproblem in dem gelernten statistischen Modell interpretiert und als solches gelöst werden.

In Folgenden werden die Beiträge dieser Arbeit kurz zusammengefasst und ein Ausblick auf mögliche zukünftige Forschungsarbeiten gegeben.

7.1 Wissenschaftliche Beiträge der Arbeit

Im Einzelnen wurden im Rahmen dieser Arbeit folgende wissenschaftliche Beiträge geleistet:

Repräsentation menschlicher Ganzkörperbewegung und deren Organisation in einer großskaligen Bewegungsdatenbank: Das in dieser Arbeit vorgestellte Verfahren zur Bewegungsgenerierung bedingt die Verfügbarkeit von größeren Mengen an aufgenommenen Ganzkörperbewegungen unterschiedlicher menschlicher Probanden. Aus diesem Grund wurden zwei neue Verfahren für das am H^2T entstandene MMM-Framework entwickelt, welche basierend auf nicht-linearer Optimierung bzw. Kalman-Filterung eine effiziente und robuste Rekonstruktion der menschlichen Ganzkörperbewegung als Gelenkwinkeltrajektorie des MMM-Referenzmodells aus gelabelten oder ungelabelten Messungen eines markerbasierten Bewegungserfassungssystems und die Übertragung der rekonstruierten Bewegungen auf die Kinematik eines humanoiden Roboters ermöglichen.

Die MMM-Repräsentation einer Bewegung legte im Anschluss die Basis zur Definition von 29 unterschiedlichen Merkmalen mit einer Gesamtzahl von 702 Dimensionen für die Modellierung menschlicher Ganzkörperbewegung mit Hilfe von HMMs. Im nächsten Schritt wurden mit einem Wrapper-Ansatz zur Merkmalsauswahl hochdeskriptive niedrigdimensionale Teilmengen dieser Merkmale identifiziert und es konnte im Zuge der Evaluation gezeigt werden, dass für lediglich 8-dimensionale Merkmalsvektoren bereits Kombinationen der Merkmale gefunden werden können, die auf dem zugrunde liegenden Problem der Bewegungsklassifikation mit einem F_1 -Score von 95,77 % alle zu Vergleichszwecken evaluierten klassischen Merkmalsvektoren übertreffen, gleichzeitig aber aufgrund ihrer niedrigen Dimensionalität eine deutlich reduzierte Trainingszeit aufweisen. Aufbauend auf der niedrigdimensionalen Repräsentation beobachteter Bewegungen wurden unterschiedliche Verfahren zur HMM-basierten Multi-Label-Klassifikation menschlicher Ganzkörperbewegungen als Problemtransformations- und Algorithmenanpassungsverfahren vorgestellt und evaluiert, wobei ein auf der Binäre-Relevanz-Methode basierendes Problemtransformationsverfahren mit logistischer Regression als nachgelagertem binären Klassifikator auf dem

Evaluationsdatensatz, bestehend aus 579 Bewegungsaufnahmen mit einer Label-Kardinalität von 2,25, einen F_1 -Score von 89,31 % erzielen konnte.

Zur Verwaltung erfasster Bewegungsdaten wurde im Rahmen dieser Arbeit die *KIT Whole-Body Human Motion Database* entwickelt, die neben den erfassten menschlichen Bewegungen in der rekonstruierten MMM-Repräsentation auch die für eine Analyse der Interaktion zwischen Mensch und Umgebung entscheidenden Posen der involvierten Umgebungsobjekte bereitstellt. Für die Organisation der Daten in der Bewegungsdatenbank wurden hierarchisch strukturierte Tags eingeführt, welche eine Spezifikation von Bewegungstyp und -modalitäten (z.B. Geschwindigkeit und Richtung) erlauben und als *Motion Description Tree* eine systematische Abfrage des Datenbestands ermöglichen. Zusätzlich zur Möglichkeit der manuellen Klassifikation verfügbarer Bewegungen durch den Benutzer wurde hierfür das zuvor beschriebene Verfahren zur Multi-Label-Klassifikation für eine datengetriebene Erzeugung von Tagging-Vorschlägen in die Bewegungsdatenbank integriert und im Rahmen einer Benutzerstudie mit 115 der generierten Vorschläge validiert.

Segmentierung menschlicher Bewegung zur Extraktion von Support-

Posen: Ausgangspunkt für die Analyse menschlicher Bewegungsaufnahmen und die sich anschließende datengetriebene Generierung neuer Bewegungen stellt die am H²T entwickelte, aus 46 Support-Posen bestehende Ganzkörperposen-Taxonomie dar. Die Darstellung menschlicher Bewegungen als Abfolge von durch die Taxonomie vorgegebenen Posen ermöglicht hierbei die erforderliche Komplexitätsreduktion des Problems und liefert die Trainingsdaten für das im Rahmen dieser Arbeit beschriebene Verfahren zur datengetriebenen Bewegungsgenerierung. Zur automatischen Segmentierung menschlicher Bewegungsdaten und ihrer Darstellung als Abfolge von durch die Taxonomie vorgegebenen Support-Posen wurde ein heuristisches Segmentierungsverfahren entwickelt, welches die durch die Bewegungsdatenbank bereitgestellte MMM-Repräsentation einer Bewegung einschließlich beteilig-

ter Umgebungsobjekte zur Bestimmung der Support-Kontakte des Menschen mit seiner Umgebung zu jedem Zeitpunkt der Bewegung nutzt.

Bei einer Validierung des entwickelten Segmentierungsverfahren mit 121 Aufnahmen unterschiedlicher menschlicher Ganzkörperbewegungen, für welche manuell die korrekten Segmentierungen vorgegeben wurden, konnte gezeigt werden, dass 98 % der durch das Segmentierungsverfahren identifizierten Posen korrekt waren und 95 % der vorgegebenen Posen durch das Segmentierungsverfahren erkannt werden konnten. Im Anschluss wurde die Segmentierung von Bewegungen für eine partielle Validierung der dieser Arbeit zugrunde liegenden Ganzkörperposen-Taxonomie genutzt. Hierfür wurde durch eine Betrachtung der bei der Segmentierung auftretenden Support-Posen eine statistische Analyse zur Auftretenshäufigkeit einzelner Posen und der Transitionen zwischen diesen Posen sowie der mit diesen Transitionen assoziierten Zeitdauern durchgeführt, und es wurde demonstriert, dass sich die Darstellung einer Bewegung als Sequenz von Support-Posen als aussagekräftige semantische Beschreibung für Multi-Kontakt-Bewegungen eignet.

Datengetriebene Generierung von Multi-Kontakt-Bewegungen für humanoide Roboter: Zum Einsatz der durch die Taxonomie vorgegebenen Support-Posen für das Problem der Bewegungsgenerierung wurde eine nachgelagerte, aus der Analyse menschlicher Bewegungsdaten abgeleitete Unterteilung dieser Posen hinsichtlich ihrer Ganzkörperkonfiguration vorgenommen, die in einer Gesamtzahl von 111 charakteristischen, auf die Kinematik humanoider Roboter übertragbaren Konfigurationsposen resultierte. Anhand dieser Posen konnte aus der segmentierten Darstellung menschlicher Bewegungsaufnahmen im Anschluss ein statistisches Modell der menschlichen Ganzkörperbewegung gelernt werden, welches auf einem N-Gramm-Modell zur Beschreibung bedingter Transitionswahrscheinlichkeiten zwischen Posen und einer Untersuchung zu den mit diesen Transitionen assoziierten Transitionen des Ganzkörper-Massenschwerpunkts basiert. Der Einsatz eines

N-Gramm-Modells entspringt hierbei einer nach Kenntnisstand des Autors neuartigen linguistischen Sichtweise auf das Problem der Generierung von Multi-Kontakt-Bewegungen, im Rahmen derer die Posen, welche eine Bewegung begründen, als *Wörter* zu einem die Bewegung repräsentierenden *Satz* verknüpft werden. Diese Sichtweise auf das Problem erlaubt eine Adaption von in der statistischen Sprachmodellierung bekannten Methoden und ermöglicht dadurch die als Wegfindungsproblem in dem Baum möglicher Posen-Transitionen aufgefasste Generierung von Multi-Kontakt-Bewegungen mit einem durch solche Methoden (*Beam Search*) inspirierten heuristischen Suchalgorithmus mit Pruning.

Im Zuge der Evaluation wurde zunächst eine Exploration des Parameter-raums der eingesetzten statistischen Modelle zur Identifikation geeigneter Parameterwerte für die Anwendung dieser Modelle vorgenommen. Im nächsten Schritt wurden exemplarische Fortbewegungsaufgaben betrachtet, für welche die Distanz der als geradlinig angenommenen Fortbewegung sowie positionsabhängig die erlaubten Support-Kontakte (z.B. mit den Händen) vorgegeben wurden, und es konnte die Eignung des vorgeschlagenen Verfahrens für die Generierung komplexer Posen-Sequenzen in diesen Szenarien demonstriert werden. Die hierfür benötigte Rechenzeit lag jeweils im Bereich weniger Sekunden und damit deutlich unter der Laufzeit von aus der Literatur bekannten Planungsverfahren für Sequenzen von Multi-Kontakt-Posen, die nicht auf einer Verwendung menschlicher Bewegungsdaten basieren.

7.2 Diskussion und Ausblick

Im Rahmen der vorliegenden Arbeit wurde ein neuartiges auf der Analyse menschlicher Ganzkörperbewegungen basierendes Verfahren zur Generierung von Posen-Sequenzen für Multi-Kontakt-Bewegungen humanoider Roboter zusammen mit den für seine Anwendung notwendigen Voraussetzungen hinsichtlich der Organisation und Repräsentation erforderlicher Bewegungsdaten vorgestellt. Für zukünftige Forschungsarbeiten ergeben sich hieraus

zahlreiche weiterführende Fragestellungen und Ansätze für eine Weiterentwicklung der in dieser Arbeit beschriebenen Methoden, die im Folgenden kurz diskutiert werden sollen.

Bei der im Zuge dieser Arbeit entwickelten und in Kapitel 4 vorgestellten *KIT Whole-Body Human Motion Database* handelt es sich mit derzeit über 28 Stunden aufgenommener Bewegungsdaten in fast zehntausend Aufnahmen bereits jetzt um eine der weltweit größten existierenden Bewegungsdatenbanken, die seit ihrer erstmaligen Vorstellung im Jahr 2014 ein fortwährendes Interesse in der wissenschaftlichen Gemeinschaft erfährt¹ und Einsatz innerhalb mehrerer EU-Forschungsprojekte fand. Die mutmaßlich ausschlaggebenden Faktoren hierfür stellen die freie Zugänglichkeit eines breiten Spektrums menschlicher Ganzkörperbewegungen, die Verfügbarkeit von Bewegungsaufnahmen in einer normalisierten Darstellungsform, ihre semantische Organisation mit entsprechenden Abfragemöglichkeiten und die Berücksichtigung von Umgebungsobjekten dar.

Es versteht sich daher von selbst, dass neben einer Erweiterung des Datenbestands auch die Bewegungsdatenbank an sich weiterentwickelt werden soll, um ihre Anwendbarkeit und Nützlichkeit für eine große Bandbreite an Forschungsfragestellungen im Bereich der menschlichen Ganzkörperbewegung zu erhöhen. Unter Beteiligung des Autors wurde ein auf Methoden des Deep Learning (*Sequence-to-Sequence Learning* mit tiefen rekurrenten neuronalen Netzen) basierendes Verfahren zur statistischen wechselseitigen Verknüpfung menschlicher Ganzkörperbewegung mit Bewegungsbeschreibungen in natürlicher Sprache entwickelt (Plappert et al. 2017), das wie in Abbildung 7.1 gezeigt nach einer Integration in die Bewegungsdatenbank automatisch natürlichsprachliche Beschreibungen zur Ergänzung der durch den *Motion Description Tree* bereitgestellten Tags erzeugen könnte. Auf

¹ Mit Stand vom Mai 2017 haben sich mehr als 300 Benutzer für einen Zugriff auf die Bewegungsdatenbank registriert und die begleitende wissenschaftliche Konferenzpublikation (Mandery et al. 2015b) wurde in weniger als zwei Jahren bereits 30 Mal zitiert (Quelle: Google Scholar).

der anderen Seite könnten für die Multi-Label-Bewegungsklassifikation zur datengetriebenen Generierung von Tagging-Vorschlägen neben HMMs auch verwandte Ansätze wie FHMMs oder PHMMs oder beispielsweise ebenfalls Methoden aus dem Bereich des Deep Learning zum Einsatz kommen. Für die hierzu erforderliche Modellierung menschlicher Ganzkörperbewegung könnte die Relevanz einzelner Gelenke oder Marker eingehender untersucht oder die Menge möglicher Merkmale erweitert werden, wofür sich etwa die in (Kapadia et al. 2013) eingeführten Merkmale basierend auf den Laban-Bewegungsstudien (Laban et al. 2011) eignen würden.

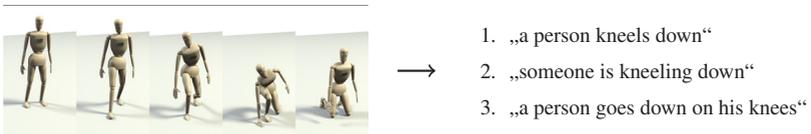


Abbildung 7.1: Automatische Generierung möglicher natürlichsprachlicher Beschreibungen für menschliche Ganzkörperbewegungen (modifiziert nach Plappert et al. 2017).

Für eine tatsächliche Ausführung von Posen-Sequenzen, die mit dem in dieser Arbeit beschriebenen datengetriebenen Verfahren erzeugt wurden, auf humanoiden Robotern in realen Umgebungen verbleiben zwei fehlende Schlüsselkomponenten, die Anfang bzw. Ende einer gedachten Verarbeitungskette repräsentieren und in der vorliegenden Arbeit nicht explizit behandelt wurden: Auf der einen Seite muss ein humanoider Roboter in der Lage sein, durch die sensorische, meist mit Kameras realisierte Erfassung seiner Umgebung autonom Support-Affordanzen in der Umgebung und somit mögliche Support-Kontakte zu identifizieren, anstatt dass diese als bekannt vorausgesetzt werden. Als Problem des maschinellen Sehens und der Szeneninterpretation geht dieser Anspruch zwar weit über den Rahmen dieser Arbeit hinaus, allerdings existieren hierzu bereits zahlreiche Ansätze (siehe Abschnitt 6.2), die sich für eine Integration anbieten könnten. Auf der anderen Seite müssen generierte Posen-Sequenzen unter Berücksichtigung von Stabilitätskriterien zu kontinuierlichen Trajektorien verbunden und ge-

eignete Regler zur tatsächlichen Ausführung dieser Trajektorien auf einem realen Roboter entworfen werden. In einer Kooperation mit dem *Laboratoire d'Informatique, de Robotique et de Microélectronique* (LIRMM) in Montpellier wurden bereits erste Versuche hierzu auf Basis quadratischer Programmierung durchgeführt und es konnte mit Posen-Sequenzen für Laufbewegungen (ohne Verwendung der Hände), welche mit dem in dieser Arbeit entwickelten Verfahren generiert wurden, die Erzeugung stabiler kontinuierlicher Trajektorien und ihre in Abbildung 7.2 dargestellte Ausführung in einer dynamischen Simulation des Roboters HRP-4 demonstriert werden.

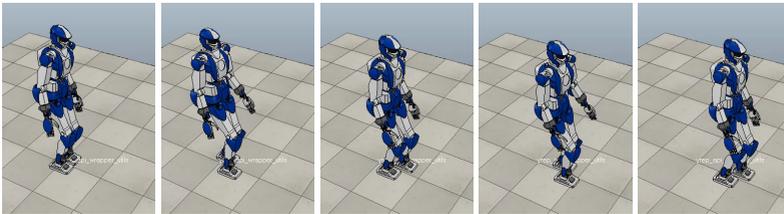


Abbildung 7.2: Ausführung einer mit dem in dieser Arbeit vorgestellten Verfahren generierten Posen-Sequenz für eine stabile Laufbewegung in einer dynamischen Simulation des Roboters HRP-4.

Darüber hinaus wurde im Kontext dieser Arbeit die Generierung von Multi-Kontakt-Bewegungen mit Fortbewegungsaufgaben (unter den zuvor beschriebenen Annahmen) lediglich für einen einzelnen Typ von Ganzkörperbewegungen thematisiert. Ohne Ausweitung auf weitere mögliche Bewegungstypen bestünde die Möglichkeit zu einer allgemeineren Modellierung dieser Fortbewegungsaufgaben, wofür beispielsweise die Annahme einer geradlinigen Fortbewegung aufgegeben und die dreidimensionale Translation des Ganzkörper-Massenschwerpunkts berücksichtigt werden könnte, um eine Bewegungsgenerierung für Fortbewegung um Hindernisse oder mit Richtungsänderungen sowie auf Treppen und Leitern zu ermöglichen. Demgegenüber stehen mögliche Erweiterungen der Bewegungsgenerierung auf über Fortbewegung hinausgehende Bewegungstypen wie

(Loko-)Manipulationsaufgaben, die sich als Intra-Posen-Transitionen ebenfalls mit Hilfe der eingesetzten Ganzkörperposen-Taxonomie beschreiben lassen. Hierfür könnte einerseits das entwickelte Segmentierungsverfahren hinsichtlich einer robusten Unterscheidung von Flächen- und Haltekontakten der Hand (wofür höchstwahrscheinlich zusätzliche Finger-Marker oder Sensoren wie Datenhandschuhe benötigt werden) weiterentwickelt werden. Andererseits existieren für die Generalisierung und Erzeugung von Manipulationsbewegungen aus der Beobachtung des Menschen bereits hochentwickelte Ansätze im Bereich des *Programmierens durch Vormachen* (Wächter et al. 2013), deren (sicher nicht triviale) Integration mit den in dieser Arbeit vorgestellten Methoden spannend erscheint.

A Bewegungsdatensätze der Evaluation in Abschnitt 3.5

Tabelle A.1: Bewegungsklassen mit der jeweiligen Anzahl an Bewegungsaufnahmen für die Evaluation der Merkmalsauswahl in Abschnitt 3.5.1 (modifiziert nach Mandery et al. 2016b, ©2016 ISIF). Die angegebenen IDs korrespondieren mit den Bewegungsexperimenten in der KIT-Bewegungsdatenbank unter der URL <https://motion-database.humanoids.kit.edu/details/motions/<ID>/>.

Bewegungsklasse	# Aufn.	ID(s) in KIT-Bew.datenbank
Laufen	49	318, 362, 395, 452, 467
Rennen	41	324, 364, 399, 426, 533
Drehung (auf der Stelle)	59	326, 327, 402, 403, 445, 446
Stoß von hinten	9	476, 477, 478
Werfen	10	573, 581
Hinknien	5	15
Verbeugung	10	582, 609
Treten	20	610, 611, 612, 613
Kniebeuge	5	616
Faustschlag	10	617, 618
Fuß-Stampfen	10	619, 620
Sprung	25	621, 622, 623, 624, 625
Golf-Putt	5	626
Golf-Drive	6	627
Tennis-Smash	10	628, 629
Tennis-Vorhand	10	630, 631
Winken	15	633, 634, 635
Gitarre spielen	11	636, 637
Violine spielen	10	638, 639
Rühren	11	640, 641
Wischen	11	642, 643
Walzer tanzen	6	644
Cha-Cha-Cha tanzen	5	645
Summe	353	

Tabelle A.2: Kombinationen von (durch Semikola getrennten) Bewegungslabels zusammen mit der jeweiligen Anzahl an Bewegungsaufnahmen für die Evaluation der Multi-Label-Bewegungsklassifikation in Abschnitt 3.5.2. Die angegebenen IDs korrespondieren mit den Bewegungsexperimenten in der KIT-Bewegungsdatenbank unter der URL <https://motion-database.humanoids.kit.edu/details/motions/<ID>/>.

Bewegungsklassen	# Aufn.	ID(s) in KIT-Bew.datenb.
Laufen; Geschw.-Normal; Richtung-Vorne	10	395
Laufen; Geschw.-Langsam; Richtung-Vorne	10	396
Laufen; Geschw.-Schnell; Richtung-Vorne	9	398
Laufen; Kurve-Rechts	10	444
Laufen; Kurve-Links	10	443
Laufen; Richtung-Kreis; Uhrzeigersinn	17	40, 58
Laufen; Richtung-Kreis; Gegen-Uhrzeigersinn	18	39, 57
Laufen; Richtung-Slalom	27	337, 405, 450
Laufen; Richtung-Aufwärts	30	410, 411, 454, 448, 1320
Laufen; Richtung-Abwärts	31	412, 413, 460, 461, 1321
Laufen; Richtung-Abwärts; Richtung-Hinten	5	416
Rennen; Richtung-Vorne	41	324, 364, 399, 426, 533
Laufen; Drehung-Rechts	31	326, 402, 446
Laufen; Drehung-Links	28	327, 403, 445
Stoß-Ausfallschritt; Richtung-Vorne	9	476, 477, 478
Stoß-Ausfallschritt; Richtung-Hinten	9	473, 474, 475
Stoß-Ausfallschritt; Richtung-Rechts	7	482, 483, 484
Stoß-Ausfallschritt; Richtung-Links	10	479, 480, 481

(Fortsetzung)

Bewegungsklassen	# Aufn.	ID(s) in KIT-Bew.datenb.
Lehnen; Hand-Beide	5	519
Werfen; Hand-Rechts	5	573
Werfen; Hand-Links	5	581
Verbeugung; Tief	5	582
Verbeugung; Flach	5	609
Treten; Hoch; Fuß-Rechts	5	610
Treten; Hoch; Fuß-Links	5	611
Treten; Tief; Fuß-Rechts	5	612
Treten; Tief; Fuß-Links	5	613
Kniebeuge	5	614
Faustschlag; Hand-Rechts	5	615
Faustschlag; Hand-Links	5	618
Fuß-Stampfen; Fuß-Rechts	5	619
Fuß-Stampfen; Fuß-Links	5	620
Sprung; Richtung-Hoch	5	621
Sprung; Richtung-Vorne	5	622
Sprung; Richtung-Hinten	5	623
Sprung; Richtung-Rechts	5	625
Sprung; Richtung-Links	5	624
Golf; Golf-Putt	5	626
Golf; Golf-Drive	6	627
Tennis; Tennis-Smash; Hand-Rechts	5	628
Tennis; Tennis-Smash; Hand-Links	5	629
Tennis; Tennis-Vorhand; Hand-Rechts	5	630
Tennis; Tennis-Vorhand; Hand-Links	5	631
Winken; Hand-Rechts	5	633
Winken; Hand-Links	5	634

(Fortsetzung)

Bewegungsklassen	# Aufn.	ID(s) in KIT-Bew.datenb.
Winken; Hand-Beide	5	635
Gitarre spielen; Hand-Rechts	5	636
Gitarre spielen; Hand-Links	6	637
Violine spielen; Hand-Rechts	5	638
Violine spielen; Hand-Links	5	639
Rühren; Hand-Rechts	5	640
Rühren; Hand-Links	6	641
Wischen; Hand-Rechts	5	642
Wischen; Hand-Links	6	643
Tanzen; Tanzen-Waltzer	6	644
Tanzen; Tanzen-Cha-Cha-Cha	5	645
Knieen; Richtung-Abwärts; Hand-Rechts	10	727
Knieen; Richtung-Abwärts; Hand-Links	10	728
Knieen; Richtung-Aufwärts; Hand-Rechts	11	729
Knieen; Richtung-Aufwärts; Hand-Links	10	730
Nehmen; Hand-Rechts	6	1234
Nehmen; Hand-Links	5	1235
Essen; Hand-Rechts	5	1242
Essen; Hand-Links	5	1243
Haare kämmen; Hand-Rechts	5	1245
Haare kämmen; Hand-Links	5	1246
Summe	579	

Abbildungsverzeichnis

1.1	Beispiele für menschliche Fortbewegung unter Nutzung des ganzen Körpers.	2
2.1	Durch das MMM-Framework realisierte Entkoppelung der Erfassung menschlicher Bewegung von ihren verschiedenen Anwendungsfeldern, wie der Übertragung der Bewegung auf humanoide Roboter, Bewegungserkennung oder -analyse.	13
2.2	Multimodale Bewegungserfassung im Labor des H ² T.	14
2.3	Kinematik des MMM-Referenzmodells für den menschlichen Körper.	15
2.4	Kategorisierung von Verfahren zur Erfassung der menschlichen Bewegung nach Field et al. (2011, 2009).	18
2.5	Positionierung von passiv-reflektierenden Markern am menschlichen Körper gemäß dem <i>KIT Reference Marker Set</i>	22
2.6	Rekonstruktion eines komplexen menschlichen Ganzkörper-Bewegungsablaufs basierend auf ungelabelten Markermessungen.	27
2.7	Mit Markern versehene Umgebungselemente und Objekte zur zeitsynchronen Erfassung der Bewegungen von Mensch und Objekten.	29
2.8	Rekonstruktion von Gelenkwinkeln für eine Bewegung.	29
2.9	Verlauf des (nicht normalisierten) Ganzkörper-Massenschwerpunkts (Merkmal <i>CoM_pos</i>) und des Ganzkörper-Drehimpulses (Merkmal <i>angular_momentum</i>) für eine exemplarische Laufbewegung mit vier Schritten.	35

2.10	Vergleich der Verläufe des ungeglätteten (links) und geglätteten (rechts) Geschwindigkeitsvektors des Ganzkörper-Massenschwerpunkts (Merkmal <i>CoM_vel</i>) für eine exemplarische Laufbewegung mit vier Schritten.	37
2.11	Rekonstruktion einer menschlichen Laufbewegung als MMM-Repräsentation und Übertragung der Bewegung auf die humanoiden Roboter ARMAR-4 und NAO.	42
3.1	Darstellung eines beispielhaften HMM mit drei versteckten Zuständen s_1 , s_2 und s_3 und durch drei Ausgabesymbole o_1 , o_2 und o_3 repräsentierte diskrete Emissionen.	52
3.2	Schematische Darstellung verschiedener Topologien für ein HMM mit vier versteckten Zuständen.	59
3.3	Multi-Class-Bewegungsklassifikation.	60
3.4	Multi-Label-Bewegungsklassifikation mit Label-Potenzmengen-Methode.	71
3.5	Multi-Label-Bewegungsklassifikation mit Binäre-Relevanz-Methode.	73
3.6	Multi-Label-Bewegungsklassifikation mit angepasstem Algorithmus.	74
3.7	Manuell ausgewählte Schlüsselbilder exemplarischer Bewegungsaufnahmen des zur Merkmalsauswahl verwendeten Datensatzes.	76
3.8	F_1 -Maße der jeweils zehn besten Merkmalsmengen für die Dimensionalitäten 1 bis 15.	80
4.1	Wichtige in der KIT-Bewegungsdatenbank verfügbare Datentypen und der Datenfluss zwischen ihnen.	102
4.2	Satz anthropometrischer Messungen, der von menschlichen Probanden in der KIT-Bewegungsdatenbank erhoben wird.	103

4.3	Erstellung von anonymisierten Fassungen hochgeladener Videoaufnahmen mit Hilfe eines Anonymisierungswerkzeugs durch menschliche Unterstützung.	105
4.4	Stark abstrahierter Aufbau der KIT-Bewegungsdatenbank und bestehende Interaktionsmöglichkeiten mit dem Benutzer.	107
4.5	UML-Klassendiagramm für die an der Modellierung von in der Datenbank verfügbaren Datenbankobjekten beteiligten Klassen und ihren Beziehungen.	109
4.6	UML-Klassendiagramm für die an der Modellierung von hochgeladenen Dateien beteiligten Klassen und ihren Beziehungen.	111
4.7	Webbasierte Oberfläche der KIT-Bewegungsdatenbank in den zwei unterschiedlichen Gestaltungsvarianten.	114
4.8	Exemplarischer Auszug aus den Slice-Definitionen zur Festlegung von Datenstrukturen und Methoden für die Programmierschnittstelle der KIT-Bewegungsdatenbank.	115
4.9	Auszug aus der Struktur des Motion Description Tree für die Klassifikation menschlicher Ganzkörperbewegung.	119
4.10	Beispiele für die Klassifikation von Bewegungsaufnahmen gemäß dem Motion Description Tree.	120
4.11	Webbasierte Oberfläche zum Sammeln natürlichsprachlicher Annotationen für menschliche Ganzkörperbewegungen durch Crowdsourcing.	123
4.12	UML-Klassendiagramm für die an der Generierung und Verwaltung von Klassifikationsvorschlägen beteiligten Klassen und ihren Beziehungen.	125
4.13	Verschiedene Bewegungstypen in Aufnahmen zweibeiniger Fortbewegung der KIT-Bewegungsdatenbank.	137
5.1	Mögliche Typen von Support-Kontakten für Arme und Beine.	153
5.2	Am H ² T entworfene Taxonomie menschlicher Ganzkörper-Support-Posen.	157

5.3	Repräsentation einer beispielhaften Ganzkörperbewegung (Treppe mit Geländer aufwärts steigen) als mit dem in Abschnitt 5.3 vorgestellten Segmentierungsverfahren ermittelte Abfolge von Support-Posen der GKP-Taxonomie.	160
5.4	Beispielhafte kinematische Instanzen von SP 4.1 und SP 4.10 für unterschiedliche Umgebungsgeometrien.	161
5.5	Bei der Segmentierung ermittelte Support-Posen für manuell ausgewählte Schlüsselbilder exemplarischer Ganzkörperbewegungen.	169
5.6	Aus der Segmentierung einer exemplarischen Bewegung (Treppe mit Geländer aufwärts steigen) resultierende Abfolge von Support-Posen.	170
5.7	Manuell ausgewählte Schlüsselbilder einiger der im Bewegungsdatensatz enthaltenen Bewegungsaufnahmen (siehe Tabelle 5.1). 173	
5.8	Durch Segmentierung der analysierten Bewegungsaufnahmen erzeugter gerichteter Graph für die Transitionen zwischen Support-Posen.	177
5.9	Verteilung der Zeitdauern als Histogramme mit einer Klassengröße von 100 ms (entsprechend 10 Frames) für spezifische Support-Posen-Transitionen.	183
6.1	Automatische Extraktion geometrischer Primitive für die Affordanzen der Umgebung aus RGB-D-Daten mit dem in (Kaiser et al. 2016) beschriebenen Verfahren.	204
6.2	Exemplarische Definition einer Fortbewegungsaufgabe.	206
6.3	Veranschaulichung des Konzepts von Konfigurationsposen durch die Darstellung exemplarischer Optionen für die Konfiguration der Arme und Beine innerhalb der SP 2.3.	209
6.4	Exemplarisch ausgewählte Konfigurationsposen verschiedener Support-Posen für das MMM-Referenzmodell und ihre Übertragung auf den humanoiden Roboter ARMAR-4.	212

6.5	Auszug aus der Repräsentation eines gelernten Trigramm-Modells im ARPA-Back-Off-Dateiformat.	217
6.6	Manuell ausgewählte Schlüsselbilder aus einigen der in Tabelle 6.2 aufgelisteten Bewegungsaufnahmen.	229
6.7	Visualisierung der drei evaluierten exemplarischen Fortbewegungsaufgaben.	232
6.8	Anzahl aktiver Pfade (links) zu verschiedenen Zeitpunkten während der Ausführung von Algorithmus 5 und die durch den jeweils besten Pfad realisierte Translation (rechts).	235
6.9	Jeweils beste Konfigurationsposen-Sequenz zu verschiedenen Zeitpunkten während der Ausführung von Algorithmus 5 für Szenario 1.	235
6.10	Überlagerte Visualisierung der generierten Konfigurationsposen-Sequenzen für ARMAR-4.	238
6.11	Generierte Konfigurationsposen-Sequenzen, wobei jede Pose horizontal entsprechend ihrer Translation aufgetragen ist.	239
7.1	Automatische Generierung möglicher natürlichsprachlicher Beschreibungen für menschliche Ganzkörperbewegungen.	251
7.2	Ausführung einer mit dem in dieser Arbeit vorgestellten Verfahren generierten Posen-Sequenz für eine stabile Laufbewegung in einer dynamischen Simulation des Roboters HRP-4.	252

Tabellenverzeichnis

2.1	Übersicht über die definierten Merkmale zur Beschreibung menschlicher Ganzkörperbewegung, zusammen mit der jeweiligen Dimensionalität der Merkmale.	31
2.2	Verwendete Gelenke des MMM-Referenzmodells (vgl. Abbildung 2.3) zur Repräsentation menschlicher Ganzkörperbewegung in den Merkmalen <i>joint_pos</i> , <i>joint_vel</i> und <i>joint_acc</i>	33
3.1	Jeweils beste gefundene Merkmalsmenge für die Dimensionalitäten 1 bis 15.	78
3.2	Vergleich von niedrigdimensionalen Merkmalsmengen mit manuell ausgewählten, häufig zur Beschreibung menschlicher Ganzkörperbewegung eingesetzten Merkmalen hinsichtlich Dimensionalität, Klassifikationsleistung und insgesamt benötigter Trainingszeit.	83
3.3	Ergebnisse der Evaluation unterschiedlicher Verfahren zur Multi-Label-Bewegungsklassifikation.	87
4.1	Vergleich ausgewählter existierender Bewegungsdatenbanken. . .	98
4.2	Ergebnisse der Benutzerstudie für die datengetrieben generierten Klassifikationsvorschläge.	130
4.3	Bewegungsexperimente und -aufnahmen der KIT-Bewegungsdatenbank, aufgeschlüsselt nach beitragenden Instituten.	131
4.4	Bewegungsdaten der KIT-Bewegungsdatenbank, aufgeschlüsselt nach ihrer Zugehörigkeit zu Forschungsprojekten.	133

4.5	Statistische Verteilung von Alter, Körpergröße und Gewicht für die in der KIT-Bewegungsdatenbank verfügbaren Probanden. . .	134
4.6	Aufnahmen zweibeiniger Fortbewegung in der KIT-Bewegungsdatenbank, aufgeschlüsselt nach unterschiedlichen Bewegungstypen.	136
5.1	Zur Evaluation eingesetzter Bewegungsdatensatz, zusammen mit den Ergebnissen für die Validierung des in Abschnitt 5.3 vorgestellten Verfahrens zur Support-Posen-Segmentierung. . .	172
5.2	Relative Auftretenshäufigkeiten spezifischer Support-Posen-Transitionen, kniende Posen ausgeschlossen.	179
5.3	Benötigte Zeitdauern für spezifische Support-Posen-Transitionen, kniende Posen ausgeschlossen.	181
6.1	Unterteilung von Support-Posen in unterschiedliche Kombinationen von Support-Kontakten, zusammen mit der jeweiligen Zahl der Vorkommen dieser Support-Posen und der Anzahl der extrahierten Konfigurationsposen.	210
6.2	Zur Evaluation eingesetzte Bewegungsaufnahmen als Trainingsdaten für die in Abschnitt 6.4 beschriebenen statistischen Modelle.	229
6.3	Berechnete Perplexität des N-Gramm-Transitionsmodells für verschiedene Ordnungen (Werte von n) und Glättungstechniken.	231
6.4	Durchgeführte Zahl an Iterationen für die Ausführung von Algorithmus 5 und die hierfür benötigte Laufzeit.	234
6.5	Generierte Konfigurationsposen-Sequenz für Szenario 1.	240
6.6	Mittelwert und Standardabweichung der Positionsveränderung von jeweils einem Fuß bzw. einer Hand zwischen aufeinanderfolgenden Konfigurationsposen.	241
A.1	Bewegungsklassen mit der jeweiligen Anzahl an Bewegungsaufnahmen für die Evaluation der Merkmalsauswahl in Abschnitt 3.5.1.	255

A.2 Kombinationen von (durch Semikola getrennten) Bewegungslabels zusammen mit der jeweiligen Anzahl an Bewegungsaufnahmen für die Evaluation der Multi-Label-Bewegungsklassifikation in Abschnitt 3.5.2.	256
---	-----

Algorithmenverzeichnis

1	N-Best Feature Subset Exploration	67
2	Berechnung von geglätteten Geschwindigkeiten der Körpersegmente	164
3	Berechnung von Distanzen zwischen Körpersegmenten und der Umgebung	165
4	Support-Posen-Segmentierung von Ganzkörperbewegungen . .	167
5	Planung von Posen-Sequenzen	224
6	Expandieren eines Pfads	225

Literaturverzeichnis

- Aggarwal, J. und Ryoo, M. (2011). „Human Activity Analysis: A Review“. In: *ACM Computing Surveys* 43.3, S. 1–43.
- Aksoy, E. E., Abramov, A., Dörr, J., Ning, K., Dellen, B. und Wörgötter, F. (2011). „Learning the Semantics of Object-Action Relations by Observation“. In: *The International Journal of Robotics Research* 30.10, S. 1229–1249.
- Aksoy, E. E., Tamosiunaite, M. und Wörgötter, F. (2015). „Model-Free Incremental Learning of the Semantics of Manipulation Actions“. In: *Robotics and Autonomous Systems* 71, S. 118–133.
- Aristidou, A. und Lasenby, J. (2013). „Real-Time Marker Prediction and CoR Estimation in Optical Motion Capture“. In: *The Visual Computer* 29.1, S. 7–26.
- Asfour, T., Azad, P., Gyarfas, F. und Dillmann, R. (2008). „Imitation Learning of Dual-Arm Manipulation Tasks in Humanoid Robots“. In: *International Journal of Humanoid Robotics* 5.2, S. 183–202.
- Asfour, T., Do, M., Welke, K., Bierbaum, A., Azad, P., Vahrenkamp, N., Gärtner, S., Ude, A. und Dillmann, R. (2011). „From Sensorimotor Primitives to Manipulation and Imitation Strategies in Humanoid Robots“. In: *Robotics Research*. Springer, S. 363–378.
- Asfour, T., Schill, J., Peters, H., Klas, C., Bücker, J., Sander, C., Schulz, S., Kargov, A., Werner, T. und Bartenbach, V. (2013). „ARMAR-4: A 63

- DOF Torque Controlled Humanoid Robot“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 390–396.
- Asfour, T., Borràs, J., Mandery, C., Kaiser, P., Aksoy, E. E. und Grotz, M. (2018). „On the Dualities Between Grasping and Whole-Body Locomanipulation Tasks“. In: *Robotics Research*. Springer, S. 305–322.
- Al-Asqhar, R. A., Komura, T. und Choi, M. G. (2013). „Relationship Descriptors for Interactive Motion Adaptation“. In: *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, S. 45–53.
- Atkeson, C., Babu, B., Banerjee, N., Berenson, D., Bove, C., Cui, X., DeDonato, M., Du, R., Feng, S., Franklin, P. et al. (2015). „No Falls, no Resets: Reliable Humanoid Behavior in the DARPA Robotics Challenge“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 623–630.
- Azad, P., Asfour, T. und Dillmann, R. (2007). „Toward an Unified Representation for Imitation of Human Motion on Humanoids“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 2558–2563.
- Babič, J., Petrič, T., Peternel, L. und Šarabon, N. (2014). „Effects of Supportive Hand Contact on Reactive Postural Control During Support Perturbations“. In: *Gait & Posture* 40.3, S. 441–446.
- Baca, A. (2012). „Methods for Recognition and Classification of Human Motion Patterns – A Prerequisite for Intelligent Devices Assisting in Sports Activities“. In: *7th Vienna International Conference on Mathematical Modelling (MATHMOD)*, S. 55–61.
- Barbič, J., Safonova, A., Pan, J.-Y., Faloutsos, C., Hodgins, J. K. und Pollard, N. S. (2004). „Segmenting Motion Capture Data into Distinct Behaviors“. In: *Graphics Interface (GI)*, S. 185–194.

- Barre, A. und Armand, S. (2014). „Biomechanical ToolKit: Open-Source Framework to Visualize and Process Biomechanical Data“. In: *Computer Methods and Programs in Biomedicine* 114.1, S. 80–87.
- Bernardin, K., Ogawara, K., Ikeuchi, K. und Dillmann, R. (2005). „A Sensor Fusion Approach for Recognizing Continuous Human Grasping Sequences Using Hidden Markov Models“. In: *IEEE Transactions on Robotics* 21.1, S. 47–57.
- Bierbaum, A., Rambow, M., Asfour, T. und Dillmann, R. (2009). „Grasp Affordances from Multi-Fingered Tactile Exploration Using Dynamic Potential Fields“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 168–174.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Secaucus, NJ, USA: Springer.
- Bitzer, S., Howard, M. und Vijayakumar, S. (2010). „Using Dimensionality Reduction to Exploit Constraints in Reinforcement Learning“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 3219–3225.
- Bitzer, S. und Vijayakumar, S. (2009). „Latent Spaces for Dynamic Movement Primitives“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 574–581.
- Blackman, S. und Popoli, R. (1999). *Design and Analysis of Modern Tracking Systems*. Norwood, MA, USA: Artech House.
- Bläsing, B., Maycock, J., Bockemühl, T., Ritter, H. und Schack, T. (2013). „Motor Synergies and Mental Representations of Grasping Movements“. In: *ICRA 2013 Workshop on Hand Synergies – How to Tame the Complexity of Grasping*.

- Bonnet, V., Daune, G., Joukov, V., Dumas, R., Fraise, P., Kulić, D., Seilles, A., Andary, S. und Venture, G. (2016). „A Constrained Extended Kalman Filter for Dynamically Consistent Inverse Kinematics and Inertial Parameters Identification“. In: *IEEE International Conference on Biomedical Robotics and Biomechanics (BioRob)*, S. 944–949.
- Borràs, J. und Asfour, T. (2015). „A Whole-Body Pose Taxonomy for Locomanipulation Tasks“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 1578–1585.
- Borràs, J., Mandery, C. und Asfour, T. (2017). „The KIT Taxonomy of Whole-Body Support Poses for the Analysis of Humanoid Robot Multi-Contact Motions (zur Veröffentlichung vorgesehen)“.
- Botvinick, M. und Toussaint, M. (2012). „Planning as Inference“. In: *Trends in Cognitive Sciences* 16.10, S. 485–488.
- Bouchard, D. und Badler, N. (2007). „Semantic Segmentation of Motion Capture Using Laban Movement Analysis“. In: *Intelligent Virtual Agents*, S. 37–44.
- Boulic, R., Capin, T., Huang, Z., Kalra, P., Lintermann, B., Magnenat-Thalmann, N., Moccozet, L., Molet, T., Pandzic, I., Saar, K., Schmitt, A., Shen, J. und Thalmann, D. (1995). „The HUMANOID Environment for Interactive Animation of Multiple Deformable Human Characters“. In: *Computer Graphics Forum*. Bd. 14. 3, S. 337–348.
- Boutin, L., Eon, A., Zegloul, S. und Lacouture, P. (2010). „An Auto-Adaptable Algorithm to Generate Human-Like Locomotion for Different Humanoid Robots Based on Motion Capture Data“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 1256–1261.

- Bouyarmane, K., Escande, A., Lamiroux, F. und Kheddar, A. (2009). „Potential Field Guide for Humanoid Multicontacts Acyclic Motion Planning“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 1165–1170.
- Bouyarmane, K. und Kheddar, A. (2010). „Static Multi-Contact Inverse Problem for Multiple Humanoid Robots and Manipulated Objects“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 8–13.
- Bouyarmane, K. und Kheddar, A. (2011a). „FEM-based Static Posture Planning for a Humanoid Robot on Deformable Contact Support“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 487–492.
- Bouyarmane, K. und Kheddar, A. (2011b). „Multi-Contact Stances Planning for Multiple Agents“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 5246–5253.
- Bouyarmane, K. und Kheddar, A. (2011c). „Using a Multi-Objective Controller to Synthesize Simulated Humanoid Robot Motion with Changing Contact Configurations“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4414–4419.
- Bouyarmane, K. und Kheddar, A. (2012). „Humanoid Robot Locomotion and Manipulation Step Planning“. In: *Advanced Robotics* 26.10, S. 1099–1126.
- Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D. und Blumberg, B. (2005). „Learning From and About Others: Towards Using Imitation to Bootstrap the Social Understanding of Others by Robots“. In: *Artificial Life* 11.1–2, S. 31–62.
- Breiman, L. (2001). „Random Forests“. In: *Machine Learning* 45.1, S. 5–32.

- Breiman, L., Friedman, J., Stone, C. J. und Olshen, R. A. (1984). *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth International Group.
- Bretl, T. und Lall, S. (2008). „Testing Static Equilibrium for Legged Robots“. In: *IEEE Transactions on Robotics* 24.4, S. 794–807.
- Brubaker, M. A., Sigal, L. und Fleet, D. J. (2009). „Estimating Contact Dynamics“. In: *IEEE International Conference on Computer Vision (ICCV)*, S. 2389–2396.
- Buchholz, B., Armstrong, T. J. und Goldstein, S. A. (1992). „Anthropometric Data for Describing the Kinematics of the Human Hand“. In: *Ergonomics* 35.3, S. 261–273.
- Bullock, I. M., Ma, R. R. und Dollar, A. M. (2013). „A Hand-Centric Classification of Human and Robot Dexterous Manipulation“. In: *IEEE Transactions on Haptics* 6.2, S. 129–144.
- Busso, C., Bulut, M., Lee, C.-C., Kazemzadeh, A., Mower, E., Kim, S., Chang, J. N., Lee, S. und Narayanan, S. S. (2008). „IEMOCAP: Interactive Emotional Dyadic Motion Capture Database“. In: *Language Resources and Evaluation* 42.4, S. 335–359.
- C3D (o.D.). *C3D: The 3D Biomechanics Data Standard*. URL: <http://www.c3d.org/> (besucht am 17.05.2017).
- Calinon, S., Guenter, F. und Billard, A. (2007). „On Learning, Representing, and Generalizing a Task in a Humanoid Robot“. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 37.2, S. 286–298.
- Čamernik, J., Potocanac, Z., Peternel, L. und Babič, J. (2016). „Holding a Handle for Balance during Continuous Postural Perturbations – Immediate and Transitional Effects on Whole Body Posture“. In: *Frontiers in Human Neuroscience* 10.

- Caron, S., Pham, Q.-C. und Nakamura, Y. (2017). „ZMP Support Areas for Multicontact Mobility Under Frictional Constraints“. In: *IEEE Transactions on Robotics* 33.1, S. 67–80.
- CGVUIInteractionDatabase (o.D.). *Edinburgh CGVU Interaction Database*. URL: <http://www.ipab.inf.ed.ac.uk/cgvu/InteractionDatabase/interactiondb.html> (besucht am 17.05.2017).
- Chen, N., Bayer, J., Urban, S. und Smagt, P. van der (2015). „Efficient Movement Representation by Embedding Dynamic Movement Primitives in Deep Autoencoders“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 434–440.
- Chen, N., Karl, M. und Smagt, P. van der (2016). „Dynamic Movement Primitives in Latent Space of Time-Dependent Variational Autoencoders“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 629–636.
- Chen, S. F. und Goodman, J. (1999). „An Empirical Study of Smoothing Techniques for Language Modeling“. In: *Computer Speech and Language* 13.4, S. 359–393.
- Chu, W.-T., Lee, Y.-L. und Yu, J.-Y. (2009). „Visual Language Model for Face Clustering in Consumer Photos“. In: *ACM International Conference on Multimedia*, S. 625–628.
- Ciocarlie, M. T. und Allen, P. K. (2009). „Hand Posture Subspaces for Dexterous Robotic Grasping“. In: *The International Journal of Robotics Research* 28.7, S. 851–867.
- Clever, D. und Mombaur, K. (2016). „An Inverse Optimal Control Approach for the Transfer of Human Walking Motions in Constrained Environment to Humanoid Robots“. In: *Robotics: Science and Systems (RSS)*.

- CMUMoCapDB (o.D.). *CMU Graphics Lab Motion Capture Database*. URL: <http://mocap.cs.cmu.edu/> (besucht am 17.05.2017).
- Cognetti, M., Mohammadi, P. und Oriolo, G. (2015). „Whole-Body Motion Planning for Humanoids Based on CoM Movement Primitives“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 1090–1095.
- Cognetti, M., Mohammadi, P., Oriolo, G. und Vendittelli, M. (2014). „Task-Oriented Whole-Body Planning for Humanoids Based on Hybrid Motion Generation“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4071–4076.
- Collette, C., Micaelli, A., Andriot, C. und Lemerle, P. (2007). „Dynamic Balance Control of Humanoids for Multiple Grasps and non Coplanar Frictional Contacts“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 81–88.
- Collette, C., Micaelli, A., Andriot, C. und Lemerle, P. (2008). „Robust Balance Optimization Control of Humanoid Robots with Multiple non Coplanar Grasps and Frictional Contacts“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 3187–3193.
- Colomé, A., Neumann, G., Peters, J. und Torras, C. (2014). „Dimensionality Reduction for Probabilistic Movement Primitives“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 794–800.
- Contini, R. (1972). „Body Segment Parameters, Part II“. In: *Artificial Limbs* 16.1, S. 1–19.
- Curran, W., Brys, T., Aha, D., Taylor, M. und Smart, W. D. (2016). „Dimensionality Reduced Reinforcement Learning for Assistive Robots“. In: *AAAI Fall Symposium on Artificial Intelligence for Human-Robot Interaction*, S. 25–31.

- Cutkosky, M. R. (1989). „On Grasp Choice, Grasp Models, and the Design of Hands for Manufacturing Tasks“. In: *IEEE Transactions on Robotics and Automation* 5.3, S. 269–279.
- Dalibard, S., Nakhaei, A., Lamiroux, F. und Laumond, J.-P. (2009). „Whole-Body Task Planning for a Humanoid Robot: A Way to Integrate Collision Avoidance“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 355–360.
- Dantam, N. und Stilman, M. (2013). „The Motion Grammar: Analysis of a Linguistic Method for Robot Control“. In: *IEEE Transactions on Robotics* 29.3, S. 704–718.
- Delp, S. L., Anderson, F. C., Arnold, A. S., Loan, P., Habib, A., John, C. T., Guendelman, E. und Thelen, D. G. (2007). „OpenSim: Open-Source Software to Create and Analyze Dynamic Simulations of Movement“. In: *IEEE Transactions on Biomedical Engineering* 54.11, S. 1940–1950.
- Dempster, W. T., Gabel, W. C. und Felts, W. J. L. (1959). „The Anthropometry of the Manual Work Space for the seated subject“. In: *American Journal of Physical Anthropology* 17.4, S. 289–317.
- Detry, R., Kraft, D., Kroemer, O., Bodenhausen, L., Peters, J., Krüger, N. und Piater, J. (2011). „Learning Grasp Affordance Densities“. In: *Paladyn, Journal of Behavioral Robotics* 2.1, S. 1–17.
- Django (o.D.). *Django Web Framework*. Django Software Foundation. URL: <https://www.djangoproject.com/> (besucht am 17.05.2017).
- Do, M., Asfour, T. und Dillmann, R. (2011). „Towards a Unifying Grasp Representation for Imitation Learning on Humanoid Robots“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 482–488.

- Do, M., Azad, P., Asfour, T. und Dillmann, R. (2008). „Imitation of Human Motion on a Humanoid Robot Using Non-Linear Optimization“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 545–552.
- Du, Y., Wang, W. und Wang, L. (2015). „Hierarchical Recurrent Neural Network for Skeleton Based Action Recognition“. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, S. 1110–1118.
- Elliott, R. J., Aggoun, L. und Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control*. New York, USA: Springer.
- Englsberger, J., Ott, C. und Albu-Schäffer, A. (2013). „Three-Dimensional Bipedal Walking Control Using Divergent Component of Motion“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 2600–2607.
- Erez, T. und Todorov, E. (2012). „Trajectory Optimization for Domains with Contacts Using Inverse Dynamics“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4914–4919.
- Ernesti, J., Righetti, L., Do, M., Asfour, T. und Schaal, S. (2012). „Encoding of Periodic and their Transient Motions by a Single Dynamic Movement Primitive“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 57–64.
- Escande, A., Kheddar, A. und Miossec, S. (2013). „Planning Contact Points for Humanoid Robots“. In: *Robotics and Autonomous Systems* 61.5, S. 428–442.
- Escande, A., Kheddar, A., Miossec, S. und Garsault, S. (2009). „Planning Support Contact-Points for Acyclic Motions and Experiments on HRP-2“. In: *International Symposium on Experimental Robotics (ISER)*, S. 293–302.

- Falco, P., Saveriano, M., Hasany, E. G., Kirk, N. H. und Lee, D. (2017). „A Human Action Descriptor Based on Motion Coordination“. In: *IEEE Robotics and Automation Letters* 2.2, S. 811–818.
- Fallon, M., Kuindersma, S., Karumanchi, S., Antone, M., Schneider, T., Dai, H., D’Arpino, C. P., Deits, R., DiCicco, M., Fourie, D. et al. (2015). „An Architecture for Online Affordance-Based Perception and Whole-Body Planning“. In: *Journal of Field Robotics* 32.2, S. 229–254.
- Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R. und Lin, C.-J. (2008). „LIBLINEAR: A Library for Large Linear Classification“. In: *Journal of Machine Learning Research* 9, S. 1871–1874.
- Farnioli, E., Gabiccini, M. und Bicchi, A. (2015). „Optimal Contact Force Distribution for Compliant Humanoid Robots in Whole-Body Locomanipulation Tasks“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 5675–5681.
- Feix, T., Pawlik, R., Schmiedmayer, H.-B., Romero, J. und Kragic, D. (2009). „A Comprehensive Grasp Taxonomy“. In: *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*.
- Feix, T., Romero, J., Schmiedmayer, H.-B., Dollar, A. M. und Kragic, D. (2016). „The GRASP Taxonomy of Human Grasp Types“. In: *IEEE Transactions on Human-Machine Systems* 46.1, S. 66–77.
- Field, M., Pan, Z., Stirling, D. und Naghdy, F. (2011). „Human Motion Capture Sensors and Analysis in Robotics“. In: *Industrial Robot: An International Journal* 38.2, S. 163–171.
- Field, M., Stirling, D., Naghdy, F. und Pan, Z. (2009). „Motion Capture in Robotics Review“. In: *IEEE International Conference on Control and Automation (ICCA)*, S. 1697–1702.

- Fod, A., Matarić, M. J. und Jenkins, O. C. (2002). „Automated Derivation of Primitives for Movement Classification“. In: *Autonomous Robots* 12.1, S. 39–54.
- Forner-Cordero, A., Levin, O., Li, Y. und Swinnen, S. P. (2005). „Principal Component Analysis of Complex Multijoint Coordinative Movements“. In: *Biological Cybernetics* 93.1, S. 63–78.
- Freeman, C. (2014). „Feature Selection and Hierarchical Classifier Design with Applications to Human Motion Recognition“. Diss. University of Waterloo.
- Freeman, C., Kulić, D. und Basir, O. (2015). „An Evaluation of Classifier-Specific Filter Measure Performance for Feature Selection“. In: *Pattern Recognition* 48.5, S. 1812–1826.
- García, N., Rosell, J. und Suárez, R. (2015). „Motion Planning Using First-Order Synergies“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 2058–2063.
- Gärtner, S., Do, M., Simonidis, C., Asfour, T., Seemann, W. und Dillmann, R. (2010). „Generation of Human-Like Motion for Humanoid Robots Based on Marker-Based Motion Capture Data“. In: *41th International Symposium on Robotics (ISR)*, S. 1–8.
- Gavrila, D. M. und Davis, L. S. (1996). „3-D Model-Based Tracking of Humans in Action: A Multi-View Approach“. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, S. 73–80.
- Ghahramani, Z. und Jordan, M. I. (1997). „Factorial Hidden Markov Models“. In: *Machine Learning* 29.2–3, S. 245–273.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin.

- Good, I. J. (1953). „The Population Frequencies of Species and the Estimation of Population Parameters“. In: *Biometrika* 40.3–4, S. 237–264.
- Grochow, K., Martin, S. L., Hertzmann, A. und Popović, Z. (2004). „Style-Based Inverse Kinematics“. In: *ACM Transactions on Graphics* 23.3, S. 522–531.
- Gross, R. und Shi, J. (2001). *The CMU Motion of Body (MoBo) Database*. Techn. Ber. CMU-RI-TR-01-18. Pittsburgh, PA: Carnegie Mellon University.
- Guerra-Filho, G. und Biswas, A. (2012). „The Human Motion Database: A Cognitive and Parametric Sampling of Human Motion“. In: *Image and Vision Computing* 30.3, S. 251–261.
- Gut, A. (2009). *The Multivariate Normal Distribution*. 2. ed. New York, USA: Springer.
- Guyon, I. und Elisseeff, A. (2003). „An Introduction to Variable and Feature Selection“. In: *Journal of Machine Learning Research* 3, S. 1157–1182.
- Hauser, K., Bretl, T., Harada, K. und Latombe, J.-C. (2008). „Using Motion Primitives in Probabilistic Sample-Based Planning for Humanoid Robots“. In: *Algorithmic Foundation of Robotics VII*. Springer, S. 507–522.
- Hauser, K., Bretl, T. und Latombe, J.-C. (2005). „Non-Gaited Humanoid Locomotion Planning“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 7–12.
- Herr, H. und Popovic, M. (2008). „Angular Momentum in Human Walking“. In: *Journal of Experimental Biology* 211.4, S. 467–481.
- Herzog, D., Ude, A. und Krüger, V. (2008). „Motion Imitation and Recognition Using Parametric Hidden Markov Models“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 339–346.

- Hirukawa, H., Hattori, S., Harada, K., Kajita, S., Kaneko, K., Kanehiro, F., Fujiwara, K. und Morisawa, M. (2006). „A Universal Stability Criterion of the Foot Contact of Legged Robots – Adios ZMP“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 1976–1983.
- Hofmann, A., Popovic, M. und Herr, H. (2009). „Exploiting Angular Momentum to Enhance Bipedal Center-of-Mass Control“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 4423–4429.
- Huang, H. J. und Ahmed, A. A. (2011). „Tradeoff Between Stability and Maneuverability During Whole-Body Movements“. In: *PLOS ONE* 6.7, S. 1–10.
- Huang, X., Acero, A. und Hon, H.-W. (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Upper Saddle River, NJ, USA: Prentice Hall.
- HumanIdentification (o.D.). *Human Identification at a Distance*. URL: <http://www.cc.gatech.edu/cpl/projects/hid/> (besucht am 17.05.2017).
- Hwang, B.-W., Kim, S. und Lee, S.-W. (2006). „A Full-Body Gesture Database for Automatic Gesture Recognition“. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, S. 243–248.
- Ice (o.D.). *ZeroC – The Internet Communications Engine (Ice)*. URL: <http://www.zeroc.com/ice.html> (besucht am 17.05.2017).
- ICS (o.D.). *ICS Action Database*. URL: <http://www.ics.t.u-tokyo.ac.jp/action/> (besucht am 17.05.2017).
- Ijspeert, A. J., Nakanishi, J., Hoffmann, H., Pastor, P. und Schaal, S. (2013). „Dynamical Movement Primitives: Learning Attractor Models for Motor Behaviors“. In: *Neural Computation* 25.2, S. 328–373.

- Ijspeert, A. J., Nakanishi, J. und Schaal, S. (2002). „Movement Imitation with Nonlinear Dynamical Systems in Humanoid Robots“. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Bd. 2, S. 1398–1403.
- Inamura, T., Toshima, I., Tanie, H. und Nakamura, Y. (2004). „Embodied Symbol Emergence Based on Mimesis Theory“. In: *The International Journal of Robotics Research* 23.4–5, S. 363–377.
- Ionescu, C., Papava, D., Olaru, V. und Sminchisescu, C. (2014). „Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36.7, S. 1325–1339.
- Janus, B. und Nakamura, Y. (2005). „Unsupervised Probabilistic Segmentation of Motion Data for Mimesis Modeling“. In: *International Conference on Advanced Robotics (ICAR)*, S. 411–417.
- Jelinek, F., Mercer, R. L., Bahl, L. R. und Baker, J. K. (1977). „Perplexity – A Measure of the Difficulty of Speech Recognition Tasks“. In: *The Journal of the Acoustical Society of America* 62.S1, S63–S63.
- Jenkins, O. C. und Mataric, M. J. (2002). „Deriving Action and Behavior Primitives from Human Motion Data“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Bd. 3, S. 2551–2556.
- Jenkins, O. C., Serrano, G. G. und Loper, M. M. (2007). „Interactive Human Pose and Action Recognition Using Dynamical Motion Primitives“. In: *International Journal of Humanoid Robotics* 4.2, S. 365–385.
- Johannsen, L., Wing, A. M. und Hatzitaki, V. (2012). „Contrasting Effects of Finger and Shoulder Interpersonal Light Touch on Standing Balance“. In: *Journal of Neurophysiology* 107.1, S. 216–225.

- Johnson, S. G. (o.D.). *The NLopt Nonlinear-Optimization Package*. URL: <http://ab-initio.mit.edu/nlopt/> (besucht am 17.05.2017).
- Jonker, R. und Volgenant, A. (1987). „A Shortest Augmenting Path Algorithm for Dense and Sparse Linear Assignment Problems“. In: *Computing* 38.4, S. 325–340.
- Ju, S. X., Black, M. J. und Yacoob, Y. (1996). „Cardboard People: A Parameterized Model of Articulated Image Motion“. In: *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, S. 38–44.
- Kaiser, P., Aksoy, E. E., Grotz, M. und Asfour, T. (2016). „Towards a Hierarchy of Loco-Manipulation Affordances“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 2839–2846.
- Kaiser, P., Grotz, M., Aksoy, E. E., Do, M., Vahrenkamp, N. und Asfour, T. (2015a). „Validation of Whole-Body Loco-Manipulation Affordances for Pushability and Liftability“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 920–927.
- Kaiser, P., Vahrenkamp, N., Schültje, F., Borràs, J. und Asfour, T. (2015b). „Extraction of Whole-Body Affordances for Loco-Manipulation Tasks“. In: *International Journal of Humanoid Robotics* 12.3.
- Kajita, S., Kanehiro, F., Kaneko, K., Fujiwara, K., Harada, K., Yokoi, K. und Hirukawa, H. (2003). „Biped Walking Pattern Generation by Using Preview Control of Zero-Moment Point“. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Bd. 2, S. 1620–1626.
- Kajita, S., Morisawa, M., Miura, K., Nakaoka, S., Harada, K., Kaneko, K., Kanehiro, F. und Yokoi, K. (2010). „Biped Walking Stabilization Based on Linear Inverted Pendulum Tracking“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4489–4496.

- Kamakura, N. (1989). *Te no katachi, te no ugoki*. Auf Japanisch verfasst. Tokio, Japan: Ishiyaku Shuppan Kabushiki Kaisha.
- Kang, C. und Lee, S.-H. (2014). „Environment-Adaptive Contact Poses for Virtual Characters“. In: *Computer Graphics Forum*. Bd. 33. 7, S. 1–10.
- Kapadia, M., Chiang, I.-k., Thomas, T., Badler, N. I. und Kider Jr, J. T. (2013). „Efficient Motion Retrieval in Large Motion Databases“. In: *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D)*, S. 19–28.
- Kapadia, M., Xianghao, X., Nitti, M., Kallmann, M., Coros, S., Sumner, R. W. und Gross, M. (2016). „PRECISION: Precomputing Environment Semantics for Contact-Rich Character Animation“. In: *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games (I3D)*, S. 29–37.
- Karčnik, T. (2003). „Using Motion Analysis Data for Foot-Floor Contact Detection“. In: *Medical and Biological Engineering and Computing* 41.5, S. 509–512.
- Khatib, O. und Chung, S.-Y. (2014). „SupraPeds: Humanoid Contact-Supported Locomotion for 3D Unstructured Environments“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 2319–2325.
- Khatib, O., Sentis, L. und Park, J.-H. (2008). „A Unified Framework for Whole-Body Humanoid Robot Control with Multiple Constraints and Contacts“. In: *European Robotics Symposium 2008*, S. 303–312.
- Kim, C., Kim, D. und Oh, Y. (2006). „Adaptation of Human Motion Capture Data to Humanoid Robots for Motion Imitation Using Optimization“. In: *Integrated Computer-Aided Engineering* 13.4, S. 377–389.

- Kohavi, R. (1995). „A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection“. In: *International Joint Conference on Artificial Intelligence (IJCAI)*, S. 1137–1145.
- Kohavi, R. und John, G. H. (1997). „Wrappers for Feature Subset Selection“. In: *Artificial Intelligence* 97.1–2, S. 273–324.
- Kohlbrecher, S., Romay, A., Stumpf, A., Gupta, A., Von Stryk, O., Bacim, F., Bowman, D. A., Goins, A., Balasubramanian, R. und Conner, D. C. (2015). „Human-Robot Teaming for Rescue Missions: Team ViGIR’s Approach to the 2013 DARPA Robotics Challenge Trials“. In: *Journal of Field Robotics* 32.3, S. 352–377.
- Krüger, V., Herzog, D. L., Baby, S., Ude, A. und Kragic, D. (2010). „Learning Actions from Observations“. In: *IEEE Robotics & Automation Magazine* 17.2, S. 30–43.
- Kuehne, H., Jhuang, H., Garrote, E., Poggio, T. und Serre, T. (2011). „HMDB: A Large Video Database for Human Motion Recognition“. In: *IEEE International Conference on Computer Vision (ICCV)*, S. 2556–2563.
- Kuffner, J., Nishiwaki, K., Kagami, S., Inaba, M. und Inoue, H. (2001). „Motion Planning for Humanoid Robots Under Obstacle and Dynamic Balance Constraints“. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Bd. 1, S. 692–698.
- Kuhn, H. W. (1955). „The Hungarian Method for the Assignment Problem“. In: *Naval Research Logistics Quarterly* 2.1–2, S. 83–97.
- Kulić, D., Ott, C., Lee, D., Ishikawa, J. und Nakamura, Y. (2012). „Incremental Learning of Full Body Motion Primitives and Their Sequencing Through Human Motion Observation“. In: *The International Journal of Robotics Research* 31.3, S. 330–345.

- Kulić, D., Takano, W. und Nakamura, Y. (2007a). „Incremental On-Line Hierarchical Clustering of Whole Body Motion Patterns“. In: *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, S. 1016–1021.
- Kulić, D., Takano, W. und Nakamura, Y. (2007b). „Representability of Human Motions by Factorial Hidden Markov Models“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 2388–2393.
- Kulić, D., Takano, W. und Nakamura, Y. (2008). „Incremental Learning, Clustering and Hierarchy Formation of Whole Body Motion Patterns Using Adaptive Hidden Markov Chains“. In: *The International Journal of Robotics Research* 27.7, S. 761–784.
- Kulić, D., Takano, W. und Nakamura, Y. (2009). „Online Segmentation and Clustering from Continuous Observation of Whole Body Motions“. In: *IEEE Transactions on Robotics* 25.5, S. 1158–1166.
- Kulvicius, T., Ning, K., Tamosiunaite, M. und Worgötter, F. (2012). „Joining Movement Sequences: Modified Dynamic Movement Primitives for Robotics Applications Exemplified on Handwriting“. In: *IEEE Transactions on Robotics* 28.1, S. 145–157.
- Laban, R. von und Ullmann, L. (2011). *The Mastery of Movement*. 4. ed. Hampshire, UK: Dance Books.
- Lan, R. und Sun, H. (2015). „Automated Human Motion Segmentation via Motion Regularities“. In: *The Visual Computer* 31.1, S. 35–53.
- Langenstein, A. (2016). „Integration von Motion Recognition in die KIT Whole-Body Human Motion Database“. Bachelorarbeit. Karlsruhe, Germany: Karlsruhe Institute of Technology (KIT).

- Lengagne, S., Terlemez, Ö., Laturnus, S., Asfour, T. und Dillmann, R. (2012). „Retrieving Contact Points Without Environment Knowledge“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 841–846.
- Lengagne, S., Vaillant, J., Yoshida, E. und Kheddar, A. (2013). „Generation of Whole-Body Optimal Dynamic Multi-Contact Motions“. In: *The International Journal of Robotics Research* 32.9–10, S. 1104–1119.
- León, B., Morales, A. und Sancho-Bru, J. (2014). „The Model of the Human Hand“. In: *From Robot to Human Grasping Simulation*. Springer, S. 123–173.
- Leva, P. de (1996). „Adjustments to Zatsiorsky-Seluyanov’s Segment Inertia Parameters“. In: *Journal of Biomechanics* 29.9, S. 1223–1230.
- Lin, J. F.-S., Joukov, V. und Kulic, D. (2014). „Full-Body Multi-Primitive Segmentation Using Classifiers“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 874–880.
- Lin, J. F.-S., Karg, M. und Kulić, D. (2016). „Movement Primitive Segmentation for Human Motion Modeling: A Framework for Analysis“. In: *IEEE Transactions on Human-Machine Systems* 46.3, S. 325–339.
- Liu, L., Yin, K., Panne, M. van de, Shao, T. und Xu, W. (2010). „Sampling-Based Contact-Rich Motion Control“. In: *ACM Transactions on Graphics* 29.4.
- Lösch, M., Schmidt-Rohr, S. R. und Dillmann, R. (2008). „Making Feature Selection for Human Motion Recognition More Interactive Through the Use of Taxonomies“. In: *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, S. 539–544.

- Lu, T.-W. und O'Connor, J. (1999). „Bone Position Estimation from Skin Marker Co-Ordinates Using Global Optimisation with Joint Constraints“. In: *Journal of Biomechanics* 32.2, S. 129–134.
- Luck, K. S., Neumann, G., Berger, E., Peters, J. und Amor, H. B. (2014). „Latent Space Policy Search for Robotics“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 1434–1440.
- Ma, Y., Paterson, H. M. und Pollick, F. E. (2006). „A Motion Capture Library for the Study of Identity, Gender, and Emotion Perception from Biological Motion“. In: *Behavior Research Methods* 38.1, S. 134–141.
- Mandery, C., Borràs, J., Jöchner, M. und Asfour, T. (2015a). „Analyzing Whole-Body Pose Transitions in Multi-Contact Motions“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*. Seoul, Korea, S. 1020–1027.
- Mandery, C., Borràs, J., Jöchner, M. und Asfour, T. (2016a). „Using Language Models to Generate Whole-Body Multi-Contact Motions“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, Korea, S. 5411–5418.
- Mandery, C., Plappert, M., Borràs, J. und Asfour, T. (2016b). „Dimensionality Reduction for Whole-Body Human Motion Recognition“. In: *19th International Conference on Information Fusion (FUSION)*. Heidelberg, Germany, S. 355–362.
- Mandery, C., Terlemez, Ö., Do, M., Vahrenkamp, N. und Asfour, T. (2015b). „The KIT Whole-Body Human Motion Database“. In: *International Conference on Advanced Robotics (ICAR)*. Istanbul, Turkey, S. 329–336.
- Mandery, C., Terlemez, Ö., Do, M., Vahrenkamp, N. und Asfour, T. (2016c). „Unifying Representations and Large-Scale Whole-Body Motion Databa-

- ses for Studying Human Motion“. In: *IEEE Transactions on Robotics* 32.4, S. 796–809.
- Matsubara, T., Hyon, S.-H. und Morimoto, J. (2011). „Learning Parametric Dynamic Movement Primitives from Multiple Demonstrations“. In: *Neural Networks* 24.5, S. 493–500.
- Matsui, D., Minato, T., MacDorman, K. F. und Ishiguro, H. (2005). „Generating Natural Motion in an Android by Mapping Human Motion“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 3301–3308.
- Maycock, J., Rohlig, T., Schroder, M., Botsch, M. und Ritter, H. (2015). „Fully Automatic Optical Motion Tracking Using an Inverse Kinematics Approach“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 461–466.
- Meier, F., Theodorou, E., Stulp, F. und Schaal, S. (2011). „Movement Segmentation Using a Primitive Library“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 3407–3412.
- Meyer, J., Kuderer, M., Müller, J. und Burgard, W. (2014). „Online Marker Labeling for Fully Automatic Skeleton Tracking in Optical Motion Capture“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 5652–5657.
- Miller, G. A. (1995). „WordNet: A Lexical Database for English“. In: *Communications of the ACM* 38.11, S. 39–41.
- Minato, T. und Ishiguro, H. (2007). „Generating Natural Posture in an Android by Mapping Human Posture in Three-Dimensional Position Space“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 609–616.

- Miossec, S., Yokoi, K. und Kheddar, A. (2006). „Development of a Software for Motion Optimization of Robots – Application to the Kick Motion of the HRP-2 Robot“. In: *IEEE International Conference on Robotics and Biomimetics*, S. 299–304.
- MoCapClub (o.D.). *Motion Capture (MoCap) Club*. URL: <http://www.mocapclub.com/> (besucht am 17.05.2017).
- MoCapData (o.D.). *Mocapdata.com*. URL: <http://www.mocapdata.com/> (besucht am 17.05.2017).
- Moeslund, T. B. und Granum, E. (2001). „A Survey of Computer Vision-Based Human Motion Capture“. In: *Computer Vision and Image Understanding* 81.3, S. 231–268.
- Moeslund, T. B., Hilton, A. und Krüger, V. (2006). „A Survey of Advances in Vision-Based Human Motion Capture and Analysis“. In: *Computer Vision and Image Understanding* 104.2, S. 90–126.
- Monzani, J.-S., Baerlocher, P., Boulic, R. und Thalmann, D. (2000). „Using an Intermediate Skeleton and Inverse Kinematics for Motion Retargeting“. In: *Computer Graphics Forum*. Bd. 19. 3, S. 11–19.
- Mordatch, I., Todorov, E. und Popović, Z. (2012). „Discovery of Complex Behaviors Through Contact-Invariant Optimization“. In: *ACM Transactions on Graphics* 31.4.
- Mörtl, A., Lawitzky, M., Kucukyilmaz, A., Sezgin, M., Basdogan, C. und Hirche, S. (2012). „The Role of Roles: Physical Cooperation Between Humans and Robots“. In: *The International Journal of Robotics Research* 31.13, S. 1656–1674.
- Müller, M., Röder, T., Clausen, M., Eberhardt, B., Krüger, B. und Weber, A. (2007). *Documentation Mocap Database HDM05*. Techn. Ber. CG-2007-2. Universität Bonn.

- Müller, M. und Röder, T. (2006). „Motion Templates for Automatic Classification and Retrieval of Motion Capture Data“. In: *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, S. 137–146.
- Myers, K. A., Wang, M., Marks, R. M. und Harris, G. F. (2004). „Validation of a Multisegment Foot and Ankle Kinematic Model for Pediatric Gait“. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 12.1, S. 122–130.
- Nakanishi, J., Morimoto, J., Endo, G., Cheng, G., Schaal, S. und Kawato, M. (2004). „Learning from Demonstration and Adaptation of Biped Locomotion“. In: *Robotics and Autonomous Systems* 47.2, S. 79–91.
- Naksuk, N., Lee, C. G. und Rietdyk, S. (2005). „Whole-Body Human-to-Humanoid Motion Transfer“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 104–109.
- Nao (o.D.). *The NAO Robot*. URL: <https://www.aldebaran.com/en/cool-robots/nao/> (besucht am 17.05.2017).
- Napier, J. R. (1956). „The Prehensile Movements of the Human Hand“. In: *Journal of Bone and Joint Surgery (British Volume)* 38.4, S. 902–913.
- Nemec, B. und Ude, A. (2012). „Action Sequencing Using Dynamic Movement Primitives“. In: *Robotica* 30.5, S. 837–846.
- Nikolić, M., Borovac, B. und Raković, M. (2017). „Dynamic Balance Preservation and Prevention of Sliding for Humanoid Robots in the Presence of Multiple Spatial Contacts“. In: *Multibody System Dynamics*, S. 1–22.
- Nori, F., Peters, J., Padois, V., Babic, J., Mistry, M. und Ivaldi, S. (2014). „Whole-Body Motion in Humans and Humanoids“. In: *Workshop on New Research Frontiers for Intelligent Autonomous Systems (NRF-IAS)*, S. 81–92.

- Novela (o.D.). *LAAS-CNRS Novela Motion Capture Database*. URL: <http://projects.laas.fr/gepetto/novela/noveladb/> (besucht am 17.05.2017).
- NUSMoCap (o.D.). *NUS Mocap*. URL: <http://animation.comp.nus.edu.sg/nusmocap.html> (besucht am 17.05.2017).
- Ogata, T., Sugano, S. und Tani, J. (2005). „Open-End Human-Robot Interaction from the Dynamical Systems Perspective: Mutual Adaptation and Incremental Learning“. In: *Advanced Robotics* 19.6, S. 651–670.
- Padois, V., Ivaldi, S., Babič, J., Mistry, M., Peters, J. und Nori, F. (2017). „Whole-Body Multi-Contact Motion in Humans and Humanoids: Advances of the CoDyCo European Project“. In: *Robotics and Autonomous Systems* 90, S. 97–117.
- Pastra, K. und Aloimonos, Y. (2012). „The Minimalist Grammar of Action“. In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 367.1585, S. 103–117.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. und Duchesnay, E. (2011). „Scikit-learn: Machine Learning in Python“. In: *Journal of Machine Learning Research* 12, S. 2825–2830.
- Petrick, R. P. und Bacchus, F. (2002). „A Knowledge-Based Approach to Planning with Incomplete Information and Sensing“. In: *International Conference on Artificial Intelligence Planning and Scheduling (AIPS)*, S. 212–222.
- Pham, T. H., Bufort, A., Caron, S. und Kheddar, A. (2016). „Whole-Body Contact Force Sensing from Motion Capture“. In: *IEEE/SICE International Symposium on System Integration (SII)*, S. 58–63.

- Philippssen, R., Nejati, N. und Sentis, L. (2009). „Bridging the Gap Between Semantic Planning and Continuous Control for Mobile Manipulation Using a Graph-Based World Representation“. In: *International Workshop on Hybrid Control of Autonomous Systems (HYCAS)*, S. 77–81.
- Plappert, M. (2015). „Classification of Human Whole-Body Motion Using Hidden Markov Models“. Bachelorarbeit. Karlsruhe, Germany: Karlsruhe Institute of Technology (KIT).
- Plappert, M., Mandery, C. und Asfour, T. (2016). „The KIT Motion-Language Dataset“. In: *Big Data* 4.4, S. 236–252.
- Plappert, M., Mandery, C. und Asfour, T. (2017). „Learning a Bidirectional Mapping Between Human Whole-Body Motion and Natural Language Using Deep Recurrent Neural Networks“. In: *arXiv preprint arXiv:1705.06400*.
- Popovic, M. und Englehart, A. (2004). „Angular Momentum Primitives for Human Walking: Biomechanics and Control“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Bd. 2, S. 1685–1691.
- Popović, M., Kraft, D., Bodenhausen, L., Bašeski, E., Pugeault, N., Kragic, D., Asfour, T. und Krüger, N. (2010). „A Strategy for Grasping Unknown Objects Based on Co-Planarity and Colour Information“. In: *Robotics and Autonomous Systems* 58.5, S. 551–565.
- Poppe, R. (2010). „A Survey on Vision-Based Human Action Recognition“. In: *Image and Vision Computing* 28.6, S. 976–990.
- Posa, M., Cantu, C. und Tedrake, R. (2014). „A Direct Method for Trajectory Optimization of Rigid Bodies Through Contact“. In: *The International Journal of Robotics Research* 33.1, S. 69–81.

- PostgreSQL (o.D.). *PostgreSQL: The World's Most Advanced Open Source Database*. PostgreSQL Global Development Group. URL: <https://www.postgresql.org/> (besucht am 17.05.2017).
- Pratt, J., Carff, J., Drakunov, S. und Goswami, A. (2006). „Capture Point: A Step Toward Humanoid Push Recovery“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 200–207.
- Pryor, W., Lin, Y.-C. und Berenson, D. (2016). „Integrated Affordance Detection and Humanoid Locomotion Planning“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 125–132.
- Puthenveetil, S. C., Daphalapurkar, C. P., Zhu, W., Leu, M. C., Liu, X. F., Chang, A. M., Gilpin-Mcminn, J. K., Wu, P. H. und Snodgrass, S. D. (2013). „Comparison of Marker-Based and Marker-Less Systems for Low-Cost Human Motion Capture“. In: *ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2013)*.
- Python (o.D.). *Python Programming Language*. Python Software Foundation. URL: <https://www.python.org/> (besucht am 17.05.2017).
- Rabiner, L. R. (1989). „A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition“. In: *Proceedings of the IEEE 77.2*, S. 257–286.
- Radford, N. A., Strawser, P., Hambuchen, K., Mehling, J. S., Verdeyen, W. K., Donnan, A. S., Holley, J., Sanchez, J., Nguyen, V., Bridgwater, L. et al. (2015). „Valkyrie: NASA's First Bipedal Humanoid Robot“. In: *Journal of Field Robotics* 32.3, S. 397–419.
- Read, J., Pfahringer, B., Holmes, G. und Frank, E. (2011). „Classifier Chains for Multi-Label Classification“. In: *Machine Learning* 85.3, S. 333–359.

- Ren, L., Shakhnarovich, G., Hodgins, J. K., Pfister, H. und Viola, P. (2005). „Learning Silhouette Features for Control of Human Motion“. In: *ACM Transactions on Graphics* 24.4, S. 1303–1331.
- Riaz, Q., Krüger, B. und Weber, A. (2016). „Relational Databases for Motion Data“. In: *International Journal of Innovative Computing and Applications* 7.3, S. 119–134.
- Richter, S. und Westphal, M. (2010). „The LAMA Planner: Guiding Cost-Based Anytime Planning with Landmarks“. In: *Journal of Artificial Intelligence Research* 39.1, S. 127–177.
- Rijsbergen, C. J. V. (1979). *Information Retrieval*. 2. ed. Newton, MA, USA: Butterworth-Heinemann.
- Rosenfeld, R. (2000). „Two Decades of Statistical Language Modeling: Where Do We Go from Here?“ In: *Proceedings of the IEEE* 88.8, S. 1270–1278.
- Rowan, T. H. (1990). „Functional Stability Analysis of Numerical Algorithms“. Diss. Department of Computer Sciences, University of Texas at Austin.
- Rumbaugh, J., Jacobson, I. und Booch, G. (2004). *The Unified Modeling Language Reference Manual*. 2. ed. Reading, MA, USA: Addison-Wesley.
- Saab, L., Mansard, N., Keith, F., Fourquet, J.-Y. und Soueres, P. (2011a). „Generation of Dynamic Motion for Anthropomorphic Systems Under Prioritized Equality and Inequality Constraints“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 1091–1096.
- Saab, L., Ramos, O., Mansard, N., Souères, P. und Fourquet, J.-Y. (2011b). „Generic Dynamic Motion Generation with Multiple Unilateral Constraints“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4127–4133.

- Saab, L., Ramos, O. E., Keith, F., Mansard, N., Soueres, P. und Fourquet, J.-Y. (2013). „Dynamic Whole-Body Motion Generation Under Rigid Contacts and Other Unilateral Constraints“. In: *IEEE Transactions on Robotics* 29.2, S. 346–362.
- Safonova, A., Hodgins, J. K. und Pollard, N. S. (2004). „Synthesizing Physically Realistic Human Motion in Low-Dimensional, Behavior-Specific Spaces“. In: *ACM Transactions on Graphics* 23.3, S. 514–521.
- Safonova, A., Pollard, N. und Hodgins, J. K. (2003). „Optimizing Human Motion for the Control of a Humanoid Robot“. In: *International Symposium on Adaptive Motion of Animals and Machines* 78.
- Salini, J., Padois, V. und Bidaud, P. (2011a). „Synthesis of Complex Humanoid Whole-Body Behavior: A Focus on Sequencing and Tasks Transitions“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 1283–1290.
- Salini, J., Padois, V., Ibanez, A., Bidaud, P. und Buendia, A. (2011b). „A Goal Driven Perspective to Generate Humanoid Motion Synthesis“. In: *International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR)*, S. 889–897.
- Sandilands, P., Choi, M. G. und Komura, T. (2013). „Interaction Capture Using Magnetic Sensors“. In: *Computer Animation and Virtual Worlds* 24.6, S. 527–538.
- Santello, M., Flanders, M. und Soechting, J. F. (1998). „Postural Hand Synergies for Tool Use“. In: *Journal of Neuroscience* 18.23, S. 10105–10115.
- Santos, L., Khoshhal, K. und Dias, J. (2015). „Trajectory-Based Human Action Segmentation“. In: *Pattern Recognition* 48.2, S. 568–579.

- Sarraf, T. A., Marigold, D. S. und Robinovitch, S. N. (2014). „Maintaining Standing Balance by Handrail Grasping“. In: *Gait & Posture* 39.1, S. 258–264.
- Schubert, T., Gkogkidis, A., Ball, T. und Burgard, W. (2015). „Automatic Initialization for Skeleton Tracking in Optical Motion Capture“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 734–739.
- Schulman, J., Duan, Y., Ho, J., Lee, A., Awwal, I., Bradlow, H., Pan, J., Patil, S., Goldberg, K. und Abbeel, P. (2014). „Motion Planning with Sequential Convex Optimization and Convex Collision Checking“. In: *The International Journal of Robotics Research* 33.9, S. 1251–1270.
- Sechidis, K., Tsoumakas, G. und Vlahavas, I. (2011). „On the Stratification of Multi-Label Data“. In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, S. 145–158.
- Sentis, L., Park, J. und Khatib, O. (2010). „Compliant Control of Multi-contact and Center-of-Mass Behaviors in Humanoid Robots“. In: *IEEE Transactions on Robotics* 26.3, S. 483–501.
- Sentis, L. und Slovic, M. (2011). „Motion Planning of Extreme Locomotion Maneuvers Using Multi-Contact Dynamics and Numerical Integration“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 760–767.
- Settimi, A., Caporale, D., Kryczka, P., Ferrati, M. und Pallottino, L. (2016). „Motion Primitive Based Random Planning for Loco-Manipulation Tasks“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 1059–1066.

- Shin, H. J., Lee, J., Shin, S. Y. und Gleicher, M. (2001). „Computer Puppetry: An Importance-Based Approach“. In: *ACM Transactions on Graphics* 20.2, S. 67–94.
- Shon, A., Grochow, K. und Rao, R. (2005). „Robotic Imitation from Human Motion Capture Using Gaussian Processes“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 129–134.
- Sidenbladh, H., Black, M. J. und Fleet, D. J. (2000). „Stochastic Tracking of 3D Human Figures Using 2D Image Motion“. In: *European Conference on Computer Vision (ECCV)*, S. 702–718.
- Sigal, L., Balan, A. O. und Black, M. J. (2010). „HumanEva: Synchronized Video and Motion Capture Dataset and Baseline Algorithm for Evaluation of Articulated Human Motion“. In: *International Journal of Computer Vision* 87.1–2, S. 4–27.
- Simonidis, C., Stelzner, G. und Seemann, W. (2007). „A Kinematic Study of Human Torso Motion“. In: *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*. Bd. 8, S. 773–780.
- Steffen, J., Pardowitz, M. und Ritter, H. (2009). „Using Structured UKR Manifolds for Motion Classification and Segmentation“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4785–4790.
- Steinbring, J., Mandery, C., Pfaff, F., Faion, F., Asfour, T. und Hanebeck, U. D. (2016a). „Real-Time Whole-Body Human Motion Tracking Based on Unlabeled Markers“. In: *IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. Baden-Baden, Germany, S. 583–590.

- Steinbring, J., Mandery, C., Vahrenkamp, N., Asfour, T. und Hanebeck, U. D. (2015). „High-Accuracy Real-Time Whole-Body Human Motion Tracking Based on Constrained Nonlinear Kalman Filtering“. In: *arXiv preprint arXiv:1511.04278*.
- Steinbring, J., Pander, M. und Hanebeck, U. D. (2016b). „The Smart Sampling Kalman Filter with Symmetric Samples“. In: *Journal of Advances in Information Fusion* 11.1, S. 71–90.
- Steinwart, I. und Christmann, A. (2008). *Support Vector Machines*. New York, USA: Springer.
- Stolcke, A. (2002). „SRILM – An Extensible Language Modeling Toolkit“. In: *International Conference on Spoken Language Processing*. Bd. 2, S. 901–904.
- Suárez, R., Rosell, J. und García, N. (2015). „Using Synergies in Dual-Arm Manipulation Tasks“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 5655–5661.
- Suleiman, W., Yoshida, E., Laumond, J.-P. und Monin, A. (2007). „On Humanoid Motion Optimization“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 180–187.
- Summers-Stay, D., Teo, C. L., Yang, Y., Fermüller, C. und Aloimonos, Y. (2012). „Using a Minimal Action Grammar for Activity Understanding in the Real World“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4104–4111.
- Takano, W. und Nakamura, Y. (2015). „Statistical Mutual Conversion Between Whole Body Motion Primitives and Linguistic Sentences for Human Motions“. In: *The International Journal of Robotics Research* 34.10, S. 1314–1328.

- Takano, W., Yamane, K., Sugihara, T., Yamamoto, K. und Nakamura, Y. (2006). „Primitive Communication Based on Motion Recognition and Generation with Hierarchical Mimesis Model“. In: *IEEE International Conference on Robotics and Automation (ICRA)*, S. 3602–3609.
- Tassa, Y., Erez, T. und Todorov, E. (2012). „Synthesis and Stabilization of Complex Behaviors Through Online Trajectory Optimization“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 4906–4913.
- Taylor, G. W. und Hinton, G. E. (2009). „Factored Conditional Restricted Boltzmann Machines for Modeling Motion Style“. In: *International Conference on Machine Learning (ICML)*, S. 1025–1032.
- Taylor, G. W., Hinton, G. E. und Roweis, S. T. (2006). „Modeling Human Motion Using Binary Latent Variables“. In: *Advances in Neural Information Processing Systems 19*. MIT Press, S. 1345–1352.
- Terlemez, Ö. (2017). „Referenzmodell des menschlichen Körpers zur Generierung und zum Transfer menschlicher Bewegungen auf humanoide Roboter“. Diss. Karlsruhe Institute of Technology (KIT).
- Terlemez, Ö., Ulbrich, S., Mandery, C., Do, M., Vahrenkamp, N. und Asfour, T. (2014). „Master Motor Map (MMM) – Framework and Toolkit for Capturing, Representing, and Reproducing Human Motion on Humanoid Robots“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*. Madrid, Spain, S. 894–901.
- Tsoumakas, G. und Katakis, I. (2007). „Multi-Label Classification: An Overview“. In: *International Journal of Data Warehousing and Mining 3.3*, S. 1–13.

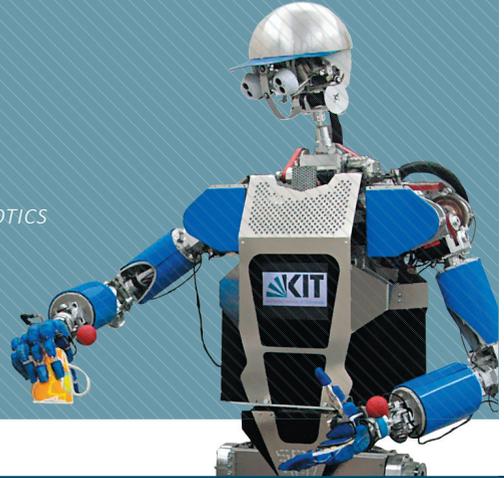
- Ude, A., Atkeson, C. G. und Riley, M. (2004). „Programming Full-Body Movements for Humanoid Robots by Observation“. In: *Robotics and Autonomous Systems* 47.2, S. 93–108.
- Ude, A., Gams, A., Asfour, T. und Morimoto, J. (2010). „Task-Specific Generalization of Discrete and Periodic Dynamic Movement Primitives“. In: *IEEE Transactions on Robotics* 26.5, S. 800–815.
- Vahrenkamp, N., Kröhnert, M., Ulbrich, S., Asfour, T., Metta, G., Dillmann, R. und Sandini, G. (2012). „Simox: A Robotics Toolbox for Simulation, Motion and Grasp Planning“. In: *International Conference on Intelligent Autonomous Systems (IAS)*, S. 585–594.
- Vahrenkamp, N., Wächter, M., Kröhnert, M., Welke, K. und Asfour, T. (2015). „The Robot Software Framework ArmarX“. In: *it – Information Technology* 57.2, S. 99–111.
- Vaillant, J., Kheddar, A., Audren, H., Keith, F., Brossette, S., Escande, A., Bouyarmane, K., Kaneko, K., Morisawa, M., Gergondet, P., Yoshida, E., Kajita, S. und Kanehiro, F. (2016). „Multi-Contact Vertical Ladder Climbing with an HRP-2 Humanoid“. In: *Autonomous Robots* 40.3, S. 561–580.
- Vens, C., Struyf, J., Schietgat, L., Džeroski, S. und Blockeel, H. (2008). „Decision Trees for Hierarchical Multi-Label Classification“. In: *Machine Learning* 73.2, S. 185–214.
- Venture, G., Ayusawa, K. und Nakamura, Y. (2008). „Motion Capture Based Identification of the Human Body Inertial Parameters“. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, S. 4575–4578.
- Venture, G., Ayusawa, K. und Nakamura, Y. (2009). „Real-Time Identification and Visualization of Human Segment Parameters“. In: *Annual*

- International Conference of the IEEE Engineering in Medicine and Biology Society*, S. 3983–3986.
- Vicon (o.D.). *Vicon Motion Capture Systems*. URL: <http://www.vicon.com/> (besucht am 17. 05. 2017).
- Vögele, A., Krüger, B. und Klein, R. (2014). „Efficient Unsupervised Temporal Segmentation of Human Motion“. In: *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, S. 167–176.
- Wächter, M. und Asfour, T. (2015). „Hierarchical Segmentation of Manipulation Actions Based on Object Relations and Motion Characteristics“. In: *International Conference on Advanced Robotics (ICAR)*, S. 549–556.
- Wächter, M., Schulz, S., Asfour, T., Aksoy, E., Wörgötter, F. und Dillmann, R. (2013). „Action Sequence Reproduction Based on Automatic Segmentation and Object-Action Complexes“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 189–195.
- Werner, A., Henze, B., Rodriguez, D. A., Gabaret, J., Porges, O. und Roa, M. A. (2016). „Multi-Contact Planning and Control for a Torque-Controlled Humanoid Robot“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 5708–5715.
- Wieber, P.-B., Billet, F., Boissieux, L. und Pissard-Gibollet, R. (2006). „The HuMANs Toolbox, a Homogenous Framework for Motion Capture, Analysis and Simulation“. In: *International Symposium on the 3D Analysis of Human Movement*.
- Wilson, A. D. und Bobick, A. F. (1999). „Parametric Hidden Markov Models for Gesture Recognition“. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21.9, S. 884–900.

- Winter, D. A. (1984). „Kinematic and Kinetic Patterns in Human Gait: Variability and Compensating Effects“. In: *Human Movement Science* 3.1–2, S. 51–76.
- Winter, D. A. (2009). *Biomechanics and Motor Control of Human Movement*. 4. ed. Hoboken, NJ, USA: Wiley.
- Witten, I. H. und Bell, T. C. (1991). „The Zero-Frequency Problem: Estimating the Probabilities of Novel Events in Adaptive Text Compression“. In: *IEEE Transactions on Information Theory* 37.4, S. 1085–1094.
- Wojtusich, J. und Stryk, O. von (2015). „HuMoD – A Versatile and Open Database for the Investigation, Modeling and Simulation of Human Motion Dynamics on Actuation Level“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 74–79.
- Wu, L., Li, M., Li, Z., Ma, W.-Y. und Yu, N. (2007). „Visual Language Modeling for Image Classification“. In: *International Workshop on Multimedia Information Retrieval (MIR)*, S. 115–124.
- Yamane, K. und Hodgins, J. (2009). „Simultaneous Tracking and Balancing of Humanoid Robots for Imitating Human Motion Capture Data“. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, S. 2510–2517.
- Yamane, K., Hodgins, J. und Brown, H. (2003a). „Controlling a Marionette with Human Motion Capture Data“. In: *IEEE International Conference on Robotics and Automation (ICRA)*. Bd. 3, S. 3834–3841.
- Yamane, K. und Nakamura, Y. (2003b). „Dynamics Filter – Concept and Implementation of Online Motion Generator for Human Figures“. In: *IEEE Transactions on Robotics and Automation* 19.3, S. 421–432.

- Yamane, K., Yamaguchi, Y. und Nakamura, Y. (2011). „Human Motion Database with a Binary Tree and Node Transition Graphs“. In: *Autonomous Robots* 30.1, S. 87–98.
- Yamato, J., Ohya, J. und Ishii, K. (1992). „Recognizing Human Action in Time-Sequential Images Using Hidden Markov Model“. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, S. 379–385.
- Yang, Y., Aloimonos, Y., Fermüller, C. und Aksoy, E. E. (2015). „Learning the Semantics of Manipulation Action“. In: *Annual Meeting of the Association for Computational Linguistics (ACL)*, S. 676–686.
- Yoshida, E., Belousov, I., Esteves, C. und Laumond, J.-P. (2005). „Humanoid Motion Planning for Dynamic Tasks“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 1–6.
- Yu, E. und Aggarwal, J. (2006). „Detection of Stable Contacts for Human Motion Analysis“. In: *ACM International Workshop on Video Surveillance and Sensor Networks (VSSN)*, S. 87–94.
- Yu, T., Shen, X., Li, Q. und Geng, W. (2005). „Motion Retrieval Based on Movement Notation Language“. In: *Computer Animation and Virtual Worlds* 16.3–4, S. 273–282.
- Zhang, M.-L. und Zhou, Z.-H. (2014). „A Review on Multi-Label Learning Algorithms“. In: *IEEE Transactions on Knowledge and Data Engineering* 26.8, S. 1819–1837.
- Zhang, Y., Luo, J., Hauser, K., Ellenberg, R., Oh, P., Park, H. A., Paldhe, M. und Lee, C. G. (2013). „Motion Planning of Ladder Climbing for Humanoid Robots“. In: *IEEE International Conference on Technologies for Practical Robot Applications (TePRA)*, S. 1–6.

- Zhao, X., Li, X., Pang, C., Zhu, X. und Sheng, Q. Z. (2013). „Online Human Gesture Recognition from Motion Data Streams“. In: *ACM International Conference on Multimedia*, S. 23–32.
- Zheng, Y. und Yamane, K. (2013). „Human Motion Tracking Control with Strict Contact Force Constraints for Floating-Base Humanoid Robots“. In: *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, S. 34–41.
- Zhou, H., You, M., Liu, L. und Zhuang, C. (2017). „Sequential Data Feature Selection for Human Motion Recognition via Markov Blanket“. In: *Pattern Recognition Letters* 86, S. 18–25.



Strategien zur effizienten und robusten Fortbewegung in menschenzentrierten Umgebungen stellen eine unabdingbare Voraussetzung für den Einsatz von humanoiden Robotern im unmittelbaren Umfeld des Menschen dar. Insbesondere Multi-Kontakt-Bewegungen, bei denen zusätzliche Support-Kontakte etwa mit den Händen genutzt werden, repräsentieren einen vielversprechenden, vom menschlichen Vorbild inspirierten Bewegungstypus, sind jedoch mit etablierten Verfahren der Bewegungsplanung bislang nicht zufriedenstellend handhabbar.

Die vorliegende Arbeit stellt einen neuartigen datengetriebenen Ansatz für die Generierung derartiger Multi-Kontakt-Bewegungen bei humanoiden Robotern vor, der auf der Beobachtung und Analyse menschlicher Ganzkörperbewegungen beruht. Hierzu wird untersucht, wie erfasste Ganzkörperbewegungen repräsentiert, mit Techniken des maschinellen Lernens klassifiziert und in einer großskaligen Bewegungsdatenbank organisiert werden können. Die Segmentierung menschlicher Bewegungen liefert im Anschluss die erforderlichen Daten zum Lernen eines statistischen Modells für die Transitionen zwischen charakteristischen Ganzkörperposen, das mit einem heuristischen Suchalgorithmus eine Generierung von Multi-Kontakt-Bewegungen für humanoide Roboter ermöglicht.

ISBN 978-3-7315-0729-1



9 783731 507291 >

ISSN 2512-0875
ISBN 978-3-7315-0729-1

Gedruckt auf FSC-zertifiziertem Papier